FIFTH EDITION

ENCYCLOPEDIA OF ENVIRONMENTAL SCIENCE and ENGINEERING

EDITED BY
JAMES R. PFAFFLIN
EDWARD N. ZIEGLER

CRC Taylor & Francis
Taylor & Francis Group

FIFTH EDITION

ENCYCLOPEDIA OF
ENVIRONMENTAL SCIENCE and ENGINEERING

VOLUME 1 A-L

FIFTH EDITION

ENCYCLOPEDIA OF ENVIRONMENTAL SCIENCE and ENGINEERING

VOLUME 1 A-L

EDITED BY
JAMES R. PFAFFLIN
EDWARD N. ZIEGLER

Taylor & Francis
Taylor & Francis Group
Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the
Taylor & Francis Group, the academic division of T&F Informa plc.

Published in 2006 by
CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2006 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group

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

International Standard Book Number-10: 0-8493-9843-6 (Hardcover)
International Standard Book Number-13: 978-0-8493-9843-8 (Hardcover)

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

No part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC) 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Catalog record is available from the Library of Congress

informa

Taylor & Francis Group is the Academic Division of Informa plc.

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

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

EDITORS

JAMES R. PFAFFLIN

EDWARD N. ZIEGLER
Polytechnic University

EDITORIAL ADVISORY BOARD

Thou ever-darting Globe! through Space and Air!
Thou waters that encompass us!
Thou that in all the life and death of us, in action or in sleep!
Thou laws invisible that permeate them and all,
Thou that in all, and over all, and through and under all, incessant!

Thou! thou! the vital, universal, giant force resistless, sleepless, calm,
Holding Humanity as in thy open hand, as some ephemeral toy,
How ill to e'er forget thee!

One thought ever at the fore—
That in the Divine Ship, the World, breasting Time and Space,
All Peoples of the globe together sail, sail the same voyage, are bound
to the same destination.

—Walt Whitman (ca 1890)

CONTENTS

VOLUME 1

VOLUME 2

FOREWORD

The editors were honored to have had the foreword to the first four editions written by the eminent thinker and renowned environmental engineer, the late Dr. Abel Wolman. His insights through the decades proved to be accurate and his overview is shared in this fifth edition as inspiration to innovators in the coming decades and in memory of his distinguished contributions to the environmental field.

The 1980s appear in a world dominated by change at an unprecedented pace. Familiar and new problems tumble over each other and are communicated globally by the second, instead of by the month. Scientific and technologic choices are spawned day by day, while search for innovation is stimulated anew by government, universities, and private industry.

Practitioners are startled by these events and try to keep apace with society's demands by pressing for relevant research, implementation of findings, and translating their significance to the public they serve. It is within this challenging setting that a fifth edition of the *Encyclopedia of Environmental Science and Engineering* is born. Its content is intended to register the striking characteristics of the changes we note to eliminate the already obsolete and to expose the new on the horizon.

In the turbulence of the sixties and seventies, policies, plans, solutions, and regulations flowed without interruption from legislative and executive halls. The eighties might appropriately be devoted to post-audit stocktaking and reorientation of both policy and action. Clarification of purpose in confrontation of the problems of the environment is overdue. Validation of our efforts, over the last two decades, should provide an arena of productivity for science and engineering to guide us through the coming decades.

As manpower and money are always in short supply, even in so-called rich countries, they must be husbanded. How to use them with maximum competence and logic, minimum risk, and least cost is a continuing obligation in the protection and control of the biosphere.

We must extricate ourselves from adversarial combat in a world of negativism and move to an orderly identification of what we know and away from the hysterical search for a doubtful Utopia. The authors in this fifth edition continue the pursuit of new knowledge, calculated to bring new fruits of health, safety, and comfort to man and his environs. The charms, as well as the subtle hazards, of the terms "conservation, preservation, and ecology" need to be crystallized so that the public and their decision-makers practice this complex art with clearer conception and perception than is apparent in recent bitter confrontations.

ABEL WOLMAN

EDITORS' PREFACE

In the editors' preface to the fourth edition it was noted that there was good news and there was bad news. It is the same for this, the fifth edition. One suspects that this will always be the case.

The 2004 Nobel Prize for Peace has been awarded to Professor Dr. Wangari Maathai. Dr. Maathai's award was based on her efforts on behalf of conservation and women's rights. These efforts were made at great personal risk.

In addition, the Kyoto Protocol has been ratified by the requisite number of countries. The bad news is that some developed nations have declined to join this global effort. It is to be hoped that, in time, these countries will change their policies.

Protection of the environment is an ongoing struggle, and it is incumbent on all citizens of the planet to join in protecting the only home that we have.

EDITORS

James R. Pfafflin holds degrees from Indiana State University, Johns Hopkins University and the University of Windsor. He is a professional engineer in Ontario, a chartered engineer in the UK and also holds the title of European Engineer (EUR ING). He is a member of the Commissioned Reserve of the US Public Health Service.

Edward N. Ziegler is Associate Professor of Chemical & Biological Engineering at Polytechnic University and former director of its Environmental Science Program. Dr. Ziegler, a Distinguished Teacher Award recipient, teaches graduate courses, one in air pollution engineering control and another in chemical reactor analysis in addition to undergraduate chemical engineering courses. He earned his BS in Ch E from CCNY (City University of New York) and MS and PhD degrees from Northwestern University. He is a member of the American Institute of Chemical Engineers (Environmental Division) and the Air & Waste Management Association (Education Committee). Dr. Ziegler is also a consultant to private industry and government. He received a US Environmental Protection Agency bronze medal (Region 2) for his technical assistance to the Agency team in a power plant pollution control settlement.

LIST OF CONTRIBUTORS

ROGER J. ALESBURY—British Petroleum. Industrial Hygiene Engineering

HAMBDY I. ALI—Ain Shams University. Physical and Chemical Treatment of Wastewaters

LARRY G. ANDERSON—University of Colorado at Denver. Atmospheric Chemistry

ALESSANDRO ANZALONE—University of South Florida. Physical and Chemical Treatment of Wastewaters. Urban Air Pollution Modeling

NEAL E. ARMSTRONG—University of Texas at Austin. Modeling of Estuarine Water Quality

STANLEY I. AUERBACH—Oak Ridge National Laboratory. Radiation Ecology

PAUL BAHAM—U.S. Merchant Marine Academy. Effects of Chemicals

JOHN E. BARDACH—University of Michigan. Pollution Effects on Fish

B. BELESSIOTIS—National Center for Scientific Research (Greece). Desalination

J. K. BEWTRA—University of Windsor. Biological Treatment of Wastewater. Physical and Chemical Treatment of Wastewaters

N. BISWAS—University of Windsor. Biological Treatment of Wastewater

MICHAEL BRUNO—Stevens Institute of Technology. Oceanography

GERARD P. CANEVARI—Cranford, New Jersey. Oil Spillage into Water—Treatment

RAUL R. CARDENAS, JR.—City College of New York. Composting

PAUL CHAN—New Jersey Institute of Technology. Groundwater Resources

S. P. CHEE—University of Windsor. Water Flow

JONATHAN CHIPMAN—University of Wisconsin. Remote Sensing

TAI-MING CHIU—Institute of Nuclear Energy Research (Taiwan). Vapor and Gaseous Pollutant Fundamentals

LEONARD L. CIACCIO—Ramapo College. Instrumentation: Water and Wastewater Analysis

JOHN D. CONSTANCE—Cliffside Park, New Jersey. Indoor Air Pollution. Prevention of Toxic Chemical Release

CHARLES C. COUTANT—Oak Ridge National Laboratory. Thermal Effects on Fish Ecology

MICHAEL DEBONIS—Federal Emergency Management Agency. Management of Solid Waste

JOHN B. DE HOFF—Cockeysville, Maryland. Community Health

DONALD A. DEIESO—Research Cottrell. Radon

E. DELYANNIS—National Center for Scientific Research (Greece). Desalination

RICHARD T. DEWLING—Dewling Associates, Inc. Hazardous Waste Management. Radon

YUAN DING—New Jersey Institute of Technology. Groundwater Resources. Management of Radioactive Wastes

P. J. DITCHETT—University of Wales. Disinfection

LEE DORIGAN—King County Department of Natural Resources. Brownfields

ROBERT DRESNACK—New Jersey Institute of Technology. Environmental Assessments and Related Impacts. Eutrophication

FRANCIS C. EVANS—University of Michigan. Ecology of Primary Terrestrial Consumers

ROBERT J. FARRELL—ExxonMobil. Coal Gasification Processes

EDWARD F. FERRAND—Edward F. Ferrand Associates. Air Pollutant Effects. Hazardous Wastes

RICHARD FIELD—Environmental Protection Agency. Urban Runoff

MARTIN FORSBERG—Harvard University. Water Chemistry. Water: Properties, Structure and Occurrence in Nature

R. BUCKMINSTER FULLER—Southern Illinois University. The Terrestrial System

JAMES GEIGER—Betz Converse Murdoch, Inc. Air Pollution Instrumentation

STEVEN GHERINI—Harvard University. Water Chemistry. Water: Properties, Structure and Occurrence in Nature

FRANK S. GILL—Hampshire, United Kingdom. Industrial Hygiene Engineering

WILLIAM GOLDFARB—Rutgers University. Environmental Law

CHARLES R. GOLDMAN—University of California, Davis. Aquatic Primary Production

EUGENE B. GOLUB—New Jersey Institute of Technology. Environmental Education

T. E. GRAEDEL—Yale University. Industrial Ecology

THOMAS GRANGER—Ebasco Services. Planning for New Processes: Environmental Aspects

HELENE N. GUTTMAN—U.S. Department of Agriculture. Microbiology

ERNST P. HALL—U.S. Environmental Protection Agency. Pollution from Mine Drainage

TODD HEPWORTH—University of Wisconsin. Geographic Information Systems

RICHARD HIRES—Stevens Institute of Technology. Oceanography

KENNETH C. HOFFMAN—Mathtech, Inc. Energy Sources—Alternatives

PHILIP H. JONES—Griffith University. Water Treatment

MOHAMMED S. KAMAL—University of Texas at Arlington. Natural Systems for Wastewater Treatment

MARCUS E. KANTZ—Environmental Protection Agency. Stack Sampling

YASUO KOUSAKA—California Institute of Technology. Aerosols

J. H. LANGE—Envirosafe Training and Consultants. Epidemiology. Toxicology

PETER B. LEDERMAN—Peter B. Lederman Associates. Management of Solid Waste

CAMERON MACINNIS—Toronto, Ontario. Municipal Wastewater

B. J. MASON—Imperial College. Greenhouse Gases Effects

JOHN M. MATSEN—Lehigh University. Particulate Emissions. Particulate Removal

COLIN A. MAWSON—Ottawa, Ontario. Management of Radioactive Wastes

J. K. MCCARTHY—Rutgers University. Composting

J. A. MCCORQUODALE—University of New Orleans. Sediment Transport and Erosion

ELIZABETH MCLOUGHLIN—PS&S Keyspan. Planning

JEHUDA MENCZEL—U.S. Environmental Protection Agency. Air Pollution Sources

ROBERT L. METCALF—University of Illinois. Pesticides

FRIEDRICH-WERNER MÖLLENKAMP—Fichtner Beratende Ingenieure. Energy Sources—Alternatives

MARK D. MUELLER—Betz Converse Murdoch. Air Pollution Instrumentation

GERALD P. NICHOLLS—New Jersey Department of Environmental Protection. Radon

EUGENE P. ODUM—University of Georgia. Ecosystem Theory

KIKUO OKUYAMA—California Institute of Technology. Aerosols

HANS A. PANOFSKY—Pennsylvania State University. Air Pollution Meteorology

CLINTON E. PARKER—University of Texas at Arlington. Industrial Waste Management

JAMES R. PFAFFLIN—Gillette, New Jersey. Effects of Chemicals. Municipal Wastewater

SHEILA M. PFAFFLIN—AT&T. Psychological Aspects of Man's Environment. Statistical Methods for Environmental Science

GREGORY A. PIKUL—Dewling Associates, Inc. Hazardous Waste Management

SYED R. QASIM—University of Texas at Arlington. Industrial Waste Management. Natural Systems for Wastewater Treatment. Planning Water Supply and Sanitation Projects in Developing Nations. Small Flow Wastewater Treatment for Domestic and Special Applications. Water and Waste Management Systems in Space

MICHAEL C. QUICK—University of British Columbia. Hydrology

ROBERT H. QUIG—Ogden Products, Inc. Planning for New Processes: Environmental Aspects

ROGER G. RAMSDELL, JR.—Rockville Center, New York. Electrostatic Precipitation

PRASANNA RATNAWEER—Open University, Sri Lanka. Water Reuse

ELROY L. RICE—University of Oklahoma. Ecology of Plants

DONALD P. ROSEMAN—David Taylor Research Center. Marine Spillage—Sources and Hazards

A. D. RUSSELL—University of Wales. Disinfection

JOSEPH A. SALVATO, JR.—Troy, New York. Environmental Health

JAMES SANDERSON—U.S. Environmental Protection Agency. Fluidized Bed Combustion

JOHN R. SCHURING, JR.—New Jersey Institute of Technology. Groundwater Resources

JOHN H. SEINFELD—California Institute of Technology. Aerosols

FRANCES PAULA SOLOMON—King County Department of Natural Resources. Limnology

GARY J. STENSLAND—Illinois Department of Natural Resources. Acid Rain

WERNER STUMM—Swiss Federal Institute of Technology. Water Chemistry. Water: Properties, Structure and Occurrence in Nature

W. MICHAEL SUTTON—New York City Department of Environmental Protection. Nitrogen Oxides Reduction

MARK A. TOMPECK—Hatch Mott MacDonald. Recycling Waste Materials. Water Treatment

TAO WANG—Yale University. Industrial Ecology

IAN WEBBER—Advisor to Government of Indonesia. PCBs and Associated Aromatics

GEORGE M. WILKENING—Bell Laboratories. Non-ionizing Radiations

CHARLES E. WILSON—New Jersey Institute of Technology. Noise

DONALD G. WRIGHT—Environmental Protection Agency. Stack Sampling

VICTOR J. YANNACONE, JR.—Patchogue, New York. Legal Aspects of the Environment

R. G. ZACHARIADIS—University of Texas at Arlington. Water and Waste Management Systems in Space

EDWARD N. ZIEGLER—Polytechnic University. Coal Gasification Processes. Fossil Fuel Cleaning Processes. Mobile Source Pollution. Nitrogen Oxides Reduction. Planning for New Processes: Environmental Aspects. Vapor and Gaseous Pollutant Fundamentals. Appendix

LIST OF DECEASED AUTHORS

JOHN D. CONSTANCE—Indoor Air Pollution. Prevention of Toxic Chemical Release

R. BUCKMINSTER FULLER—The Terrestrial System

PHILIP H. JONES —Water Treatment

HANS A. PANOFSKY—Air Pollution Meteorology

WERNER STUMM—Water Chemistry. Water: Properties, Structures and Occurrence in Nature

GEORGE M. WILKENING—Non-Ionizing Radiations

A

ACID RAIN

OVERVIEW OF THE PROBLEM

Acid rain is the general and now popular term that pertains to both acid rain and acid snow. This article discusses the physical and chemical aspects of the acid rain phenomenon, presents results from a U.S. monitoring network to illustrate spatial and seasonal variability, and discusses time trends of acid rain during recent decades. A chemical equilibrium model is presented to emphasize that one cannot measure only pH and then expect to understand why a particular rain or melted snow sample is acidic or basic. Monitoring networks are now in operation to characterize the time trends and spatial patterns of acid rain. Definitions, procedures, and results from such measurement programs are discussed. The monitoring results are necessary to assess the effects of acid rain on the environment, a topic only briefly discussed in this article.

Chemicals in the form of gases, liquids, and solids are continuously deposited from the air to the plants, soils, lakes, oceans, and manmade materials on the earth's surface. Water (H_2O) is the chemical compound deposited on the earth's surface in the greatest amount. The major atmospheric removal process for water consists of these steps: (1) air that contains water vapor rises, cools, and condenses to produce liquid droplets, i.e., a visible cloud; (2) in some clouds the water droplets are converted to the solid phase, ice particles; (3) within some clouds the tiny liquid droplets and ice particles are brought together to form particles that are heavy enough to fall out of the clouds as rain, snow, or a liquid–solid combination. When these particles reach the ground, a precipitation event has occurred. As water vapor enters the base of clouds in an air updraft in step (1) above, other solid, liquid, and gaseous chemicals are also entering the clouds. The chemicals that become incorporated into the cloud water (liquid or ice) are said to have been removed by in-cloud scavenging processes often called rainout. The chemicals that are incorporated into the falling water (liquid or ice) below the cloud are said to be removed by below-cloud scavenging, often called washout.

Carbon dioxide gas, at the levels present in the atmosphere, dissolves in pure water to produce a carbonic acid solution with a pH of about 5.6. Therefore, this value is usually considered to be the neutral or baseline value for rain and snow. Measurements show that there are always additional chemicals in rain and snow. If a salt (sodium chloride) particle in the air is scavenged (captured) by a raindrop or snow flake, it does not alter the acidity. If an acid particle, such as one composed of sulfuric acid, is scavenged, then the rain or snow becomes more acid. If a basic particle, such as a dust particle composed of calcium carbonate, is scavenged then the rain or snow becomes more basic. It is important that both pH as well as the major chemicals that alter the pH of rain and snow be included in routine measurement programs. The adverse or beneficial effects of acid rain are not related only to the hydrogen ion concentration (a measure of acidity level), but also to the other chemicals present.

In following the cycle of chemicals through the atmosphere one considers (1) the natural and manmade sources emitting chemicals to the atmosphere, (2) the transport and transformation of the chemicals in the atmosphere, and (3) the removal of the chemicals from the atmosphere. Therefore, when one regularly measures (monitors) the quantity of chemicals removed from the atmosphere, indirect information is obtained about the removal rates and processes, the transport/transformation rates and processes, and the source characteristics.

A great number of projects have been carried out to measure various chemicals in precipitation. For example, Gorham (1958) reported that hydrochloric acid should be considered in assessing the causes of rain acidity in urban areas. Junge (1963) summarized research discussing the role of sea salt particles in producing rain from clouds. Even as far back as 1872, Robert Anges Smith discussed the relationship between air pollution and rainwater chemistry in his remarkable book entitled *Air and Rain: The Beginnings of A Chemical Climatology* (Smith, 1872). These three examples indicate that the measurement of chemicals in precipitation is not just a recent endeavor. Certainly one reason for the large number of studies is the ease of collecting samples, i.e., the ease of collecting rain or snow. Over time and from project to project during a given time period, the purpose for the rain and snow chemistry measurements has varied, and thus the methods and the chemical parameters being measured have varied greatly.

The surge of interest in the 1980s in the acidity levels of rain and snow was strongly stimulated by Scandinavian studies reported in the late 1960s and early 1970s. These studies reported that the pH of rain and snow in Scandinavia during the period from 1955 to 1965 had decreased dramatically. The Scandinavians also reported that a large number of lakes, streams, and rivers in southern Norway and Sweden were devoid or becoming devoid of fish. The hypothesis was that this adverse effect was primarily the result of acid rain, which had caused the the lakes to become increasingly more acidic.

Later studies with improved sampling and analysis procedures, confirmed that the rain and snow in southern Norway and Sweden were quite acid, with average pH values of about 4.3. The reports sometimes considered the idea that changes in the acidity of the lakes were partially the result of other factors including landscape changes in the watershed, but usually the conclusion was that acid rain was the major cause of the lake acidification and that the acid rain is primarily the result of long-range transport of pollutants from the heavily industrialized areas of northern Europe.

The rain and snow in portions of eastern Canada and the eastern United States are as acid as in southern Scandinavia, and some lakes in these areas also are too acid to support fish. Studies have confirmed that many of the lakes sensitive to acid rain have watersheds that provide relatively small inputs of neutralizing chemicals to offset the acid rain and snow inputs.

Any change in the environment of an ecological system will result in adjustments within the system. Increasing the acid inputs to the system will produce changes or effects that need to be carefully assessed. Effects of acid rain on lakes, row crops, forests, soils, and many other system components have been evaluated. Evans *et al.* (1981) summarized the status of some of these studies and concluded that the acid rain effects on unbuffered lakes constituted the strongest case of adverse effects, but that beneficial effects could be identified for some other ecological components.

During the 1980s a tremendous amount of acid rain research was completed. More than 600 million dollars was spent by United States federal agencies on acid rain projects. The federal effort was coordinated through the National Acid Precipitation Assessment Program (NAPAP). This massive acid rain research and assessment program was summarized in 1990 in 26 reports of the state of science and technology which were grouped into four large volumes (NAPAP, 1990): Volume I—Emissions, Atmospheric Processes, and Deposition; Volume II—Aquatic Processes and Effects; Volume III—Terrestrial, Materials, Health, and Visibility Effects; and Volume IV—Control Technologies, Future Emissions, and Effects Valuation. The final assessment document (NAPAP, 1991) was a summary of the causes and effects of acidic deposition and a comparison of the costs and effectiveness of alternative emission control scenarios. Since adverse effects of acid rain on fish have been of particular interest to the general public, it is appropriate to note the following NAPAP (1991, pages 11–12) conclusions on this subject:

- Within acid-sensitive regions of the United States, 4 percent of the lakes and 8 percent of the streams are chronically acidic. Florida has the highest percentage of acidic surface waters (23 percent of the lakes and 39 percent of the streams). In the mid-Atlantic Highlands, mid-Atlantic Coastal Plain, and the Adirondack Mountains, 6 to 14 percent of the lakes and streams are chronically acidic. Virtually no (<1 percent) chronically acidic surface waters are located in the Southeastern Highlands or the mountainous West.
- Acidic lakes tended to be smaller than nonacidic lakes; the percentage of acidic lake area was a factor of 2 smaller than the percentage of acidic lakes based on the numbers.
- Acidic deposition has caused some surface waters to become acidic in the United States. Naturally produced organic acids and acid mine drainage are also causes of acidic conditions.
- Fish losses attributable to acidification have been documented using historical records for some acidic surface waters in the Adirondacks, New England, and the mid-Atlantic Highlands. Other lines of evidence, including surveys and the application of fish response models, also support this conclusion.

In future years the effects on materials such as paint, metal and stone should probably be carefully evaluated because of the potentially large economic impact if these materials undergo accelerated deterioration due to acid deposition.

DEFINITIONS

Some widely used technical terms that relate to acid rain and acid rain monitoring networks are defined as follows:

1) **pH** The negative logarithm of the hydrogen ion activity in units of moles per liter (for precipitation solutions, concentration can be substituted for activity). Each unit decrease on the pH scale represents a 10-fold increase in acidity. In classical chemistry a pH less than 7 indicates acidity; a pH greater than 7 indicates a basic (or alkaline) solution; and a pH equal to 7 indicates neutrality. However, for application to acid rain issues, the neutral point is chosen to be about 5.6 instead of 7.0 since this is the approximate equilibrium pH of pure water with ambient outdoor levels of carbon dioxide.
2) **Precipitation** This term denotes aqueous material reaching the earth's surface in liquid or solid form, derived from the atmosphere. Dew, frost,

and fog are technically included but in practice are poorly measured, except by special instruments. The automatic devices currently in use to sample precipitation for acid rain studies collect rain and "wet" snow very efficiently; collect "dry" snow very inefficiently; and collect some fog water, frost and dew, but these usually contribute very little to the annual chemical deposition at a site.

3) **Acid Rain** A popular term with many meanings; generally used to describe precipitation samples (rain, melted snow, melted hail, etc.) with a pH less than 5.6. Recently the term has sometimes been used to include acid precipitation, ambient acid aerosols and gases, dry deposition of acid substances, etc., but such a broad meaning is confusing and should be avoided.
4) **Acid Precipitation** Water from the atmosphere in the form of rain, sleet, snow, hail, etc., with a pH less than 5.6.
5) **Wet Deposition** A term that refers to: (a) the amount of material removed from the atmosphere by rain, snow, or other precipitation forms; and (b) the process of transferring gases, liquids, and solids from the atmosphere to the ground during a precipitation event.
6) **Dry Deposition** A term for (a) all materials deposited from the atmosphere in the absence of precipitation; and (b) the process of such deposition.
7) **Atmospheric (or Total) Deposition** Transfer from the atmosphere to the ground of gases, particles, and precipitation, i.e., the sum of wet and dry deposition. Atmospheric deposition includes many different types of substances, non-acidic as well as acidic.
8) **Acid Deposition** The transfer from the atmosphere to the earth's surface of acidic substances, via wet or dry deposition.

PROCEDURES AND EQUIPMENT FOR WET DEPOSITION MONITORING

For data comparability it would be ideal if all wet deposition networks used the same equipment and procedures. However, this does not happen. Therefore, it is important to decide which network characteristics can produce large differences in the databases. The following discussion outlines procedures and equipment which vary among networks, past and present.

Site Location

Sites are selected to produce data to represent local, regional, or remote patterns and trends of atmospheric deposition of chemicals. However, the same site may produce a mixture of data. For example, the measured calcium concentrations at a site might represent a local pattern while the sulfate concentrations represent a regional pattern.

Sample Containers

The containers for collecting and storing precipitation must be different, depending on the chemical species to be measured. Plastic containers are currently used in most networks in measuring acidic wet deposition. Glass containers are considered less desirable for this purpose because they can alter the pH: For monitoring pesticides in precipitation, plastic containers would be unacceptable.

Sampling Mode

There are four sampling modes:

Bulk Sampling A container is continuously exposed to the atmosphere for sampling and thus collects a mixture of wet and dry deposition. The equipment is simple and does not require electrical power. Thus bulk sampling has been used frequently in the past, and it is still sometimes used for economic reasons. For many studies an estimate of total deposition, wet plus dry, is desired, and thus bulk sampling may be suitable. However, there is a continuing debate as to precisely what fraction of dry deposition is sampled by open containers. The fraction collected will probably depend on variables such as wind speed, container shape and chemical species. The continuously exposed collectors are subject to varying amounts of evaporation unless a vapor barrier is part of the design. When one objective of a study is to determine the acidity of rain and snow samples, bulk data pH must be used with great caution and ideally in conjunction with adequate blank data. For wet deposition sites that will be operated for a long time (more than one year), the labor expenses for site operation and the central laboratory expenses are large enough that wet-only or wet-dry collectors should certainly be purchased and used instead of bulk collectors in order to maximize the scientific output from the project.

Wet-Only Sampling There are a variety of automatic wet-only samplers in use today that are open only during precipitation events. Side-by-side field comparison studies have documented differences in the reaction time for the sensors, in the reliability of the instruments, and in the chemical concentrations in the samples from the different sampling devices. Wet-only sampling can also be achieved by changing bulk samples *immediately* (within minutes) at the beginning and end of precipitation events, but this is very labor-intensive if done properly.

Wet-Dry Sampling With this device, one container is automatically exposed during dry periods and the second container is exposed during precipitation periods. If the sample in the dry deposition container is not analyzed, the device becomes a wet-only collector.

Sequential Sampling A series of containers are consecutively exposed to the atmosphere to collect wet deposition samples, with the advance to a new container being triggered on a time basis, a collected volume basis, or both. These devices can be rather complicated and are usually operated only for short time periods during specific research projects.

Sample Handling

Changes in the chemicals in the sample over time are decreased through (1) the addition of preservatives to prevent biological change, (2) refrigeration, (3) aliquoting, and (4) filtering. Filtering is more effective than refrigeration for stabilizing samples for some species such as calcium and magnesium. For species such as organic acids, only chemical preservatives are certain to prevent change.

Analytical Methods

Several analytical methods are available to adequately measure the major ions found in precipitation, but special precautions are necessary because the concentrations are low and thus the samples are easily contaminated. Measurement of the chemical parameter pH, although deceptively easy with modern equipment, requires special care in order to arrive at accurate results because of the low ionic strength of rain and snow samples. Frequent checks with low ionic strength reference solutions are required to avoid the frequent problem of malfunctioning pH electrodes. The ions SO_4^{2-}, NH_4^+, Ca^{2+}, etc., are measured in modern laboratories by ion chromatography, automated colorimetry, flame atomic absorption, and other methods.

Quality Assurance/Quality Control

The chemical analysts actually performing measurements should follow documented procedures, which include measurements of "check" or "known" solutions to confirm immediately and continuously that the work is "in control" and thus is producing quality results. At an administrative level above the analysts, procedures are developed to "assure" that the results are of the quality level established for the program. These quality assurance procedures should include the submission of blind reference samples to the analysts on a random basis. Quality assurance reports should routinely be prepared to describe procedures and results so that the data user can be assured (convinced) that the data are of the quality level specified by the program. In the past, insufficient attention has been given to quality assurance and quality control. As a minimum, from 10 to 20% of the cost of a monitoring program should be devoted to quality assurance/quality control. This is especially true for measurements on precipitation samples that have very low concentrations of the acid-rain-related species and thus are easily contaminated.

CALCULATING PRECIPITATION pH

This section describes the procedures for calculating the pH of a precipitation sample when the concentrations of the major inorganic ions are known (Stensland and Semonin, 1982). Granat (1972), Cogbill and Likens (1974), and Reuss (1975) demonstrated that the precipitation pH can be calculated if the major ion concentrations are known. The procedure described below is analogous to that used by these previous workers but is formulated somewhat differently.

Three good reasons to have a method to calculate the pH are that:

1) The pH can be calculated for older data sets when pH was not measured but the major inorganic ions were measured (e.g., the Junge (1963) data set),
2) The trends or patterns of pH can be interpreted in terms of trends or patterns in the measured inorganic ions such as sulfate or calcium, and
3) The calculated pH can be compared with the measured pH to provide an analytical quality control check.

Gases (e.g., SO_2 and CO_2) and aerosols (e.g., NaCl and $(NH_4)_2SO_4$) scavenged by precipitation can remain as electrically neutral entities in the water solution or can participate in a variety of chemical transformations, including simple dissociation, to form ions (charged entities). The basic premise that the solution must remain electrically neutral allows one to develop an expression to calculate pH. Stated another way, when chemical compounds become ions in a water solution, the quantity of positive ions is equal to the quantity of negative ions. This general concept is extremely useful in discussing acid precipitation data.

As a simple example, consider a solution of only water and sulfuric acid (H_2SO_4). The solution contains H^+, OH^-, and ions. At equilibrium

$$(H^+)(OH^-) = 10^{-14}(m/L)^2$$

if the ion concentrations are expressed in moles/liter (m/L). Assuming $pH = 4$, then from the defining relation $pH = -\log(H^+)$ it follows that

$$(H^+) = 10^{-4}\ m/L$$

Therefore $(OH^-) = 10^{-10}$ m/L and thus (OH^-) is so small that it can be ignored for further calculations. Since the dissociation of the sulfuric acid in the water gives one sulfate ion for each pair of hydrogen ions, it follows that

$$(SO_4^{2-}) = 1/2(H^+) = 0.5 \times 10^{-4} m/L$$

It is useful to convert from moles/liter (which counts particles) to equivalents/liter (eq/L), as this allows one to count electrical charge and thus do an "ion balance." The conversion is accomplished by multiplying the concentration in m/L by the valance (or charge) associated with each ion. The example solution contains

$$(0.5 \times 10^{-4}\ m/L) \times (2) = 10^{-4}\ eq/L = 100\ \mu eq/L$$

of sulfate and

$$(1 \times 10^{-4}\ m/L) \times (1) = 10^{-4}\ eq/L = 100\ \mu eq/L$$

of hydrogen ion. Thus the total amount of positive charge (due to H^+ in this example) is equal to the total amount of negative charge (due to SO_4^{2-}) when the concentrations are expressed in eq/L (or μeq/L).

For most precipitation samples, the major ions are those listed in Eq. (1):

$$\begin{aligned}&\left(H^+\right)+\left(Ca^{2+}\right)+\left(Mg^{2+}\right)+\left(NH_4^+\right)+\left(Na^+\right)+\left(K^+\right)\\&\quad=\left(SO_4^{2-}\right)+\left(NO_3^-\right)+\left(Cl^-\right)+\left(OH^-\right)+\left(HCO_3^-\right)\end{aligned} \tag{1}$$

with each ion concentration expressed in μeq/L. In practice, if the actual measurements are inserted into Eq. (1), then agreement within about 15% for the two sides of the equation is probably acceptable for any one sample. Greater deviations indicate that one or more ions were measured inaccurately or that an important ion has not been measured. For example, in some samples Al^{3+} contributes a significant amount and therefore needs to be included in Eq. (1). It should be noted that assumptions concerning the parent compounds of the ions are not necessary. However, if one did know, for example, that all Na^+ and all Cl^- resulted from the dissolution of a single compound such as NaCl, then these two ions would not be necessary in Eq. (1) since they cancel out on the two sides of the equation.

There are actually two useful checks as to whether or not all the major ions have been measured. First, one compares to see that the sum of the negative charges is approximately equal to the sum of the positive charges. If all the sodium and chloride ions come entirely from the compound NaCl, then this first check would produce an equality, even if these major ions were not measured. The second check is whether the calculated conductivity is equal to the measured conductivity. The calculated conductivity is the sum of all the ions (in Eq. (1)) multiplied by the factors listed in Table 1. For low pH samples of rain or melted snow (i.e., pH < 4.5), H^+ is the major contributor to the calculated conductivity because of the relatively large value of its factor in Table 1.

TABLE 1
Conductance Factors at 25°C[a]

Ion	μS/cm per μeq/L
H^+	0.3500
HCO_3^-	0.0436
Ca^{2+}	0.0520
Cl^-	0.0759
Mg^{2+}	0.0466
NO_3^-	0.0710
K^+	0.0720
Na^+	0.0489
SO_4^{2-}	0.0739
NH_4^+	0.0745

[a] From *Standard Methods for the Examination of Water and Wastewater,* American Public Health Association, Inc., Wash., D.C., 13th Edition.

For precipitation samples, bicarbonate concentration is usually not measured. Thus both (HCO_3^-) and (OH^-) must be calculated from the measured pH. To calculate (OH^-) and (HCO_3^-) the following relationships for the dissociation of water and for the solubility and first and second dissociations of carbon dioxide in water are used:

Chemical Reaction

$$H_2O \qquad OH^- + H^+ \tag{2a}$$

$$Pco_2 \qquad H_2O\cdot CO_2 \tag{2b}$$

$$H_2O\cdot CO_2 \qquad H^+ + HCO_3^- \tag{2c}$$

$$HCO_3^- \qquad H^+ + CO_3^{2-} \tag{2d}$$

Equilibrium Relationship

$$K_W = (OH^-)(H^+) \tag{3}$$

$$K_H = \frac{\left(H_2O\cdot CO_2\right)}{Pco_2} \tag{4}$$

$$K_1 = \frac{\left(H^+\right)\left(HCO_3^-\right)}{\left(H_2O\cdot CO_2\right)} \tag{5}$$

$$K_2 = \frac{\left(H^+\right)\left(CO_3^{2-}\right)}{\left(HCO_3^-\right)} \tag{6}$$

For 25°C, $K_W = 10^{-2}$ (μeq $L^{-1})^2$, $K_H = 0.34 \times 10^{+6}$ μeq L^{-1}, $K_1 = 4.5 \times 10^{-1}$ μeq L^{-1}, and $K_2 = 9.4 \times 10^{-5}$ μeq L^{-1}.

$$\frac{\left(HCO_3^-\right)}{\left(CO_3^{2-}\right)} = \frac{\left(H^+\right)}{K_2} \tag{7a}$$

For T = 25°C and pH = 8, (H^+) = 0.01 μeq/L and thus:

$$\frac{\left(HCO_3^-\right)}{\left(CO_3^{2-}\right)} = \frac{0.01}{9.4\times10^{-5}} = 106 \tag{7b}$$

Thus the concentration of HCO_3^- is much greater than that of CO_3^{2-}. For lower pH values, HCO_3^- dominates CO_3^{2-} even more, and so CO_3^{2-} is not included in applications related to precipitation samples (i.e., Eq. (1)).

From Eqs. (4) and (5)

$$\left(HCO_3^-\right)\left(H^+\right) = K_H K_1 Pco_2 \tag{8}$$

From Eqs. (3) and (8)

$$\frac{\left(HCO_3^-\right)}{\left(OH^-\right)} = \frac{K_H K_1 Pco_2}{K_W} \tag{9}$$

where it is convenient to define

$$K = \frac{K_H K_1 Pco_2}{K_W} \tag{10}$$

Equation (1) is now rearranged to give

$$\begin{aligned}\left(H^+ - OH^- - HCO_3^-\right) &= \left(SO_4^{2-} + NO_3^- + C1^-\right) \\ &\quad - \left(Ca^{2+} + Mg^{2+} + Na^+ + K^+ + NH_4^+\right)\end{aligned} \tag{11}$$

With the definition

$$\begin{aligned}\text{Net Ions } &= \left(SO_4^{2-} + NO_3^- + Cl^-\right) \\ &\quad - \left(Ca^{2+} + Mg^{2+} + Na^+ + K^+ + NH_4^+\right)\end{aligned} \tag{12}$$

Eq. (11) becomes

$$\left(H^+ - OH^- - HCO_3^-\right) = \left(\text{Net Ions}\right) \tag{13}$$

With Eqs. (3), (9), and (10), Eq. (13) becomes the quadratic equation

$$(H^+)^2 - (\text{Net Ions})(H^+) - K_w(K + 1) = 0 \tag{14}$$

Solving for the concentration of H^+ gives

$$2(H^+) = (\text{Net Ions}) \pm [(\text{Net Ions})^2 + 4K_W(K + 1)]^{1/2} \tag{15}$$

The quantity in brackets in Eq. (15) is always positive and greater than (Net Ions), and therefore only the plus sign in front of the bracketed term provides non-negative and therefore physically realistic solutions for (H^+).

Equation (15) is rewritten in terms of pH as

$$\begin{aligned}pH = +6 - \log_{10}\{\{(\text{Net Ions}) + [(\text{Net Ions})^2 \\ + 4K_H K_1 Pco_2 + 4K_w]^{0.5}\}/2\}\end{aligned} \tag{16}$$

Equation (16) is plotted in Figure 1. If the major ions have been measured for a precipitation sample such that (Net Ions) can be determined with Eq. (12), then line B on the graph allows one to read the calculated pH. Any additional ion measured, besides those listed on the right side of Eq. (12), are simply added to Eq. (12) to make the determination of (Net Ions) just that much more accurate. If the water sample being considered is pure water in equilibrium with ambient carbon dioxide, then (Net Ions) = 0.0 and curve B indicates that the pH is less than or equal to 5.65.

The precipitation sample concentrations of HCO_3^-, OH^-, and H^+ are also shown in Figure 1, where the absolute value of the ordinate is used to read off these concentrations. It is seen that the HCO_3^- and H^+ curves approach curve B. That is, at low pH, $(H^+) \sim$ (Net Ions) and at high pH, $(HCO_3^-) \sim$ (Net Ions).

If $Pco_2 = 0$ (as it would be if one bubbled an inert gas such as nitrogen through the precipitation sample as the pH was being measured), then $K = 0$ in Eq. (10), and Eq. (16) is modified and provides the curves marked accordingly in Figure 1. In this case, with no present (cf. Eq. (8)), the asymptotic limit at high pH is provided by the OH^- curve.

The sensitivity of the pH prediction via Eq. (16) to the assumed equilibrium conditions of temperature and Pco_2 is displayed in Figure 1 by curves A to D (and of course the $Pco_2 = 0$ curve as the extreme case). At T = 25°C and $Pco_2 = 316 \times 10^{-6}$ atm, $K = 483$. Therefore at pH = 8, where $(OH^-) = 1$ μeq/L, $(HCO_3^-) = 483$ μeq/L, and this procedure explains the spacing between curves A to D and the OH^- curve in Figure 1. If the temperature is kept constant, K is proportional to Pco_2. So if we double the CO_2 level (e.g., move from curve B to C), the pH = 8 intercept for HCO_3^- jumps up to (2)(483) = 966. Curves A, B, C, and D (which are plots of Eq. (16) only at high (Net Ion) values) thus graphically demonstrate the sensitivity of pH to temperature and Pco_2. As a specific example consider that with curve B and at (Net Ions) = −49, the pH = 7; when Pco_2 is doubled (curve C), the same (Net Ion) value gives pH = 6.69; if the temperature is lower (curve D), then the pH = 6.15.

Figure 1 also demonstrates that a bimodal pH distribution would be expected if both high and low pH values are present in a particular data set. For example, assume all (Net Ion) values between +45 and −45 are equally likely. From (Net Ion) = 45 to 15, $\Delta pH = 0.48$; from (Net Ion) = 15 to −15, $\Delta pH = 1.65$; and from (Net Ion) = −15 to −45, $\Delta pH = 0.48$.

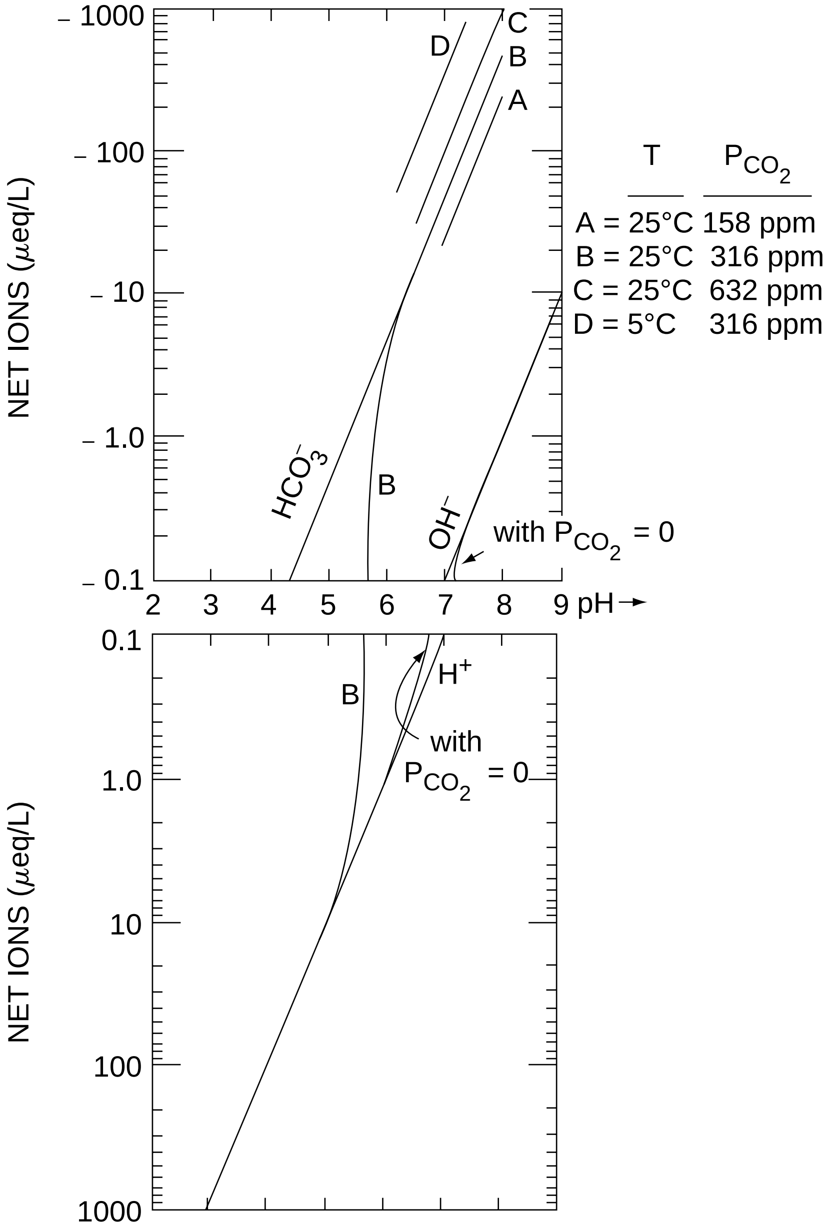


FIGURE 1 The concentration of Net Ions versus pH for precipitation samples with different values of *T* (temperature) and P_{CO_2}.

Therefore the pH will most frequently be either very large or very small, giving a bimodal distribution.

To calculate (HCO_3^-), for charge balance calculations, it is also useful to note that from equation (8),

$$\left(HCO_3^-\right) = \frac{\left(0.0153 \times 10^6\right) Pco_2}{\left(H^+\right)} \tag{17}$$

Thus, for $Pco_2 = 316 \times 10^{-6}$ atm,

$$\left(HCO_3^-\right) = \frac{4.84}{\left(H^+\right)} \tag{18}$$

Therefore, at pH = 5, $(H^+) = 10$ μeq L^{-1}, and (HCO_3^-) is only about 5% as large as (H^+).

In summary it should simply be noted that the measured ions can be combined according to Eq. (12) to produce the quantity called Net Ions, which can then be used with Eq. (16) or Figure 1 to predict the sample pH.

U.S. PRECIPITATION CHEMISTRY DATA

Many precipitation chemistry networks are being operated in the United States. Some of the networks include sites in many states, while other networks are limited to sites within a single state. For this discussion, example data from the National Atmospheric Deposition Program/National Trends Network (NADP/NTN) will be used.

The NADP/NTN began operation in 1978 with about 20 sites. By 1982 it had grown to approximately 100 sites, and by the late 1980s about 200 sites were in operation, with only the states of Rhode Island, Connecticut, and Delaware not having sites. American Samoa, Puerto Rico, and Canada each had one site. As of 1996 about 200 sites are operating. Even though the publicity about acid rain has decreased in the 1990s, the NADP/NTN has not decreased in size as some had expected. The NADP/NTN has six noteworthy characteristics:

1) The site locations were generally selected to provide precipitation chemistry data that will be representative of a *region* as opposed to a *local* area that might be dominated by a few pollution sources or by an urban area.
2) Sites are fairly long-term, operating for a minimum of five years and ideally for much longer.
3) Each site collects samples with the same automatic wet-dry collector. Sites are also equipped with a recording rain gage, an event recorder, a high-quality pH meter, a high-quality conductivity meter, and a scale to weigh the samples before they are sent to the laboratory.
4) Each site is serviced every Tuesday. The collecting bucket from the wet-side of the sampler is sent to the central laboratory each week.
5) There is a single Central Analytical Laboratory. This laboratory measures the chemical parameters for each rain and snow sample and returns clean sampling containers to the field sites. Since the inception of the program, this central laboratory has been at the Illinois State Water Survey in Champaign, Illinois.
6) Only the soluble portion of the constituents (sulfate, calcium, potassium, etc.) are measured. All NADP/NTN samples are filtered shortly after arriving at the central laboratory and this step operationally defines solubility. The fraction of the chemical species that is separated from the liquid sample and remains on the filter or remains on the inside surfaces of the collecting bucket is operationally defined as the insoluble fraction and is not measured by the NADP/NTN program. For species like sulfate, nitrate, and ammonium, the insoluble fraction is negligible while for potassium perhaps only 50 percent is soluble.

Data shown in Table 2 from the NADP/NTN weekly wet deposition network provide a quantitative chemical characterization of precipitation. Average results for the year 1984 for four sites are shown. Median ion concentrations, in units of microequivalents per liter (μeq/L), are listed. Bicarbonate (HCO_3^-) for the precipitation samples is calculated with the equations from the previous section by assuming that the samples are in equilibrium with atmospheric carbon dioxide at a level of 335×10^{-6} atm. Hydrogen ion (H^+) is calculated from the median pH for the weekly samples. The ions listed in Table 2 constitute the major ions in precipitation; this fact is supported by noting that the sum of the negatively charged ions (anions) is approximately equal to the sum of the positively charged ions (cations) for each of the four sites.

Sulfate, nitrate, and hydrogen ions predominate in the samples from the New Hampshire and Ohio sites, with levels being higher (and pH lower) at the Ohio site. For these two sites, about 70% of the sulfate plus nitrate must be in the acid form in order to account for the measured acidity (H^+). At the Nebraska site, sulfate and nitrate are higher than at the New Hampshire site, but H^+ is only 2 μeq/L (median pH = 5.80). Notice that for the Nebraska site the weighted average pH, which is a commonly reported type of average pH, is much smaller than the median pH. This indicates that one should be consistent in using the same averaging procedure when comparing pH for different data sets. If the sulfate and nitrate at the Nebraska site were in the form of acid compounds when they entered the rain, then the acidity was neutralized by bases before the rain reached the laboratory. However, irrespective of the details of the chemical processes, the net effect is that at the Nebraska site, ammonium (NH_4^+) and calcium (Ca^{2+}) are the dominant positive ions counterbalancing the dominant negative ions, sulfate (SO_4^{2-}) and nitrate (NO_3^-). For the Florida coastal site, sodium (Na^+) and chloride (Cl^-) are dominant ions derived from airborne sea salt particles that have been incorporated into the raindrops. Sulfate and nitrate are lower at the Florida site than at the other three sites. Finally, the ion concentrations for drinking water (the last column in Table 2) for one city in Illinois are much higher than for precipitation except for nitrate, ammonium, and hydrogen ion.

In summary, the data in Table 2 demonstrate that:

(a) Sulfate, or sulfate plus nitrate, is not always directly related to acidity (and inversely to pH) in precipitation samples;
(b) All the major ions must be measured to understand the magnitude (or time trends) of acidity of a sample or a site; and

TABLE 2
Median Ion Concentrations for Drinking Water and for Wet Deposition at Four NADP/NTN Sites in Four States for 1984

	New Hampshire[a]	Ohio[b]	Nebraska[c]	Florida[d]	Drinking Water[e]
Number of Samples	35	37	41	46	5
Ions			(μeq/L)		
SO_4^{2-} (Sulfate)	37	69	43	21	650
NO_3^- (Nitrate)	23	32	28	10	3
Cl^- (Chloride)	4	7	3	27	234
HCO_3^- (Bicarbonate)	0.1[f]	0.1[f]	3[f]	0.7[f]	2044[f]
Sum (rounded off)	**64**	**108**	**77**	**59**	**2931**
NH_4^+ (Ammonium)	7	16	36	3	28
Ca^{2+} (Calcium)	4	9	22	9	624
Mg^{2+} (Magnesium)	2	4	5	6	905
K^+ (Potassium)	0.4	0.6	1	1	61
Na^+ (Sodium)	4	3	4	24	1444
H^+ (Hydrogen)[g]	41	71	2	7	$<.1$
Sum (rounded off)	**58**	**104**	**70**	**50**	**3062**
Median pH	4.39	4.15	5.80	5.14	About 8.6
Weighted pH[h]	4.41	4.16	5.07	5.05	—
Calculated pH	4.33	4.12	5.17	4.93	—

[a] A site in central New Hampshire.

[b] A site in southeastern Ohio.

[c] A site in east-central Nebraska.

[d] A site in the southern tip of Florida.

[e] Levels in treated municipal well water (tap water) for a city of 100,000 in Illinois.

[f] Calculated with equation: $HCO_3^- = 5.13$ divided by H^+ for $Pco_2 = 335 \times 10^{-6}$ atm.

[g] Calculated from median pH.

[h] Sample volume weighted hydrogen ion concentration, expressed as pH. Some western sites have differences in weighted and median pH values of as much as 1 unit.

(c) Precipitation samples are relatively clean or pure as compared to treated well water used for drinking.

SPATIAL PATTERNS. The spatial distribution of five of the chemical parameters measured in the NADP/NTN weekly precipitation chemistry samples are shown in Figures 2–6. The "+" symbol indicates the location of the 180 sampling sites included in the analysis. A relatively long time period (1990–1993) was chosen for analysis in order to have sufficient data to produce stable patterns, but not so long that emissions of the major sources of the chemical parameters would have changed substantially. Samples for weeks with total precipitation less than two hundredths of an inch of equivalent liquid precipitation were not included. Every sample was required to pass rigorous quality assurance standards which included checks to assure that the proper sampling protocol was followed and that visible matter in the samples was not excessive and did not produce abnormally high concentrations of the chemical species measured. The nine sites at elevations greater than 3,000 meters were not included due to concerns about their representativeness. Completeness of data for each of the sites was judged in two ways. First, sites that started after January 1, 1990, or ceased operating before December 31, 1993, were excluded from the analysis if they operated

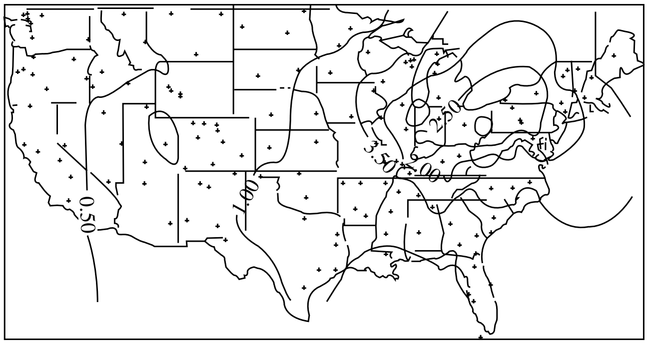


FIGURE 2 Median concentration (mg/L) of sulfate in precipitation for 180 NADP/NTN sites for the period 1990–1993.

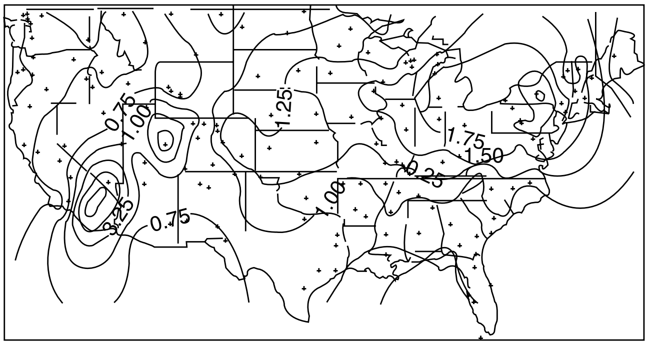


FIGURE 3 Median concentration (mg/L) of nitrate in precipitation for 180 NADP/NTN sites for the period 1990–1993.

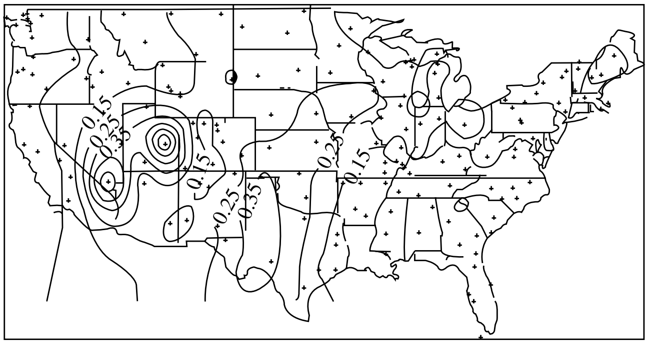


FIGURE 4 Median concentration (mg/L) of calcium in precipitation for 180 NADP/NTN sites for the period 1990–1993.

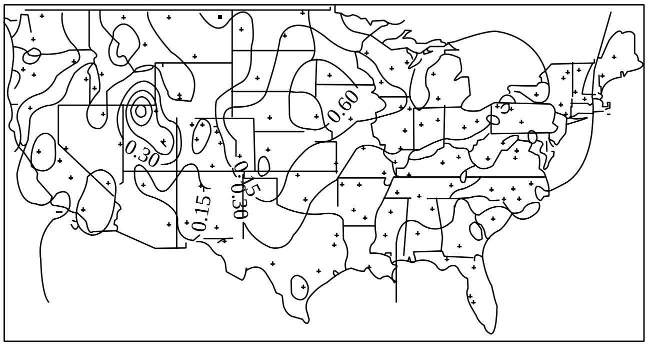


FIGURE 5 Median concentration (mg/L) of ammonium in precipitation for 180 NADP/NTN sites for the period 1990–1993.

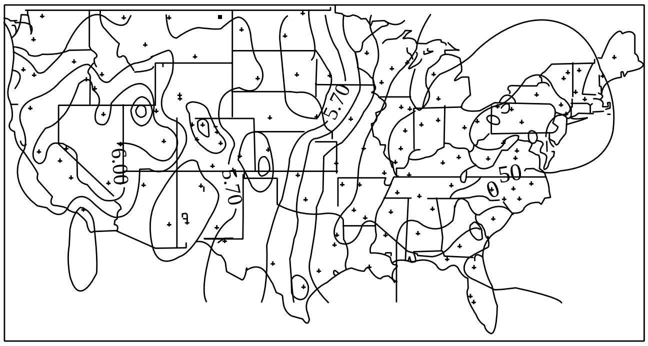


FIGURE 6 Median pH in precipitation for 180 NADP/NTN sites for the period 1990–1993.

less than 80 percent of the four-year interval (98 percent or 176 of the 180 selected sites operated for more than 95 percent of the interval). Second, sites with a low number of valid weekly samples were excluded. That is, if at least two hundredths of an inch of liquid precipitation would have fallen every week and if valid chemical measurements were obtained for each weekly sample, then 205 samples would have been available. In fact for the semi-arid western states, a large fraction of the weekly samples are completely dry. A decision was made to include in the analysis only those western sites with at least 100 valid samples and those eastern sites with at least 129 valid samples. For the 180 sites meeting all of the selection criteria, the median number of valid samples was 152.

Shown in Figures 2–6 are lines (isopleths) of median ion concentration or median pH. The isopleths are computer generated and include some automatic smoothing, but are very similar to hand-drawn contours. The concentrations are for the ion, i.e., for sulfate it is milligrams per liter of sulfate, not sulfur.

Sulfate concentrations in precipitation, shown in Figure 2, are highest in the Northeast with values exceeding 2.5 mg/L at sites in eastern Illinois, Indiana, Ohio, and western Pennsylvania. This is consistent with known high emissions to the atmosphere of sulfur from coal burning electrical power plants in this region. The sulfate levels decrease to the west of this area, with West Coast values being less than 0.5 mg/L.

The major anthropogenic sources for the nitrogen precursors which become nitrate in precipitation are high temperature combustion sources, which includes power plants and automobiles. The known locations for these sources are consistent with the observed nitrate concentrations in precipitation shown in Figure 3. Nitrate concentrations are high in the Northeast, from Illinois to New York. The high values of nitrate in southern California are reasonable considering the high density of people and automobiles in this area. The lack of high sulfate values in this California area reflects the lack of intensive coal combustion in the area.

Figure 4 shows the concentrations of calcium in precipitation. With respect to sources of the calcium, Gillette *et al.* (1989) have indicated that dust from soils and dust from traffic on unpaved roads are the major sources of calcium in the atmosphere. Dust devils in the southwestern states, wind erosion of agricultural fields, and crop production activities in areas with intensive agriculture are the major dust generation processes for soils. The elevated levels of calcium shown in Figure 4 in the Midwestern, plains, and western states are due to a combination of the location of the mentioned dust generating sources as well as the generally more arid conditions in these areas. The higher amounts and frequency of precipitation in the East, Southeast, and Northwest effectively shut off the dust sources by both keeping soil and road material damp and by causing dense vegetation to protect soil surfaces from erosion.

The ammonium concentration pattern shown in Figure 5 is similar to that for calcium but for different reasons. The high values in the Midwestern, plains, and western states are likely due to the emissions of ammonia from livestock feedlots. The 0.45 mg/L isopleth in the central United States encloses the region of large cattle feedlots. Emissions related to agricultural fertilizers may also be important. The site in northern Utah near Logan is in a small basin surrounded by mountains. This terrain and the relatively high density of livestock in the basin likely explains the very high ammonium levels there.

The median pH is shown in Figure 6. As was demonstrated with the data in Table 2, the pH can be understood only by considering all the major acidic and basic constituents. For example notice that a 4.2 pH isopleth encloses sites in Pennsylvania and New York while the maximum sulfate isopleth in Figure 2, with a value of 2.50 mg/L, is shifted further west. The other major acidic anion, nitrate, has its maximum further to the east than sulfate and the two basic cations shown in Figures 4 and 5 have decreasing concentrations from Ohio eastward. Therefore the location of the pH maximum isopleth becomes reasonable when all the major ions are considered.

The pH values in Figure 6 increase westward of Ohio with maximum values of about 6 for sites from southeastern South Dakota to the panhandle of Texas. Continuing westward, the pH values decrease to values less than 5.4 for Rocky Mountain sites in Wyoming, Colorado, and New Mexico, then increase again to values of 6 or higher for many sites in Utah and Nevada, and finally decrease again to values less than 5.4 for sites in the extreme northwestern United States.

The pH values shown in Figure 6 result from measurements made shortly after the samples arrive at the Central Analytical Laboratory in Illinois. During the interval of time between when samples are collected at the field site and until the pH is measured in Illinois, some acid neutralization occurs. In fact the pH determined at the local field site laboratory would be a couple hundredths of a pH unit lower (more acid) for samples with pH values in the 4s and several tenths lower for samples with pH values in the 5s or 6s. Therefore, a map showing the median of field pH values will be somewhat different than Figure 6. The use of other pH averaging procedures (e.g. weighted averages) can also produce substantial differences (for some locations) from values of the median pH shown in Figure 6.

TEMPORAL PATTERNS. In addition to determining the spatial patterns of chemicals in rain and snow, it is important to determine the temporal patterns. Research in the 1970s showed that the sulfate and hydrogen ion concentrations in precipitation in the northeastern United States were higher during the warm season than the cold season. A study by Bowersox and Stensland (1985) showed that this seasonal time dependence was more general, applying to other regions and other ions. For this 1985 study, NADP/NTN data for 1978–1983 were grouped by site into warm-period months (May–September) and cold-period months (November–March). Rigorous data selection criteria were applied, including a stipulation that at least ten valid concentration values be available for each site for each period. Median concentrations were calculated by site for each period. Then the ratios of the warm- to cold-period concentrations were calculated for each site. The means of the resulting site ratios for four regions are presented in Table 3. Sodium and chloride have ratio values *less* than 1.0 for three of the regions, probably because increased storm activity during the cold period injects greater quantities of sea salt into the air in the cold months than is injected in the warm months. Detailed explanations for ratio values being *greater than* or equal to 1.00 for the other ions, in all regions, have not been established. The interannual variation of photochemical conversion rates is certainly an important factor for some ions such as sulfate and hydrogen, while ground cover and soil moisture content are likely to be important factors for the dust-related ions. Meteorological features, such as stagnation conditions and typical wind direction, may also be important factors to explain the seasonality effect shown in Table 3.

For making pollution abatement decisions, the time trends of acid rain, on the scale of years, are important. There has been considerable debate in the literature with respect to the long-term time trends of chemicals in precipitation. Precipitation chemistry sampling locations, equipment, and procedures have varied in the last 30–40 years, producing inconsistent data sets that in turn have led to flawed interpretations and have resulted in controversy. A report from the National Research Council (1986) critically reviews much of the relevant literature. There is quite general agreement that over the last 100 years, the large increase of sulfur emissions to the atmosphere over the United States has increased the levels of sulfate in precipitation. The problem is in trying to quantify the changes for specific regions with enough precision to provide a database sufficient for policy decisions.

The reported changes in precipitation acidity since the mid-1950s are probably the result of three phenomena: the acidity differences related to changes in dust emissions from wind erosion of soils and traffic on unpaved roads; the acidity differences due to changes in sampling techniques; and the acidity differences due to changes in acidic emissions from combustion pollution. Since the combined effect of the first two components is large, the increases in acidity due to changes in sulfur and nitrogen emissions in the

TABLE 3

Seasonality of Ion Concentrations in Precipitation as Shown By Average Ratio Values (Warm Period/Cold Period Precipitation Concentrations) for Four Regions of the United States

		**********Mean ± 2 Std. Dev. of Period Ratios**********				
Region[a]	N[b]	SO_4^{2-}	NO_3^-	NH_4^+	Ca^{2+}	H^+
MW	20	1.35 ± 0.64	1.00 ± 0.47	1.67 ± 1.45	1.63 ± 1.02	1.03 ± 0.88
SE	15	1.52 ± 0.60	1.73 ± 0.92	1.87 ± 0.92	1.57 ± 0.62	1.52 ± 0.87
NE	23	2.19 ± 0.80	1.36 ± 0.88	2.45 ± 1.48	1.44 ± 0.72	1.89 ± 0.64
RM	16	2.15 ± 1.11	2.63 ± 2.87	2.65 ± 1.54	2.39 ± 1.30	2.58 ± 2.37

		**********Mean ± 2 Std. Dev. of Period Ratios**********			
Region[a]	N	Mg^{2+}	K^+	Na^+	Cl^-
MW	20	1.40 ± 0.67	1.55 ± 0.68	0.79 ± 0.58	0.92 ± 1.21
SE	15	1.23 ± 0.69	1.53 ± 0.54	0.95 ± 0.73	0.87 ± 0.51
NE	23	1.17 ± 0.65	1.43 ± 0.67	0.67 ± 0.53	0.64 ± 0.36
RM	16	1.82 ± 0.90	2.67 ± 1.58	1.30 ± 0.84	1.51 ± 1.05

[a] MW is Midwest, SE is Southeast, NE is Northeast, and RM is Rocky Mountain.

[b] N is the number of sites in the region used in the analysis. States bordering the Pacific Ocean and states in the Great Plains were not included in this analysis.

Midwest and Northeast (or other regions) cannot be precisely quantified on the basis of the historical precipitation chemistry data.

The longest continuous precipitation chemistry record is for the Hubbard Brook site in New Hampshire, where the record began in 1963 (Likens *et al.*, 1984). The sampling method was to continuously expose a funnel and bottle, i.e. bulk sampling. From 1964 to 1982 sulfate decreased quite regularly, which seems to be consistent with the trend of combustion sulfur emissions for this area of the country. Values for pH did not show a significant change. The National Research Council (1986) tabulated the published trends for the Hubbard Brook data set to indicate that the results are sometimes sensitive to the specific type of analysis. For example, one publication indicated that nitrate increased from 1964 to 1971, and then remained steady through 1980. A second publication included the nitrate data for 1963 to 1983, and found no significant *overall* trend. A third publication, including data for 1964 to 1979, found a significant overall increase in nitrate. Bulk data should not generally be compared with wet-only data, however, comparisons have shown that the dry deposition component is relatively small for the Hubbard Brook site and thus it appears valid to suggest that the bulk trends are probably representative of wet-only trends.

The NADP/NTN weekly wet deposition data provides the best data set for trend analysis because of the comprehensive quality assurance program for the network and because of the good spatial coverage across the 48 states. Lynch *et al.* (1995) reported the most recent comprehensive summary of temporal trends in precipitation chemistry in the United States using data from 58 NADP/NTN sites from 1980 through 1992. Results showed widespread declines in sulfate concentrations accompanied by significant decreases in all of the base cations, most noticeably calcium and magnesium. As a result of the decreases in both acids and bases, only 17 of the 42 sites with significantly decreasing sulfate trends had concurrent significant decreasing trends in hydrogen ion (acidity). The decline in precipitation sulfate during this period is consistent with the known declines in sulfur dioxide emissions from electric power plants. The decline in base cations does not yet have a definitive explanation since the strengths of the various emission sources are not well known.

Phase I of Title IV of the 1990 Clean Air Act Amendments required specific reductions in sulfur dioxide emissions on or before 1 January 1995 at selected electric utility plants, the majority of which are located in states east of the Mississippi River. As a result of this legislation, large reductions in sulfur dioxide emissions were likely to have occurred in 1995, which should have affected sulfate and hydrogen ion concentrations in precipitation in this region. Lynch *et al.* (1996) compared the 1995 concentrations to those expected from the 1983–1994 trends and indeed found that sulfate and hydrogen ion decreased much more than expected due to just the 1983–1994 trends. Thus they concluded that acid rain in the eastern United States had decreased as a result of the Phase I emission reductions. Additional major emission reductions in sulfur dioxide are required in Phase II by the year 2000 so it will be important to look for corresponding additional reductions in acid rain.

REMOTE SITE PH DATA

Acid precipitation is also being measured at remote sites. pH data for more than 1700 daily or three-day samples collected in the Hawaiian Islands were reported by Miller and Yoshinaga (1981). The observed pH for the Hawaiian samples ranged from about 3.6 to 6.0. The average pH for about 800 daily samples collected at three sites in the Hilo, Hawaii area was 4.7. The pH decreased with altitude, with an average pH of 4.3 for 92 samples collected at a site at an altitude of 3400 meters. To check for the possibility of local volcanic emissions being the dominant source, samples were collected on the island of Kauai, which has no volcanic emissions and is 500 km north of the big island of Hawaii where all the other sampling took place. For the Kauai site, the average pH was 4.79, which is similar to the pH for the Big Island.

Galloway *et al.* (1982) have measured the chemistry of precipitation for several sites remote from manmade pollution. An important feature documented by these investigators is that the pH of samples from these remote sites increased significantly between the time of field collection and the time of sample receipt at the laboratory in Virginia. However, the pH of the samples remained stable when a chemical was added to stop bacterial activity in the samples. It was established that organic acids (from natural sources) are an important acid component in samples from the remote sites and without the pH stabilization procedure, the organic acids were lost during shipment and only the strong mineral acids and the elevated pH values were detected. For three remote sites in Australia, in Venezuela, and on Amsterdam Island, the weighted average pH values for stabilized samples were 4.8, 4.8, and 4.9 respectively.

The detection of acid rain at locations remote from manmade pollution has led researchers to suggest that departures of precipitation pH below 5.0, instead of the commonly used level of 5.6 or 5.7, would better indicate the local and regional manmade modulations to the natural global background. That is, perhaps we should define acid rain to be samples where pH is less than 5.0. However, since pH is in fact the balance of a group of ions, it is scientifically better to use the levels of these ions, and not just pH, to characterize samples as acid rain.

RECOMMENDATIONS FOR THE FUTURE

This discussion has focused on results of wet deposition measurements. However, both wet and dry deposition must be measured so that eventually a mass balance can be evaluated to account, year by year, for the pollutants put into the air. Therefore:

1) Wet deposition measurements across the United States should be continued indefinitely, just as we continue to monitor emissions, air quality, and weather variables such as precipitation amount and type, and
2) Dry deposition measurement techniques need continued development and evaluation, and a long-term monitoring network must become available to provide data for calculating total deposition (wet and dry).

REFERENCES

Bowersox, V.C. and G.J. Stensland (1985), Seasonal patterns in the chemistry of precipitation in the United States. *In* Proceedings of the 78th Annual Meeting, Air Pollution Control Association, Pittsburgh, PA, Paper No. 85–6.A.2.

Cogbill, C.V. and O.E. Likens (1974), Acid precipitation in the northeastern United States. *Wat. Resources Res.*, **10**, 1133–1137.

Evans, L.S., G.R. Hendrey, G.J. Stensland, D.W. Johnson, and A.J. Francis (1981), Acidic precipitation: considerations for an air quality standard. *Water, Air, and Soil Pollution,* **16**, 469–509.

Galloway, J.N., G.E. Likens, W.C. Keene, and J.M. Miller (1982), The composition of precipitation in remote areas of the world. *J. Geophys. Res.*, **87**, 8771–8786.

Gillette, D.A., G.J. Stensland, A.L. Williams, P.C. Sinclair, and T.Z. Marquardt (1992), Emissions of alkaline elements calcium, magnesium, potassium, and sodium from open sources in the contiguous United States. *Global Geochemical Cycles,* **6**, 437–457.

Gorham, E. (1958), Atmospheric pollution by hydrochloric acid. *Quart. J. Royal Meterol. Soc.,* **84**, 274–276.

Granat, L. (1972), On the relationship between pH and the chemical composition in atmospheric precipitation. *Tellus,* **24**, 550–560.

Junge, C.E. (1963), *Air Chemistry and Radioactivity.* Academic Press, New York, 382 pp.

Likens, G.E., F.H. Borman, R.S. Pierce, J.S. Eaton, and R.E. Munn (1984), Long-term trends in precipitation chemistry at Hubbard Brook, New Hampshire. *Atmos. Environ.,* **18**, 2641–2647.

Lynch, J.A., V.C. Bowersox, and J.W. Grimm (1996), Trends in precipitation chemistry in the United States, 1983–94: An analysis of the effects in 1995 of phase I of the Clean Air Act Amendments of 1990, Title IV. Open-File Report 96-0346 (http://h20.usgs.gov/public/pubs/acidrain), U.S. Geological Survey, Reston, VA.

Lynch, J.A., J.W. Grimm, and V.C. Bowersox (1995), Trends in precipitation chemistry in the United States: A national perspective, 1980–1992. *Atmos. Environ.,* **29**, 1231–1246.

Miller, J.M. and A.M. Yoshinaga (1981), The pH of Hawaiian precipitation—A preliminary report. *Geophys. Res. Letters,* **7**, 779–782.

National Acid Precipitation Assessment Program (1990), *Acidic Deposition: State of Science and Technology,* Volumes I–IV, Supt. of Documents, Government Printing Office, Washington, DC.

National Acid Precipitation Assessment Program (1991), *The U.S. National Acid Precipitation Assessment Program 1990 Integrated Assessment Report,* NAPAP Office, Washington, DC, 520 pp.

National Research Council (1986), Acid deposition—long-term trends. Wash. DC, National Academy Press, 506 pp.

Reuss, J.O. (1975), *Chemical/Biological Relationships Relevant to Ecological Effects of Acid Rainfall.* U.S. EPA Report EPA-660/3-75-032, 46 pp.

Seinfeld, J.H. (1986), *Atmospheric Chemistry and Physics of Air Pollution.* John Wiley & Sons, New York, 738 pp.

Smith, R.A. (1872), *Air and Rain: The Beginnings of a Chemical Climatology.* Longmans, Green, and Co., London, England.

Stensland, G.J. and R.G. Semonin (1982), Another interpretation of the pH trend in the United States. *Bull. Amer. Meteorol. Soc.,* **63**, 1277–1284.

OTHER GENERAL REFERENCES

Graedel, T.E. and P.J. Crutzen (1993), *Atmospheric Change—An Earth System Perspective*. W.H. Freeman and Company, New York, 446 pp.

Graedel, T.E. and P.J. Crutzen (1995), *Atmosphere, Climate, and Change.* W.H. Freeman and Company, New York, 196 pp.

Hidy, G.M. (1994), *Atmospheric Sulfur and Nitrogen Oxides—Eastern North American Source-Receptor Relationships.* Academic Press, New York, 447 pp.

Mohnen, V.A. (1988), The challenge of acid rain. *Scientific American,* **259**(2), 30–38.

National Atmospheric Deposition Program Data Reports. Available from the NADP Program Office, Illinois State Water Survey, 2204 Griffith Drive, Champaign, IL 61820 (http://nadp.sws.uiuc.edu).

GARY J. STENSLAND
State Water Survey Division
Illinois Department of Natural Resources

ACOUSTICS OF THE ENVIRONMENT: see NOISE

AEROSOLS: see also PARTICULATE EMISSIONS; PARTICULATE REMOVAL

AEROSOLS

An aerosol is a system of tiny particles suspended in a gas. Aerosols or particulate matter refer to any substance, except pure water, that exists as a liquid or solid in the atmosphere under normal conditions and is of microscopic or submicroscopic size but larger than molecular dimensions. There are two fundamentally different mechanisms of aerosol formation:

- nucleation from vapor molecules (photochemistry, combustion, etc.)
- comminution of solid or liquid matter (grinding, erosion, sea spray, etc.)

Formation by molecular nucleation produces particles of diameter smaller than 0.1 μm. Particles formed by mechanical means tend to be much larger, diameters exceeding 10 μm or so, and tend to settle quickly out of the atmosphere. The very small particles formed by nucleation, due to their large number, tend to coagulate rapidly to form larger particles. Surface tension practically limits the smallest size of particles that can be formed by mechanical means to about 1 μm.

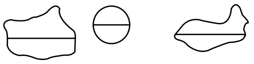
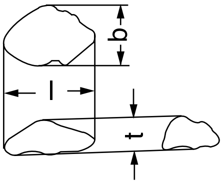
PARTICLE SIZE DISTRIBUTION

Size is the most important single characterization of an aerosol particle. For a spherical particle, diameter is the usual reported dimension. When a particle is not spherical, the size can be reported either in terms of a length scale characteristic of its silhouette or of a hypothetical sphere with equivalent dynamic properties, such as settling velocity in air.

Table 1 summarizes the physical interpretation for a variety of characteristic diameters. The Feret and Martin diameters are typical geometric diameters obtained from particle silhouettes under a microscope.

TABLE 1

Measures of particle size

		Definition of characteristic diameters	Physical meaning and corresponding measuring method
	geometric size	$(b+l)/2, (b+l+t)/3, (blt)^{1/3}, 3/(1/l+1/b+1/t), \sqrt{lb}, \sqrt{\{(2lb+2bt+2lt/6)\}}$	
		Feret diam. 	unidirectional diameter: diameter of particles at random along a given fixed line, no meaning for a single particle.
		Martin diam. 	unidirectional diameter: diameter of particles as the length of a chord dividing the particle into two equal areas.
	equivalent diam.	equivalent projection area diam. (Heywood diam.)	diameter of the circle having the same area as projection area of particle, corresponding to diam. obtained by light extinction.
		equivalent surface area diam. (specific surface diam.) $(s/\pi)^{1/2}$	diameter of the sphere having the same surface as that of a particle, corresponding to diam. obtained by absorption or permeability method.
		equivalent volume diam. $(6v/\pi)^{1/3}$	diameter of the sphere having the same volume as that of a particle, corresponding to diam. obtained by Coulter Counter.
breadth: *b* length: *l*		Stokes diam.	diameter of the sphere having the same gravitational setting velocity as that of a particle, $D_{st} = [18\ \mu v_t/g(\rho_p - \rho_f)C_c]^{1/2}$, obtained by sedimentation and impactor.

(*continued*)

TABLE 1 (*continued*)
Measures of particle size

Definition of characteristic diameters		Physical meaning and corresponding measuring method
thickness: t volume: v	aerodynamic diam.	diameter of the sphere having unit specific gravity and having the same gravitational setting velocity as that of a particle, $D_{ae} = [18\ \mu u_t/gC_c]^{1/2}$, obtained by the same methods as the above.
surface area: s	electrical mobility equivalent diam.	diameter of the sphere having the same electrical mobility as that of a particle, $D_e = n_p e C_c/3\pi\mu B_e$, obtained by electrical mobility analyzer.
	equivalent diffusion diam.	diameter of the sphere having the same penetration as that of a particle obtained by diffusion battery.
	equivalent light scattering diam.	diameter of the sphere having the same intensity of light scattering as that of a standard particle such as a PSL particle, obtained by light scattering method.

When particles, at total number concentration N, are measured based on a certain characteristic diameter as shown in Table 1 and the number of particles, dn, having diameters between D_p and $D_p + \mathrm{d}D_p$ are counted, the normalized particle size distribution $f(D_p)$ is defined as follows:

$$f(D_p) = \frac{1}{N}\frac{\mathrm{d}n}{\mathrm{d}D_p}, \tag{1}$$

where

$$\int_0^\infty f(D_p)\,\mathrm{d}D_p = 1.$$

The discrete analog which gives a size distribution histogram is

$$f(D_p) = \frac{1}{N}\frac{\Delta n}{\Delta D_p} \tag{2}$$

where Δn is the particle number concentration between $D_p - \Delta D_p/2$ and $D_p + \Delta D_p/2$.

The cumulative number concentration of particles up to any diameter D_p is given as

$$F(D_p) = \int_0^{D_p} f(D'_p)\,\mathrm{d}D'_p = 1 - \int_{D_p}^\infty f(D'_p)\,\mathrm{d}D'_p$$

$$\frac{\mathrm{d}F}{\mathrm{d}D_p} = f(D_p). \tag{3}$$

The size distribution and the cumulative distribution as defined above are based on the number concentration of particles. If total mass M and fractional mass dm are used instead of N and dn, respectively, the size distributions can then be defined on a mass basis.

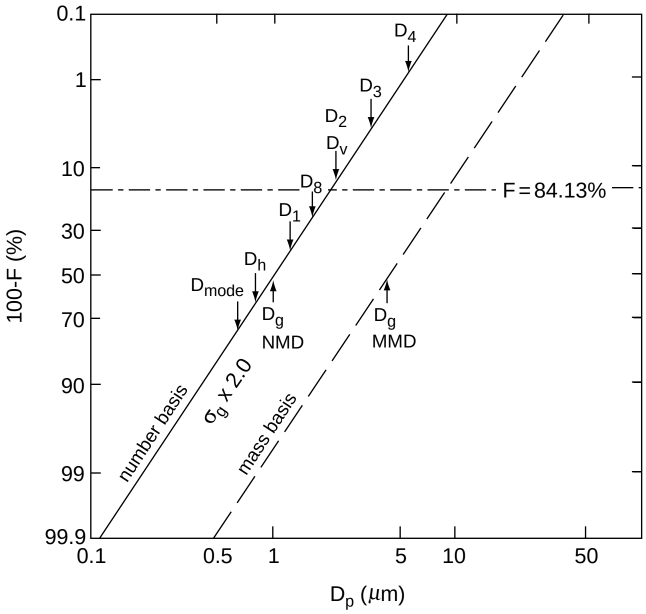


FIGURE 1 Log-normal size distribution for particles with geometric mean diameter of 1 μm and geometric standard deviation of 2.0. The different average particle diameters for this distribution are defined in Table 2.

Many particle size distributions are well described by the normal or the log-normal distributions. The normal, or Gaussian, distribution function is defined as,

$$f(D_p) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(D_p - \overline{D}_p\right)^2}{2\sigma^2}\right) \tag{4}$$

where $\overline{D}_p$ and σ are, respectively, the mean and standard deviation of the distribution. The mean diameter $\overline{D}_p$ is defined by

$$\overline{D}_p = \int_{-\infty}^{\infty} D_p f\left(D_p\right)\mathrm{d}D_p \tag{5}$$

and the standard deviation, indicating the dispersion of the distribution, is given by

$$\sigma^2 = \int_{-\infty}^{\infty} \left(D_p - \overline{D}_p\right)^2 f\left(D_p\right)\mathrm{d}D_p. \tag{6}$$

In the practical measurement of particle sizes, $\overline{D}_p$ and σ are determined by

$$\begin{gathered}\overline{D}_p = \frac{\sum n_i D_{pi}}{N} \\ \sigma = \left(\frac{n_i\left(D_{pi} - \overline{D}_p\right)^2}{N}\right)^{1/2}\end{gathered} \tag{7}$$

where n_i is the number of particles with diameter D_{pi} and N is the total particle number measured.

TABLE 2

Names and defining equations for various average diameters

	Defining equations	
	General case	In the case of log-normal distribution
number mean diam. D_1	$\frac{\Sigma\Delta nD_p}{N}$	$\ln D_1 = A + 0.5C = B - 2.5C$
length mean diam. D_2	$\frac{\Sigma\Delta nD_p^2}{\Sigma\Delta nD_p}$	$\ln D_2 = A + 1.5C = B - 1.5C$
surface mean, Sauter or mean volume-surface diam. D_3	$\frac{\Sigma\Delta nD_p^3}{\Sigma\Delta nD_p^2} = \frac{\Sigma\Delta sD_p}{S}$	$\ln D_3 = A + 2.5C = B - 0.5C$
volume or mass mean diam. D_4	$\frac{\Sigma\Delta nD_p^4}{\Sigma\Delta nD_p^3} = \frac{\Sigma\Delta mD_p}{M}$	$\ln D_4 = A + 3.5C = B + 0.5C$
diam. of average surface D_s	$\sqrt{\frac{\Sigma\Delta nD_p^2}{N}}$	$\ln D_s = A + 1.0C = B - 2.0C$
diam. of average volume or mass D_v	$\sqrt[3]{\frac{\Sigma\Delta nD_p^3}{N}}$	$\ln D_v = A + 1.5C = B - 1.5C$
harmonic mean diam. D_h	$\frac{N}{\Sigma(\Delta/D_p)}$	$\ln D_h = A - 0.5C = B - 3.5C$
number median diam. or geometric mean diam. NMD	$\exp\left[\frac{\Sigma\Delta n \ln D_p}{N}\right]$	NMD
volume or mass median diam. MMD	$\exp\left[\frac{\Sigma\Delta nD_p^3 \ln D_p}{\Sigma\Delta nD_p^3}\right]$ $= \exp\left[\frac{\Sigma\Delta m \ln D_p}{M}\right]$	$\ln \mathrm{MMD} = A + 3C$

$A = \ln \mathrm{NMD}$, $B = \ln \mathrm{MMD}$, $C = (\ln \sigma_g)^2$

N(total number) $= \Sigma\Delta n$, S(total surface) $= \Sigma\Delta s$, M(total mass) $= \Sigma\Delta m$

The log-normal distribution is particularly useful for representing aerosols because it does not allow negative particle sizes. The log-normal distribution function is obtained by substituting $\ln D_p$ and $\ln \sigma_g$ for D_p and σ in Eq. (4),

$$f\left(\ln D_p\right) = \frac{1}{\sqrt{2\pi}\ln \sigma_g}\exp\left(-\frac{\left(\ln D_p - \ln \overline{D}_p\right)^2}{2\ln^2 \sigma_g}\right). \quad (8)$$

The log-normal distribution has the following cumulative distribution,

$$F = \frac{1}{\sqrt{2\pi}\ln \sigma_g}\int_0^{D_p} \exp\left(-\frac{\left(\ln D_p - \ln D_g\right)^2}{2\ln^2 \sigma_g}\right) \mathrm{d}\left(\ln D_p\right). \quad (9)$$

The geometric mean diameter D_g, and the geometric standard deviation σ_g, are determined from particle count data by

$$\begin{aligned} \ln D_g &= \left(\sum n_i \ln D_{pi}\right)/N \\ \ln \sigma_g &= \left[\sum n_i \left(\ln D_{pi} - \ln D_g\right)^2 /N\right]^{1/2}. \end{aligned} \quad (10)$$

Figure 1 shows the log-normal size distribution for particles having $D_g = 1$ μm and $\sigma_g = 2.0$ on a log-probability graph, on which a log-normal size distribution is a straight line. The particle size at the 50 percent point of the cumulative axis is the geometric mean diameter D_g or number median diameter, NMD. The geometric standard deviation is obtained from two points as follows:

$$\sigma_g = \frac{D_p \text{ at } F = 84.13\%}{D_p \text{ at } F = 50\%} = \frac{D_p \text{ at } F = 50\%}{D_{po} \text{ at } F = 15.7\%}.$$

The rapid graphical determination of the geometric mean diameter D_g as well as the standard deviation σ_g is a major advantage of the log-normal distribution. It should be emphasized that the size distribution on a number basis shown by the solid line in Figure 1 differs significantly from that on a mass basis, shown by the dashed line in the same figure. The conversion from number median diameter (NMD) to mass median diameter (MMD) for a log-normal distribution is given by

$$\ln(\mathrm{MMD}) = \ln(\mathrm{NMD}) + 3(\ln \sigma_g)^2. \quad (11)$$

If many particles having similar shape are measured on the basis of one of the characteristic diameters defined in Table 1, a variety of average particle diameters can be calculated as shown in Table 2. The comparison among these diameters is shown in Figure 1 for a log-normal size distribution. Each average diameter can be easily calculated from σ_g and NMD (or MMD).

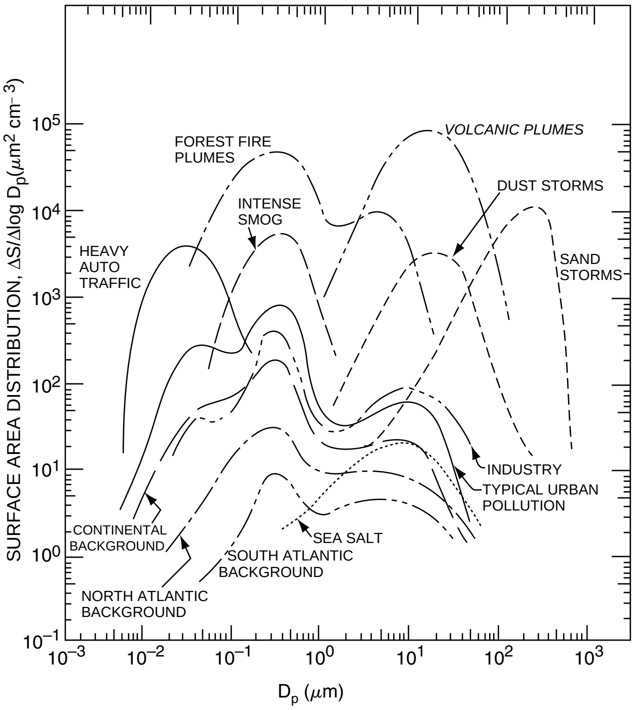


FIGURE 2 Surface area distributions of natural and anthropogenic aerosols.

Figure 2 indicates approximately the major sources of atmospheric aerosols and their surface area distributions. There tends to be a minimum in the size distribution of atmospheric particles around 1 μm, separating on one hand the coarse particles generated by storms, oceans and volcanoes and on the other hand the fine particles generated by fires, combustion and atmospheric chemistry. The comminution processes generate particles in the range above 1 μm and molecular processes lead to submicron particles.

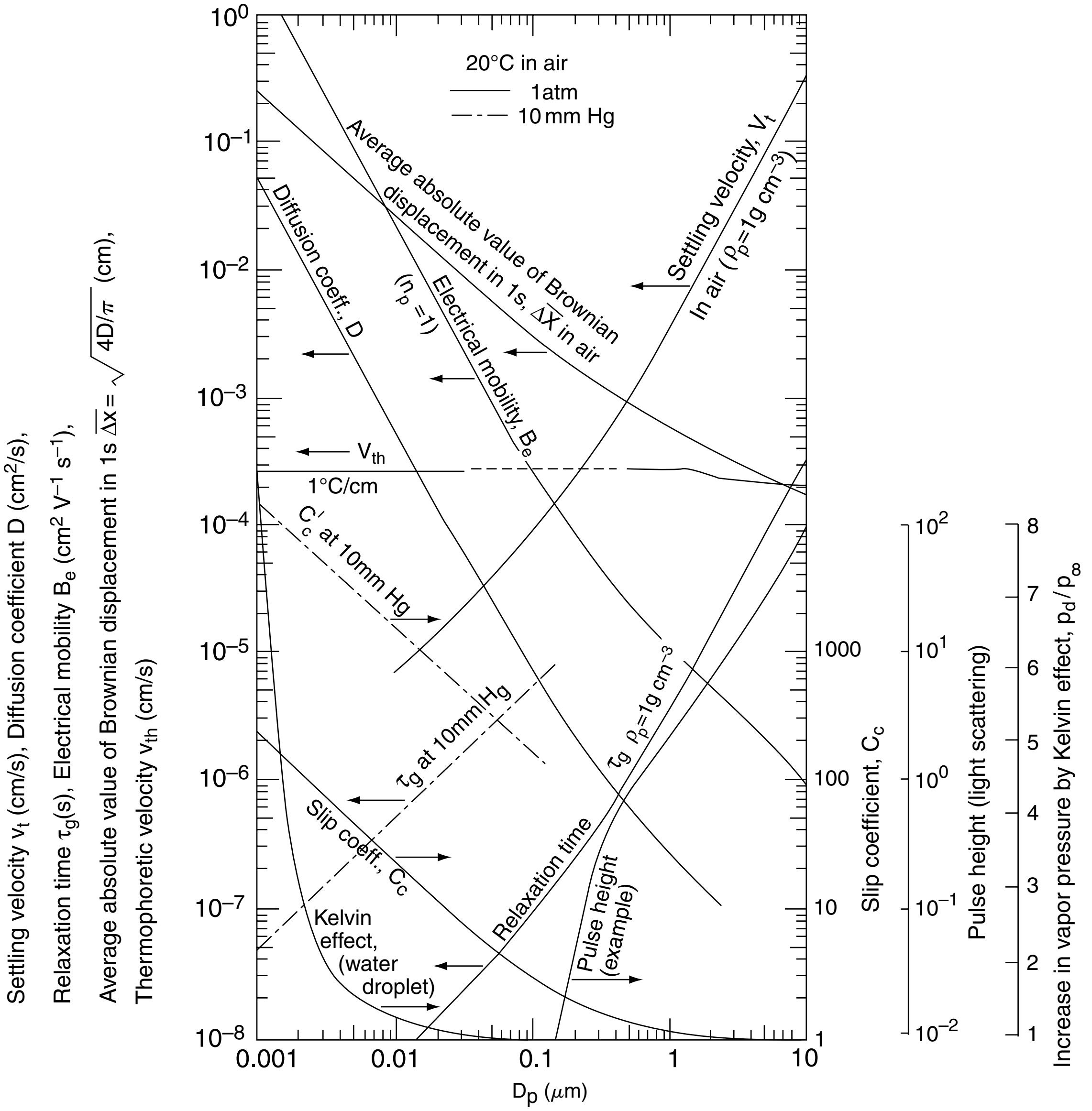
PARTICLE DYNAMICS AND PROPERTIES

Typical size-dependent dynamic properties of particles suspended in a gas are shown in Figure 3 together with defining equations (Seinfeld, 1986). The solid lines are those at atmospheric pressure and the one-point dashed lines are at low pressure. The curves appearing in the figure and the related particle properties are briefly explained below.

Motion of Large Particles

A single spherical particle of diameter D_p with a velocity u in air of density ρ_f experiences the following drag force,

$$F_d = C_D A_p (\rho_f u^2/2) \quad (12)$$



$$V_t = \frac{(\rho_p - \rho_f) g D_p^2 C_c}{18\mu} \quad (3.1) \qquad C_c = 1 + 2.514\frac{\lambda}{D_p} + 0.80\frac{\lambda}{D_p}\exp\left(-0.55\frac{D_p}{\lambda}\right) \quad (3.2)$$

$C_c' = 1 + (2/pD_p)\,[6.32 + 2.01 \exp(-0.1095 pD_p)]$ p in cm Hg, D_p in μm (3.3)

$$\overline{\Delta x} = \sqrt{\frac{4Dt}{\pi}} \quad (3.4) \qquad D = \frac{kTC_c}{3\pi\mu D_p} \quad (3.5)$$

$$\tau_g = \frac{\rho_p D_p^2 C_c}{18\mu} \quad (3.6) \qquad B_e = \frac{n_p\, e\, C_c}{3\pi\mu D_p} \quad (3.7)$$

$$P_d/P_\infty = \exp\left(\frac{4M\sigma}{RT\rho_l D_p}\right) \quad (3.8)$$

FIGURE 3 Fundamental mechanical and dynamic properties of aerosol particles suspended in a gas.

where A_p is the projected area of the particle on the flow ($= \pi D_p^2/4$), and C_D is the drag coefficient of the particle. The drag coefficient C_D depends on the Reynolds number,

$$\mathrm{Re} = u_r D_p \rho_f / \mu \tag{13}$$

where u_r is the relative velocity between the particle and air ($= |u - v|$, u = velocity of air flow, v = particle velocity), and μ is the viscosity of the fluid.

The motion of a particle having mass m_p is expressed by the equation of motion

$$m_p \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \sum \mathbf{F} \tag{14}$$

where **v** is the velocity of the particle and **F** is the force acting on the particle, such as gravity, drag force, or electrical force. Table 3 shows the available drag coefficients depending on

TABLE 3
Motion of a single spherical particle

	$\mathrm{Re}_p < 1$ (Stokes)	$1 < \mathrm{Re}_p < 10^4$	$10^4 < \mathrm{Re}_p$ (Newton)
drag coefficient, C_D	$24/\mathrm{Re}_p$	$\left(0.55 + \frac{4.8}{\sqrt{\mathrm{Re}_p}}\right)^2$	0.44
drag force, $R_f = C_D A_p \frac{\rho_f v^2}{2}$	$3\pi\mu D_p v$	$\left(0.55\sqrt{\frac{vD_p\rho_f}{\mu}} + 4.8\right)^2 \frac{\pi\mu D_p v}{8}$	$0.055\pi\rho_f (vD_p)^2$
gravitational settling equation of motion	$m_p \frac{\mathrm{d}v}{\mathrm{d}t} = m_p\left(1 - \frac{\rho_f}{\rho_p}\right)g - R_f$ or, $\frac{\mathrm{d}v}{\mathrm{d}t} = \left(1 - \frac{\rho_f}{\rho_p}\right)g - \frac{3\rho_f}{4\rho_p D_p} C_D v^2$		
terminal velocity, v_t ($\mathrm{d}v/\mathrm{d}t = 0$)	$\frac{D_p^2(\rho_p - \rho_f)g}{18\mu}$	$\left(\frac{\sqrt{A_1^2 + A_2} - A_1}{1.1}\right)^2$ $A_1 = 4.8\sqrt{\frac{\mu}{\rho_f D_p}}$ $A_2 = 2.54\sqrt{\frac{\rho_p - \rho_f}{\rho_f} g D_p}$	$\left(\frac{3D_p(\rho_p - \rho_f)g}{\rho_f}\right)^{1/2}$
unsteady motion time, t velocity, v	$t = \tau_g \ln\left(\frac{v_0 - v_t}{v - v_t}\right)$	$t = 24\tau_g \int_{\mathrm{Re}_{p0}}^{\mathrm{Re}_p} \frac{\mathrm{d}\,\mathrm{Re}_p}{C_{Dt}\,\mathrm{Re}_t^2 - C_D\,\mathrm{Re}_p^2}$	not simple because of $\mathrm{Re}_p << 10^4$ at initiation of motion
falling distance, S $S = \int_0^t v\mathrm{d}t$	$v_t t = \tau_g (v_t - v_0)\left[\exp\left(-\frac{t}{\tau_g}\right) - 1\right]$	$v_t \tau_g \int_0^t \mathrm{Re}_p\,\mathrm{d}t$ $t = t/\tau_g, \mathrm{Re}_p = \mathrm{Re}_p/\mathrm{Re}_{p0}$	

$\mathrm{Re}_p = \frac{vD_p\rho_f}{\mu}$, $\tau_g = \frac{\rho_p D_p^2}{18\mu}$, v_0: initial velocity, v_t: terminal velocity

Re_{p0}, Re_{pt}: Re_p at v_0 and at v_t respectively, C_{Dt}: drag coefficient at terminal velocity

Reynolds number and the basic equation expressing the particle motion in a gravity field.

The terminal settling velocity under gravity for small Reynolds number, v_t, decreases with a decrease in particle size, as expressed by Eq. (3.1) in Figure 3. The distortion at the small size range of the solid line of v_t is a result of the slip coefficient, C_c, which is size-dependent as shown in Eq. (3.2). The slip coefficient C_c increases with a decrease in particle size suspended in a gaseous medium. It also increases with a decrease in gas pressure p as shown in Figure 3. The terminal settling velocities at other Reynolds numbers are shown in Table 3.

τ_g in Figure 3 is the relaxation time and is given by Eq. (3.6). It characterizes the time required for a particle to change its velocity when the external forces change. When a particle is projected into a stationary fluid with a velocity v_o, it will travel a finite distance before it stops. Such a distance called the stop-distance and is given by $v_0\tau_g$. Thus, τ_g is a measure of the inertial motion of a particle in a fluid.

Motion of a Small Diffusive Particle

When a particle is small, Brownian motion occurs caused by random variations in the incessant bombardment of molecules against the particle. As the result of Brownian motion, aerosol particles appear to diffuse in a manner analogous to the diffusion of gas molecules.

The Brownian diffusion coefficient of particles with diameter D_p is given by

$$D = C_c kT/3\pi\mu D_p \tag{15}$$

where k is the Boltzmann constant ($=1.38 \times 10^{-16}$ erg/K) and T the temperature [K]. The mean square displacement of a particles $\overline{\Delta x^2}$ in a certain time interval t, and its absolute value of the average displacement $\overline{\Delta x}$, by the Brownian motion, are given as follows

$$\begin{aligned} \overline{\Delta x^2} &= 2Dt \\ \overline{\Delta x} &= \sqrt{4Dt/\pi} \end{aligned} \tag{16}$$

The number concentration of small particles undergoing Brownian diffusion in a flow with velocity $\mathbf{u}$ can be determined by solving the following equation of convective diffusion,

$$\frac{\partial N}{\partial t} + \nabla \cdot \mathbf{u}N = D\nabla^2 N - \nabla \cdot \mathbf{v}N \tag{17}$$

$$\mathbf{v} = \tau_g \sum \mathbf{F}/m_p \tag{18}$$

where N is the particle number concentration, D the Brownian diffusion coefficient, and $\mathbf{v}$ the particle velocity due to an external force $\mathbf{F}$ acting on the particle.

The average absolute value of Brownian displacement in one second, $\overline{\Delta x}$, is shown in Figure 3, which is obtained from $t = 1$s in Eq. (3.4). The intersection of the curves $\overline{\Delta x}$ and v_t lies at around 0.5 μm at atmospheric pressure. If one observes the settling velocity of such a small particle in a short time, it will be a resultant velocity caused by both gravitational settling and Brownian motion.

The local deposition rate of particles by Brownian diffusion onto a unit surface area, the deposition flux j (number of deposited particles per unit time and surface area), is given by

$$j = -D\nabla N + vN + uN. \tag{19}$$

If the flow is turbulent, the value of the deposition flux of uncharged particles depends on the strength of the flow field, the Brownian diffusion coefficient, and gravitational sedimentation.

Particle Charging and Electrical Properties

When a charged particle having n_p elementary charges is suspended in an electrical field of strength E, the electrical force F_e exerted on the particle is $n_p eE$, where e is the elementary charge unit ($e = 1.6 \times 10^{-19}$C). Introducing F_e into the right hand side of the equation of particle motion in Table 3 and assuming that gravity and buoyant forces are negligible, the steady state velocity due to electrical force is found by equating drag and electrical forces, $F_d = F_e$. For the Stokes drag force ($F_d = 3\pi\mu v_e D_p/C_c$), the terminal electrophoretic velocity v_e is given by

$$v_e = n_p eEC_c/3\pi\mu D_p. \tag{20}$$

B_e in Figure 3 is the electrical mobility which is defined as the velocity of a charged particle in an electric field of unit strength. Accordingly, the steady particle velocity in an electric field E is given by Eb_e. Since B_e depends upon the number of elementary charges that a particle carries, n_p, as seen in Eq. (3.7), n_p is required to determine B_e. n_p is predictable with aerosol particles in most cases, where particles are charged by diffusion of ions.

The charging of particles by gaseous ions depends on the two physical mechanisms of diffusion and field charging (Flagan and Seinfeld, 1988). Diffusion charging arises from thermal collisions between particles and ions. Charging occurs also when ions drift along electric field lines and impinge upon the particle. This charging process is referred to as field charging. Diffusion charging is the predominant mechanism for particles smaller than about 0.2 μm in diameter. In the size range of 0.2–2 μm diameter, particles are charged by both diffusion and field charging. Charging is also classified into bipolar charging by bipolar ions and unipolar charging by unipolar ions of either sign. The average number of charges on particles by both field and diffusion charging are shown in Figure 4. When the number concentration of bipolar ions is sufficiently high with sufficient charging time, the particle charge attains an equilibrium state where the positive and negative charges in a unit volume are approximately equal. Figure 5 shows the charge distribution of particles at the equilibrium state.

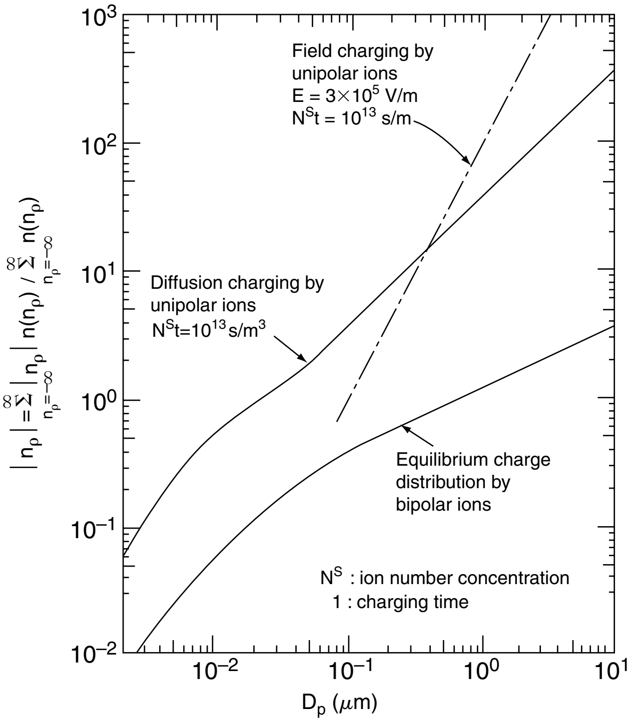


FIGURE 4 The average number of charges on particles by both field and diffusion charging.

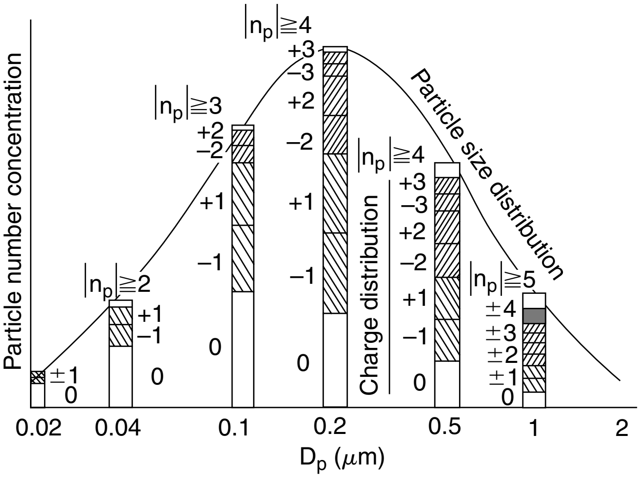


FIGURE 5 Equilibrium charge distribution through bipolar ion charging. The height of each section corresponds to the number concentration of particles containing the indicated charge..

Brownian Coagulation

Coagulation of aerosols causes a continuous change in number concentration and size distribution of an aerosol with the total particle volume remaining constant. Coagulation can be classified according to the type of force that causes collision. Brownian coagulation (thermal coagulation) is a fundamental mechanism that is present whenever particles are present in a background gas.

In the special case of the initial stage of coagulation of a monodisperse aerosol having uniform diameter D_p, the particle number concentration N decreases according to

$$\begin{aligned} dN/dt &= -0.5K_0N^2 \\ K_0 &= K\left(D_p, D_p\right) \end{aligned} \tag{21}$$

where $K(D_p, D_p)$ is the coagulation coefficient between particles of diameters D_p and D_p.

When the coagulation coefficient is not a function of time, the decrease in particle number concentration from N_0 to N can be obtained from the integration of Eq. (21) over a time period from 0 to t,

$$N = N_0/(1 + 0.5K_0N_0t). \tag{22}$$

The particle number concentration reduces to one-half its initial value at the time $2(K_0N_0)^{-1}$. This time can be considered as a characteristic time for coagulation.

In the case of coagulation of a polydisperse aerosol, the basic equation that describes the time-dependent change in the particle size distribution $n(v, t)$, is

$$\begin{aligned} \frac{\partial n(v,t)}{\partial t} &= \frac{1}{2}\int_0^v K(v', v-v')\,n(v',t)\,n(v-v',t)\,dv' \\ &\quad - n(v,t)\int_0^\infty K(v,v')\,n(v',t)\,dv' \end{aligned} \tag{23}$$

The first term on the right-hand side represents the rate of formation of particles of volume v due to coagulation, and the second term that rate of loss of particles of volume v by coagulation with all other particles.

The Brownian coagulation coefficient is a function of the Knudsen number $Kn = 2\lambda/D_p$, where λ is the mean free path of the background gas. Figure 6 shows the values of the Brownian coagulation coefficient of mono-disperse particles, $0.5\,K(D_p, D_p)$, as a function of particle diameter in

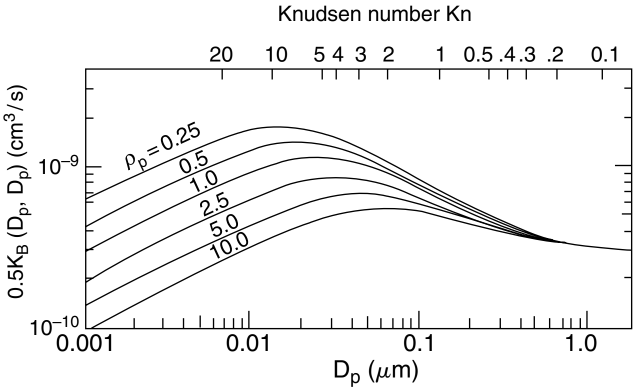


FIGURE 6 Brownian coagulation coefficient for coagulation of equal-sized particles in air at standard conditions as a function of particle density.

air at atmospheric pressure and room temperature. There exist distinct maxima in the coagulation coefficient in the size range from 0.01 μm to 0.01 μm depending on particle diameter. For a particle of 0.4 μm diameter at a number concentration of 10^8 particles/cm^3, the half-life for Brownian coagulation is about 14 s.

Kelvin Effect

p_d/p_∞ in Figure 3 indicates the ratio of the vapor pressure over a curved droplet surface to that over a flat surface of the same liquid. The vapor pressure over a droplet surface increases with a decrease in droplet diameter. This phenomenon is called the Kelvin effect and is given by Eq. (3.8). If the saturation ratio of water vapor S surrounding a single isolated water droplet is larger than p_d/p_∞, the droplet grows. If $S < p_d/p_\infty$, that is, the surrounding saturation ratio lies below the curve p_d/p_∞ in Figure 3, the water droplet evaporates. Thus the curve p_d/p_∞ in Figure 3 indicates the stability relationship between the droplet diameter and the surrounding vapor pressure.

Phoretic Phenomena

Phoretic phenomena refer to particle motion that occurs when there is a difference in the number of molecular collisions onto the particle surface between different sides of the particle. Thermophoresis, photophoresis and diffusiophoresis are representative phoretic phenomena.

When a temperature gradient is established in a gas, the aerosol particles in that gas are driven from high to low temperature regions. This effect is called thermophoresis. The curve v_{th} in Figure 3 is an example (NaCl particles in air) of the thermophoretic velocity at a unit temperature gradient, that is, 1 K/cm. If the temperature gradient is 10 K/cm, v_{th} becomes ten times higher than shown in the figure.

If a particle suspended in a gas is illuminated and nonuniformly heated due to light absorption, the rebound of gas molecules from the higher temperature regions of the particle give rise to a motion of the particle, which is called photophoresis and is recognized as a special case of thermophoresis. The particle motion due to photophoresis depends on the particle size, shape, optical properties, intensity and wavelength of the light, and accurate prediction of the phenomenon is rather difficult.

Diffusiophoresis occurs in the presence of a gradient of vapor molecules. The particle moves in the direction from higher to lower vapor molecule concentration.

OPTICAL PHENOMENA

When a beam of light is directed at suspended particles, various optical phenomena such as absorption and scattering of the incident beam arise due to the difference in the refractive index between the particle and the medium. Optical phenomena can be mainly characterized by a dimensionless parameter defined as the ratio of the particle diameter D_p to the wavelength of the incident light λ,

$$\alpha = \pi D_p/\lambda. \tag{24}$$

Light Scattering

Light scattering is affected by the size, shape and refractive index of the particles and by the wavelength, intensity, polarization and scattering angle of the incident light. The theory of light scattering for a uniform spherical particle is well established (Van de Hulst, 1957). The intensity of the scattered light in the direct θ (angle between the directions of the incident and scattered beams) consists of vertically polarized and horizontally polarized components and is given as

$$I = I_0 \frac{\lambda^2}{8\pi^2 r^2}\left(i_1 + i_2\right) \tag{25}$$

where I_0 denotes the intensity of the incident beam, λ the wavelength and r the distance from the center of the particle, i_1 and i_2 indicate the intensities of the vertical and horizontal components, respectively, which are the functions of θ, λ, D_p and m.

The index of refraction m of a particle is given by the inverse of the ratio of the propagation speed of light in a vacuum k_0 to that in the actual medium k_1 as,

$$m = k_1/k_0 \tag{26}$$

and can be written in a simple form as follows:

$$m = n_1 - in_2. \tag{27}$$

The imaginary part n_2 gives rise to absorption of light, and vanishes if the particle is nonconductive.

Light scattering phenomena are sometimes separated into the following three cases: (1) Rayleigh scattering (molecular scattering), where the value of α is smaller than about 2, (2) Mie scattering, where α is from 2 to 10, and (3) geometrical optics (diffraction), where α is larger than about 10. In the Rayleigh scattering range, the scattered intensity is in proportion to the sixth power of particle size. In the Mie scattering range, the scattered intensity increases with particle size at a rate that approaches the square of particle size as the particle reaches the geometrical optics range. The amplitude of the oscillation in scattered intensity is large in the forward direction. The scattered intensity greatly depends on the refractive index of the particles.

The curve denoted as pulse height in Figure 3 illustrates a typical photomultiplier response of scattered light from a particle. The intensity of scattered light is proportional to the sixth power of the particle diameter when particle size is smaller than the wavelength of the incident light (Rayleigh scattering range). The curve demonstrates the steep decrease in intensity of scattered light from a particle.

Light Extinction

When a parallel beam of light is passed through a suspension, the intensity of light is decreased because of the scattering and absorption of light by particles. If a parallel light beam of intensity I_0 is applied to the suspension, the intensity I at a distance l into the medium is given by,

$$I = I_0 \exp(-\gamma l) \tag{28}$$

where γ is called the extinction coefficient,

$$\gamma = \int_0^{\infty} C_{\text{ext}} n\left(D_p\right) \mathrm{d}D_p \tag{29}$$

$n(D_p)$ is the number distribution function of particles, and C_{ext} is the cross sectional area of each particle.

For a spherical particle, C_{ext} can be calculated by the Mie theory where the scattering angle is zero. The value of C_{ext} is also given by

$$C_{\text{ext}} = C_{\text{sca}} + C_{\text{abs}} \tag{30}$$

where C_{sca} is the cross sectional area for light scattering and C_{abs} the cross sectional area for light absorption. The value of C_{sca} can be calculated by integrating the scattered intensity I over the whole range of solid angles.

The total extinction coefficient γ in the atmosphere can be expressed as the sum of contributions for aerosol particle scattering and absorption and gaseous molecular scattering and absorption. Since the light extinction of visible rays by polluted gases is negligible under the usual atmospheric conditions and the refractive index of atmospheric conditions and the refractive index of atmospheric aerosol near the ground surface is $(1.33 \sim 1.55) - (0.001 \sim 0.05)i$ (Lodge *et al.*, 1981), the extinction of the visible rays depends on aerosol particle scattering rather than absorption. Accordingly, under uniform particle concentrations, the extinction coefficient becomes a maximum for particles having diameter 0.5 μm for visible light.

VISIBILITY

The visible distance that can be distinguished in the atmosphere is considerably shortened by the light scattering and light extinction due to the interaction of visible light with the various suspended particles and gas molecules. To evaluate the visibility quantitatively, the visual range, which is defined as the maximum distance at which the object is just distinguishable from the background, is usually introduced. This visual range is related to the intensity of the contrast C for an isolated object surrounded by a uniform and extensive background. The brightness can be obtained by integrating Eq. (28) over the distance from the object to the point of observation. If the minimum contrast required to just distinguish an object from its background is denoted by C^*, the visual range L_v for a black object can be given as

$$L_v = -(1/\gamma)\ln(-C^*) \tag{31}$$

where γ is the extinction coefficient. Introduction of the value of -0.02 for C^* gives the well known Koschmieder equation,

$$L_v = 3.912/\gamma \tag{32}$$

For aerosol consisting of 0.5 μm diameter particles ($m = 1.5$) at a number concentration of 10^4 particles/cm^3, the extinction coefficient γ is 6.5×10^{-5} cm and the daylight visual range is about 6.0×10^4 cm (=0.6 km). Since the extinction coefficient depends on the wavelength of light, refractive index, aerosol size and concentration, the visual range greatly depends on the aerosol properties and atmospheric conditions.

MEASUREMENT OF AEROSOLS

Methods of sizing aerosol particles are generally based upon the dynamic and physical properties of particles suspended in a gas (see Table 4).

Optical Methods

The light-scattering properties of an individual particle are a function of its size, shape and refractive index. The intensity of scattered light is a function of the scattering angle, the intensity and wavelength of the incident light, in addition to the above properties of an individual particle. An example of the particle size-intensity response is illustrated in Figure 3. Many different optical particle sizing devices have been developed based on the Mie theory which describes the relation among the above factors. The principle of one of the typical devices is shown in Figure 7.

The particle size measured by this method is, in most cases, an optical equivalent diameter which is referred to a calibration particle such as one of polystyrene latex of known size. Unless the particles being measured are spheres of known refractive index, their real diameters cannot be evaluated from the optical equivalent diameters measured. Several light-scattering particle counters are commercially available.

Inertial Methods (Impactor)

The operating principle of an impactor is illustrated in Figure 8. The particle trajectory which may or may not collide with the impaction surface can be calculated from solving the equation of motion of a particle in the impactor flow field. Marple's results obtained for round jets are illustrated in Figure 8 (Marple and Liu, 1974), where the collection efficiency at the impaction surface is expressed in terms of the Stokes number, *Stk*, defined as,

$$Stk = \frac{\rho_p C_c D_p^2 u_0}{18\mu\left(W/2\right)} = \tau \frac{u_0}{W/2} \tag{33}$$

where

$$\tau = \frac{\rho_p D_p^2 C_c}{18\mu} \tag{34}$$

$$C_c = 1 + 2.514\frac{\lambda}{D_p} + 0.80\frac{\lambda}{D_p}\exp\left(-0.55\frac{D_p}{\lambda}\right) \tag{35}$$

TABLE 4
Methods of aerosol particle size analysis

Quantity to be measured		Method or instrument	Media	Detection	Approx size range	Concentration	Principle
		microscope	gas	number	>0.5 μm		
length		electron microscope	vacuum	number	>0.001		
	absorbed gas	adsorption method, BET	gas	–	>0.01		BET
area			liquid				
	permeability	permeability method	gas	–	>0.1		Kozeny-Carman's equation
volume	electric resist.	Coulter Counter	liquid	number	>0.3	low	
	gravitational	(individual) ultramicroscope	gas	number	>1	low	Stokes equation
	settling	(differential conc.)	liquid	mass	>1	high	Stokes equation
motion in fluid	velocity	(cumulative conc.)	liquid	mass	>1	high	Stokes equation
	centrifugal	(differential conc.)	liquid	area mass	>0.05	high	Stokes equation
	settling velocity	spiral centrifuge, conifuge	gas	number mass	>0.05–1	high–low	Stokes equation
	inertial collection	impactor, acceleration method	gas	mass number	>0.5	high–low	relaxation time
	inertial motion	impactor, aerosol beam method	gas	number	>0.05	high–low	in low pressure
	diffusion loss	diffusion battery and CNC	gas	mass number	0.002–0.5	high–low	Brownian motion
	Brownian motion	photon correlation	liquid	number	0.02–1	high	
		integral type (EAA)	gas	number (current)	0.005–0.1	high–low	
	electric mobility						
		differential type (DMA)	gas	number (current)	0.002–0.5	high–low	
intensity of scattered light	light scattering	gas liquid	number	>0.1	low	Mie theory	
	light diffraction	gas liquid	number	1	high–low		

ρ_p is the particle density, μ the viscosity and λ is the mean free path of the gas. The remaining quantities are defined in Figure 8.

The value of the Stokes number at the 50 percent collection efficiency for a given impactor geometry and operating condition can be found from the figure, and it follows that the cut-off size, the size at 50 percent collection efficiency, is determined.

If impactors having different cut-off sizes are appropriately connected in series, the resulting device is called a cascade impactor, and the size distribution of aerosol particles can be obtained by weighing the collected particles on each impactor stage. In order to obtain an accurate particle size distribution from a cascade impactor, the following must be taken into account: 1) data reduction considering cross sensitivity between the neighboring stages, 2) rebounding on the impaction surfaces, and 3) particle deposition inside the device.

Various types of impactors include those using multiple jets or rectangular jets for high flow rate, those operating under low pressure (Hering *et al.*, 1979) or having microjets for particles smaller than about 0.3 μm and those having a virtual impaction surface, from which aerosols are sampled, for sampling the classified aerosol particles (Masuda *et al.*, 1979).

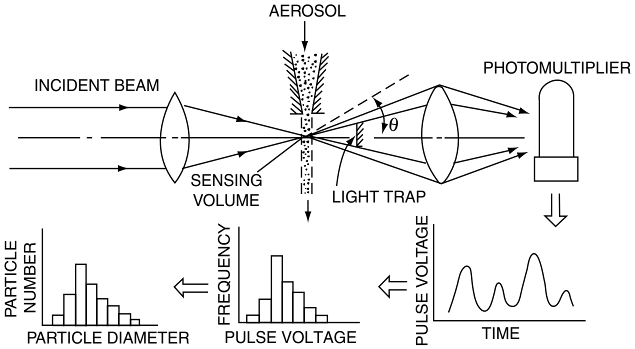


FIGURE 7 Measurement of aerosol particle size by an optical method.

(Other Inertial Methods)

Other inertial methods exist for particles larger than 0.5 μm, which include the particle acceleration method, multi-cyclone (Smith *et al.*, 1979), and pulsation method (Mazumder *et al.*, 1979). Figure 9 illustrates the particle acceleration method where the velocity difference between

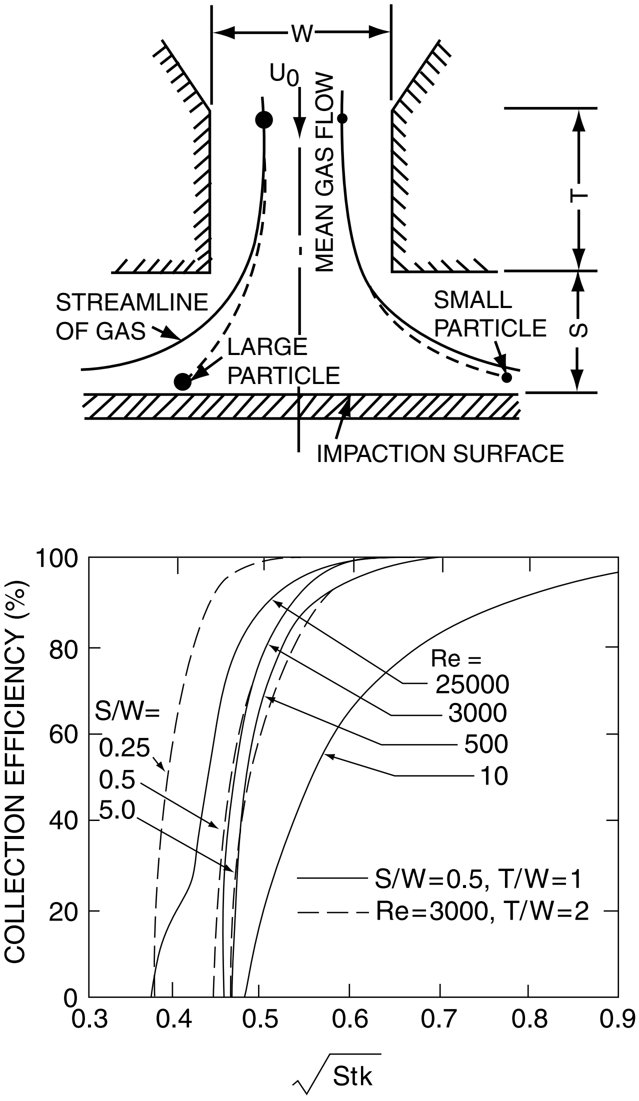


FIGURE 8 Principle of operation of an impactor. Collection efficiency of one stage of an impactor as a function of Stokes number, Stk, Reynolds number, Re, and geometric ratios.

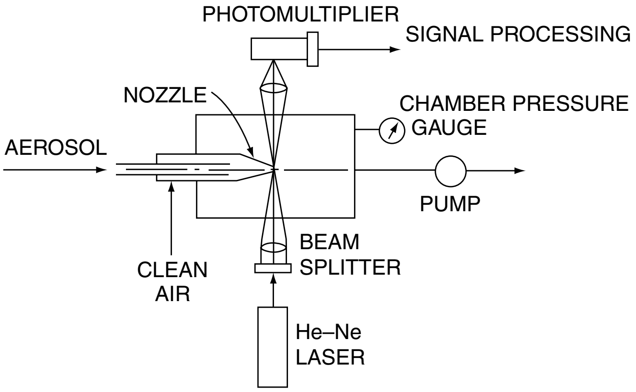


FIGURE 9 Measurement of aerosol particle size by laser-doppler velocimetry.

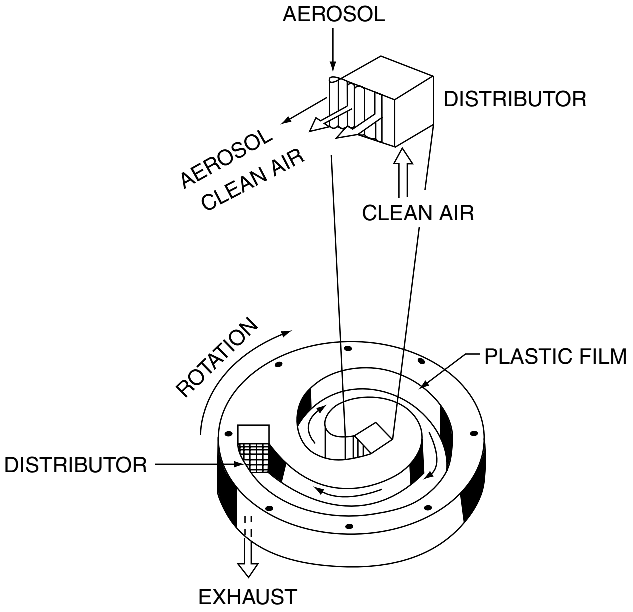


FIGURE 10 Spiral centrifuge for particle size measurements.

a particle and air at the outlet of a converging nozzle is detected (Wilson and Liu, 1980).

Sedimentation Method

By observing the terminal settling velocities of particles it is possible to infer their size. This method is useful if a TV camera and He–Ne gas laser for illumination are used for the observation of particle movement. A method of this type has been developed where a very shallow cell and a TV system are used (Yoshida *et al.*, 1975).

Centrifuging Method

Particle size can be determined by collecting particles in a centrifugal flow field. Several different types of centrifugal chambers, of conical, spiral and cylindrical shapes, have been developed for aerosol size measurement. One such system is illustrated in Figure 10 (Stöber, 1976). Particle shape and chemical composition as a function of size can be analyzed in such devices.

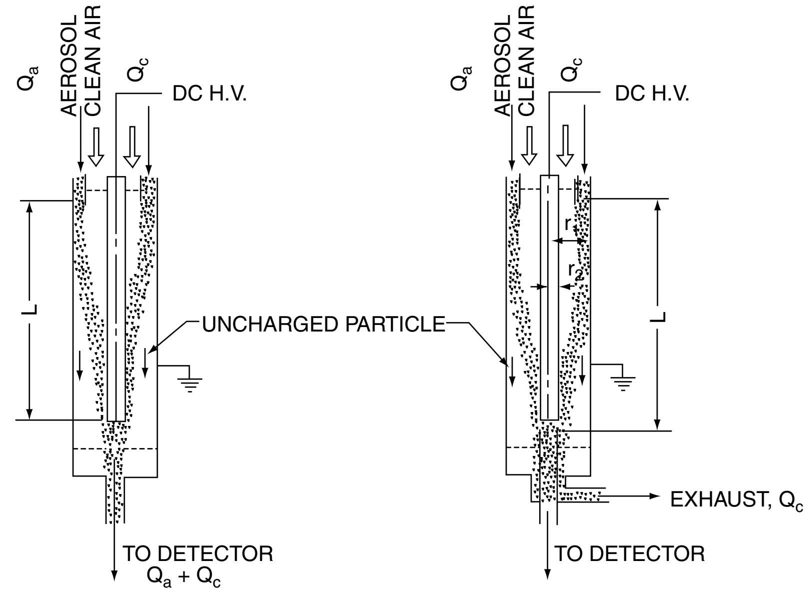
Electrical Mobility Analyzers

The velocity of a charged spherical particle in an electric field, v_e, is given by Eq. (20). The velocity of a particle having unit charge (n_p = 1) in an electric field of 1 V/cm is illustrated in Figure 3. The principle of electrical mobility analyzers is based upon the relation expressed by Eq. (20). Particles of different sizes are separated due to their different electrical mobilities.

a) Corona discharge (unipolar ions)

b) Radioactive source (bipolar ions)

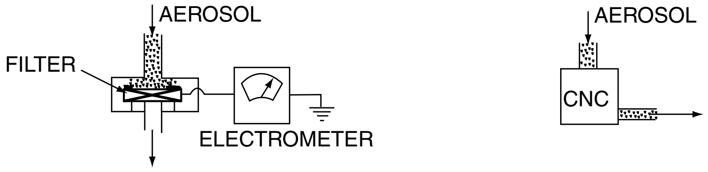
(a) Charging section for particles



a) Integration type

b) Differential type

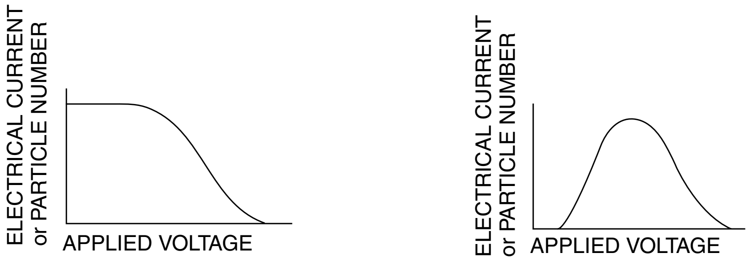
(b) Main section



a) Electrometer

b) CNC or Electrometer

(c) Detection of charged particles



a) Integration type

b) Differential type

(d) Response curve

FIGURE 11 Two types of electrical mobility analyzers for determining aerosol size. Charging, classification, detection and response are shown for both types of analyzers.

Two different types of electrical mobility analyzers shown in Figure 11 have been widely used (Whitby, 1976). On the left hand side in the figure is an integral type, which is commercially available (EAA: Electrical Aerosol Analyzer). That on the right hand side is a differential type, which is also commercially available (DMA: Differential Mobility Analyzer). The critical electrical mobility B_{ec} at which a particle can reach the lower end of the center rod at a given operating condition is given, respectively, for the EAA and DMA as

$$B_{ec} = \frac{(Q_a + Q_c)}{2\pi LV} \ln\left(\frac{r_1}{r_2}\right) \tag{36}$$

$$B_{ec} = \frac{Q_c}{2\pi LV} \ln\left(\frac{r_1}{r_2}\right), \Delta B_e = \frac{Q_a}{\pi LV} \ln\left(\frac{r_1}{r_2}\right) \tag{37}$$

B_{ec} can be changed by changing the electric voltage applied to the center rod. A set of data of the particle number concentration or current at every B_{ec} can be converted into a size distribution by data reduction where the number distribution of elementary charges at a given particle size is taken into account.

Electrical mobility analyzers are advantageous for smaller particles because v_e in Eq. (20) increases with the decrease in particle size. The differential mobility analyzer has been increasingly utilized as a sizing instrument and a monodisperse aerosol generator of particles smaller than 1 μm diameter (Kousaka *et al.*, 1985).

Diffusion Batteries

The diffusion coefficient of a particle D is given by Eq. (15). As shown in Figure 3, D increases with a decrease in particle size. This suggests that the deposition loss of particles onto the surface of a tube through which the aerosol is flowing increases as the particle size decreases. The penetration (=1–fractional loss by deposition) η_p for a laminar pipe flow is given as (Fuchs, 1964),

$$\begin{aligned} \eta_p &= 0.8191\exp(-3.657\beta) + 0.00975\exp(-22.3\beta) \\ &\quad + 0.0325\exp(-57\beta), \beta = \pi DL/Q \geq 0.0312 \end{aligned} \tag{38}$$

$$\eta_p = 1 - 2.56\beta^{2/3} + 1.2\beta + 0.177\beta^{4/3}, \beta < 0.0312 \tag{39}$$

where L is the pipe length and Q is the flow rate. A diffusion battery consists of a number of cylindrical tubes, rectangular ducts or a series of screens through which the gas stream containing the particles is caused to flow. Measurement of the penetration of particles out the end of the tubes under a number of flow rates or at selected points along the distance from the battery inlet allows one to obtain the particle size distribution of a polydisperse aerosol. The measurement of particle number concentrations to obtain penetration is usually carried out with a condensation nucleus counter (CNC), which detects particles with diameters down to about 0.003 μm.

REFERENCES

Flagan, R.C., Seinfeld, J.H. (1988) *Fundamentals of Air Pollution Engineering.* Prentice Hall, Englewood Cliffs, NJ.

Fuchs, N.A. (1964) *The Mechanics of Aerosols.* Pergamon Press, New York, 204–205.

Hering, S.V., Friedlander, S.K., Collins, J.J., Richards, L.W. (1979) Design and Evaluation of a New Low-Pressure Impactor. 2. *Environmental Science & Technology,* **13,** 184–188.

Kousaka, Y., Okuyama, K., Adachi, M. (1985) Determination of Particle Size Distribution of Ultra-Fine Aerosols Using a Differential Mobility Analyzer. *Aerosol Sci. Technology,* **4,** 209–225.

Lodge, J.P., Waggoner, A.P., Klodt, D.T., Grain, C.N. (1981) Non-Health Effects of Particulate Matter. *Atmospheric Environment,* **15,** 431–482.

Marple, V.A., Liu, B.Y.H. (1974) Characteristics of Laminar Jet Impactors. *Environmental Science & Technology,* **8,** 648–654.

Masuda, H., Hochrainer, D. and Stöber, W. (1979) An Improved Virtual Impactor for Particle Classification and Generation of Test Aerosols with Narrow Size Distributions. *J. Aerosol Sci.,* **10,** 275–287.

Mazumder, M.K., Ware, R.E., Wilson, J.D., Renninger, R.G., Hiller, F.C., McLeod, P.C., Raible, R.W. and Testerman, M.K. (1979). SPART analyzer: Its application to aerodynamic size distribution measurement. *J. Aerosol Sci.,* **10,** 561–569.

Seinfeld, J.H. (1986) *Atmospheric Chemistry and Physics of Air Pollution.* Wiley, New York.

Smith, W.B., Wilson, R.R. and Harris, D.B. (1979). A Five-Stage Cyclone System for *In Situ* Sampling. *Environ. Sci. Technology,* **13,** 1387–1392.

Stöber, W. (1976) Design, Performance and Application of Spiral Duct Aerosol Centrifuges, in "Fine Particles", edited by Liu, B.Y.H., Academic Press, New York, 351–397.

Van de Hulst, H.C. (1957) *Light Scattering by Small Particles.* Wiley, New York.

Whitby, K.T. (1976) Electrical Measurement of Aerosols, in "Fine Particles" edited by Liu, B.Y.H., Academic Press, New York, 581–624.

Wilson, J.C. and Liu, B.Y.H. (1980) Aerodynamic Particle Size Measurement by Laser-Doppler Velocimetry. *J. Aersol Sci.,* **11,** 139–150.

Yoshida, T., Kousaka, Y., Okuyama, K. (1975) A New Technique of Particle Size Analysis of Aerosols and Fine Powders Using an Ultramicroscope. *Ind Eng. Chem. Fund.,* **14,** 47–51.

KIKUO OKUYAMA
YASUO KOUSAKA
JOHN H. SEINFELD
University of Osaka Prefecture and California Institute of Technology

---

AGRICULTURAL CHEMICALS: see PESTICIDES

AIR POLLUTANT EFFECTS

AIR POLLUTANTS

Air pollutants fall into two main categories: (1) those that are pervasive throughout areas because they are the products of daily-life activities such as transportation, power generation, space and water heating, and waste incineration, and (2) those generated by activities such as chemical, manufacturing, and agricultural processing whose pollutant byproducts tend to be localized in nearby areas or are spread long distances by tall stacks and prevailing winds.

Air pollutants are also categorized by their emission characteristics: (1) point sources, such as power plants, incinerators, and large processing plants; (2) area sources, such as space and water heating in buildings; and (3) mobile sources, mainly cars and trucks, but also lawn mowers and blowers and airplanes.

The United States has established National Ambient Air Quality Standards (NAAQS) for seven pollutants that are pervasive and are threats to public health and welfare. The Clean Air Act, which initiated this program, was passed in 1963 and last amended in 1990. The primary standards are intended to protect health, and the secondary standards protect public-welfare interests such as visibility and danger to animals, crops, and buildings.

The standards reflect, for the most part but not always, a conservative approach in favor of the protection of health. It is notable that the public, who in the final analysis must pay the cost, appears to be firmly committed to enforcement of the standards without overwhelming concern for costs.

The act requires the states to determine the status of their air quality and to find and introduce the controls that will enable them to meet these standards. Their proposal describing how and when the standards will be met is submitted to the EPA (U.S. Environmental Protection Agency) as an implementation plan for approval. Meeting target dates for air-quality standards has been problematic because the complex system that has to be managed includes important socioeconomic and political factors. For example, the close connection between air quality and daily activities such as transportation, waste disposal, and the heating of homes and workplaces requires education of the population to obtain their support for alternative and perhaps costly lifestyle choices in the vehicles they purchase, the packaging of articles they choose, and the type and cost of the fuels they use—choices they may be reluctant to make, even if they will improve the quality of their air environment. Choices benefiting air quality that carry disadvantages for important sectors of the economy are usually skillfully discouraged by some of those sectors.

CONTROL OF CRITERIA POLLUTANTS

Control of the criteria pollutants requires a measurement program to determine the daily and short-term patterns of the ambient concentrations, identification of the emitting sources, and design and implementation of strategies for their control. A detailed inventory of the sources causing the pollution is prepared. The effectiveness of control technology and potential regulatory strategies are evaluated and their availability determined with consideration given to the economic and political restraints on their implementation. In other words, the total system to be managed and its interactions have to be detailed and understood in order to evaluate the potential for successful control of the air pollution in an area.

The amount of exposure to the pollutants from independent or grouped sources depends upon the intensity of the activities producing the emissions, the effectiveness of the controls, and the quality of the surveillance instituted to ensure the continued proper use and maintenance of the controls. A factor that can be overwhelming is the pattern of the local meteorology and its effectiveness in dispersing emitted pollutants. The effects of dispersions from one area upon downwind areas should also be considered.

Detailed analysis of data accumulated over many years using unchanging analytical methods has shown that very significant changes in an area's air pollution can take place from year to year without significant changes in controls, primarily as the result of changes in the local weather patterns. The combination of 10 years of data at three sampling sites in New York City showed that its sulfur-dioxide pollution problems was clearly related to the sulfur content of the fuel that was burned in the city. The data for a 10-year period were combined on a week-by-week basis, with the result that the shape of the 10-year curve for ambient sulfur-dioxide concentrations and the long-term temperature curve for the city could be superimposed with significant success. Therefore, the sometimes great variations found between years when little change occurred in controls were caused by variations in the local atmosphere, demonstrating that the success or failure of control strategies cannot be evaluated with security over short intervals of time.

Pollutant	Primary Stds.	Averaging Times	Secondary Stds.
Carbon monoxide	9 ppm (10 mg/m^3)	8-hour[1]	None
	35 ppm (40 mg/m^3)	1-hour[1]	None
Lead	1.5 μg/m^3	Quarterly Average	Same as primary
Nitrogen dioxide	0.053 ppm (100 μg/m^3)	Annual (arith. mean)	Same as primary
Particulate matter (PM_{10})	50 μg/m^3	Annual[2] (arith. mean)	Same as primary
	150 μg/m^3	24-hour[1]	Same as primary
Particulate matter ($PM_{2.5}$)	15.0 μg/m^3	Annual[3] (arith. mean)	Same as primary
	65 μg/m^3	24-hour[4]	—
Ozone	0.08 ppm	8-hour[5]	Same as primary
	0.12 ppm	1-hour[6]	Same as primary
Sulfur oxides	0.03 ppm	Annual (arith. mean)	—
	0.14 ppm	24-hour[1]	—
	—	3-hour[1]	0.5 ppm (1300 μg/m^3)

---

1. Not to be exceeded more than once per year.
2. To attain this standard, the expected annual arithmetic mean PM_{10} concentration at each monitor within an area must not exceed 50 μg/m^3.
3. To attain this standard, the 3-year average of the annual arithmetic mean $PM_{2.5}$ concentrations from single or multiple community-oriented monitors must not exceed 15.0 μg/m^3.
4. To attain this standard, the 3-year average of the 98th percentile of 24-hour concentrations at each population-oriented monitor within an area must not exceed 65 μg/m^3.
5. To attain this standard, the 3-year average of the fourth-highest daily maximum 8-hour average ozone concentrations measured at each monitor within an area over each year must not exceed 0.08 ppm.
6. (a) The standard is attained when the expected number of days per calendar year with maximum hourly average concentrations above 0.12 ppm is ≤1, as determined by appendix H.

   (b) The 1-hour NAAQS will no longer apply to an area one year after the effective date of the designation of that area for the 8-hour ozone NAAQS. The effective designation date for most areas is June 15, 2004. (40 CFR 50.9; see Federal Register of April 30, 2004 (69 FR 23996).)

FIGURE 1 National Ambient Air Quality Standards.

The primary standards to protect health and the secondary standards to protect welfare, Figure 1, have improved with increasing knowledge about the effects of exposures and measurement technology.

EPIDEMIOLOGY

Epidemiology is the study of the occurrence and distribution of disease within a population as opposed to its study on an individual basis. An epidemiologist who undertakes to determine the acute and chronic effects caused by exposures of a population to a particular component of local air pollution faces complex problems that can be itemized as follows:

- In a community study the subjects under scrutiny are subjected to pollutants, known and unknown, other than the ones being investigated.
- Supporting clinical studies guiding the investigation are seldom based upon human data, but must depend upon studies using surrogate species that were exposed to much higher doses without the contaminants that may contribute to the effects found in the epidemiological study.

- The true dose is not always the simple product of the measured concentration and the duration of exposure, because of the complexity that can exist between exposure and response—the biologically active dose can be quite different.
- Individuals whose exposure and symptoms are being correlated very often spend the major part of their time indoors or traveling, where they may be subjected to different pollutants and different concentrations.
- Different pollutants will disperse and interact differently with the surroundings, introducing a location factor caused by the relationship of exposed individuals to the measurement site—for example, sulfur-dioxide concentrations will not vary as much as ozone concentrations, because the higher reactivity of ozone with structural materials and other compounds will affect its concentration at the receptor.

INFORMING THE PUBLIC

The aerometric networks established by cities and states have been gathering and analyzing data about air pollutants for many years. During these years, attempts were made to inform the public about the quality of its air environment, which can change from day to day and even hour to hour, and about the possible impact that local concentrations are having upon their health. The relationship between raw air-pollution data and its health-impact significance is complex; therefore, the attempt is made to present the information in a simplified manner that is understandable to the public. Toward this goal, the EPA has developed an Air Quality Index for daily reporting about what has been found in the air together with some indication of its potential effects on health. Important considerations are the variability in the susceptibility of the exposed population, meaning that what may have little or no effect on one group can be a serious concern for others, and that personal patterns of behavior of the exposed can affect the amounts of pollutants that they breathe. Individuals whose lifestyle requires them to move throughout an area (indoors, outdoors, and in vehicles) will receive very different exposures from those who stay at home, depending upon the pervasiveness of the pollutants. In particular, exposures to carbon monoxide will be much greater for those whose daily activities requires them to be in the vicinity of motor vehicles than for those who stay indoors or travel on railroads and subways.

The Air Quality Index designed by EPA reports the daily levels of ozone, particulate matter, carbon monoxide, sulfur dioxide, and nitrogen dioxide on a scale of 0 to 500. The range corresponds to six different categories of health concern that are also characterized by colors Table 1.

RISK REDUCTION

Air pollution affects people primarily through the respiratory system; therefore, the logical way to start minimizing risk is by avoidance of activities that increase one's inhalation of polluted air. When air-pollution levels are high, activities that cause increases in breathing rate should be minimized as much as possible, depending upon the importance and necessity of the activity and the seriousness of the pollution episode. As an example, jogging in the vicinity of vehicles where local ventilation is poor, as in the canyon streets of cities, should be avoided because of the high concentrations of carbon monoxide and other pollutants usually found in those areas. This is of special importance to people with asthma or heart diseases such as angina.

Children who spend their time playing outdoors should be restrained from overexerting themselves when ozone levels are high during warm-weather episodes, as should individuals with asthma or other respiratory diseases or those who are hypersensitive to ozone.

EFFECTS OF EXPOSURE TO CRITERIA POLLUTANTS

Respiratory-System Overview

An elementary understanding of studies describing the adverse health effects caused by the inhalation of gaseous or particulate air pollutants requires at least an elementary familiarity with respiratory-tract anatomy and dynamics. The respiratory tract can be considered to include three sections:

- Nasopharynx—nose and mouth down to epiglottis and larynx

TABLE 1

Air Quality Index	Health Concerns
0–50 Good (green)	Little or no risk
51–100 Moderate (yellow)	Concern for unusually sensitive people
101–150 Unhealthy for sensitive groups (orange)	The general public is unaffected, but people with health problems such as lung and heart disease may be affected
151–200 Unhealthy (red)	Everyone is affected to some degree, especially those in sensitive groups
201–300 Very unhealthy (purple)	A health alert exists; everyone should take precautions, especially those in sensitive groups
301–500 Hazardous (maroon)	Everyone is affected and everyone should take precautions

- Tracheobronchial—bronchi down to terminal bronchiole
- Pulmonary—respiratory bronchiole, alveoli ducts, and alveoli

The trachea divides into left and right bronchi, which divide many times into smaller and smaller tubes down to the respiratory bronchioles. These feed about 65,000 lobules, each containing approximately 5,000 thin-walled air sacs called alveoli. Thus, in an adult there are approximately 300 million alveoli whose thin walls, totaling 70 m^2 in area, contain hundreds of miles of tiny capillaries. Oxygen is added to and carbon dioxide is removed from the blood through the walls of these capillaries. The transfer of toxic chemicals into the blood also can takes place in the alveoli.

Starting in the nose, where the air is conditioned for proper temperature and humidity, the direction of airflow is changed many times, thereby causing the impaction and deposition of particles on the surfaces of the branching airways. These surfaces contain hairlike ciliary cells whose rapid, wavelike motion, over 15 times per second, carry impacted particles on a mucus layer upward into the trachea for subsequent ingestion.

The velocity of the airflow decreases from about 150 cm/sec at the start to almost zero in the alveoli; the smaller the particles, the greater the ease with which they turn corners, thus escaping impaction to penetrate to the alveoli, where they are collected via sedimentation. The larger particles and soluble gases will be trapped in the upper airways, where tissues and their defense mechanisms can be damaged reversibly or irreversibly depending upon the nature, intensity, and duration of the attack.

The amounts of a water-soluble gas or suspended particles that reach the pulmonary region are strongly dependent upon their inhalation pathway into the body. When inhaled through the nose instead of the mouth, they experience a number of chances of removal by impaction. In the case of sulfur dioxide this process is greatly enhanced by its very rapid solution in the watery fluids on the surface of nasal tissues. The greater tendency for mouth breathing combined with the greater intake of air that accompanies increased exertions contraindicates strenuous activity wherever pollutant levels are high.

Particle removal by deposition along the upper and lower respiratory system is strongly dependent upon particle size. Particles with an aerodynamic diameter above 10 μm are removed in the convoluted, moist passages of the nose and tracheobronchial region. While almost all those below 2 μm reach the pulmonary region, intermediate sizes tend to distribute themselves along both regions. When the particles are insoluble, they are removed in a few days from the upper respiratory system by mucociliary action; however, those that penetrate down farther can remain for many months or even years. Removal of particles also occurs by phagocytosis through the scavenging action of macrophages.

The size distribution of particles suspended in the atmosphere exhibits a log-normal behavior. The distribution by mass tends to separate into a fine and a coarse group depending principally upon whether they are formed by condensation of very small precursors, such as those produced in combustion, or are produced from larger particles by mechanical breakdown processes.

OZONE

Ozone is a very reactive chemical that readily attacks other molecules, including those in the tissues of the respiratory system. Exertions that increase the need for oxygen will increase air intake and allow ozone molecules to penetrate and damage the sensitive areas of the lungs. Ozone can aggravate asthma attacks by making individuals more sensitive to allergens that promote the attacks and more susceptible to respiratory infections. Lung tissue can be scarred by continued exposure to ozone over the years. Researchers at Johns Hopkins found that an increase of 10 ppb in weekly ozone levels in cities whose average level was 26 ppb was associated with a 0.52% daily increase in deaths the following week. They calculated that a 10-ppb reduction in daily ozone levels could save nearly 4,000 lives throughout the 95 urban communities included in the study. Out of 296 metropolitan areas, 36 have significant upward trends in the criteria pollutants; however, of these, only trends involving ozone had values over the level of air-quality standards.

The presence of ozone and other photochemical pollutants depends upon atmospheric conditions, notably temperature, as experience shows that this type of pollution is associated with warm temperatures. The precursors that are affected by elevated temperatures are volatile organic compounds (VOCs) and nitric oxide. Natural sources for these compounds are a less important factor than the emissions produced by human activities, but the long-range transport of the precursors while atmospheric conditions are converting them to photochemical oxidants means that that there is a possibility of picking up precursor material from natural sources en route. Control of this type of air pollutant is focused on controlling emissions of VOCs and nitrogen oxides. It should be noted that ambient concentrations of the criteria pollutant nitrogen dioxide have been found to be generally below the levels considered to be health-damaging; therefore, efforts to control its presence in the atmosphere is driven by the need to control ozone. The combustion of fuels and other materials provides sufficient energy to cause the nitrogen and oxygen in the air to react to form nitric oxide. The slow air oxidation of nitric oxide to nitrogen dioxide results in a mixture described as nitrogen oxides (NO_x).

The chemical reactions involved in the formation of photochemical oxidants from these precursors is complex. The basic reactions are:

$$NO_2 + h\nu = NO + O$$
$$O + O_2 + M = O_3 + M^*$$

where hν represents a photon and M and M^* represent material before and after absorbing energy from the ozone-formation reaction. In the absence of other molecules capable of reacting with the nitric oxide, the ozone is removed by the rapid reaction

$$NO + O_3 = NO_2 + O_2$$

Therefore, concentrations of ozone will remain quite small unless there is a competing reaction for rapid removal of the nitric oxide.

Many organic compounds can play the role of nitric-oxide remover in forming photochemical oxidants such as peroxyacetylnitrate (PAN, $CH_3COO_2NO_2$). VOCs possessing varying reactivities are able to remove nitric oxide and thus make possible the buildup of ozone:

$$VOC + NO \rightarrow NO_2 + \text{organic nitrates}$$

Although ozone is the major component, peroxynitrates, peracids, hydroperoxides, aldehydes, and a variety of other compounds are found in photochemical smog. Among the major sources releasing reactive organic compounds are automobile engines and tailpipes; gasoline stations; the use of solvents, paints, and lacquers; and a variety of industrial operations. Thus, the control of ozone is complicated by the variety of sources and the distances that can occur between high-ozone areas and the sources. Suburban and rural areas downwind of urban sources will often have higher ozone levels than source areas because of the transport that occurs while ozone is being formed. Both ozone and PAN cause serious injury to vegetation, but PAN does so at much lower concentrations.

PARTICULATE MATTER

"Fine particles" are less than 2.5 μm in size and require electron microscopy for detection; nevertheless, they are much larger than molecules such as ozone and other gaseous pollutants, which are thousands of times smaller and cannot be seen even with electron microscopy. Fine particles are formed by the condensation of molecules into solids or liquid droplets, whereas larger particles are mostly formed by mechanical breakdown of material. "Coarse particles" are between 2.5 and 10 μm in diameter and cannot penetrate as readily as fine particles; nevertheless, they have been found to cause serious deterioration of health. The severity of effects will vary with the chemical nature of the particles; however, since their nature can be so varied and difficult to determine, coarse and fine particles are considered in terms of what epidemiological studies have shown.

The inhalation of particles has been linked with illnesses and deaths from heart and lung disease as a result of both short- and long-term exposures. People with heart disease may experience chest pain, palpitations, shortness of breath, and fatigue when exposed to particulate-matter pollutants. Exposures have been linked to cardiac arrhythmias and heart attacks. Inhalation of particulate matter can increase susceptibility to respiratory infections such as asthma and chronic bronchitis. The EPA has found that nearly 100 million people in the United States live in areas that have not met the standard for particulate matter with a diameter less than 2.5 μm. It estimates that compliance by 2010 will prevent 15,000 premature deaths, 75,000 cases of chronic bronchitis, 20,000 cases of acute bronchitis, 10,000 hospital admissions for respiratory and cardiovascular disease, and the loss of 3.1 million days worked.

Emissions from diesel-fuel combustion in vehicles and equipment are a special problem, especially for those individuals breathing in close proximity to the exhausts. Cars, trucks, and off-road engines emit more than half a million tons of diesel particulate matter per year. Emissions of 2.5-μm particles have decreased in the United States from 2.3 million tons in 1990 to 1.8 million tons in 2003.

SULFUR DIOXIDE

The combustion of sulfur-containing fuels is the main source of sulfur-dioxide air pollution. The oil and coal burned to heat homes and water and to produce electrical power are the main sources that affect the general population, but individuals who live near metal smelting and other industrial processes can be heavily exposed. Sulfur-dioxide exposures are usually accompanied by exposures to particulate matter, which together exacerbate the effects.

Emissions of sulfur compounds from motor vehicles have increased in importance as those from oil and coal burning have been reduced. The diesel fuel used in vehicles can contain up to 500 ppm by weight of sulfur. California, which has the unfortunate combination of high emissions and poor atmospheric ventilation, hopes to reduce the allowable sulfur content of fuels to 15 ppm by 2007. It must be noted that California is the only state that is not preempted by the federal government in controlling pollution, because its efforts anteceded those of the federal government. Emissions of sulfur dioxide in the United States decreased from 31 million tons in 1970 to 16 million tons in 2003.

The defense mechanisms of the lung are challenged by sulfur dioxide; however, its rapid solution in water irritates tissues but reduces the concentrations that reach the deeper parts of the lung. Inhalation of particulate matter together with sulfur dioxide increases the hazard to the lungs. Asthmatic children and active adults can experience breathing difficulties in high concentrations of sulfur dioxide, and individuals with cardiovascular disease can have their symptoms exacerbated. The conversion in the atmosphere of sulfur dioxide into sulfite and sulfate acidic aerosol particles increases its threat to health.

Sulfur dioxide harms the body's defense system against particulate pollution and the ingress of bacteria into the body through the respiratory system. It also increases the harmful effects of ozone when both of these gases are present. Asthmatics, the elderly, and those already suffering from respiratory problems are affected at lower concentrations than the general population. Studies have shown that in the 1950s and 1960s, when ambient concentrations were sometimes higher than 1 ppm and mixed with particulate matter, the occurrence of lasting atmospheric inversions resulted in thousands of excess deaths.

CARBON MONOXIDE

Carbon monoxide has afflicted the human race since the discovery of fire. Nature contributes very significant quantities, but it does so in such a highly dispersed fashion that human exposures from this source are insignificant. Nature has provided sinks for this insoluble, relatively unreactive gas; otherwise background concentrations would rise much more rapidly as human contributions added their burden. The oceans, which at one time were believed to be a major sink, are now considered to be a source, because certain marine organisms release enough carbon monoxide to supersaturate the surface layer. The important removal mechanism is believed to be the action of microorganisms that live in soils and plants and the reaction of carbon monoxide with hydroxyl radicals in the atmosphere.

The rapid growth in the use of internal combustion engines has created an outdoor problem as indoor problems were decreased by improvements in space-heating equipment. The problem is concentrated in urban areas where traffic congestion is combined with canyonlike streets. Emissions of carbon monoxide in the United States decreased from 197 million tons in 1970 to 94 million tons in 2003.

With the exception of exposures resulting from the breakdown or misuse of indoor heating equipment that produces fatalities or serious injuries, carbon-monoxide exposures of significance occur in the vicinity of congested traffic. People whose occupation requires them to be near such traffic receive the highest exposures, as do those who jog or bicycle in these areas. Malfunctions in the exhaust system of vehicles also can result in high exposures to their occupants. Exposure to carbon monoxide results in the buildup of carboxyhemoglobin in the blood, which will interfere with the transport of oxygen to cells in the body.

Carbon-monoxide molecules attach themselves to the hemoglobin molecules in the blood with much greater tenacity than do oxygen molecules. The Haldane equation attempts to approximate this competition.

$$\frac{(\mathrm{HbCO})}{(\mathrm{HbO_2})} = 210\,\frac{\mathrm{PCO}}{\mathrm{PO_2}}$$

(HbCO) and (HbO_2) are the concentrations of carboxyhemoglobin and oxyhemoglobin, and PCO and PO_2 are the partial pressures of carbon monoxide and oxygen. Inspiration of air containing high concentrations of carbon monoxide results in its preferential absorption in the blood, thereby interfering with oxygen delivery to the cells in the body. Exposure to carbon monoxide causes a gradual increase in the percentage of carboxyhemoglobin in the blood until an equilibrium value dependent upon the ambient air concentration is reached. The rate of intake is dependent upon the breathing rate; therefore, equilibrium is reached more quickly the greater the exertion. Up to 50 ppm, the equilibrium values of carboxyhemoglobin corresponding to different concentrations of inspired carbon monoxide can be estimated from the equation

$$\%\mathrm{HbCO} = 0.4 + \frac{\mathrm{PPM\ CO}}{7}$$

The 0.4 constant is in the equation to account for the endogenous carbon monoxide, that is, the carboxyhemoglobin that results from the body's own production of carbon monoxide.

Graphic representations of the conversion of hemoglobin to carboxyhemoglobin in the presence of different concentrations of ambient carbon monoxide and the effect of various levels of activity on the rate of uptake are presented in Figure 2 and Figure 3.

The level of HbCO in the blood (Table 2) is the important measurement in the evaluation of carbon-monoxide pollution. High levels of HbCO are associated with cigarette smokers, firemen, garage workers, foundry workers, and individuals who spend extended periods of time in heavy congested traffic or in vehicles with faulty exhaust systems. Ambient carbon-monoxide measurements at a monitoring site can be very misleading as an index of exposure, because study populations are usually mobile and carbon-monoxide concentrations can vary significantly, both horizontally and vertically, throughout an urban area.

Exposures to the high concentrations of carbon monoxide sometimes encountered in community atmospheres, even those well above the national standards, are not believed to be sufficient to initiate cardiopulmonary disease; however, individuals whose pulmonary functions are already significantly impaired because of anemia or damage to the heart, vascular system, or lungs can suffer adverse health effects from such exposures.

In order to maintain normal function, the tissues of the body must receive oxygen at a rate that depends upon their nature and functions. Those with a high rate of oxygen demand are more susceptible to the oxygen-depriving action of carbon monoxide. For example, studies of the brain and liver show a decrease in oxygen pressure at those sites even at levels as low as 2% carboxyhemoglobin. Cardiopulmonary-system abnormalities, such as shunts that have developed that allow venous blood to mix directly with arterial blood, cause the individuals affected to be explicitly sensitive to carbon monoxide. Angina-pectoris patients who experienced exposures that raised their carboxyhemoglobin level to 2.5%—that is, approximately to the level produced by an 8-hour exposure at the concentration set as the air-quality standard—suffered the onset of chest pain from exercise significantly sooner than did other angina patients not similarly exposed. The reduction in risk of heart attack that is observed soon after the cessation of the cigarette-smoking habit indicates that carbon monoxide may be an important factor in precipitating heart attacks. The inhalation of carbon monoxide during pregnancy is a special concern because a higher concentration of carboxyhemoglobin is generated in the fetus than in the mother, and the elimination of carbon monoxide after exposure is slower in the

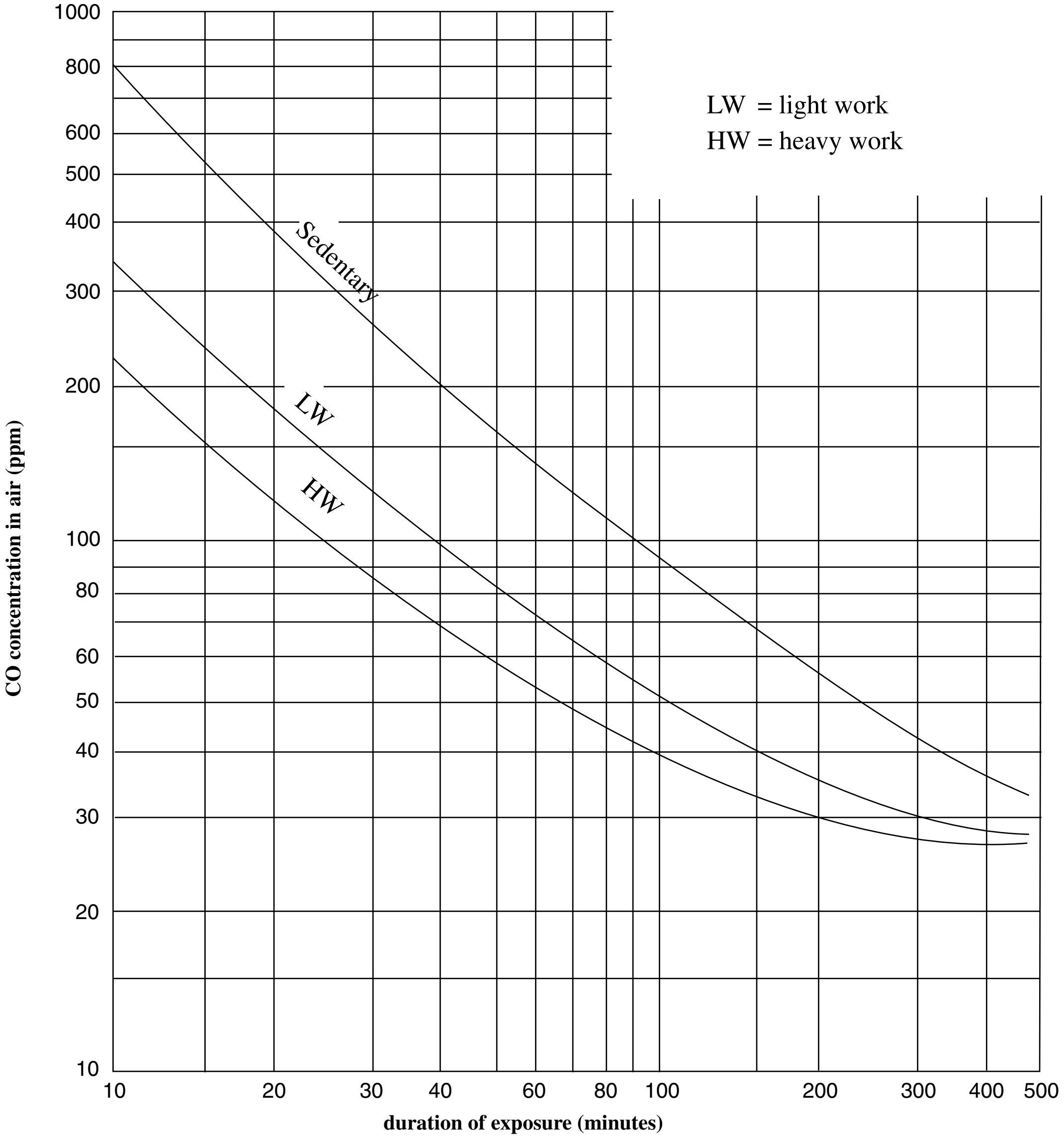


FIGURE 2 Criteria for a Recommended Standard Occupational Exposure to Carbon Monoxide. NIOSH, 1972.

fetus. The effects of combining exposure to carbon monoxide with sudden significant changes in altitude or the intake of drugs or alcohol upon the performance of body functions should be considered and avoided.

NITROGEN OXIDES (NO_x)

"Nitrogen oxides" refers to the mixtures of nitric oxide and nitrogen dioxide that are formed when combustion causes the nitrogen and oxygen in the atmosphere to combine to form nitric oxide, some of which then oxidizes further to nitrogen dioxide; combustion gases contain about 5 to 10% nitrogen dioxide mixed with nitric oxide. The mechanism for the process is believed to be

$$O_2 = 2O$$
$$N_2 + O = NO + N$$
$$N + O_2 = NO + O$$
$$N + OH = NO + H$$

The overall reaction for the formation of nitrogen dioxide is

$$2NO + O_2 = 2NO_2$$

Nitric oxide is oxidized rapidly by ozone; therefore, ozone levels tend to be lower in the vicinity of nitric-oxide sources, such as the tailpipes of vehicles.

Nitrogen dioxide, the most toxic of the nitrogen oxides, causes damage to lung tissues at concentrations higher than usually found in ambient atmospheres. Exposures above the

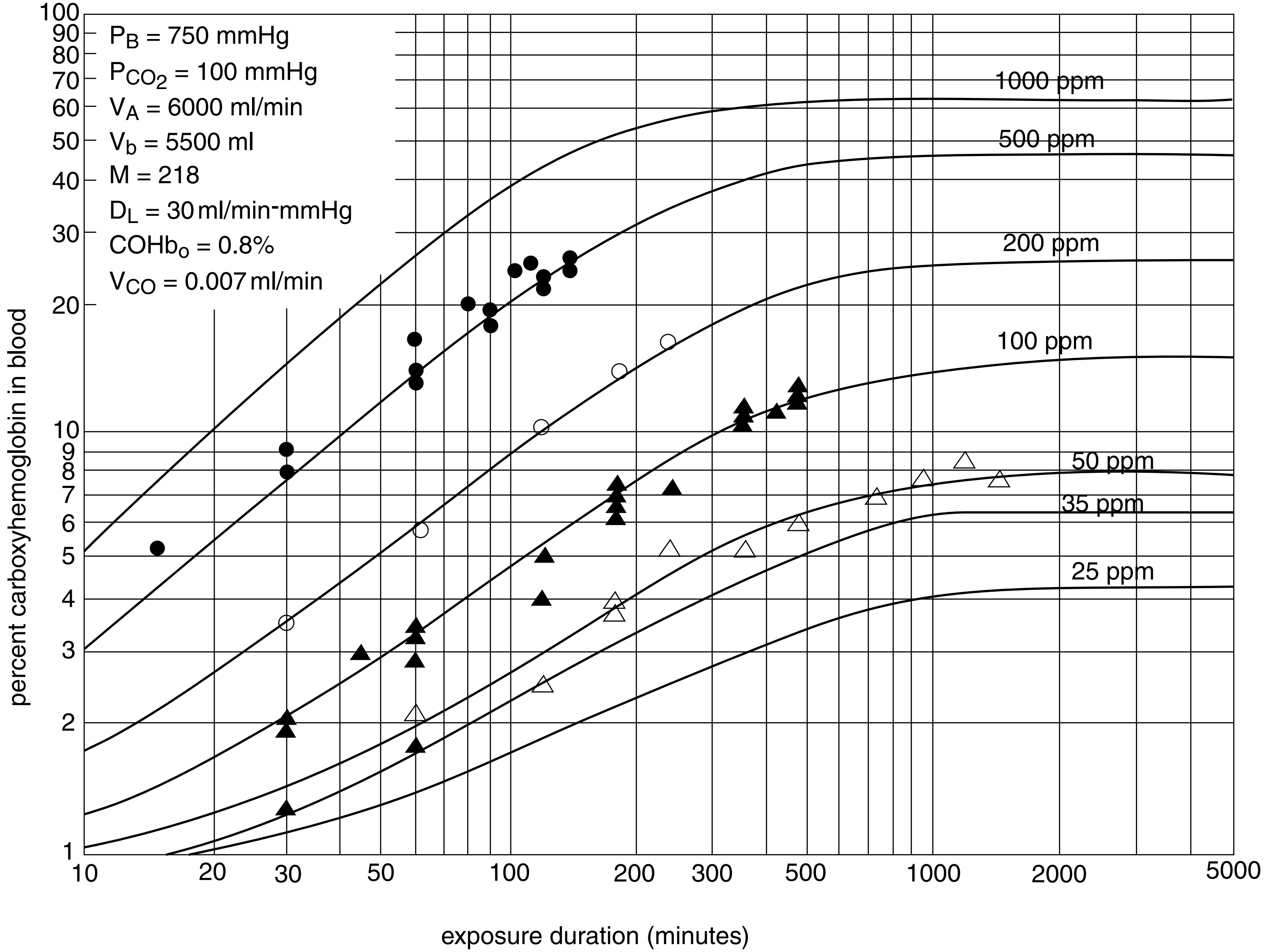


FIGURE 3 Criteria for Recommended Standard Occupational Exposure to Carbon Monoxide. NIOSH, 1972.

TABLE 2
Carboxyhemoglobin levels resulting from steady-state exposure to increasing concentrations of CO in ambient air

CO in Atmosphere (ppm)	COHb in Blood (%)	Signs and Symptoms
10	2	Asymptomatic
70	10	No appreciable effect, except shortness of breath on vigorous exertion; possible tightness across the forehead; dilation of cutaneous blood vessels
120	20	Shortness of breath on moderate exertion; occasional headache with throbbing in temples
220	30	Headache; irritable; easily fatigued; judgment disturbed; possible dizziness; dimness of vision
350–520	40–50	Headache; confusion; collapse; fainting on exertion
800–1220	60–70	Unconsciousness; intermittent convulsion; respiratory failure; death if exposure is long continued
1950	80	Rapidly fatal

Source: Ellenhorn's Medical Toxicology, 2nd Ed, Baltimore, MD: Lippincott Williams & Wilkins.

national standard of 0.053 ppm are rare; therefore, with the exception of activities in the vicinity of industrial sources, nitrogen oxides have not been found to be a cause for community concern. An important consideration in the case of significant exposures is the delay that can occur between exposure and sensations of distress, which may delay prompt treatment. An important effect is the increased susceptibility to pathogens that may result from the destruction of macrophages and general injury to the lung's defense mechanisms.

LEAD

The major source of lead in the air environment has been motor vehicles; therefore, levels have decreased dramatically as regulations have mandated the elimination of lead from gasoline because of its health effects and its detrimental action on the catalytic converters in vehicles. Metal processing, such as in lead smelters, is currently responsible for most of the lead in the air, but waste incinerators and lead-acid-battery manufacturing also contribute.

The chief cause of concern about lead is its effect on children. Lead damages the brain, particularly the cerebellum, and the kidneys, liver, and other organs, and can lead to osteoporosis and reproductive disorders. Its effects upon fetuses and young children produces learning disabilities and lowers IQ. Lead exposures result in high blood pressure and can lead to anemia.

The exposure of children occurs not only through the air but also through accidental or intentional eating of paint chips and contaminated food or water.

TOXIC AIR POLLUTANTS

The Clean Air Act of 1977 required that emission standards be imposed upon air pollutants considered hazardous because they have been found to increase illness or mortality. The complexities encountered in attempts to control pollutants by declaring them to be criteria pollutants and setting air-quality standards resulted in the choice of emission controls instead of air-quality standards for toxic materials. The EPA has listed 188 pollutants whose emissions must be reduced. Examples are benzene (gasoline), perchlorethylene (used in dry cleaning), and methylene chloride (a solvent and paint stripper), as well as toluene, dioxin, asbestos, cadmium, mercury, and chromium.

The effects of significant exposures to toxic pollutants may be cancer, neurological effects, damage to the immune system, and reproductive effects. The risk of cancer associated with exposure to toxic pollutants in the air for a population is calculated on the basis of two factors. One describes the potency of the air contaminant, the other the magnitude and duration of the exposure, which is commonly assumed to be a lifetime of 70 years. The potency of a hazardous material can be expressed as a unit risk value. The unit risk value for an air pollutant is the increased lifetime cancer risk occurring in a population in which all individuals are exposed continuously from birth (70 years). The following discussion is based on a relatively simple version of risk assessment compared to the more sophisticated methods that are now in use.

The unit risk values are used to compare the potency of carcinogens with each other and to make crude estimates of the risk to populations whose exposures are known or assumed. The unit risk values are calculated so as to represent plausible upper bounds that are unlikely to be higher but could be appreciably lower. The units of unit risk values are $(\mu g/m^3)^{-1}$. The product of the unit risk value and the ambient concentration is the individual risk, and the product with the population exposed is the aggregate risk. Division of the individual or aggregate risk by 70 results in the corresponding annual risks. The maximum average concentration of the hazardous material in the ambient atmosphere is used in order to be conservative. Thus, if a maximum value of cadmium in the atmosphere in the vicinity of a copper smelter is $0.3\ \mu g/m^3$, and the unit risk value of cadmium is $2.3 \times 10^{-3}\ (\mu g/m^3)^{-1}$, then the probability of cancer (i.e., the maximum individual risk from the inhalation of cadmium) is

$$2.3 \times 10^{-3} \times 0.3 = 0.69 \times 10^{-3}$$

and the aggregate risk is

$$1000/70 = 14.3$$

Risk assessment has become increasingly important but also more complex as the basis for the management of exposures. The EPA issues guidelines for assessing the risks of carcinogens, mutagens, developmental toxicants, and chemical mixtures together with guidelines for estimating exposures. The Integrated Risk Information System (IRIS) is an electronic database maintained by EPA that contains information on the human health effects that can result from exposure to hazardous pollutants. The EPA provides telephone, fax, and e-mail contacts for obtaining information about hazardous pollutants.

DIOXIN

The term "dioxin" refers to a group of compounds that cause similar adverse health effects. They belong to three classes of chemicals: chlorinated dibenzo-p-dioxins (CDDs), chlorinated dibenzofurans (CDFs), and polychlorinated biphenyls (PCBs). Studies to date indicate that the compound 2,3,7, 8-tetrachlorodibenzo-p-dioxin (TCDD) is the more toxic substance. CDDs and CDFs are not created on purpose but result as by-products of certain activities; PCBs were produced for use in transformers and other purposes, but their use has now been prohibited. Combustion of certain materials, chlorine bleaching of pulp and paper, and certain chemical manufacturing processes all may create small amounts of dioxins.

Dioxins are characterized as likely human carcinogens, with the compound TCDD considered a human carcinogen on the basis of available human and animal data. The cancer risk to the population from exposures to dioxins is estimated to be 1 in 1000, with the likelihood that the risk may be much lower. Adverse health effects have been associated with personnel exposed to Agent Orange in Vietnam because of its dioxin content. Based upon available data, there is no clear indication that the general population is suffering health diseases from exposure to dioxins.

INDOOR AIR

Indoor air quality became important to those responsible for protection against adverse health effects caused by the inhalation of pollutants when it was realized that most individuals spend 90% of their time indoors and that indoor air quality is deteriorated by a large variety of sources. Four organizations—the American Lung Association, EPA, Consumer Product Safety Commission, and American Medical Association—prepared a document titled *Indoor Air Pollution* in 1989 that presents a summary of information for health professionals about the causes and effects of indoor air pollution. Figure 4 from this document provides an overview of the effects of air pollutants and their causes.

From a practical standpoint, the most important factor in the control of indoor air pollution is the quality of the ventilation of occupied space. The reduction of energy costs by cutting down on forced ventilation can lead to "sick building syndrome," the term applied to outbreaks of complaints as a result of poorly ventilated indoor spaces. The National Institute for Occupational Safety and Health has investigated many cases of indoor air-quality health hazards and has published guidelines for such investigations.

In certain cases, air-quality standards are met outdoors but not indoors. For example, an investigator who measured indoor versus outdoor levels of suspended particulate matter found that he spent 84% of his time indoors, and that 82.3% of his exposure was attributable to indoor air. The average indoor levels of nitrogen dioxide of 95 homes in rural Wisconsin was higher than the outdoor level, sometimes exceeding the ambient air-quality standard.

SECONDHAND SMOKE

The mixture of combustion products from the burning end of tobacco products and the smoke exhaled by smokers is referred to as "environmental tobacco smoke" or "secondhand smoke." It contains more than 4,000 chemicals, more than 50 of which are cancer-causing agents. It is associated with an increase in lung cancer and coronary heart disease and is particularly dangerous to the not yet fully developed lungs of young children, increasing their risk for sudden infant death syndrome, asthma, bronchitis, and pneumonia. An estimated 3,000 lung-cancer deaths and 35,000 coronary-heart-disease deaths occur annually among adult nonsmokers in the United States as a result of exposure to secondhand smoke. In children it is estimated that 8,000 to 26,000 new asthma cases and 15,000 to 300,000 new cases of bronchitis and pneumonia for those less than 18 months are the result of inhaling secondhand smoke.

INDOOR RADON LEVELS

Next to cigarette smoking, the inhalation of radon gas and the products of its radioactive disintegration are considered the most significant cause of lung cancer. The EPA has estimated that 20,000 of the lung-cancer deaths expected annually can be ascribed to radon, and the surgeon general has attributed 85% of lung-cancer deaths to smoking.

Radon-222, an odorless, colorless radioactive gas, is one of the products in the chain of decay of elements starting with uranium-238 in the soil, which after radon goes on to produce polonium isotopes 218 and 214. Their alpha-particle emissions dissipate their energy while destroying lung tissue,

Signs and Symptoms	Environmental Tobacco Smoke	Other Combustion Products	Biological Pollutants	Volatile Organics	Heavy Metals	Sick Bldg. Syndrome
Respiratory						
Rhinitis, nasal congestion	■	■	■	■		■
Epistaxis				■[1]		
Pharyngitis, cough	■	■	■	■		■
Wheezing, worsening asthma	■	■		■		■
Dyspnea	■[2]		■			■
Severe lung disease						■[3]
Other						
Conjunctival irritation	■	■	■	■		■
Headache or dizziness	■	■	■	■	■	■
Lethargy, fatigue, malaise		■[4]	■[5]	■	■	■
Nausea, vomiting, anorexia		■[4]	■	■	■	
Cognitive impairment, personality change		■[4]		■	■	■
Rashes			■	■	■	
Fever, chills			■[6]		■	
Tachycardia		■[4]			■	
Retinal hemorrhage		■[4]				
Myalgia				■[5]		■
Hearing loss				■		

1. Associated especially with formaldehyde. 2. In asthma. 3. Hypersensitivity pneumonitis, Legionnaires' disease. 4. Particularly associated with high CO levels.
5. Hypersensitivity pneumonitis, humidifier fever. 6. With marked hypersensitivity reactions and Legionnaires' disease.

FIGURE 4 Diagnostic quick reference. From American Lung Association, EPA, Consumer Product Safety Commission, and American Medical Association, *Indoor Air Pollution* (1989).

which leads to lung cancer. The radioactivity attributable to radon in the air is measured in picocuries per liter, which correspond to two atoms decaying per liter per minute. The concentration of the decay products is measured in "working level units," where 1 working level unit of decay products is released from approximately 200 pCi/l of radon.

The only certain method of determining the presence of radon is by testing, which should be performed whenever a dwelling is brought or sold. The EPA has prepared the "Home Buyer's and Seller's Guide to Radon," which provides information about what can be done to protect against this problem.

BIOLOGICAL CONTAMINANTS

The New York Academy of Medicine issued a resolution in 1983 (Health Aspects of Indoor Air, New York Academy of Medicine [1981] 57, No.10) expressing its concern about the damage to health caused by the inhalation of biologic agents that points out that "by far the most important substances in indoor air that affect human health are infectious agents, primarily viruses and bacteria in the form of aerosols or as part of droplets or particles. These cause more than 60,000 deaths and 250 million disabling illnesses in the United States each year." In addition to viruses and bacteria, the indoor air can be contaminated with pollens, fungal sores, algae, amoebas, actinomycetales, arthropod fragments, and droppings and dander from humans and animals. The airborne transfer of disease can involve pathogenic, toxicogenic, or allergenic agents via short-range direct person-to-person transfer or long-range dispersion throughout rooms sharing a common ventilation system. An important preventive strategy is to isolate other rooms from contamination by the air from rooms containing sources of microbial pollution.

Secondary sources of biological contamination can be established indoors by viable organisms that find friendly environments and sources of nourishment in soils, plants, and stagnant water. Pathogens, toxins, and allergens can be brought indoors through air intakes and shoes, clothing, or tools. The human body—mainly from the nose and mouth but also from other parts of the body—is a primary source of biological air contaminants and of nourishment for the growth of microorganisms. For example, it has been reported that the body sheds skin scales at a rate, dependent upon activity level, that averages 7 million scales per minute with an average of four bacteria per scale. Dust mites, dermatophagoides, feed on these scales and in turn produce fecal pellets in the respirable size range that can strongly affect sensitive individuals.

Almost all surfaces containing organic material such as cloth fabrics, paper, wood, leather, adhesive ceiling tiles, paint, soaps, and greases under proper conditions can sustain the growth of fungi, bacteria, acarids, and other microbes. Locations indoors where water can stagnate or cause continuous dampness of surfaces—for example, chilled-water air-conditioning systems, refrigerator drip pans, bathrooms, flooded basements, hot tubs, and saunas—are possible sources of biological pollutants.

Aerosol formation originating from water reservoirs is an important mechanism for airborne transmission of disease. The "jet-drop" phenomenon that occurs in washbasins, bathtubs, toilet bowls, and urinals can generate aerosolized microorganisms, as can the high-speed water-cooled drills used by dentists. Humidifiers that generate steam by other than thermal means can be dangerous sources of microorganisms that have grown inside the reservoirs; therefore, they should be sterilized before use and will always form droplets containing the minerals in the water being used. Filters in ventilating systems, if not properly cleaned and maintained, can also become significant sources of infectious diseases.

Some organisms are not able to survive for long periods outside a host and thus require rapid person-to-person transfer to cause serious problems, but others can remain viable for long periods and thus carry the diseases that are recognized as transmittable through air. The following are some of the major diseases that are classified as human airborne infections.

Tuberculosis
Rheumatic fever
Scarlet fever
German measles
Smallpox
Chicken pox
Hemolytic streptococci
Pneumococcus pneumonia
Mycoplasma pneumonia
Influenza
Diphtheria
Meningococcus meningitis
Psittacosis
Measles
Whooping cough
Poliomyelitis
Systemic mycosis
Mumps
The common cold

Certain lesser-known diseases have been brought into focus by investigators of illnesses attributed to indoor air pollution. Some of these have been associated with contaminated water reservoirs in humidifiers, chilled-water air-conditioning systems, and other adjuncts to indoor ventilation and climatization.

The most notable infection is Legionnaires' disease, which has caused a number of epidemics, including the one at the Legionnaires Convention in Philadelphia in 1977, which resulted in 182 cases, 25 of which were fatal. The infection is caused by a not-uncommon soil bacterium, *Legionella pneumophilia,* which survives very well in water and can thus contaminate cooling towers, air conditioners, and other potential sources of aerosols. Drinking water in a hospital has also been implicated in the transmission of this infection. Since its discovery, studies have claimed that 10 to 15% of pneumonia cases in hospitals are attributable to legionella.

A milder, self-limiting disease with flulike symptoms called Pontiac fever is also attributed to legionella.

Tuberculosis is caused by a virulent microorganism, mycobacterium tuberculosis, which can survive for long periods is encapsulated form and has been shown to be readily transmitted through ventilation systems. The gradual disappearance of this disease has been reversed in areas where AIDS is prevalent, therefore, its importance as a public-health problem is increasing.

Hypersensitivity pneumonitis and humidifier fever also are associated with aerosols from contaminated water. The former, which is the more serious, is an interstitial lung disease that in its acute form causes fever, chills, cough, and dyspnea 4 to 6 hours after exposure. The more commonly blamed agents are the thermoactinomycetes or the thermophylic micropolyspora. Humidifier fever is self-limited, with symptoms resembling the flu.

FUNGAL DISEASES

Fungi can cause disease and can also generate highly potent toxins. Treating the diseases is a special problem because available drugs are few in number because of the difficulty of finding ones that will preferentially attack fungal cells in the presence of mammalian cells.

Histoplasma capsulatum and other molds and yeast organisms such as blastomyces, cryptococcus, and coccidiodes are pathogenic to humans. The latter is responsible for valley fever, which is causing concern in hot, dry areas such as the Southwest, where it thrives. Cryptococcus, often found in soil contaminated by pigeon droppings, is a killer of AIDS patients, attacking the central nervous system and causing meningitis. *Aspergillus fumigatus* is a common fungus that invades the lung. When the spores are stirred up and inhaled they can grow fungus balls in the lung. One condition is called “farmer’s lung,” but in hospitals where transplant or AIDS patients have suppressed immune systems, the fungus can produce severe infections that must be treated rapidly. Detection of the fungus is difficult because instead of using the bloodstream it attacks by traveling from tissue to tissue. *Aspergillus flavus* produces aflatoxin, one of the most potent carcinogens known.

As in the case of tuberculosis, immuno-compromised individuals are readily susceptible to the acquisition of disease via inhalation. *Nosocomial aspergillosis* from sources such as contaminated air conditioners should be a special concern in hospitals.

ASBESTOS

“Asbestos” is the generic term for silicate materials that occur in fibrous form. A fibrous form that is classified as a serpentine is called “chrysotile” and is the type of asbestos most common in the United States. The length and flexibility of its fibers allow it to be spun and woven. Other types of asbestos are amphiboles that include amosite, crocidolite, tremolite, antophyllite, and actionide. Their fibers do not lend themselves to weaving, but in general they have higher heat resistance than chrysotile. Asbestos is used in many products, but most of it is used in construction—in floor tiles, cements, roofing felts, and shingles—and was used in very large quantities to protect steel structures from weakening during fires.

Measurements made of asbestos concentrations in the ambient air of New York City led to its banning in 1971 in sprayed material, a ban rapidly adopted nationally and worldwide. Nevertheless, many tons of asbestos will have to be handled with great care when the many buildings that used it to protect their steel structure from fire are finally demolished.

A study by Selikoff et al. of mortality among 17,800 asbestos workers from 1967 to 1976 found 675 excess deaths from cancer out of a total of 995. The ratio of observed to expected deaths in such a population was 3.11 for all cancers, 4.60 for lung cancers, and greater than 1 for several other

TABLE 3
Relative sensitivity of plants to SO_2

	Crops	
Sensitive		Tolerant
Alfalfa	Sweet potatoes	Celery
Buckwheat	Lettuce	Corn
Barley	Beans	Onion
Cotton	Broccoli	Potato
Red clover	Squash	Cabbage
Oats	Wheat	—
Peas	Carrot	—
Spinach	Tomato	—
	Flowers	
Sensitive		Tolerant
Sweet pea	Sweet William	—
Four-o’clock	Gladiola	—
Cosmos	Tulip	—
Bachelor button	Violet	—
Aster	Zinnia	—
	Trees	
Sensitive		Tolerant
Trembling aspen	Alder	Cedar
Jack pine	Red pine	Citrus
White pine	Austrian pine	Maple
White birch	Hazel	Linden
Larch	Apple	—
Large-toothed aspen	Douglas fir	—
Willow	Ponderosa pine	—
	Garden plants	
Sensitive		Tolerant
Peas	Cucumber	Muskmelon
Rhubarb	Lettuce	Corn
Spinach	Radish	Onion
Swiss chard	Squash	Potato
Beans	Tomato	Cabbage
Beets	Cultivated mustard	Broccoli
Turnips	Kohlrabi	—
Carrot		—

types of cancer. They found 175 deaths from mesothelioma, a cancer rarely found in the general population, and 168 from asbestosis, a form of pneumoconiosis characterized by fibrosis of the lung that often continues after exposure ceases.

Cigarette smoking was found to have a strong multiplicative effect when combined with exposure to asbestos. Based upon a mortality rate of 1 for a nonsmoking, nonexposed comparison group, the mortality ratio for nonsmoking asbestos workers, nonexposed cigarette smokers, and workers exposed to asbestos who also smoked were 5, 11, and 53.

The methods of asbestos analysis differed in that occupational exposures were originally measured by counting the number of fibers longer than 5 μm in a given volume of air using optical microscopy. In the ambient air the submicroscopic fibers, called "fibrils," which are the ones that can penetrate deeply into the respiratory system, had to be measured; therefore, the Environmental Science Laboratory, led by Irving Selikoff, at Mount Sinai School of Medicine developed an electron-microscopy method that measured the mass of fibrils per cubic meter of air.

Asbestos air concentrations are now reported in terms of nanograms per cubic meter—asbestos background concentrations are usually in the range of 0 to 10 ng/m^3.

Asbestos in schools has been a major concern because of the prevalence of its use and the nature of the population involved. A study of chrysotile asbestos in schools known to contain damaged asbestos insulation measured 9 to 1950 ng/m^3. Measurements in schools selected at random found a mean of 179 ng/m^3, compared to average outdoor concentrations of 6. The excess risk of death from lung cancer caused by asbestos exposure is proportional to the intensity of the exposure and its duration. One estimate of the cancer-incidence rate from asbestos reported by the EPA was that continuous exposure to 0.01 fibrils per ml (about 300 ng/m^3) will cause 28 mesotheliomas per million females and 19 per million males, and 5 excess lung cancers per million females and 17 per million males.

PLANT EFFECTS

The effects upon the plant surfaces of the contact of vegetation by suspended particulate matter via dry or wet deposition are not well documented. Direct entry of sulfur-dioxide gas molecules through the plant stomata, on the other hand, produces effects that are better understood and appear to depend, to a great extent, upon the rate of conversion to sulfate, a natural plant nutrient. As might be expected, the response of a plant to a particular exposure incident is dependent upon the concentration and duration of exposure and, because of the opening and closing patterns of stomata, also on the time in its daily cycle that exposures occur. Data on the injury threshold of 31 species of forest and agricultural plants were plotted to show their relative sensitivity to

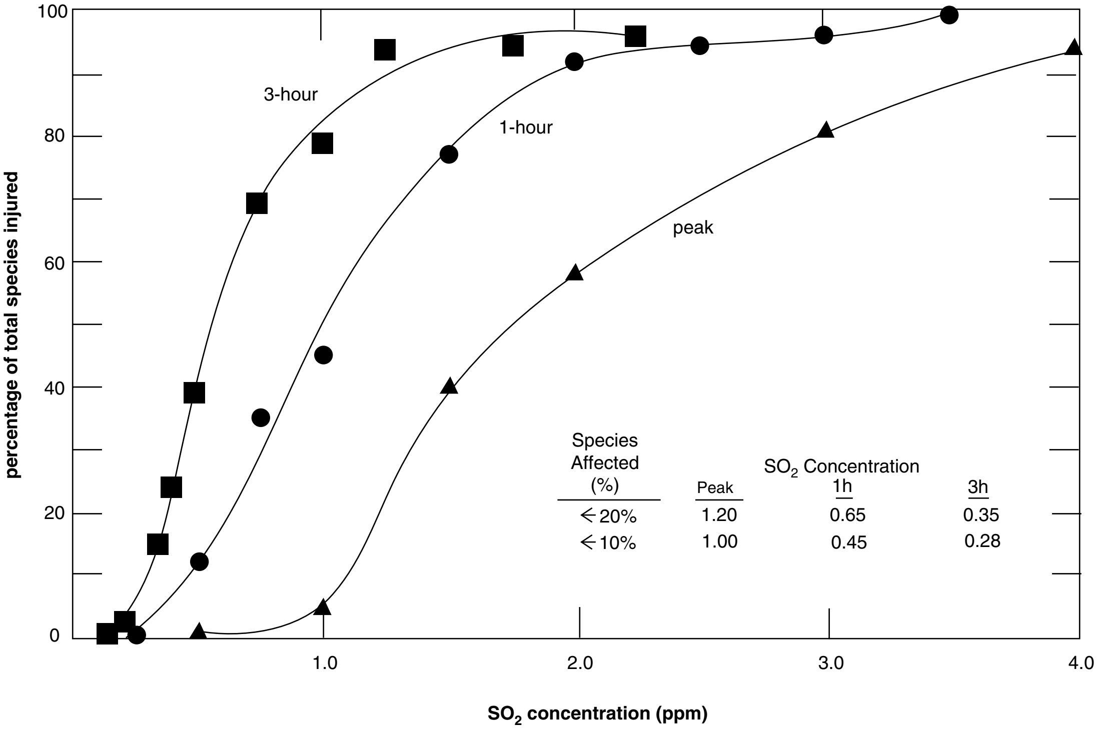


FIGURE 5 Percentage of plant species visibly injured as a function of peak 1-hour and 3-hour SO$_2$ concentrations. *Source:* McLaughlin and Lee (1974); McLaughlin (1980). (Air Quality Criteria for Particulate Matter and Sulfur Oxide, 1981.)

sulfur dioxide (Figure 5). The relative sensitivity of specific plants to SO_2 is indicated in Table 3.

REFERENCES

Assessing Levels of Airborne Bioallergens in New York City, *Environmental Health Perspective* (1981), **37:** 171–178.

Carbon Monoxide (1984)—Addendum to the 1979 Air Quality Criteria Document for Carbon Monoxide—Revised Evaluation of Health Effects Associated with Carbon Monoxide Exposure, EPA-600/8–83–033F.

EPA OAR Publications Annual Report and Basic Facts.

Fisher, D.C., Polluted Coastal Waters: The Role of Acid Rain, Environmental Defense Fund Report (1988).

A Guide to Indoor Air Quality (1995), EPA-402-K-93–007.

Health Aspects of Indoor Air, *New York Academy of Medicine* (1981), **57,** No. 10.

Health Effects of Acid Precipitation, Report of NIEHS Conference in Environmental Health Perspectives (Nov. 1984).

Lead (1986)—Air Quality Criteria Document for Lead II, EPA-600/8–83/028bF.

National Air Toxics Information Clearinghouse, Lung Cancer Risk from Passive Smoking, *Environment International* (1985), **11:** 3–22.

National Research Council, Committee on Indoor Pollutants, *Indoor Pollutants* (1981), Washington, D.C., National Academy Press.

National Research Council, *Carbon Monoxide* (1977); *Ozone* (1977); *Lead* (1972); *Nitrogen Oxides* (1977); *Polycyclic Organic Matter* (1972), Washington, D.C., National Academy Press.

Nitrogen Oxides (1982)—Air Quality Criteria Document for Oxides of Nitrogen, EPA-600/8–82.

The Plain English Guide to the Clean Air Act (1993), EPA-400-K-93–001.

Repace, J.L., Indoor Air Pollution, *Environment International* (1982), **8:** 21–36.

Repace, J.L. and Lourey, A.H., A Quantitative Estimate of Nonsmokers Lung Cancer Risk from Passive Smoking.

Samet, J.M., Marbury, M.C., and Spegler, J.D., Health Effects and Sources of Indoor Air Pollution, *Am. Rev. Respir. Dis.* (1988), **137:** Part I, 1486–1508, Part II, 221–242.

Second Addendum (1986)—Assessment of Newly Available Health Information (1986), EPA-600/8–86/02 OF.

Selikoff, et al. Asbestos Lung Disease—A Primer for Patients, Physicians and Lawyers.

Spengler, J., Hallowell, C., Moschandreas, D., and Fanger, O., Indoor Air Pollution, *Environment International* (1982), **8:** 1–534.

Stewart, R.D., et al., Experimental Human Exposure to Carbon Monoxide. *Archives of Environmental Health* (1970), 21:154–164.

Surgeon General National Health Advisory on Radon HHS 1/13/05.

Symposium on Health Aspects of Indoor Air Pollution, *Bull. N.Y. Acad. Med.* (1981), **57:** 825–1084.

Task Group on Lung Dynamics, Deposition and Retention Models for Internal Pasimetry of the Human Respiratory Tract, *Health Physics* (1966), **12:** 173–207.

USEPA Airborne Asbestos Health Assessment Update (1985), USEPA-600/8–84–003F.

USEPA Air Quality Criteria Documents Particulate Matter and Sulfur Oxides (1982), EPA-600/8–82–029 AF-CF.

USEPA Environmental Progress and Challenges: EPA's Update (1988), USEPA-230–07–88–033.

USEPA EPA Indoor Air Quality Implementation Plan (1987), USEPA-600/8–87/014, Appendix A, Preliminary Indoor Air Pollution Information Assessment.

Waldbott, G.L., *Health Effects on Environmental Pollutants* (1973), St. Louis, C.V. Masby Co.

Waloh, M.P., Car Lines (2004), No.6.

World Watch Institute Annual Reports and Featured Publications.

EDWARD F. FERRAND
Edward F. Ferrand Associates

AIR POLLUTANT REMOVAL: see ELECTROSTATIC PRECIPITATION; INDOOR AIR POLLUTION; PARTICULATE REMOVAL; MOBILE SOURCE POLLUTION

AIR POLLUTION INSTRUMENTATION: see also STACK SAMPLING; REMOTE SENSING

AIR POLLUTION INSTRUMENTATION

Advances in pollutant monitoring techniques and instrumentation have progressed significantly in the past several years to the point where noteworthy confidence in instrumentation-generated data has been established. Most systems, whether manual or automatic, have been designed to determine the quality of pollutants present per unit volume of air. These determinations can be made either in a duct/stack or in the outside atmosphere for the purpose of either ensuring adherence to published air quality regulations or for defining air pollution control device design criteria.

Air pollution instrumentation is utilized in the two major air quality areas of source and ambient monitoring with a further breakdown in the source monitoring category to the manual, and continuous type of instrumentation. Each of the three categories presents various types of monitoring problems which must be overcome, and each program is performed for a different reason.

MANUAL MONITORING—SOURCE INSTRUMENTATION

Instrumentation is utilized in this type of monitoring to provide a means to determine compliance with existing regulations and for developing design criteria for control of air pollution. The systems must be durable and allow for a certain amount of flexibility due to the many different situations which are encountered during the evaluations of stationary sources. This is emphasized by the fact that there are more than 25 different Environmental Protection Agency (EPA) sampling methods which, all except Method 9, require some type of instrumentation to complete the evaluation. (See *Stack Sampling.*) The most familiar procedure, "Method 5," has been the basis for the development of most of the other procedures. Method 5, shown in Figure 1, utilizes the basic equipment and sampling procedures (i.e., impingers, dry gas meter, pump, etc.) which with minor modifications or additions can be utilized to perform other EPA methods. As an example, Method 17 shown in Figure 2, utilizes the same components from the EPA 5 "sampling train" with a change in the filter location, and Method 8 for sulfuric acid mists presents even other changes. The EPA methodology and equipment were developed specifically to improve the accuracy of the source testing programs, and supplement the "ASME Train" (the "ASME" or American Society of Mechanical Engineers Train was developed to allow the determination of stack gas concentrations primarily in the electric utility area). The major reason for the testing methodology change (ASME to EPA) was to improve the capture efficiency of smaller (submicron) particles, and allow for gaseous concentration determinations. Due to thimble porosity the alundum thimble utilized in the "ASME Train" did not allow for the determination of representative data when sampling after a control device for either the determination of regulatory compliance or control efficiency. In addition, the use of an entirely glass-lined sampling system ("EPA Five Train") minimized the introduction of outside contamination to the samples being collected.

Such theories as isokinetic sampling for particulates ≥ 1.0 micron and proportional testing for gaseous determinations remain unchallenged. Only the equipment for manual monitoring has changed and will continue to change as improvements are made.

Manual source testing (stack sampling) equipment is available through several manufacturers, some of which are presented in the "Product Line Profiles" developed by *Pollution Equipment News*.

CONTINUOUS MONITORING—SOURCE INSTRUMENTATION

There are two basic types of source emission monitoring instrumentation—opacity monitors and gaseous emission monitors. Opacity monitors, which measure the transmittance of light through the gas stream, are called transmissometers, and are of two basic types; single-pass systems and double pass systems. Gaseous emission monitors on the other hand, are utilized to measure the concentration of specific gaseous components of the exhaust gas, such as O_2, CO, SO_2, NO_2, NO_x, etc. These instruments are divided into three basic types: extractive systems, *in-situ* systems, and remote systems. The *in-situ* systems can be further sub-divided into either in-stack or cross-stack monitors.

Opacity Monitors

Opacity monitors, also known as transmissometers, serve the purpose of providing information relative to combustion conditions, or control device efficiency. Transmissometers, although required by the EPA for various stationary sources

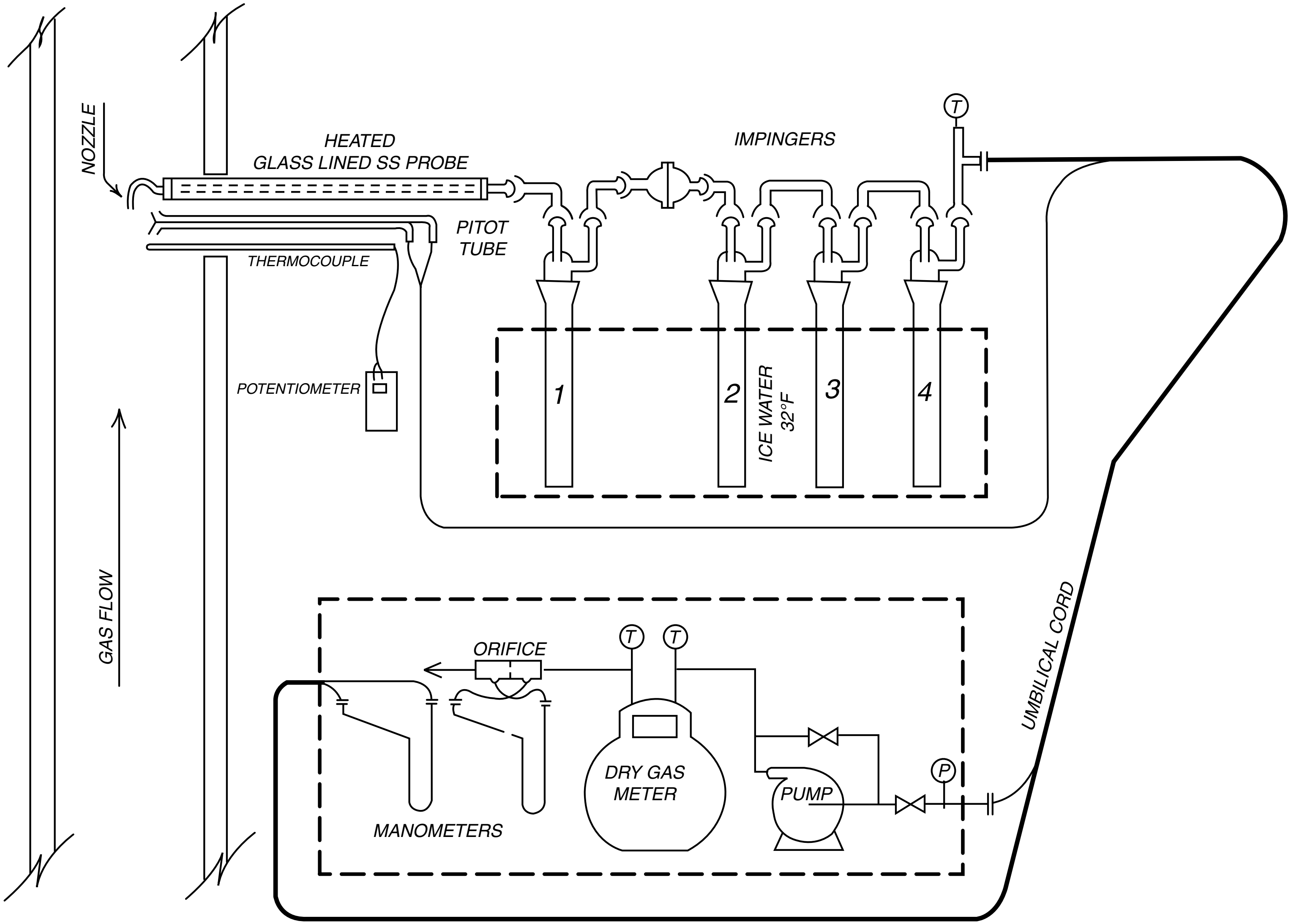
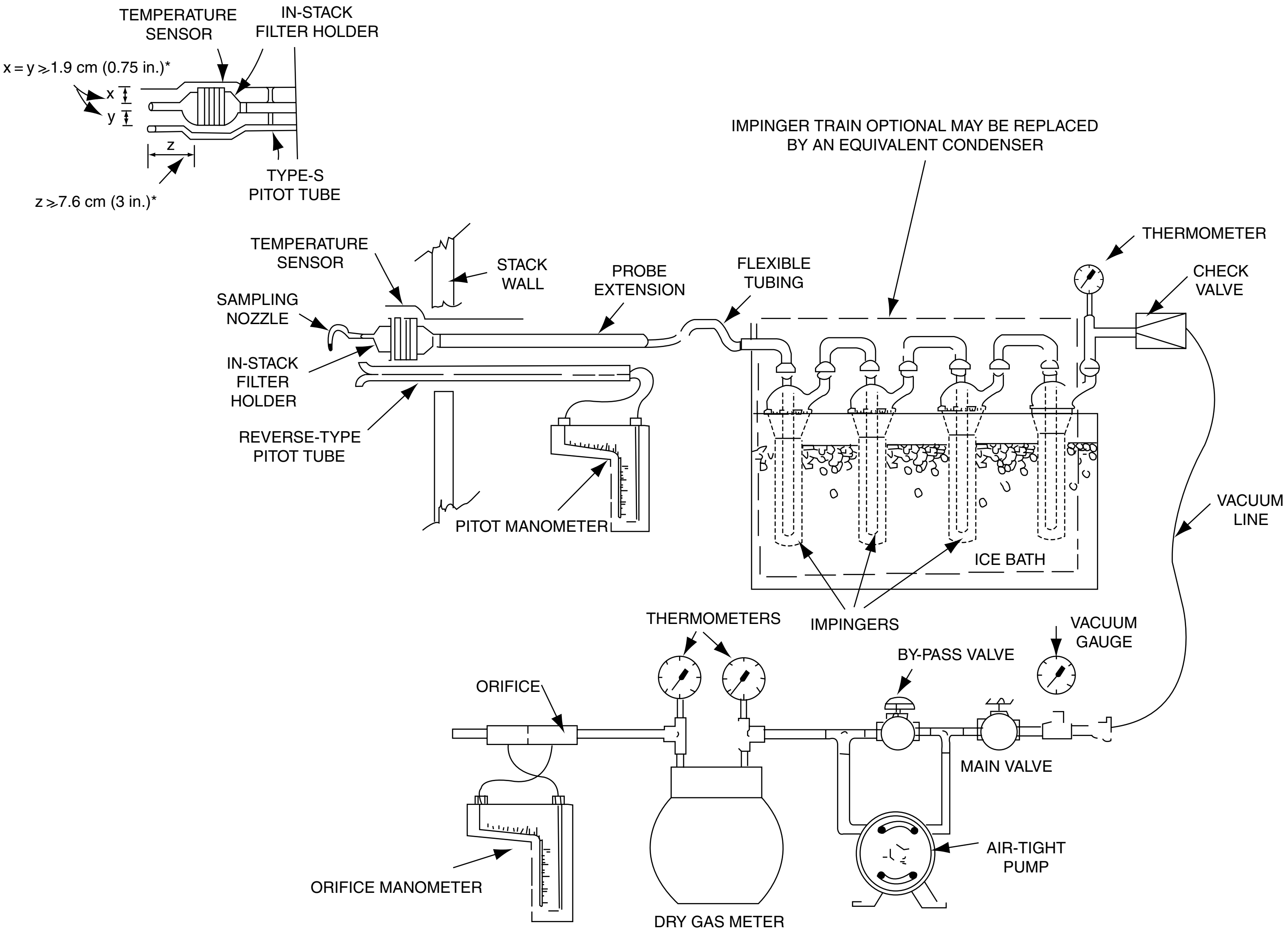


FIGURE 1 Method 5 schematic.



* SUGGESTED (INTERFERENCE-FREE) SPACINGS

FIGURE 2 Method 17 schematic.

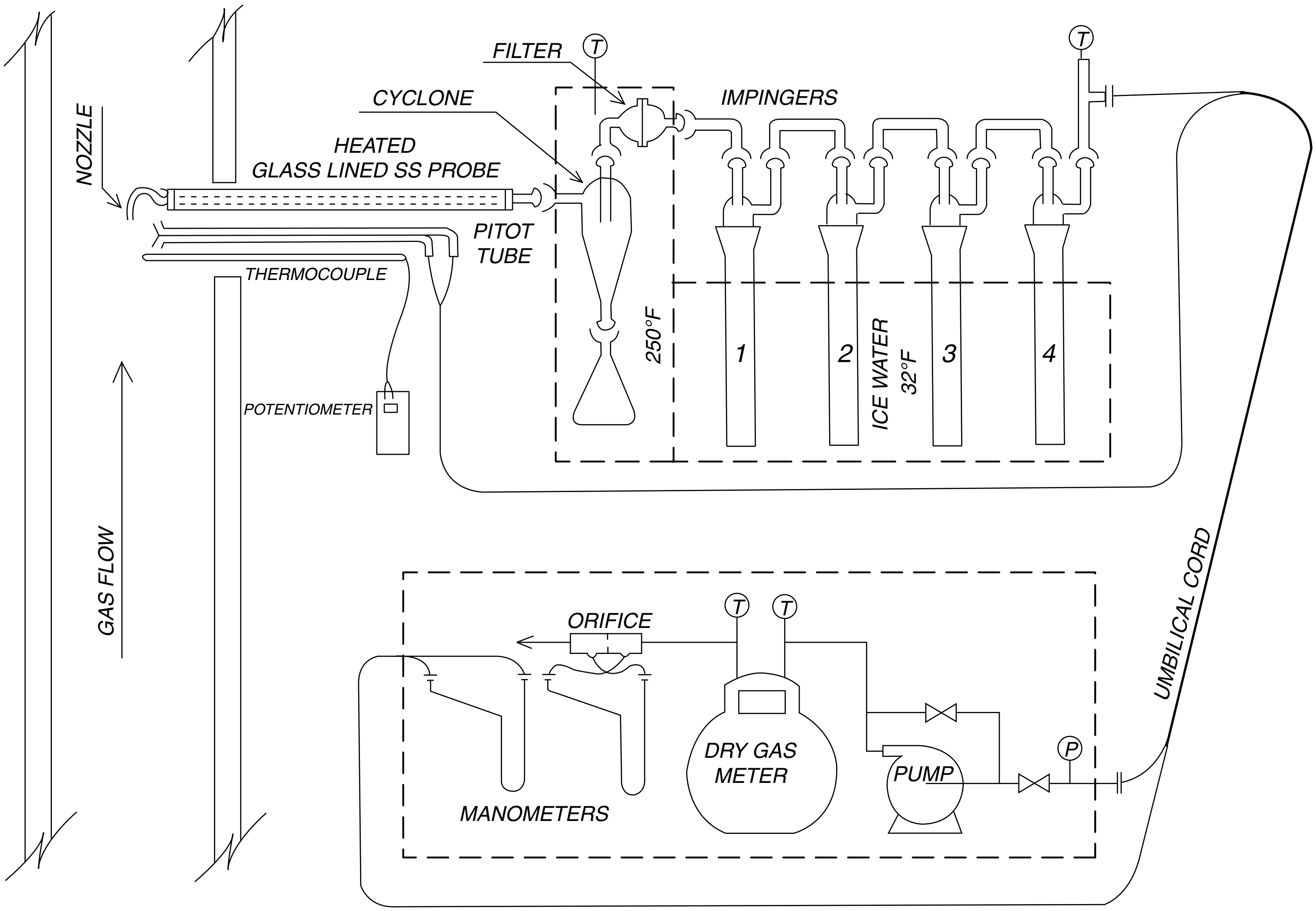


FIGURE 3 Method 8 schematic.

for the above-described purpose, are not considered to be enforcement tools. This requirement is fulfilled by the EPA certified Visual Emissions Observer as is specified in EPA Reference Method 9.

There are two basic types of transmissometers, single-pass and double-pass systems. The single-pass system incorporates a light source on one side of the stack and a detector on the opposite side. Although this is the more economical of the two systems, it does not meet the EPA requirements for system zero and calibration checks without complete process shutdown every 24-hours. It is better applied in a situation where direct compliance with the EPA criteria is not a factor, such as process control or baghouse filter bag breakage detection.

The double-pass system houses both the light source and detector with attendant calibration and zero-check instrumentation on the same side of the stack with only a reflecting mirror on the opposite side. Therefore, most of the double-pass systems satisfy the EPA design criteria. Refer to Table 1 for a list of vendors of either single-pass or double-pass transmissometers. The fraction of light lost in crossing the stack is used to calculate opacity and its value is related to the amount of dust or smoke passing through the light path. The cost per unit including control options is about $20,000–40,000 (1996$). The lower figure is for a quantity of more than 30 units; the higher figure is for a single installation.

An acid dew point meter is a related instrument produced by Land Combustion (see address above). It is useful in estimating SO_3/H_2SO_4 concentration.

Gaseous Emissions Monitoring

Stationary sources that are required by the EPA to install a continuous gaseous emissions monitor must match their specific process, and source emissions to the capabilities of the continuous monitor types available. Most instrumentation will fall into two categories, extractive systems and *in-situ* systems. A third category, remote monitors, utilizes concepts such as lasers and advanced spectroscopic methods to monitor gaseous emissions at distances from 500 to 100 meters away from the source.

EXTRACTIVE MONITORS

The basic principle behind an extractive monitor is the withdrawal of a gas sample from the main exhaust stream into the analyzer. This withdrawal must be conducted such that a representative sample is selected, and then appropriate interferents (particulates, water vapor, etc.) must be removed dependent upon analytical methodology. Extractive monitor types can be subdivided into three general categories: absorption spectrometers, luminescence analyzers, and electroanalytical monitors. Specialized extractive methods that do not fit into these three categories include paramagnetism and thermal conductivity.

Absorption Spectrometers

Spectroscopic analyzers utilized as continuous emissions monitors include two basic types: non-dispersive infrared analyzers (NDIR), and non-dispersive ultraviolet analyzers (NDUV).

NDIR detectors can monitor SO_2, NO_x, CO, CO_2 and hydrocarbons. As the gas travels through the instrument and is exposed to the infrared light source, light energy absorption occurs which is subsequently detected in comparison with a reference gas. Different gases are characterized by differing absorption characteristics, and are thereby identified and quantified.

NDUV detectors are used primarily to monitor SO_2 and NO_2. These instruments use light in the ultraviolet and visible portions of the spectrum. They are similar to NDIR monitors except that they do not use a reference gas for comparison. Instead, they use a reference wavelength with minimal absorption capabilities. NDUV analysis, also known as differential absorption, is also utilized in *in-situ* and remote sensing systems.

Luminescence Analyzers

Luminescence analyzers measure the emission of light from an excited molecule. Dependent on the mode of molecule excitement, molecules can exhibit photoluminescence (fluorescence), chemiluminescence or flame luminescence. Fluorescence occurs when a molecule is excited by light energy of a given wavelength, and light energy of a second wavelength is emitted. Fluorescence analyzers are utilized for SO_2 analysis.

Chemiluminescence analyzers are used for NO_x and NO_2 determinations, and operate on the principle of the emission of light energy resulting from a chemical reaction. In the case of chemiluminescence analyzers, the reaction involves ozone (O_3) and nitric oxide (NO).

Flame photometric analyzers use the principle of luminescence through molecule/flame interaction. These analyzers detect sulfur compounds, and are specific to sulfur alone.

Electroanalytical Monitors

Four distinct types of electroanalytical monitors are used in continuous source monitoring. These instruments rely on the methods of polarography, electrocatalysis, amperometric analysis, and conductivity.

Polarographic analyzers, also known as voltametric analyzers or electrochemical transducers, are capable of detecting SO_2, NO_2, CO, O_2, H_2S and other gases dependent on instrument setup. The analytical basis is a self-contained electrochemical cell in which a chemical reaction takes place involving the pollutant molecule. As a result of the chemical reaction, a current change through a specific electrode indicates pollutant concentration.

Electrocatalytic analyzers are utilized for O_2 determinations. These analyzers use a solid catalytic electrolyte and are available in both extractive and *in-situ* models.

TABLE 1
Continuous monitor equipment

Vendors of single-pass transmissometers		
Bailey Meter 29801 Euclid Avenue Wickliffe, (H44092)		Leeds & Northrup Sumneytown Pike North Wales, PA 19454
Cleveland Controls, Inc. 1111 Brookpark Road Cleveland, OH 44109		Photomation, Inc. 270 Polatis Avenue Mountain View, CA 94042
De-Tec-Tronic Corp. 2512 N. Halsted Street Chicago, IL 60614		Preferred Utilities Mfg. 11 South Street Danbury, CT 06810
Electronics Corp. of America 1 Memorial Drive Cambridge, MA 02142		Reliance Instr. Mfg. 164 Garibaldi Avenue Lodi, NJ 07644
HABCO 85 Nutmeg Lane Glastonbury, CN 06033		Robert H. Wager Passiac Avenue Chatham, NJ 07928
Vendors of double-pass transmissometers		
Environmental Data Corp. 608 Fig Avenue Monrovia, CA 91016		Land Combustion International 2525-B Pearl Buck Road Bristol, PA 19007
Lear Siegler, Inc. 74 Inverness Drive East Englewood, CO 80110		
Research Appliance Co. Chemed Corp. Route 8 Gibsonia, PA 15044	Same instrument	Contraves Goerz Corp. 301 Alpha Drive Pittsburgh, PA 15238
Dynatron, Inc. 57 State Street North Haven, CT 06473	Same instrument	Western Precipitation Div. Joy Manufacturing Co. PO Box 2744 Terminal Annex Los Angeles, CA 90051
Datatest, Inc. 1117 Cedar Avenue Croydon, PA 19020		

Amperometric analyzers, also called coulometric analyzers, measure the current in an electrochemical reaction. They are susceptible to various interferents; however, they are useful for SO_2, H_2S, and mercaptan analyses.

Conductimetric analyzers for SO_2 determinations measure the change in the electrical conductivity in water after a soluble substance is dissolved in it. This is a non-specific method, therefore interfering gases must be removed prior to introduction to the monitor.

TABLE 2
Lists of extractive instrument manufacturers

Fluorescence source analyzers

Thermo Electron Corporation
Environmental Instruments Div.
108 South Street
Hopkinton, MA 01748

Chemiluminescence analyzers

Beckman Instruments, Inc.
Process Instruments Division
2500 Harbor Boulevard
Fullerton, CA 92634

Bendix Corporation
Process Instruments Division
PO Drawer 831
Lewisburg, WV 24901

McMillan Electronics Corporation
7327 Ashcroft
Houston, TX 77036

Meloy Laboratories, Inc.
6715 Electronic Drive
Springfield, VA 22151

Monitor Labs
4202 Sorrento Valley Boulevard
San Diego, CA 92121

Scott Environmental Systems Division
Environmental Tectonics Corp.
County Line Industrial Park
Southampton, PA 18966

Source Gas Analyzers, Inc.
7251 Garden Grove Boulevard
Garden Grove, CA 92641

Thermo Electron Corporation
Environmental Instruments Div.
108 South Street
Hopkinton, MA 01748

Flame photometric analyzers

Tracor, Inc., Meloy Laboratories, Inc.
Analytical Inst.
1101 State Road
6500 Tracor Lane
Princeton, NJ 08540

Process Analyzers, Inc.
6715 Electronic Drive
Springfield, VA 22151

Polarographic analyzers

Dynasciences (Whitaker Corp.)
Township Line Road
Blue Bell, PA 19422

IBC/Berkeley Instruments
2700 DuPont Drive
Irvine, CA 92715

Western Precipitation Division
Joy Manufacturing Company
PO Box 2744 Terminal Annex
Los Angeles, CA 90051
(Portable models—not designed for continuous stack application)

Interscan Corp.
20620 Superior Street
Chatsworth, CA 91311

Theta Sensors, Inc.
Box 637
Altadena, CA 91001
(will provide systems)

Teledyne Analytical Instruments
333 West Mission Drive
San Gabriel, CA 91776
(O_2 only—micro-fuel cell)

TABLE 2 (*continued*)
Lists of extractive instrument manufacturers

Beckman Instruments, Inc. Process Instruments Division 2500 Harbor Boulevard Fullerton, CA 92634 (O_2 only)	Lynn Products Company 400 Boston Street Lynn, MA 01905 (O_2 only)
Gas Tech, Inc. Johnson Instrument Division 331 Fairchild Drive Mountain View, CA 94043 (O_2 only)	
Electrocatalytic oxygen analyzers	
Westinghouse Electric Corporation Computer and Instrument Division Orrville, OH 44667 (*in situ*)	Mine Safety Appliances Instrument Division 201 Penn Center Boulevard Pittsburgh, PA 15235 (extractive)
Lear Siegler, Inc. Environmental Technology Divisions Englewood, CO 80110 (*in situ*)	Thermox Instruments, Inc. 6592 Hamilton Avenue Pittsburgh, PA 15206
Dynatron, Inc. Barnes Industrial Park Wallingford, CT 06492	Cleveland Controls, Inc. 1111 Brookpark Road Cleveland, OH 44109
Teledyne Analytical Instruments 333 West Mission Drive San Gabriel, CA 91776	Corning Glass Works Ceramic Products Division Corning, NY 14803 (designed for glass furnaces)
Astro Resources Corp. Instrument Division PO Box 58159 Houston, TX 77573	Hays-Republic Milton Roy Company 4333 South Ohio Street Michigan City, IN 46360
Amperometric analyzers	
Barton ITT Process Instruments and Controls 580 Monterey Pass Road Monterey Park, CA 91754	International Ecology Systems 4432 North Ecology Systems Chicago, IL 60625 (combined colorimetric method)
NDIR monitors	
Positive filtering instruments	*Negative filtering instruments*
Beckman Instruments, Inc.	Bendix Corporation

(*continued*)

TABLE 2 (*continued*)
Lists of extractive instrument manufacturers

2500 Harbor Boulevard
Fullerton, CA 92634

Calibrated Instruments, Inc.
731 Saw Mill River Road
Ardsley, NY 10502

CEA Instruments (Peerless)
555 Madison Avenue
New York, NY 10022

Horiba Instruments, Inc.
1021 Duryea Avenue
Santa Ana, CA 92714

Infrared Industries
PO Box 989
Santa Barbara, CA 93102

Process Instrument Division
PO Drawer 831
Lewisburg, WV 24901

Esterline Angus
19 Rozel Road
Princeton, NJ 08540

Leeds & Northrop
Sumneytown Pike
North Wales, PA 19454

MSA Instrument Division
Mine Safety Appliances
201 Penn Center Boulevard
Pittsburgh, PA 15208

Teledyne-Analytical Instruments
333 West Mission Drive
PO Box 70
San Gabriel, CA 91176

Extractive differential absorption analyzers

Teledyne-Analytical Instruments
333 West Mission Drive
PO Box 70
San Gabriel, CA 91776

CEA Instruments
555 Madison Avenue
New York, NY 10022

Western Research and Development Ltd
Marketing Department
No. 3. 1313 44th Avenue NE
Calgary, Alberta T2E GL5

DuPont Company
Instrument Products
Scientific & Process Division
Wilmington, DE 19898

Esterline Angus
19 Rozel Road
Princeton, NJ 08540

Extractive Analyzers—Other Methods

There are a few special methods that do not fit into the three general classifications of absorption spectrometers, luminescence analyzers or electroanalytical methods. Paramagnetism is used in some O_2 analyzers, and thermal conductivity is used in some SO_2 continuous monitors.

Paramagnetic analyzers rely on the fact that O_2 molecules are attracted by a magnetic field, and this attraction can be quantified. However, it should be noted that NO and NO_2 are also paramagnetic, and in high enough concentrations can interfere in the analysis.

Thermal conductivity analyzers utilize a heated wire which undergoes resistance changes as gases flow over it. CO_2, SO_2 and other gases may be continuously monitored by thermal conductivity.

Please refer to Table 2 for list of some extractive instrument manufacturers.

IN-SITU ANALYZERS

Unlike an extractive monitor, an *in-situ* monitoring system will directly measure gas concentrations in the stack without

TABLE 3
Manufacturers of *in-situ* monitors

Cross-stack

Environmental Data Corporation
608 Fig Avenue
Monrovia, CA 91016

Contraves Goerz Corporation
610 Epsilon Drive
Pittsburgh, PA 15238

In-stack

Lear Siegler, Inc.
Environmental Technology Division
74 Inverness Drive East
Englewood, CO 80110

Oxygen monitors only

Westinghouse Electric Corporation
Computer and Instrument Division
Orville, OH 44667

Corning Glass Works
Ceramic Products Division
Corning, NY 14803

Dynatron, Inc.
Barnes Industrial Park
Wallingford, CT 06492

Hays-Republic
Milton Roy Company
4333 South Ohio Street
Michigan City, IN 46360

Cleveland Controls, Inc.
1111 Brookpart Road
Cleveland, OH 44109

modifying the flue gas composition. This can even be accomplished in the presence of particulate matter. Three techniques, differential absorption, gas filter correlation and second derivative spectroscopy, eliminate the problems associated with a reduction in light transmission due to the presence of particulates.

Two types of *in-situ* monitors exist: cross-stack and in stack. Cross-stack monitors, which can either be single-pass or double-pass systems (like transmissometers), measure the gas concentration across the entire, or a majority, of the stack diameter. In-stack systems (or short-path monitors) have a shorter path length of 5 centimeters to a meter.

In-Situ *Cross-Stack Analyzers*

Cross-stack analyzers use either the principle of differential absorption spectroscopy or gas-filter correlation spectroscopy. Differential absorption analyzers utilize a technique similar to that used by NDUV extractive analyzers; however, they operate *in-situ* and eliminate the particulate matter interference. CO_2, SO_2, and NO can be monitored in this manner.

Gas-filter correlation spectroscopy, used for CO, CO_2, SO_2, and NO analysis, is an NDIR *in-situ* method which, like the differential absorption technique, eliminates particulate interference.

In-Situ *In-Stack Analyzers*

In-stack analyzers utilize second-derivative spectroscopy to measure NH_3, SO_2, and NO concentrations. They are also known as in-stack point, or short-path monitors. Ultraviolet light is transmitted through the probe and the sensing area, to a reflector and back.

Please refer to Table 3 for a list of some of the manufacturers of the various *in-situ* instruments available.

AMBIENT INSTRUMENTATION

Ambient monitoring requires the use of instrumentation ranging in sophistication from the standard high volume particulate sampler to electronic systems incorporating several different gaseous detectors and data loggers all maintained in a temperature and humidity controlled remote sampling station. The reasons for performance of an ambient monitoring program are presented below:

1. Collection of background air quality data for preparation of air permits.
2. Verification of the reduction of specific air quality impacts resulting from emission control programs.
3. Verification of groundlevel downwind concentrations as determined by computer modeling.
4. To validate and refine models.

In each of the above cases, instrumentation is selected based on the requirements and the length of the program. Monitors are available for all parameters for which the National Ambient Air Quality Standards (NAAQS) have been set.

1. Suspended Particulates
2. Sulfur Dioxide
3. Oxides of Nitrogen (as NO_2)
4. Total Hydrocarbons (Non Methane)
5. Photochemical Oxidants
6. Carbon Monoxide

In each case, an EPA reference test method has been established and to monitor for any of the above, an instrument employing that approved reference method must be utilized. If a parameter is chosen and no reference method is available, then direct contact with the EPA or the governing regulatory body is highly recommended before proceeding with the monitoring program.

The identification of continuous Reference or Equivalent Methods for NAAQS parameters can be obtained by writing the Environmental Monitoring and Support Laboratory, Department E, US Environmental Protection Agency, Research Triangle Park, NC 27711.

In addition to the NAAQS parameters, selected meteorological data will also be required for the analysis and unless the proposed test site is near a National Climatic Recording Station, a meteorological recording station must also be established to collect at a minimum, wind speed and direction, humidity, and temperature. This can be accomplished through the use of a highly portable compact system or very sophisticated meteorological monitoring system employing such items as a 30 meter tower with monitors at different heights, daily weather balloon releases, etc.

Please refer to the “Product Line Profiles” included in this chapter for a brief description of the equipment provided by some of the equipment manufacturers.

PRODUCTION LINE PROFILES

Anderson Samplers, Inc. EPA Method 5 and ASME In-Stack Filter Method) stack sampler equipment is available. Full compliance, double pass opacity monitors are also available. A complete line of cascade impactors may be used for in-stack and ambient particle-sizing applications. Alarm systems are produced which detect faulty control equipment performance or broken bags in a baghouse. Anderson Samplers Inc., 4215-C Wendell Drive, Atlanta, GA 30336, 404-691-1910 or 800-241-6898.

BGI Inc. Samplers have been developed for airborne particulates, liquid droplet clouds, volatile gases and toxic materials. Cascade impactors are featured with four stages, 2 × 5 in. binderless glass fiber filters, wind vanes, suction pumps and still air adaptors. High volume air samplers are continuous or intermittent. Cascade centripeters accommodate flowrates of 30 l/min. Detector tubes have protective holders and continuous pumps. BGI Inc., 58 Guinan Street, Waltham, MA 02154, 617-891-8380.

Baird Corp. Instrumentation is available for measuring sodium concentrations in ambient air for gas turbine applications. Continuous Model LSM-30 ean detect and record ppm concentrations in gas turbine inlet air and fuel. Baird Corp., 125 Middlesex Turnpike, Bedford, MA 01730, 617-276-6131.

Beckman Instruments, Process Instrument Division Air quality monitoring instrumentaion includes a chemiluminescence O_3 analyzer, a fluorescence SO_2 analyzer, an infrared CO analyzer, and an NO_2 analyzer. These four instruments have received reference or equivalent method designation from the EPA. Each instrument is a totally self-contained system that offers precise interference-free measurement, convenient interface with peripherals, minimum expendables for ease of maintenance and choice of mounting options. Beckman Instruments PID, 2500 Harbor Blvd., Fullerton, CA 92634, 714-871-4848.

Bendix Corp. Chemiluminescent, chromatographic, infrared, and photometric apparatus may be specified for ambient and stack analysis of NO, NO_2, NO_x, CO, CO_2, SO_2, H_2S, TRS, THC, benzene, and other gases associated with the environmental and process industries. Bendix Environmental and Process Instrument Division, Box 831, Lewisburg, WV 24901, 304-647-4358.

Berkeley Controls Series 300 and 3300 semiportable continuous stack monitoring systems can be applied to source monitoring, ambient air monitoring, scrubber development, combustion studies, pollution research and OSHA standards. The cartridge sensor is an electrochemical membrane type polarographic gas detector. The 3300 series provides a complete integral sampling system consisting of filters, condensate removal and heat trace line controls. Berkeley Controls, 2825 Laguna Canyon Road, Laguna Beach, CA 92652, 714-494-9401 or 800-854-2436.

Byron Instruments Air quality instruments analyze hydrocarbons, organics, methane, carbon monoxide and carbon dioxide at source and ambient levels. The total emission analyzer oxidation/reduction system insures accurate readings. The analyzer gives part-per-million carbon readings on non-methane hydrocarbons from 10 ppmc full scale to 50,000 ppmc (5%) full scale. The instrument also provides readings on total hydrocarbons, methane, carbon monoxide, and carbon dioxide each analytical cycle. Byron Instruments Inc., 520-1/2 S. Harrington Street, Raleigh, NC 27601, 919-832-7501.

CEA Instruments Inc. The model 555 air monitor provides for the analysis of acrylonitrile, ammonia, bromine, chlorine, fluorine, formaldehyde, hydrazine, hydrogen chloride, hydrogen cyanide, hydrogen fluoride, hydrogen sulfide, nitrogen dioxide, oxides of nitrogen sulfur dioxide, and total oxidants. The RI 550 single channel IR analyzer provides for the analysis of CO, CO_2, methane, ethylene, ethane, propane, and butane or total hydrocarbons in the 0–2 up to 0–100% range, CEA Instruments Inc., 15 Charles Street, Westwood, NJ 07675, 201-664-2300.

Climatronics Corp. Equipment includes portable instruments and turn-key systems for meteorological testing. Units can be specified with sensors, data transmission and acquisition equipment, and upper air sounding systems. Climatronics Corp., 1324 Motor Parkway, Hauppauge, NY 11787, 516-234-2772.

Columbia Scientific Chemiluminescent laboratory and on-site automated apparatus measure NO, NO_2, NO_x, and ozone. Performance specifications are guaranteed over the range of +10 to +40°C. The equipment is capable of at least seven days of unattended operation for prolonged on-site monitoring. Columbia Scientific Inds., Box 9908, Austin, TX 78766, 412-258-5191 or 800-431-5003.

Contraves-Goerz Corp. Infrared electronic equipment can be selected for monitoring stationary source emissions. Total source control packages are offered for power generation and process industries. Portable and *in-situ* instruments are available. Contraves-Goerz Corp., 610 Epsilon Drive, Pittsburgh, PA 15238, 412-782-7700.

D and S Instruments Sampling cannisters are precleaned and prepared for part per trillion level sampling. Another product is a cryotrap suitable for EPA Method 25 sampling of volatile hydrocarbons. D and S Instruments Ltd., SE 1122 Latah Street, Pullman, WA 99163. 509-332-8577.

Daco Products Inc. Wet impingement samplers can be specified for gases, fumes, and vapors. Custom packages are available for sampling toxic substances. Colorimetric, electrochemical, photometric, and wet chemical analyzers are available. Daco Products Inc., 12 S. Mountain Avenue, Monclair, NJ 07042, 201-744-2453.

Datametrics Hot-wire anemometer-type air velocity and flow meters are used in conjunction with air sampling. The instrumentation is used to determine sample size. Other equipment analyzes air composition. Datametrics, 340 Fordham Road, Wilmington, MA 01887, 617-658-5410.

Datatest Inc. Instruments are designed for the continuous monitoring of particulate emissions. The instruments use photometric techniques and are *in-situ.* The emission is continuously recorded on a strip chart or circular recorder. Relay contacts are provided for controlling external equipment such as dampers and air flow valves. Datatest Inc., 1410 Elkins Avenue, Levittown, PA 19057, 215-943-0668.

Davis Instrument Manufacturing Co. Stack monitoring devices include optical probes for smoke density measurement and dust samplers for quantitative analysis of particulate emissions. Davis Instrument Manufacturing Company, Inc., 513 E. 36th Street, Baltimore, MD 21218, 301-243-4301.

Delta F. Corp. Factory calibrated oxygen analyzer is capable of monitoring oxygen in gas streams containing "acid" gases as well as combustibles. Trace and percent analyzers are available in panel mount configurations, battery operated models and remove sensor versions. Delta F Corp., One Walnut Hill Park, Woburn, 01801, 617-935-6536.

Digicolor Inc. Automatic and manual analyzers are available for the determination of ammonia, halogens, acidic sulfur, as well as most organic gases. Samplers may be specified as grab samplers, intermittent samplers, or continuous samplers. All have the option of filtration, or wet or dry impingement separation techniques. Metering is either rate or volume control. Digicolor Inc., 2770 E. Main Street, Columbus, OH 43209, 614-236-1213.

Dionex Corp. The ion chromatograph is used in a variety of air quality applications. Among these are ambient aerosols and SO_2 levels, carbon dioxide analysis, ammonia, sulfur species, halogens and nitrogen oxides in auto exhausts and other sources. Flue gas desulfurization analysis is also done by ion chromatograph. Toxicology applications include sulfate and oxalate ions in industrial environments, chloroacetyl chloride and formaldehyde at trace levels, and ambient levels of SO_2. The ion chromatograph is also extensively used in acid rain analysis. Dionex Corp., 1228 Titan Way, Sunnyvale, CA 94086, 408-737-0700.

Dupont Company Source monitoring equipment may be specified for the determination of SO_2, NO_x, H_2S, and ammonia as well as halogens and aromatics. All equipment features photometric detectors and has the ability to measure multiple sources. Dupont Company, Scientific and Process Instrument Division, Concord Plaza, Wilmington, DE 19898, 302-772-5500.

Dynasciences Continuous electrochemical apparatus may be used for EPA compliance monitoring for inspection and testing. Instrumentation monitors oxides of nitrogen, sulfur, and oxygen. Turn-key installations, as well as engineering assistance and field support are offered. Dynasciences Env. Products Division, Township Line Road, Blue Bell, PA 19422, 215-643-0250.

Dynatron Inc. Air pollution monitoring systems include a complete line of *in-situ* stack gas measurement and analysis equipment. Opacity monitoring systems offer digital displays, automatic EPA calibration, and direct optical density readout. Dynatron Inc., Box 745, Wallingford, CT 06492, 203-265-7121.

Edwards Engineering Hydrocarbon vapor analyzer is used for the continuous check of operation and emission percentage from hydrocarbon vapor recovery units. It is designed to mount directly within the vapor recovery control room. It features automatic replenishing of charcoal absorption chambers and a constant meter indicator with a strip chart recorder. Edwards Engineering Corp., 101 Alexander Avenue, Pompton Plains, NJ 07444, 201-835-2808.

Energetics Science Instruments are available for measurement of toxic gas, combustible gas/oxygen and oxygen deficiency in ambient air and in process control. Toxic gas capability includes the measurement of carbon monoxide, hydrogen sulfide, nitric oxide, nitrogen dioxide, hydrazine, and sulfur dioxide. The combustible gas detector is a catalytic filament type and the oxygen sensor uses a polarographic sensor. Energetics Science Inc., 85 Executive Blvd., Elmsford, NY 10523, 914-572-3010.

Enmet Corp. Monitors are offered for detecting dangerous levels of toxic or combustible bases. Monitors can be specified with meters as well as integral lights and audible alarms, with external signal capabilities, actuated when gas concentrations exceed safe levels. Portable O_2 deficiency detectors are available, featuring push-button checks for alarms and batteries. Automatic CO respiratory air line monitors may be specified for detection of concentrations as low as 10 ppm. Enmet Corp., 2308 S. Industrial Highway, Ann Aror, MI 48104, 313-761-1270.

Environmental Measurements Ambient air sampling systems are designed to collect and store pollutant and tracer gases in sampling bags for their subsequent analysis. Also available are heavy and regular duty Tedlar air sampling bags. These sampling bags are available in sizes from 0.5 to 300 liters. An automatic sequential or radon bag sampler designed for unattended gas collection may be programmed by the user to collect up to 8, 12, 16, or 24 hourly samples over a 96-hour period. This self-contained battery operated system is operated in real time and is designed for use in harsh environments. Environmental Measurements, 215 Leidesdorff Street, San Francisco, CA 94111, 408-734-0713.

GCA Environmental A fibrous aerosol monitor provides a real-time count of airborne fibers in the presence of large concentrations of nonfibrous particles. A respirable dust monitor tells how much respirable dust is in the air being breathed. A recording dust monitor is designed for short and extended time monitoring of respirable dust. GCA/Environmental Instruments, Burlington Road, Bedford, MA 01730, 617-275-9000.

Gastech Inc. Portable and fixed detectors and alarms measure combustible and toxic gases, oxygen deficiency, hydrocarbons, and airborne halogens. Calibration kits are also available for a variety of gases. Gastech, Inc., Johnson Instrument Div., 331 Fairchild Drive, Mountain View, CA 94043, 415-967-6794.

General Metal Works High-volume filtration samples feature continuous flow controllers and recorders, weatherproof housings, and stainless steel inlet tubing materials. Integrated packages can be ordered which comply with federal and state monitoring regulations. Related equipment which can be specified includes timers, flowmeters, impactors, and calibration units. General Metal Works Inc., 8368 Bridgetown Road, Cleves, OH 45002, 513-941-2229.

General Monitors Single-channel system for continuous monitoring of combustible gas concentrations consists of a remote sensing assembly and a solid state controller. Control electronics include an analog meter scaled from 0–100% LEL and four vertically arranged LEDs. The LEDs indicate high and low alarm status, normal operation and any malfunction conditions. A digital display from 0–99% LEL is one of many user-selected options available. Relay options include a choice of normally energized or de-energized high and low alarms, latching or non-latching high and low alarms, or sealed relays. General Monitors Inc., 3019 Enterprise Street, Costa Mesa, CA 92626, 714-540-4895.

Horiba Instruments Instruments and integrated single or multi-gas monitoring systems are for stack gas and ambient air applications. Sampling and continuous equipment is available. Instruments include NDIR gas analyzers free of interference from water vapor and carbon dioxide, and flame ionization analyzers for measuring total hydrocarbons. Systems are custom designed and may include remote computerized operation and automatic calibration. Horiba Instruments Inc., 1021 Duryea Avenue, Irvine, CA 92714, 714-540-7874.

Interscan Corporation Toxic gas monitors available from pocket alarm units and dosimeters to complete plant-scale multipoint systems. The line includes analyzers for CO, SO_2, H_2S, $NO/NO_2/NO_x$, and Cl_2. Additionally, systems which may use products manufactured by others, specified by the customer, are available. Examples of this are multi-gas and source monitoring systems. Interscan Corp., 9614 Cozycroft Avenue, Chatsworth, CA 91311, 213-882-2331.

Kurz Instruments Constant flow air samplers are produced for low volume air sampling, for sampling of organics or particulates. Flow ranges are available from 30 sccm to 150 slpm. They are mass flow controlled and referenced to EPA standards of 25°C and 760 mmHG. Vacuum capability as high as 20″ Hg and control accuracy of ±3% reading over a wide temperature range are standard. Higher sampling requirements are handled by the high volume air samplers, which sample from 20–60 scfm. Kurz Instruments Co., Box 849, Carmel Valley, CA 93924, 408-659-3421.

Lamotte Chemical Portable air quality sampling and measurement outfit provides individual tests for 14 major contaminants. Tests are provided for ammonia, bromine, cadmium, carbon monoxide, chlorine, cyanide, hydrogen peroxide, hydrogen sulfide, iodine, lead, nitrogen dioxide, ozone, phenols, and sulfur dioxide. The outfit features a portable air sampling pump with calibrated flowmeter. The absorbing solutions, which are contained in the impinger, collect the air to be tested. Driven by 4 standard "D" cell batteries, the pump will sample up to 2.5 liters per minute at 6 volts and is capable of maintaining flow rate for 48 hours of continuous sampling. An adjustable flow-meter regulates and indicates the air sampling rate. The sampling pump is furnished with flow-meter, impinger holder, batteries, and connection tubing. Lamotte Chemical Products Co., Box 329, Chestertown, MD 21620, 301-778-3100.

Mast Development Company Portable and online instruments measure ambient and work place TLV levels of ozone, chlorine, fluorine, bromine, and iodine. Appropriate calibration devices are also available, including automatic bubble meter for the determination of low air flow in devices using miniature pumps. Mast Development Co., 2212 E. 12th Street, Davenport, IA 52803, 319-326-0141.

Met One Equipment, systems, technical, and engineering assistance provided for complete meterological environmental monitoring, measuring, and control. Systems provide statistics on wind direction and velocity, ambient air temperatures, relative humidity, precipitation and solar radiation. Portable and permanent systems available. Met One, Box 60279, Sunnyvale, CA 94088, 408-733-9100.

Mine Safety Appliances Analyzers and sampling systems are supplied for the measurement of contaminants such as SO_2, NO_2, CO, CO_2, hydrocarbons, and oxygen. Applications include monitoring power plant stacks, metallurgical processes, combustion control, and solvent recovery beds. Mine Safety Appliance Company, 600 Penn Center Boulevard, Pittsburgh, PA 15235, 412-273-5101.

Monitor Labs Ambient air analyzers, calibrators, data loggers, telemetry systems, and computer-based monitoring networks are offered for ozone, sulfur dioxide, total sulfur, and oxides of nitrogen. Calibration sources for nitrogen dioxide and sulfur dioxide are supplied with certificates of traceability to NBS. Data loggers accept up to 20 parameters. Monitor Labs Inc., 10180 Scripps Ranch Boulevard, San Diego, CA 92131, 714-578-5060.

Napp Inc. Model 31 manual stack sampling system is lightweight, modular equipment designed for compliance testing and performance evaluation of industrial stack emissions. The molecular design allows selection of equipment for sampling all EPA Methods (1–17) except 7, 9, 14, and 16. A standard Method 7 system is also offered. Method 16 is constructed for individual applications. Napp. Inc., 8825 N. Lamar, Austin, TX 78753, 512-836-5110.

National Draeger Portable personnel monitors are used for the determinations of TLV levels of over 140 different gases and vapors. Grab sampling is available. National Draeger Inc., 401 Parkway View Drive, Pittsburgh, PA 15205, 412-787-1131.

Nutech Corp. Assay, chromatographic and wet chemical equipment may be used in the determinations of most organic gases, as well as oxides of nitrogen and solid and liquid particulates. Grab samplers can be specified for aerosols, gases or particulates. Nutech Corp., 2806 Cheek Road, Durham, NC 27704, 919-682-0402.

Pollution Measurement Corp. Non-absorbent sample bags of Tedler, Teflon or Mylar are available in eight sizes from 0.4 to 70 liter. Gas sample spheres are available in sizes from 0.5 to 14.5 liter with vacuums of 22 inches of mercury. Special packages are available for meeting EPA and OSHA requirements. Pollution Measurement Corp., Box 6182, Chicago, IL 60680, 312-383-7794.

Rader Company High-volume samplers measure solid particulates emitted from stacks and other stationary sources. Equipment is for manual or automatic operation, and can be specified with a variety of accessories. Rader Company Inc., Box 20128, Portland, OR 97220, 503-255-5330.

Research Appliance Company RAC designs, manufactures and supplies diversified lines of precision environmental instruments and laboratory apparatus. The product mix includes instruments and systems that sample/monitor ambient air and process emissions, laboratory and testing apparatus, certified reagents for wet chemical gas sampling/analyzing, meteorological indicating/recording instruments and a broad range of related accessories. Research Appliance Company, Moose Lodge Road, Cambridge, MD 21613, 301-228-9505.

Sierra Instruments Instruments are available for particulate sampling and size fractionating in ambient air quality monitoring, stack sampling, OSHA applications, and aerosol research. Instruments include dichotomous samplers, cascade impactors, cyclone samplers, flow-controlled high volume air samplers, flow-controlled low volume air samplers, hi-vol size selective inlets, and cotton-dust samplers. Sierra Instruments Inc., Box 909, Carmel Valley, CA 93924, 408-659-3177.

Sierra Misco Inc. Grab, intermittent, and continuous samplers are available for the sampling of aerosols, particulates and gases. Glass, stainless steel, and Teflon inlet tubing are also available. Samplers are AC or battery operated. Separation collection techniques include filtration, charcoal, and wet and dry impinging. Sierra Misco Inc., 1825 Eastshore Highway, Berkeley, CA 94710, 415-843-1282.

Teledyne Analytical Equipment is suitable for continuous interference free monitoring of such pollutants as H_2S, SO_2 and hydrocarbons. Analyzers are designed for permanent location and continuous operation with minimal maintenance. Teledyne Analytical, Box 70, San Gabriel. CA 91776, 213-576-1633. www.teledyne-api.com

Thermo Electron Air pollution monitoring instrumentation can be specified for NO_x and SO_2 in ambient air, stack gases, and automotive emissions. The chemiluminescence principle is used for NO_x, while SO_2 is determined by pulsed fluorescence. All instrumentation meets or exceeds federal and state performance requirements. Thermo Electron Corp., Environmental Instruments, 27 Forge Parkway, Franklin, MA 02038, USA Tel +1 (508) 520 0430, Toll free +1 (866) 282 0430.

Varian Associates Gas chromatographs are offered for research as well as monitoring applications, and may be specified with special options for total hydrocarbon, vinyl chloride, and ppb sulfur gas analyses. Atomic absorption devices, with optical microsamplers, are also available and are especially useful for measuring trace levels of metal pollutants. Varian Associates, Instruments Group, 611 Hansen Way, Palo Alto, CA 94303, 415-493-4000.

VICI Metronics H_2S detection system is based upon card mounted, sensitized pads that visibly change color when exposed to H_2S. Applications range from odor surveys and area wide transport studies to worker dosage monitoring and work area testing. VICI Metronics, 2991 Corvin Drive, Santa Clara, CA 95051, 408-737-0550.

Wedding & Associates Critical Flow Device A high volume sampler and volumetric flow controller is offered which meets federal standards of volumetric flow rate at ambient conditions. Size specific inlets such as PM_{10} systems employing fractionating devices whose performance depends on air velocity may experience substantial variations in sampler performance values if operated using mass flow controllers. Also, the value for total sampled volume of air used in the denominator when calculating ambient concentration levels will bear little resemblance to the actual volume sampled if the ambient sample does not utilize a volumetric flow controller.[9]

R. M. Young Company The portable recording wind set provides continuous analog chart records of wind speed and wind direction side by side on a single 6″ wide chart. The windvane and 3 cup anemometer are generally used where analog records of wind speed and wind direction are required. A wind run anemometer can be substituted where a record of total wind passage is desired. The propvane provides signal characteristics in the range of 0–10 mph. R. M. Young Company, 2801 AeroPark Drive, Traverse City, MI 49684, 916-946-3980.

For additional monitoring and testing sources, including those outside the US, the reader is referred to the environmental expert home page.[10]

REFERENCES

1. American Society of Mechanical Engineers, Determining dust concentration in a gas stream, Power Test Code Nos. 21 and 27. December 1941 and April 1957.
2. Environmental Protection Agency, Research Triangle Park, NC. Ambient monitoring guidelines for prevention of significant deterioration, (PSD), EPA 450/2-78-019.
3. Los Angeles Air Pollution Control District, Los Angeles, CA. *Source Sampling Manual*, November 1963.
4. Pollution Equipment News, Guide to selecting air quality monitoring and sampling equipment, June 1980.
5. Powals, Richard J., Zaner, Loren J., and Sporek, Karel F., *Handbook of Stack Sampling and Analysis,* 1978.
6. United States Environmental Protection Agency, Continuous air pollution source monitoring systems, EPA 625/6-79-005, June 1979.
7. United States Environmental Protection Agency, Standards of performance for new stationary sources, Title 40, Part 60, Federal Register, August 18, 1977.
8. United States Environmental Protection Agency, Industrial guide for air pollution control, EPA 625/6-78-004, June 1978.
9. Wedding, J.B., Weigand, M.A., Kim, Y.J., Swift, D.L., and Lodge, J.P., A critical flow device for accurate PM-10 sampling, *Jnl. of Air. Poll. Cont. Assoc.*, **37,** 254 (1988).
10. http://www.environmental-expert.com/air.htm (2005)

JAMES GEIGER
MARK D. MUELLER
Betz Converse Murdoch Inc.

AIR POLLUTION METEOROLOGY

EFFECTS OF WEATHER ON POLLUTION

Introduction

As the world's population and industrialization grow, air pollution (Figure 1) becomes a progressively more serious problem. The control of air pollution requires the involvement of scientists from many disciplines: physics, chemistry and mechanical engineering, meteorology, economics, and politics. The amount of control necessary depends on the results of medical and biological studies.

The state of the atmosphere affects, first, many types of pollution. For example, on a cold day, more fuel is used for space heating. Also, solar radiation, which is affected by cloudiness, has an influence as smog production. Second, atmospheric conditions determine the behavior of pollutants after they leave the source or sources until they reach receptors, such as people, animals, or plants. The question to be answered is: given the meteorological conditions, and the characteristics of the source or sources, what will be the concentration of the pollutants at any distance from the sources? The inverse question also is important for some applications: given a region of polluted air, where does the pollution originate?

Finally, the effect of the pollution on the receptor may depend on atmospheric conditions. For example, on a humid day, sulfur dioxide is more corrosive than on a dry day.

Meteorological information is needed in three general areas of air pollution control:

(1) In planning control measures, wind climatology is required. Pollution usually must be reduced to a point where the air quality is substantially better than the existing quality. In order to assure improved quality, certain standards are set which prescribe maximum concentrations of certain pollutants.

In order to reach such standards, the points of origin of the pollution must first be located; traditionally, everybody blames everybody else for the unsatisfactory air quality. Given possible pollution sources, tracing of air trajectories coupled with estimates of atmospheric dispersion will give the required answers. Once the relative importance of different pollution sources is known, strategies have to be developed to determine the degree to which each source must reduce its effluent.

FIGURE 1 Air pollution in New York City prior to SO_2 and particulate restriction.

The most economical way to cut concentration of some pollutant may not be to cut the effluent of each emitter by the same amount. In order to find the best strategy, city models must be constructed, separately for each pollutant and for different meteorological conditions, which show how the air pollution climate of an urban region is affected by the existing distribution of sources, and what change would be produced when certain sources are controlled. The construction of such models will be discussed later, and requires a fairly sophisticated handling of meteorological data. The same models then also help in planning future growth of housing and industry.

Of course, not all problems of air pollution meteorology are as complex as those involving urban areas. The planning of individual plants, for example, must be based in part on the air pollution to be expected from the plant under various atmospheric conditions; meteorological calculations may show whether expensive techniques for cleaning the effluent before leaving the stack may be required.

(2) Meteorological forecasts can be used to vary the effluent from day to day, or even within a 24 hour period. This is because at different times the atmosphere is able to disperse contaminants much better than at other times; purer fuels must be used, and operation of certain industries must be stopped completely in certain areas when the

mixing ability of the atmosphere is particularly bad.

(3) Meteorological factors have to be taken into account when evaluating air pollution control measures. For example, the air quality in a region many improve over a number of years—not as a result of abatement measures, but because of gradual changes in the weather characteristics. If the effects of the meteorological changes are not evaluated, efforts at abatement will be relaxed, with the result of unsupportable conditions when the weather patterns change again.

Effects Between Source and Receptor

The way in which the atmospheric characteristics affect the concentration of air pollutants after they leave the source can be divided conveniently into three parts:

(1) The effect on the "effective" emission height.
(2) The effect on transport of the pollutants.
(3) The effect on the dispersion of the pollutants.

Rise of Effluent

To begin with the problem of effluent rise, inversion layers limit the height and cause the effluent to spread out horizontally; in unstable air, the effluent theoretically keeps on rising indefinitely—in practice, until a stable layer is reached. Also, wind reduces smoke rise.

There exist at least 40 formulae which relate the rise of the meteorological and nonmeteorological variables. Most are determined by fitting equations to smoke rise measurements. Because many such formulae are based only on limited ranges of the variables, they are not generally valid. Also, most of the formulae contain dimensional constants suggesting that not all relevant variables have been included properly.

For a concise summary of the most commonly used equations, the reader is referred to a paper by Briggs (1969). In this summary, Briggs also describes a series of smoke rise formulae based on dimensional analysis. These have the advantage of a more physical foundation than the purely empirical formulae, and appear to fit a wide range of observed smoke plumes. For example, in neutrally stable air, the theory predicts that the rise should be proportional to horizontal distance to the 2/3 power which is in good agreement with observations. The use of dimensionally correct formulae has increased significantly since 1970.

Given the height of effluent rise above a stack, an "effective" source is assumed for calculation of transport and dispersion. This effective source is taken to be slightly upwind of a point straight above the stack, by an amount of the excess rise calculated. If the efflux velocity is small, the excess rise may actually be negative at certain wind velocities (downwash).

Transport of Pollutants

Pollutants travel with the wind. Hourly wind observations at the ground are available at many places, particularly airports. Unfortunately, such weather stations are normally several hundred kilometers apart, and good wind data are lacking in between. Further, wind information above 10 meters height is even less plentiful, and pollutants travel with winds at higher levels.

Because only the large-scale features of the wind patterns are known, air pollution meteorologists have spent considerable effort in studying the wind patterns between weather stations. The branch of meteorology dealing with this scale—the scale of several km to 100 km—is known as mesometeorology. The wind patterns on this scale can be quite complex, and are strongly influenced by surface characteristics. Thus, for instance, hills, mountains, lakes, large rivers, and cities cause characteristic wind patterns, both in the vertical and horizontal. Many vary in time, for example, from day to night. One of the important problems for the air pollution meteorologist is to infer the local wind pattern on the mesoscale from ordinary airport observations. Such influences are aided by theories of sea breezes, mountain-valley flow, etc.

In many areas, local wind studies have been made. A particularly useful tool is the tetroon, a tetrahedral balloon which drifts horizontally and is followed by radar. In some important cities such as New York and Chicago, the local wind features are well-known. In general, however, the wind patterns on the mesoscale are understood qualitatively, but not completely quantitatively. Much mesoscale numerical modeling is in progress or has been completed.

Atmospheric Dispersion

Dispersion of a contaminant in the atmosphere essentially depends on two factors: on the mean wind speed, and on the characteristics of atmospheric "turbulence." To see the effect of wind speed, consider a stack which emits one puff per second. If the wind speed is 10 m/sec, the puffs will be 10 m apart; if it is 5 m/sec, the distance is 5 m. Hence, the greater the wind speed, the smaller the concentration.

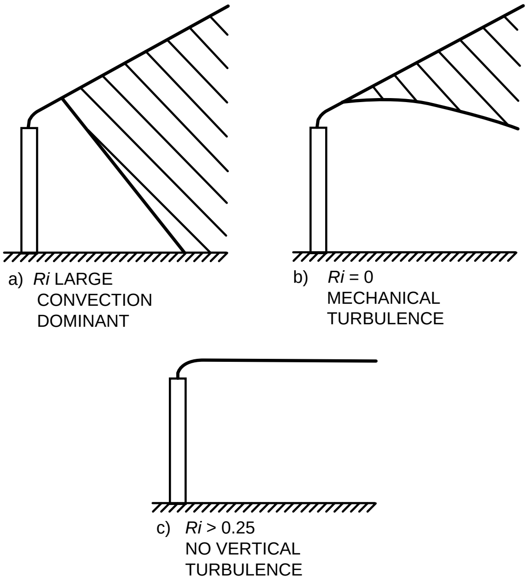
Atmospheric "turbulence" consists of horizontal and vertical eddies which are able to mix the contaminated air with clean air surrounding it; hence, turbulence decreases the concentration of contaminants in the plume, and increases the concentration outside. The stronger the turbulence, the more the pollutants are dispersed.

There are two mechanisms by which "eddies" are formed in the atmosphere: heating from below and wind shear. Heating produces convection. Convection occurs whenever the temperature decreases rapidly with height—that is, whenever the lapse rate exceeds 1°C/100 m. It often penetrates into regions where the lapse rate is less. In general, convection occurs from the ground up to about a thousand meters elevation on clear days and in cumulus-type clouds.

The other type of turbulence, mechanical turbulence, occurs when the wind changes with height. Because there is no wind at ground level, and there usually is some wind above the ground, mechanical turbulence just above the ground is common. This type of turbulence increases with increasing wind speed (at a given height) and is greater over rough terrain than over smooth terrain. The terrain roughness is usually characterized by a "roughness length" z_0 which varies from about 0.1 cm over smooth sand to a few meters over cities. This quantity does not measure the actual height of the roughness elements; rather it is proportional to the size of the eddies that can exist among the roughness elements. Thus, if the roughness elements are close together, z_0 is relatively small.

The relative importance of heat convection and mechanical turbulence is often characterized by the Richardson number, *Ri.* Actually, –*Ri* is a measure of the relative rate of production of convective and mechanical energy. For example, negative Richardson numbers of large magnitude indicate that convection predominates; in this situation, the winds are weak, and there is strong vertical motion. Smoke leaving a source spreads rapidly, both vertically and laterally (Figure 2). As the mechanical turbulence increases, the Richardson number approaches zero, and the angular dispersion decreases. Finally, as the Richardson number becomes positive, the stratification becomes stable and damps the mechanical turbulence. For Richardson numbers above 0.25 (strong inversions, weak winds), vertical mixing effectively disappears, and only weak horizontal eddies remain.

Because the Richardson number plays such an important role in the theory of atmospheric turbulence and dispersion, Table 1 gives a qualitative summary of the implication of Richardson numbers of various magnitudes.



a) *Ri* LARGE CONVECTION DOMINANT

b) *Ri* = 0 MECHANICAL TURBULENCE

c) *Ri* > 0.25 NO VERTICAL TURBULENCE

FIGURE 2 Average vertical spread of effluent from an elevated source under different meteorological conditions (schematic).

It has been possible to describe the effect of roughness length, wind speed, and Richardson number on many of the statistical characteristics of the eddies near the ground quantitatively. In particular, the standard deviation of the vertical wind direction is given by an equation of the form:

$$\sigma_\theta = \frac{f(Ri)}{ln\ z/z_0 - \psi(Ri)}. \tag{1}$$

Here z is height and $f(Ri)$ and $\psi(Ri)$ are known functions of the Richardson number which increase as the Richardson number decreases. The standard deviation of vertical wind direction plays an important role in air pollution, because it determines the initial angular spread of a plume in the vertical. If it is large, the pollution spreads rapidly in the vertical. It turns out that under such conditions, the contaminant also spreads rapidly sideways, so that the central concentrations decrease rapidly downstream. If σ_θ is small, there is negligible spreading.

Equation 1 states that the standard deviation of vertical wind direction does not explicitly depend on the wind speed, but at a given height, depends only on terrain roughness and Richardson number. Over rough terrain, vertical spreading is faster than over smooth terrain. The variation with Richardson number given in Eq. (1) gives the variation of spreading with the type of turbulence as indicated in Table 1: greatest vertical spreading with negative *Ri* with large numerical values, less spreading in mechanical turbulence ($Ri = 0$), and negligible spreading on stable temperature stratification with little wind change in the vertical.

An equation similar to Eq. (1) governs the standard deviation of horizontal wind direction. Generally, this is somewhat larger than σ_θ. For light-wind, stable conditions, we do not know how to estimate σ_θ. Large σ_θ are often observed, particularly for $Ri > 0.25$. These cause volume meanders, and are due to gravity waves or other large-sclae phenomena, which are not related to the usual predictors.

In summary, then, dispersion of a plume from a continuous elevated source in all directions increases with increasing roughness, and with increasing convection relative to mechanical turbulence. It would then be particularly strong on a clear day, with a large lapse rate and a weak wind, particularly weak in an inversion, and intermediate in mechanical turbulence (strong wind).

TABLE 1

Turbulence characteristics with various Richardson numbers

$0.24 < Ri$	No vertical mixing
$0 < Ri < 0.25$	Mechanical turbulence, weakened by stratification
$Ri = 0$	Mechanical turbulence only
$-0.03 \leq Ri < 0$	Mechanical turbulence and convection but mixing mostly due to the former
$Ri < -0.04$	Convective mixing dominates mechanical mixing

Estimating Concentration of Contaminants

Given a source of contaminant and meteorological conditions, what is the concentration some distance away? Originally, this problem was attacked generally by attempting to solve the diffusion equation:

$$\frac{d\chi}{dt} = \frac{\partial}{\partial x} K_x \frac{\partial \chi}{\partial x} + \frac{\partial}{\partial y} K_y \frac{\partial \chi}{\partial y} + \frac{\partial}{\partial z} K_z \frac{\partial \chi}{\partial z}. \tag{2}$$

Here, χ is the concentration per unit volume; *x, y,* and *z* are Cartesian coordinates, and the *K*'s are diffusion coefficients, not necessarily equal to each other.

If molecular motions produced the dispersion, the *K*'s would be essentially constant. In the atmosphere, where the mixing is produced by eddies (molecular mixing is small enough to be neglected), the *K*'s vary in many ways. The diffusion coefficients essentially measure the product of eddy size and eddy velocity. Eddy size increases with height; so does *K.* Eddy velocity varies with lapse rate, roughness length, and wind speed; so does *K.* Worst of all, the eddies relevant to dispersion probably vary with plume width and depth, and therefore with distance from the source. Due to these complications, solutions of Eq. (2) have not been very successful with atmospheric problems except in some special cases such as continuous line sources at the ground at right angles to the wind.

The more successful methods have been largely empirical: one assumes that the character of the geometrical distribution of the effluent is known, and postulates that effluent is conserved during the diffusion process (this can be modified if there is decay or fall-out), or vertical spread above cities.

The usual assumption is that the distribution of effluent from a continuous source has a normal (Gaussian) distribution relative to the center line both in the vertical direction, *z* (measured from the ground) and the direction perpendicular to the wind, *y.* The rationalization for this assumption is that the distributions of observed contaminants are also nearly normal.[†] Subject to the condition of continuity, the concentration is given by (including reflection at the ground).

$$\chi = \frac{Q}{2\pi V \sigma_y \sigma_z}\left(\exp - \frac{y^2}{2\sigma y^2}\right) \times \left(\exp - \frac{(z-H)^2}{2\sigma_z^2} + \exp \frac{-(z+H)^2}{2\sigma_z^2}\right). \tag{3}$$

Here, *H* is the "effective" height of the source, given by stack height plus additional rise, σ is the standard deviation of the distribution of concentration in the *y* and *z*-direction, respectively, and *V* is the wind speed, assumed constant. *Q* is the amount of contaminant emitted per unit time.

The various techniques currently in use differ in the way σ_y and σ_z are determined. Clearly, these quantities change with downwind distance *x* (Figure 3) as well as with roughness and Richardson number.

Quantitative estimation of the Richardson number requires quite sophisticated instrumentation; approximately, the Richardson number can be estimated by the wind speed, the time of the day and year, and the cloudiness. Thus, for example, on a clear night with little wind, the Richardson number would be large and positive, and σ's in Eq. (3) are small; on the other hand, with strong winds, the Richardson numbers are near zero, and the dispersion rate as indicated by the σ would be intermediate.

For many years, standard deviations were obtained by Sutton's technique, which is based on a very arbitrary selection for the mathematical form of Lagrangian correlation functions. More popular at present is the Pasquill–Gifford method in which σ_y and σ_z as function of *x* are determined by empirical graphs (Figure 4). Note that the dependence of the standard deviations on *x* varies with the "stability category" (from A to F). These categories are essentially Richardson number categories, judged more or less subjectively. Thus, A (large dispersion) means little wind and strong convection; D is used in strong winds, hence strong mechanical turbulence and less dispersion; F applies at night in weak winds.

One drawback of the Pasquill–Gifford method is that it does not allow for the effect of terrain roughness; the empirical curves were actually based on experiments over smooth terrain, and therefore underestimate the dispersion over cities and other rough regions. Some users of the method suggest allowing for this by using a different system of categories over rough terrain than originally recommended.

This difficulty can be avoided if fluctuations of wind direction and vertical motion are measured. Taylor's diffusion theorem at right angles to the mean wind can be written approximately,

$$\sigma_y = \sigma_0 \times F\left(\frac{t}{T_L}\right). \tag{4}$$

Here *F* is a function which is 1 for small diffusion time, *t.* For larger *t, F* decreases slowly; its behavior is fairly well known. T_L is a Lagrangian time scale which is also well known.

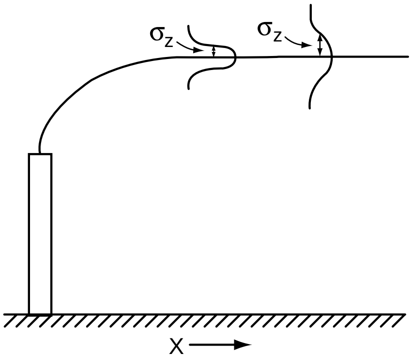


FIGURE 3 Change of vertical effluent distribution downstream.

[†] Note added in proof: It now appears that this assumption is not satisfactory for vertical dispersion, especially if the source is near the surface.

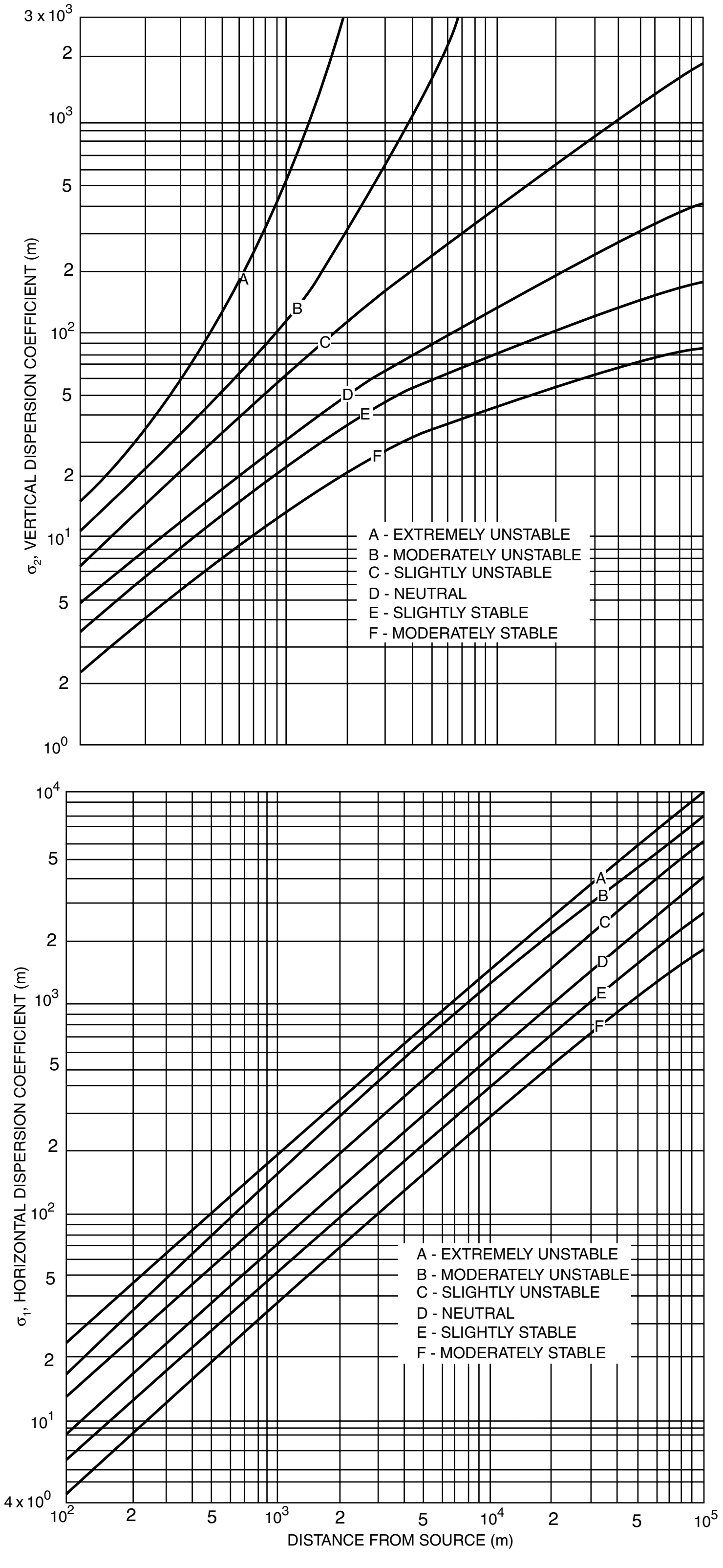


FIGURE 4 Pasquill–Gifford scheme for estimating vertical and lateral plume width as function of downwind distance and meteorological conditions.

An equation similar to (4) also exists for vertical spreading; however, it is theoretically less valid, since turbulence is not homogeneous in the vertical.

As the plume expands vertically, the vertical distribution cannot remain normal indefinitely. At the bottom, the plume is limited by the ground. At the top, the plume will be limited by an elevated inversion layer. Eventually, the vertical distribution becomes uniform. In that case, the concentration is given by the equation:

$$x = \frac{Q}{\sqrt{2\pi}VD\sigma_y}\exp - \frac{y^2}{2\sigma_y^2} \tag{5}$$

where *D* is the height of the inversion layer, which is also the thickness of the "mixed layer." Note that the concentration is inversely proportional to *VD,* the "ventilation factor," which is the product of *D,* and *V,* the average wind in the mixed layer.

The lateral spread is often limited by topography. In a valley of width *W,* the factor $(\exp - y^2/2\sigma_y^2)/(\sqrt{2\pi}\sigma_y)$ in Eqs. (3) and (5) is replaced by 1/*W,* after the contaminant concentration fills the valley uniformly in the *y*-direction (the direction perpendicular to the valley). The effect of this change is that relatively large concentrations are maintained at large distances from the sources.

Although the Pasquill–Gifford graphs are still popular in practical applications, evaluation in diffusion experiments have suggested serious deficiencies. Thus, the research community is groping for alternate methods. In particular, vertical distributions are far from Gaussian, particularly for ground sources. Significant progress has been made only for the important case of light-wind, sunny conditions. Then, the basic predictors are the thickness of the planetary boundary layer (PBL), z_i; another important predictor is a vertical-velocity parameter, $w*$ which is proportional to $(z_i H)^{1/3}$ where *H* is the vertical heat flux at the surface. *H* is not usually measured, but must be estimated independently; fortunately, it is raised to the 1/3 power. Lateral dispersion is still Gaussian, but with σ_y given by

$$\sigma_y/z_i = f(tw*/z_i) = f(X) \qquad \text{where } X = tw*/z_i$$

f is presumably universal and fairly well known.

The vertical distribution is definitely not Gaussian; for example, the center line of the plume rises for ground sources. More important, the center line comes down toward the surface for elevated sources, unless the sources are buoyant.

If vertical diffusion is normalized by the new variables, it depends on z/z_i, *X* and h/z_i where *h* is stack height. The distributions have been measured for different h/z_i, and complicated formulas exist to fit the observations. The results are believed to be quite reliable, because numerical models, laboratory experiments and full-scale observations are all in satisfactory agreement.

The results of this research should be used in practical applications, but have not been. For more detail, see Panofsky and Dutton, 1984.

City Models

These different methods give the pollutant concentrations downwind from a single source. In order to obtain the total picture of air pollution from a city, the concentrations resulting from all sources must be added together, separately for all different wind directions, different meteorological conditions, and for each contaminant. Such a procedure is expensive, even if carried out with an electronic computer, and even if, as is usually done, all small sources in square-mile areas are combined. Therefore, complete city models of air pollutant concentrations have only been constructed for very few locations. It is necessary, however, to have city models in order to understand the distribution of contaminants; only then it is possible to determine the most economical strategy to reduce the pollution, and to evaluate the effects of expansion of housing and industry.

Because the construction of a complete city model is so expensive, city models are often simplified. For example, if the city is represented by a series of parallel line sources, the computations are greatly reduced. Many other simplifications have been introduced; for a summary of many city models now in existence, see Stern (1968).

Diurnal Variation of Air Pollution

Equation (5) which shows that concentrations at considerable distances from individual sources are inversely proportional to the ventilation factor (*VD*), can be used to explain some of the variations in air pollution caused by meteorological factors. First, we shall consider the diurnal variation of air pollution. Of course, the actual variation of pollution may be different if the source strength varies systematically with time of day. The diurnal variation is different in cities and in the country. Consider typical vertical temperature distributions as seen in Figure 5. During the day, both over cities and country, the ground temperature is high, giving a deep mixed layer. After sunset, the air temperature near the surface in the country falls, producing an inversion reaching down to the ground. After air moves from the country out over the relatively warmer and rougher city, a thin mixed layer is formed near the ground. The thickness of this mixed

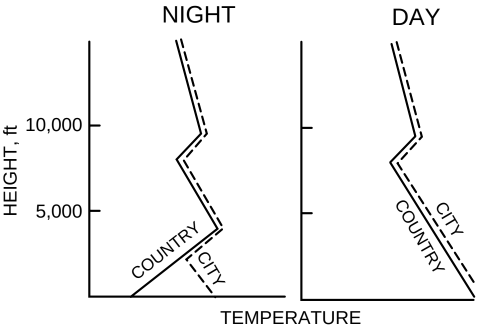


FIGURE 5 Vertical temperature distribution (schematic) over city and country, day and night.

layer varies with the size of the city, and depends on how long the air has moved over the city. In New York, for example, the mixed layer is typically 300 m thick; in Johnstown, Pa., an industrial valley city with just under 100,000 population, it is only a little over 100 m.

Figure 6 indicates how the temperature changes shown in Figure 5 influence the diurnal variation of pollution due to an elevated source in the country; at night, vertical mixing is negligible and the air near the ground is clean. Some time shortly after sunrise, the mixed layer extends to just above the source, and the elevated polluted layer is mixed with the ground air, leading to strong pollution (also referred to as "fumigation"), which may extend many kilometers away from the source. Later in the morning and early afternoon, the heating continues and thickens the mixed layer. Also, the wind speed typically increases, and the pollution decreases.

In the city, many sources usually exist in the thin nighttime mixed layer. Since this layer is so thin, and the wind usually weak, dense pollution occurs at night. Right after sunrise, the pollution at first increases somewhat, as the effluent from large, elevated sources is brought to the ground. As the mixed layer grows, the concentrations diminish, and, in the early afternoon, they are often less than the nighttime concentrations (see Figure 7).

Thus, the main difference between air pollution climates in the city and country is that country air near industrial sources is usually clean at night, whereas the city air is dirtier at night than in the middle of the day. These differences are most pronounced during clear nights and days, and can be obliterated by diurnal variations of source strengths. Figure 7 shows the characteristic behavior only because the sources of pollution at Johnstown, Pa., are fairly constant throughout.

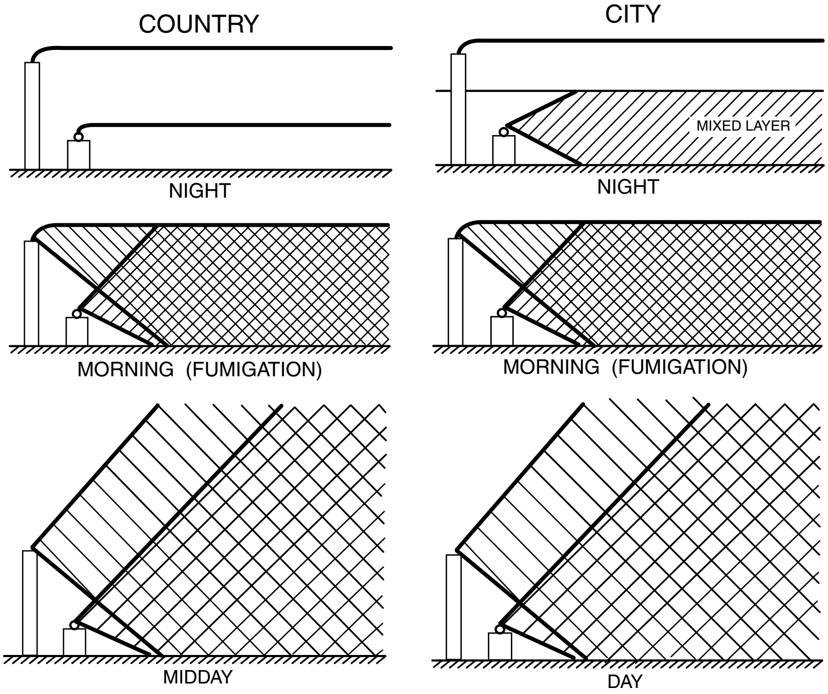


FIGURE 6 Concentrations of effluent (schematic) as function of time of day, over city and country.

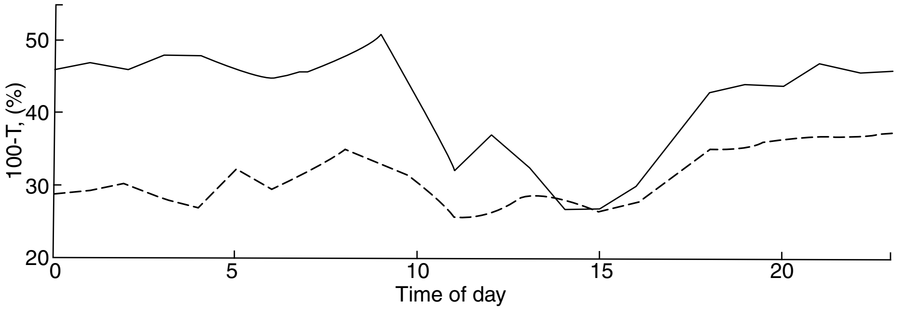


FIGURE 7 Concentrations of air pollution (100-T%), as function of time of day, on clear day (solid line) and cloudy day (dashed line), at Johnstown, Pa.

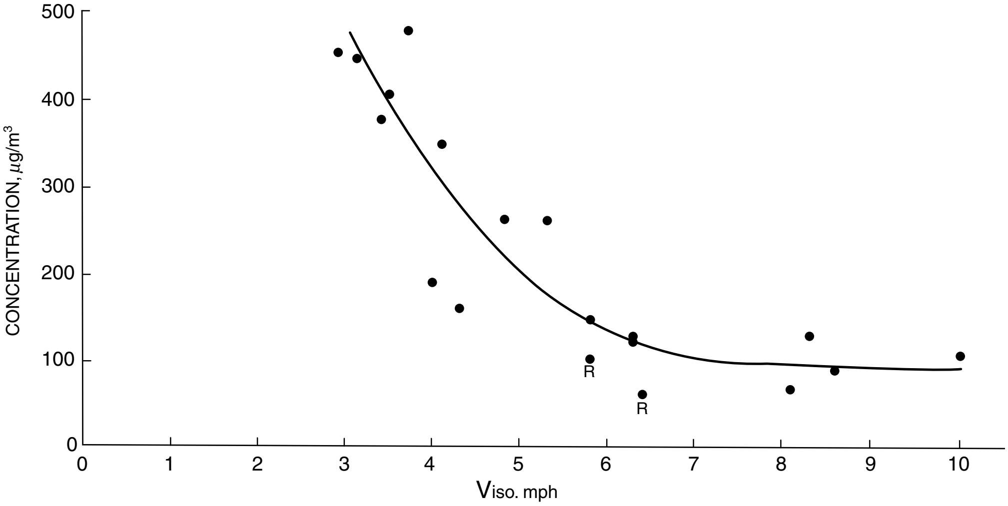


FIGURE 8 Dependence of 24-hour average particle concentrations at Johnstown on wind speed at 150 ft. R denotes rain.

Day-to-day Variations in Air Pollution

Equation (5) shows that, other things being equal, the concentration of contaminants is inversely proportional to the wind speed. Figure 8 shows this effect on 24-hr total particulate concentration at Johnstown, for cases where the source strengths were roughly the same, during the fall of 1964.

Conditions of particularly bad air pollution over wide areas and for extended periods are accompanied not only by light winds and calms, but also by unusually small mixing depths (*D*) so that the ventilation factor is usually small. Such conditions occur within large high-pressure areas (anticyclones). In such areas, air is sinking. Sinking air is warmed by compression. Thus, in an anticyclone (high-pressure area), an elevated warm layer forms, below which there is room only for a relatively thin mixed layer (Figure 9). The inversion on top of the mixed layer prevents upward spreading of the pollution, and when mountains or hills prevent sideways spreading the worst possible conditions prevail. A particularly bad situation arose in the industrial valley town of Donora, Pa., in which many people were killed by air pollution in 1948.

Cities in California, like Los Angeles, are under the influence of a large-scale anticyclone throughout the summer, and an elevated inversion at a few hundred meters height occurs there every day; that is why Los Angeles had air pollution problems as soon as pollutants were put into the atmosphere to any large extent. In the United States outside the West Coast, stagnant anticyclones occur only a few times per year, usually in the fall.

So far, relatively little use has been made in the USA of forecast changes in air pollution potential from day to day. As air pollution problems become more severe, more use will be made of such forecasts. Already, this type of information has proved itself in air pollution management in some European countries.

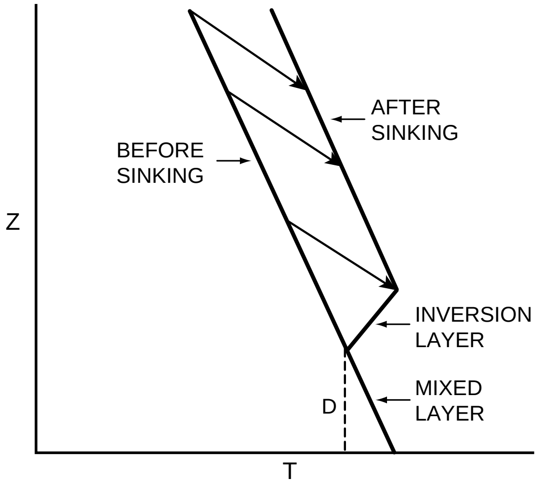


FIGURE 9 Effect of sinking on vertical temperature distribution (schematic).

Not much has been said about the influence of wind direction on air pollution. When pollution is mainly due to many, relatively small sources, as it is New York, the pollution is surprisingly insensitive to changes in wind direction. Even in Johnstown, Pa., wind direction is unimportant except for the case of easterly winds, when a single, huge steel plant adds significantly to the contaminant concentration.

In contrast, wind direction plays a major role when most of the pollution in a given area is due to a single or a few major plants or if an industrial city is nearby. Also, there are special situations, in which wind direction is particularly important; for example, in Chicago, which has no pollution sources east of the city, east winds bring clean air.

The main difference between the effects of lapse rate, mixing depth, and wind speed on the one hand, and wind direction on the other, is that the wind direction has different effects at various sites, depending on the location of the sources; the other factors have similar effects generally.

EFFECT OF AIR POLLUTION ON LOCAL AND REGIONAL WEATHER

Visibility

The most obvious effect of air pollution is to reduce visibility. This effect has been studied frequently by comparing visibility in different parts of a city, or the visibility in a city with visibility in the country. For a summary of many such investigations, see Peterson, 1969.

To give some examples: Around London and Manchester, afternoon visibility less than 6¼ miles occurs on more than 200 days; in Cornwall in SW England, the number is less than 100. In central London, there are 940 hours a year with visibilities less than ⅝ mile; in SE England, only 494.

In many cities, visibilities have recently improved probably due to control of particle emissions; however, as mentioned before, some of this change may be due to changes in large-scale weather patterns.

Although decreased visibility is usually associated with industrial or automobile pollution, considerable attention has been paid recently to decreased visibilities due to the "contamination" of the atmosphere by water droplets by industry. This problem arises because many processes generate excess heat; if this is added to streams and lakes, undesirable effects ensue; hence, progressively more and more heat is used to evaporate water which is then emitted into the atmosphere, and later condenses to form water plumes.

There are many unpublished studies estimating the effect of cooling towers on visibility. This varies greatly with meteorological conditions, but is particularly serious in winter, then the air is nearly saturated and little additional vapor is required to produce liquid drops. Under those conditions, water plumes from industries produce clouds and fog which may reach over a hundred miles from the sources. Automobile accidents have been blamed on such fogs, particularly when the particles freeze and make roads slippery, adding to the visibility hazard.

Sunshine Intensity

"Turbidity" is an indicator of the reduction of light due to haze, smoke and other particles. Turbidity is now being monitored at many places in the world. It is quite clear that it is larger over cities than over the country; it has been suggested that the average decrease of sunshine over cities is 15 to 20% due to pollution. The effect is even larger if only ultraviolet light is considered.

Control of smoke emission in cities such as London has caused a very noticeable increase of sunshine intensity: for example the hours of "Bright sunshine" increased by 50% after control measures had become effective. Again, for a summary of some of these studies, the reader is referred to Peterson, 1969.

Precipitation Amount

There have now been several studies suggesting that precipitation is increased downstream of industrial centers. The investigations are of a statistical nature, and it is not known whether the effects are due to increased convection (increased heat), increased condensation nuclei or increased water vapor. Further, the reliability of the statistics has been questioned.

For example, Changnon (1969) found a large precipitation anomaly at La Porte (Indiana) just downwind of large industrial complexes of Northwestern Indiana. But change in observational techniques of rainfall and other uncertainties have thrown doubt on the results. Hobbs *et al.* (1970) have compared rainfall distribution in Western Washington before and after the construction of industries and found an increase by 30% or so; but some of this increase may have been due to "normal" climatic change. For a summary of these and other studies see Robinson (1970). It becomes quite clear from this summary that more, careful investigations of this type are needed before effects of air pollution on precipitation patterns can be definitely proven.

A large study (Metromex) found strong enhancement of precipitation downwind of St Louis. But this may be due to the St Louis heat sources rather than to pollution.

Acid Rain

There is no question that acid rain is produced by atmospheric pollution. The acidity of rainfall is large only when the wind direction suggests industrial or urban sources. Most important is sulphuric acid, produced by power plants or smelters, the effluent from which contains SO_2. Also important is nitric acid, which is formed mostly from nitrogen oxides in car exhausts. Acid rain has done important damage to lakes and forests; but there is controversy how to deal with the problem. For example, the relation between acidity and SO_2 may be nonlinear, so that substantial reduction of SO_2 may not effect acid rain significantly.

GLOBAL EFFECTS OF AIR POLLUTION

Natural Climatic Changes

We will assess the effect of some atmospheric pollutants as to their ability to change the earth's climate. In doing so, we are hampered by the fact that the present climate is produced by a multitude of interacting factors; if one factor is changed, others will too, and a complex chain reaction will ensue. These reactions can be studied by complex mathematical models of the atmosphere, which so far have been quite successful in describing the existing climate. But, as yet these models contain assumptions which make it impossible at this time to assess accurately the effects of changes in some of the factors affecting climate. Until such models are improved, then, we cannot really estimate quantitatively climatic changes produced by pollutants.

The concentration of CO_2 is about 340 parts per million (ppm). According to observations at Mauna Loa in Hawaii, over the last forty years or so, it has increased at the rate of 0.7% per year. This is less than half the amount put into the atmosphere by industry. The other half goes into the ocean or into vegetation; but it is not known how much goes into each. Further, we do not know whether the same fraction can disappear out of the atmosphere in the future—e.g., the amount going into the ocean is sensitive to temperature, and the amount going into vegetation may be limited by other factors. However, a reasonable guess is that the fraction of CO_2 in the atmosphere will double in the middle of the 21st century.

The basic effect of CO_2 on climate is due to the fact that it transmits short-wave radiation from the sun, but stops a part of the infrared radiation emitted by the earth. Hence, the more CO_2, the greater the surface temperature. This is known as the greenhouse effect. Also, since CO_2 increases the radiation into space, the high atmosphere is cooled by increasing CO_2.

The heating rate at the ground expected with a doubling of CO_2 has been calculated by many radiation specialists. The answers differ, depending on how many other variables (such as cloud cover) are allowed to change as the CO_2 changes. The best current estimates are that doubling CO_2 would increase the surface temperature about 2°C, and decrease the temperature aloft a little more. But these estimates do not treat changes of cloud cover and oceanic effects realistically, and these estimates may yet be corrected. Still, if we expect only a 20% change in CO_2 by the end of the century, the climatic change due to this factor should be small. However, a serious problem could arise in the next century, particularly because it is difficult to see how a trend in CO_2 concentration can be reversed. It is therefore of great importance to continue monitoring CO_2 concentration accurately.

As of 1987, it appears likely that increases of concentration of other trace gases (e.g. fluorocarbons) may, in combination, have as strong a warming effect at the surface as CO_2. So far, no significant warming has been detected.

Ozone

Ozone (O_3) is an important part of photochemical smog; originating mostly from the effect of sunlight on automobile exhaust. The concentration is critically dependent on chemical reactions as well as on diffusion. Chemistry is beyond the scope of this paper as O_3 and ozone pollution near the ground will not be discussed further.

More important, 90% of the ozone exists in the stratosphere (above about 11 km). Its concentration even there is small (usually less than 10 ppm). If all ozone were to be brought to the surface of the ground, its thickness would average about 0.3 cm.

Most of the ozone occurs at high latitudes, and there is a spring maximum. The great importance of stratospheric ozone is due to its ability to absorb ultraviolet (UV) light, particularly in the UVB region (290–320 μm) where human skin is extremely sensitive. Thus, decreased ozone would increase skin cancer.

We now realize that small fractions (10^{-9}) of certain gases can destroy ozone by catalytic reactions. The most important are oxides of nitrogen and chlorine. Nitrogen oxides could originate for example, from supersonic transports. However calculations show that, unless the number of SSTs is increased significantly, this problem is not serious.

More important is the problem of chlorofluoromethanes (CFM) the use of which has been rapidly increasing. They are used in sprays, foams and refrigeration, CFMs are so stable that most of them are still in the atmosphere. Eventually, however, CFMs will seep into the stratosphere (about 1%/year). In the high stratosphere, UV will dissociate CFMs producing chlorine, which destroys ozone.

A slow decrease of ozone in the stratosphere has indeed been indicated by recent satellite observations. For total ozone, the results are much more controversial. Chemical–meterological models show only a very small decrease so far, too small to isolate from the "noisy" observations. However, the accuracy of the models can be questioned, particularly since new relevant reactions have been discovered every few years, so that model results have varied.

Of special interest has been the recognition of an "ozone hole," centered at the South Pole, and lasting a month or so in the Southern Spring. Total column ozone falls to about half its normal value. The phasing out of chlorofluorocarbons, or CFCs began in 1989 with the implementation of the Montreal Protocol.

Editors Notes: Scientists at NASA and various U.S. universities have been studying satellite data taken over the past 2 decades. They found the rate of ozone depletion in the upper stratosphere is slowing partially as a result of a reduction in the use of CFCs (see Newchurch, et al., 2005).

In the troposphere, aerosol formation from the combustion of fossil fuels and biomass is a precursor to the formation of brown clouds, which are intricately linked to climate changes (Ramanathan and Ramana, 2003). Ozone, a component of smog, also forms in the troposphere, when NO_x combines with volatile organic compounds in the presence of sunlight. There is growing scientific evidence that the intercontinental transport (ICT) of aerosols and ozone influences surface air quality over downwind continents (Fiore, *et al.*, 2003). For example during the dust storm events in Asia in April of 2001, the ground level aerosol concentrations in the western U.S. and Canada increased by as much as 40 μg/m^3 resulting from the ICT of aerosols. Fiore, *et al.* found there are global dimensions to the aerosol and ozone problems.

It has also been suggested that ozone changes can produce climate changes, but these appear rather unimportant at present, except that they may worsen slightly the CO_2 greenhouse effect.

Summary

In summary, increasing air pollution can modify the climate in many ways. There is no evidence that any significant change has occurred so far; but eventually, large effects are likely.

REFERENCES

1. Briggs, G.A. (1969), Plume Rise. USAEC critical review series, TID-25075, Clearinghouse for federal scientific and technical information.
2. Changnon, S.H. (1968), The LaPorte Weather Anomaly, fact or fiction, *Bull. Amer. Met. Soc.,* **49,** pp. 4–11.
3. Fiore, A.T. Holloway and M. Galanter, *Environmental Manager,* pp. 13–22, Dec. 2003.
4. Hanna, S.R., G. Briggs, J. Deardorff, B.E. Egan, F.A. Gilfford, and F. Pasquill (1977), AMS workshop on stability classification schemes and sigma curves, *Bull. Amer. Met. Soc.,* **58,** pp. 1305–1309.
5. Hobbs, P.V., L.F. Radke, and S.E. Shumway (1970), Cloud condensation nuclei from industrial sources and their apparent influence on precipitation in Washington State, *Jour. Atmos. Sci.,* **27,** pp. 81–89.
6. Newchurch, M.J., *Journal of Geophysical Research,* V110, 2005.
7. Panofsky, H.A. and F.X. Dutton (1984), *Jour. Atmos. Sci.,* **41,** pp. 18–26.
8. Peterson, J.T. (1969), The Climate of Cities: A Survey of Recent Literature. US Dept. HEW, NAPCA, Pub. No. AP-50.
9. Ramanathan, V. and M.V. Ramana, *Environmental Manager,* pp. 28–33, Dec. 2003.
10. Robinson, G.D. (1970), Long-Term Effects of Air Pollution, Center for Environment and Man, Hartford, CEM 4029–400.
11. Schoeberl, M.A. and A.J. Krueger (1986), *Geoph. Res. Paper's Suppl.* **13,** No. 12.
12. Stern, A.C. (1976), *Air Pollution,* Vol. 1, Academic Press, New York. 3rd Ed.

HANS A. PANOFSKY (DECEASED)
Pennsylvania State University

AIR POLLUTION MODELING—URBAN: see URBAN AIR POLLUTION MODELING

AIR POLLUTION SOURCES

Air pollution may be defined as the presence in the atmosphere of any substance (or combination of substances) that is detrimental to human health and welfare; offensive or objectionable to man, either externally or internally; or which by its presence will directly or indirectly adversely affect the welfare of man. ("Air Pollution," Homer W. Parker, 1977.) The substances present in the atmosphere which cause this detriment to health and welfare are the air pollutants.

A considerable quantity of air pollution occurs naturally as a consequence of such processes as soil erosion and volcanic eruptions. However, those pollutants which pose a threat to human health and cause extensive damage to property are primarily derived from activities associated with the development of community living, as well as with the growth of affluence and living standards in industrial societies. These activities include the burning of fuel for heat and power, the processing of materials for food and goods, and the disposal of wastes.

Much of the materials which pollute our atmosphere represent valuable resources which are being wasted. We have available today the technological means of controlling most sources of air pollution. The cost of control however has been estimated on the order of 10 to 20 percent of the world's gross national product. Moreover, full implementation of the control measures that would be necessary to achieve healthful air quality in many of our large centers of population would require significant changes in lifestyle in those areas.

POLLUTANT CLASSIFICATIONS

Air pollutants are numerous, each with its own peculiar characteristics. Therefore it is usual to have these pollutants classified by some design. Classification allows for the study of pollutants in subgroups on the basis of some characteristic of interest or concern and also provides an ordering which makes it easier to formulate air pollution control programs. Accordingly, the classification of air pollutants may be based on:

1. How the pollutants are borne into the atmosphere.
2. The physical state of the pollutant.
3. The molecular composition of the pollutants.
4. The nature of the problem or health threat associated with the pollutants.

Classification According to the Method of Entry into the Atmosphere

This classification contains two categories: (1) Primary and (2) secondary.

Primary Pollutants Primary air pollutants are emitted into the atmosphere directly from identifiable sources whether from mechanical or chemical reaction processes. Examples of such direct discharge from an identifiable source into the atmosphere include the complete and incomplete combustion of carbonaceous fuels from industrial processes and automobile engines yielding carbon monoxide and carbon dioxide.

Secondary Pollutants These pollutants are those which are formed as a result of some reaction in the atmosphere. This reaction may occur between any combination of air pollutants (including primary pollutants) and natural components of the atmosphere. Some of these reactions require the presence of sunlight and are called photo-chemical reactions. An example of such a reaction is the formation of ozone from the interaction of organic and nitrous compounds in the presence of sunlight.

Classification According to the Physical State of the Pollutant

According to their state of matter, pollutants may be classified as: (1) gaseous and (2) particulate.

Gaseous Pollutants Most air pollutants exhibit gaseous properties in that they tend to obey gas laws, for example, there is a predictable interrelationship between their pressure, volume and temperature. In many ways these pollutants behave like air itself and do not tend to settle out or condense over long periods.

However, they almost always undergo some form of chemical transformation while resident in the atmosphere. Approximately 90% of air pollutants are gaseous.

Particulate Pollutants Any pollutant that is not gaseous is defined as a particulate pollutant or particulate whether they exist in the form of finely divided solids or liquids. The larger particulates after having been introduced into the air tend to settle out quickly and affect lives and property near the source. The smaller and lighter particles travel further away, and eventually settle out great distances from the source. The very smallest particulates exhibit certain gaseous characteristics, remaining suspended in the atmosphere for long periods of time and are readily transported by wind currents.

Classification According to Chemical Composition

Pollutants may also be classified according to their chemical structure. The basic classifications are (1) organic and (2) inorganic.

Organic Pollutants Organic compounds may be defined as those which contain carbon, hydrogen, and may contain other elements. By this definition we exclude the very simple carbon monoxide and carbon dioxide. These contain carbon, but no hydrogen.

Inorganic Pollutants Inorganic pollutants may be defined as compounds which do not contain compounds of carbon, with the exception of carbon oxides, like CO and CO_2, and carbon disulfide. Many of the most commonly encountered pollutants are inorganic. You might be asking yourself why CO_2 is considered a pollutant. Isn't CO_2 beneficial in the maintenance of the earth's ecological system by providing a source of energy for manufacturing plants? The answer is yes, but the earth's ecosystem can utilize only so much carbon dioxide. The surplus of CO_2 in the atmosphere is believed to be one of the contributors to the "Greenhouse Effect." Excesses of this gas are believed to cause the global heating that is now being experienced. The long-term outlook for this phenomenon is the melting of the polar icecaps resulting in the oceans' levels rising and threatening population areas that are located at the coastline.

Classification According to the Nature of the Problem or Health Threat Posed by the Pollutant

Under the Clean Air Act, the Congress of the United States established a classification system which recognized two distinct categories of air pollutants: those air pollutants which because of their universal nature or ubiquity, presented a threat to public health and welfare (called criteria pollutants); and those pollutants, while not widespread, contribute to higher mortality rates in humans (called hazardous pollutants).

Criteria Pollutants These are air pollutants for which a national ambient air quality standard has been established. In the selection of these standards, certain criteria are established using observed levels of air pollution and the associated impacts on human health, vegetation and materials relating air quality level to health and welfare effects. Six specific

TABLE 1
Classification of Pollutants

Major Classes	Sub-classes	Typical Members of Sub-classes
Organic	Alkanes	Ethane
Gases	Alkenes	Ethylene
(Hydrocarbons)	Alkynes	Acetylene
	Alkyl Halides	Ethylenedichloride
	Aldehydes	Formaldehyde
	Ketones	Acetone
	Amines	Methyl Amine
	Alcohols	Ethanol
	Aromatics	Benzene
Inorganic	Photochemical Oxidants	Ozone
Gases	Oxides of Nitrogen	Nitrogen Dioxide, Nitric Oxide
	Oxides of Sulfur	Sulfur Dioxide, Sulfur Trioxide
	Oxides of Carbon	Carbon Monoxide, Carbon Dioxide
	Halides	Chlorine, Flourine
	Miscellaneous	Ammonia, Hydrogen Sulfides
Particulates	Solid Particulates	Dust, Smoke
	Liquid Particulates	Mist, Spray
		Heavy Metals

Other Pollutants Include:

—Radioactive Substances
—Pesticides
—Aeroallergens

pollutants (nitrogen dioxide, sulfur dioxide, hydrocarbons, carbon monoxide, particulate matter and ozone) were identified in 1971 as the most "universal" within the United States and the most significant pollutants contributing to the degradation of the lower atmosphere or troposphere. Once national air quality standards were established each state was given the responsibility to make sure that emissions from sources of air pollution in that state and neighboring states do not violate these air quality standards by developing and implementing creative plans for reducing source emissions. Recognizing that hydrocarbons in the atmosphere did not, as a class of pollutants, create a singular and internally consistent ambient air quality problem, the class term was dropped and lead was added as a new pollutant class.

Hazardous Pollutants These are air pollutants for which no air quality standard has been established but nevertheless cause or contribute to an increase in the mortality rate or serious irreversible or incapacitating illness. The hazardous pollutants listed by January 1988 are: asbestos, beryllium, mercury, vinyl chloride, radionuclides, coke oven emissions, benzene and inorganic arsenic.

In November of 1990, the U.S. Congress passed Clean Air Act amendments (CAAA) into law which greatly expand the list of regulated chemicals—Hazardous Air Pollutants (HAPs)– to about 190. The EPA's mandate is to promulgate standards for the control of HAP emissions from about 100 source categories, employing maximum achievable control technology (MACT). To date greater than 95% of MACT standards have been published.

Source: http://www.epa.gov/ttn/atw/eparules.html

SOURCE CLASSIFICATIONS

The management and control of air pollution is generally achieved through the regulation and control of air pollution sources. For convenience, sources of air pollutants may be classified according to the size or the nature of the pollutant activity and source type characteristics.

Classification According to Magnitude

For convenience of analysis, air pollution sources are divided into two classes (1) major sources and (2) minor sources.

Major sources are sources whose emissions quantities are large enough to cause them to have a dominant role in the pollution potential of an area. Prior to the 1990 CAAA, the U.S. Environmental Protection Agency classified all sources that emitted or had the potential for emitting 100 tons/year of any single pollutant as a major source.

Today, the definition has been revised and made more stringent. Depending upon an area's air quality, emissions of as little as 10 tons/year would constitute a major source.

Major sources are fixed (stationary) and commonly occupy a limited area relative to a community. They include:

1. Major industrial and manufacturing plants.
2. Steam—Electric power plants.
3. Industrial and Municipal Incinerators.
4. Facilities that use solvents (surface coating, degreasing, dry cleaning, plastics manufacture, rubber manufacture) and lose petroleum products by evaporation.
5. Facilities that lose petroleum product from storage and marketing (tank farms, service stations) operations.
6. Motor vehicles, aircraft, ships and railroads in which the combustion of fuels for transportation occurs.
7. Dumps, incinerators, etc. in which combustion of wastes occur.
8. Facilities or units in which the decomposition of organic wastes occur.
9. Sewage treatment plants.

Industrial plants constitute a highly varied and complex chemical system, each industry presenting a unique air pollution problem. The characteristics of the emissions produced are directly related to the peculiarities of the operation in question, that is, on the raw materials, the fuels, the process method, the efficiency of the chosen process, the method and the type of air pollution control measures applied.

Minor sources are those which cannot be cataloged practically on a source-by-source basis. They may be stationary or mobile and are commonly spread throughout the community. These sources are associated with:

1. Combustion of fuels in residences and commercial buildings and institutions for personal comfort and convenience.
2. Service industries such as laundries, dry-cleaning plants, repair services, etc.
3. Animal processing.
4. Handling and use of paints, lacquers and other surface coatings containing organic solvents.
5. Food processing in restaurants, grills, coffee roasting, etc.

Classification According to Nature of Emissions

The U.S. Environmental Protection Agency classifies sources depending on both the quantitative and qualitative nature of the emissions. The source categories are:

1. NSPS (New Source Performance Standard) sources. These are sources for which national emissions standards have been established. All sources built subsequent to the date of establishment of these emissions standards must meet NSPS requirements.
2. SIP (State Implementation Plan) sources. These are sources built prior to the establishment of the new source standards. These older SIP sources have no national emissions standards to follow per se, but rather their level of emissions is determined on a source-by-source basis and depend on the air quality of the area in which they are located. If the

air quality is particularly poor, stricter operating requirements are imposed.

3. NESHAP (National Emission Standards for Hazardous Air Pollutants) sources. These are sources which emit any of the nine hazardous pollutants which were discussed in the section on air pollutant classification. These sources also have operating standards imposed on the equipment.
4. Transportation sources. These are sources of air pollution which do not necessarily remain stationary but are mobile, and include cars, trucks, buses, airplanes, railroad locomotives and marine vessels. These sources' main emissions are carbon monoxide, carbon dioxide, nitrogen dioxide and lead and result from the internal combustion of fuel in their engines.

TABLE 2

Summary of National Emissions (thousand short tons, 1.1 million short tons equals 1 million metric tons)

Year	Carbon Monoxide	Nitrogen Oxides	Volatile Organic Compounds	Sulfur Dioxide	Particulate Matter (PM-10) (w/o) fugitive dust	Fugitive Dust (PM-10)*	Lead (short tons)
1900**	NA***	2,611	8,503	9,988	NA	NA	NA
1905**	NA	3,314	8,850	13,959	NA	NA	NA
1910**	NA	4,102	9,117	17,275	NA	NA	NA
1915**	NA	4,672	9,769	20,290	NA	NA	NA
1920**	NA	5,159	10,004	21,144	NA	NA	NA
1925**	NA	7,302	14,257	23,264	NA	NA	NA
1930**	NA	8,018	19,451	21,106	NA	NA	NA
1935**	NA	6,639	17,208	16,978	NA	NA	NA
1940	93,615	7,374	17,161	19,953	15,956	NA	NA
1945****	98,112	9,332	18,140	26,373	16,545	NA	NA
1950	102,609	10,093	20,936	22,358	17,133	NA	NA
1955****	106,177	11,667	23,249	21,453	16,346	NA	NA
1960	109,745	14,140	24,459	22,227	15,558	NA	NA
1965****	118,912	17,424	30,247	26,380	14,198	NA	NA
1970*****	128,079	20,625	30,646	31,161	13,044	NA	219,471
1975	115,110	21,889	25,677	28,011	7,617	NA	158,541
1980	115,625	23,281	25,893	25,905	7,050	NA	74,956
1984	114,262	23,172	25,572	23,470	6,220	NA	42,217
1985******	114,690	22,860	25,798	23,230	4,094	40,889	20,124
1986	109,199	22,348	24,991	22,442	3,890	46,582	7,296
1987	108,012	22,403	24,778	22,204	3,931	38,041	6,857
1988	115,849	23,618	25,719	22,647	4,750	55,851	6,513
1989	103,144	23,222	23,935	22,785	3,927	48,650	6,034
1990*******	100,650	23,038	23,599	22,433	3,882	39,451	5,666
1991*******	97,376	22,672	22,877	22,068	3,594	45,310	5,279
1992*******	94,043	22,847	22,420	21,836	3,485	40,233	4,899
1993*******	94,133	23,276	22,575	21,517	3,409	39,139	4,938
1994*******	98,017	23,615	23,174	21,118	3,705	41,726	4,956

Note(s):

* Fugitive dust emissions not estimated prior to 1985. They include miscellaneous-agriculture and forestry, miscellaneous-fugitive dust, and natural sources-wind erosion.

** NAPAP historical emissions.[3,4]

*** NA denotes not available.

**** Combination of revised transportation values and NAPAP historical emissions.

***** There is a change in methodology for determining on-road vehicle and non-road sources emissions (see chapter 6).

****** There is a change in methodology in all sources except on-road vehicles and non-road sources and all pollutants except lead, as reflected by the dotted line.

******* 1990 through 1994 estimates are preliminary. The emissions can be converted to metric tons by multiplying the values by 0.9072.

The NSPS, SIP and NESHAP sources are further classified depending on their actual and potential emissions.

Presuming that a certain area's major-source cutoff is 100 tons/year, for that area:

1. Class A sources are sources, which actually or potentially, can emit greater than 100 tons per year of effluent.
2. Class SM sources, can emit less than 100 tons per year of effluent, if and only if the source complies with federally enforceable regulations.
3. Class B sources are sources, which at full capacity, can emit less than 100 tons per year of effluent, products, and by-products.

Miscellaneous

The group is used to include such air environmental problems as aeroallergens, biological aerosols, odorous compounds, carbon dioxide, waste heat, radioactive emissions, and pesticides. In many cases they are not normally characterized as air pollutants.

The remainder of this chapter is divided into two parts.

Part 1 deals with emissions from three major classes of pollutants: hydrocarbons, inorganic gases and particulates. Typical pollutants in these major classes are described, along with their sources and the method of abatement or control.

Part 2 discusses the nature of the activity and the types of air pollutant problems associated with sources identified under standard categories of industries.

Part 1. Pollutant Emissions

Pollutant types	Sources and abundance	Abatement and control
A. HYDROCARBONS: Hydrocarbons are compounds containing the elements of carbon and hydrogen. The gaseous compounds of carbon found in nature and polluted atmospheres make up a broad spectrum of the compounds of organic chemistry. Carbon atoms bond readily to one another to form the stable carbon–carbon link. It is this link which forms the great number of organic molecules in existence (>1,000,000). By linking together in various ways, carbon atoms form a great assortment of chain and ring molecules (Aliphatics and Aromatics). The most significant hydrocarbons when considering air pollutants are known as volatile compounds (VOCs), that exist in the atmosphere primarily as gases because of their low vapor pressures. However, it is important to note that solid hydrocarbons can cause an environmental and health threat as well. For example, Benzo-(a)-pyrene, a well known carcinogen, exists in the air as a fine particulate. Hydrocarbons by themselves in air have relatively low toxicity. They are of concern because of their photochemical activity in the presence of sunlight and oxides of nitrogen (NO_x). They react to form photochemical oxidants. The primary pollutant is ozone, however, other organic pollutants like peroxyacetal nitrate (PAN) have been identified as the next highest component. Table 11 shows ozone levels generated in the photochemical oxidation of various hydrocarbons with oxides of nitrogen. The immediate health effects associated with ozone is irritation to the eyes and lungs. Long-term health effects include scarring of the lung tissue. The long-term welfare effects include damage to architectural surface coatings as well as damage to rubber products. Ozone can also damage plants and reduce crop yields.	More hydrocarbons (HC) are emitted from natural sources than from the activities of man. The one in greatest abundance is methane which has an average background concentration of 1.55 ppm. This is produced in the decomposition of dead material, mostly plant material. Methane is joined by a class of compounds of a more intricate molecular structure known as terpenes. These substances are emitted by plants, and are most visible as the tiny aerosol particulates or the "blue haze" found over most forested areas. Other hydrocarbons found in large concentrations in the ambient air besides methane (CH_4), are Ethane (C_2H_6), Propane (C_3H_8), acetylene (C_3H_4), butane and isopentane. Methane gas is one of the major greenhouse gases See Greenhouse Gases Effects, B.J. Mason. As can be inferred from Table 3, landfill emissions are the primary source of methane. About 15 percent of all atmospheric hydrocarbon is due to man's activity. However, the impact of man-made hydrocarbons to human health is out of proportion to their abundance since they are emitted in urban areas which have a high population concentration.	FROM MOBILE SOURCES: Emissions resulting from the evaporation of gasoline from fuel tanks and carburetors can be limited by storage of the vapors (within the engine itself or in a carbon canister which absorbs the fuel vapors) and then routs the vapors back to the tanks where they will be burned. Controls also exist in the refueling of automobiles and other mobile sources. These controls usually involve pressurized vacuum hoses and tighter seals at the filler pipe. FROM STATIONARY SOURCES: a) Design equipment to use or consume completely the processed material. b) In the surface coating industry, use a higher percent solids paint to reduce the amount of VOC. c) Use materials which have a higher boiling point or are less photochemically active. d) Use control equipment and recycling or organic solvents to reduce emissions. e) Control by adsorption, absorption and condensation.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
1. *Oxygenated Hydrocarbons:* Like hydrocarbons, these compounds make up an almost infinite array of compounds which include alcohols, phenols, ethers, aldehydes, ketones, esters, peroxides, and organic acids, like carboxylic acids. Oxygenated hydrocarbons are very commonly used in the paint industry as solvents, and in the chemical industry as reactants for many chemical products and intermediates. Oxygenated hydrocarbons have a two-fold environmental problem. First, they are very reactive thus readily form photochemical oxidants in the presence of sunlight (light energy) and oxides of nitrogen; thus adding to the tropospheric ozone problem.	Small amounts of oxygenated hydrocarbons are emitted by industrial processes such as spray paint coating, chemical and plastics industry. The large majority of emissions of these chemicals are associated with the internal combustion engine. Table 6 shows some typical concentrations, (parts per million), of simple hydrocarbon fuels. The aldehydes are the predominant oxygenates (these compounds will be discussed in greater detail in the following section) in emissions, but are emitted in minor amounts when compared to aliphatics and aromatics, carbon dioxide, carbon monoxide, and nitrogen oxide emissions.	FROM MOBILE SOURCES: Emissions resulting from the evaporation of gasoline from fuel tanks and carburetors can be limited by storage of the vapors (within the engine itself or in a carbon canister which absorbs the fuel vapors) and then routs the vapors back to the tanks where they will be burned. Controls also exist in the refueling of automobiles and other sources. These controls usually involve pressurized vacuum hoses and tighter seals at the filler pipe.

TABLE 3

Summary of U.S. Methane Emissions by Source Category, 1990 to 1994 Preliminary Estimates (thousand short tons)

Source Category	1990	1991	1992	1993	1994
WASTE					
Landfills	10,900	11,100	10,900	11,000	11,200
Wastewater	200	200	200	200	200
AGRICULTURE					
Cattle	6,000	6,000	6,100	6,200	6,300
Other	300	300	300	300	300
Animal Waste					
Dairy	900	900	900	900	1,000
Beef	200	200	200	200	200
Swine	1,100	1,100	1,200	1,100	1,300
Poultry	300	300	300	300	200
Other	40	40	40	40	40
Agricultural Waste Burning	100	100	100	100	100
Rice Cultivation	500	500	500	500	600
Total Agriculture	9400	9,500	9,700	9,700	10,200
FUGITIVE FUEL EMISSIONS					
Coal Mining	4,900	4,700	4,500	4,000	4,400
Oil and Gas Systems	3,600	3,600	3,600	3,600	3,600
MOBILE SOURCE COMBUSTION	300	300	300	300	100
STATIONARY COMBUSTION	700	800	800	700	700
Total Emissions	29,900	30,100	30,000	29,500	30,600

Note(s): Totals presented in this table may not equal the sum of the individual source categories due to rounding.

Source(s): Inventory of U.S. Greenhouse Gas Emissions and Sinks, 1900–1994. Draft Report, U.S. Environmental Protection Agency. September 1995.

TABLE 4
Total National Emissions of Volatile Organic Compound Emissions, 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
FUEL COMB. -ELEC UTIL	2	9	9	30	45	36		
FUEL COMB. -INDUATRIAL	108	98	106	150	157	135		
FULE COMB. -OTHER	1,867	1,336	768	541	848	749		
Residential Wood	1,410	970	563	460	809	718	698	684
CHEMICAL and ALLIED PRODUCT MFG	884	1,324	991	1,341	1,595	1,526		
Organic Chemical Mfg	58	110	245	629	884	554	562	567
METALS PROCESSING	325	442	342	394	273	72		
PETROLIUM and RELATED INDUSTRIES	571	548	1,034	1,194	1,440	643		
OTHER INDUSTRIAL PROCESSES	130	184	202	270	237	401		
SOLVENT UTILIZATION	1,971	3,679	4,403	7,174	6,584	5,975		
Surface Coating	1,058	2,187	2,128	3,570	3,685	2,619	2,687	2,773
Nonindustrial	490	NA	1,189	1,674	1,002	1,900	1,982	2,011
consumer solvents	NA	NA	NA	NA	NA	1,083	1,116	1,126
Bulk Terminals and Plants	185	361	528	599	517	658	614	606
area source: gasoline	158	307	449	509	440	560	512	501
WASTE DISPOSAL and RECYCLING	990	1,104	1,546	1,984	758	2,262		
ON ROAD VEHICLES	4,817	7,251	10,506	12,972	8,979	6,854		
Light-Duty Gas Vehicles and Motorcycles	3,647	5,220	8,058	9,193	5,907	4,285	3,812	3,921
light-duty gas vehicles	3,646	5,214	8,050	9,133	5,843	4,234	3,777	3,884
Light-Duty Gas Trucks	672	1,101	1,433	2,770	2,059	1,769	1,647	1,664
Heavy-Duty Gas Vehicles	498	908	926	743	611	470	326	393
Diesels	NA	22	89	266	402	330	318	317
heavy-duty diesel vehicles	NA	22	89	266	392	316	301	299
NON-ROAD SOURCES	778	1,213	1,215	1,542	1,869	2,120		
Non-Road Gasoline	208	423	526	1,284	1,474	1,646	1,704	1,730
lawn and garden	NA	NA	NA	574	655	728	753	761
MISCELLANEOUS	4,079	2,530	1,573	1,101	1,134	1,069		
Other Combustion	—	—	—	1,101	1,134	1,068	515	684
wildfires	3,420	1,510	768	770	739	768	212	379
TOTAL ALL SOURCES	17,161	20,936	24,459	30,646	25,893	23,599	22,575	23,174

Note(s): Categories displayed below Tier 1 do not sum to Tier 1 totals because they are intended to show major contributors.
1994 emission estimates are preliminary and will be updated in the next report.
Tier 1 source categories and emissions are shaded.

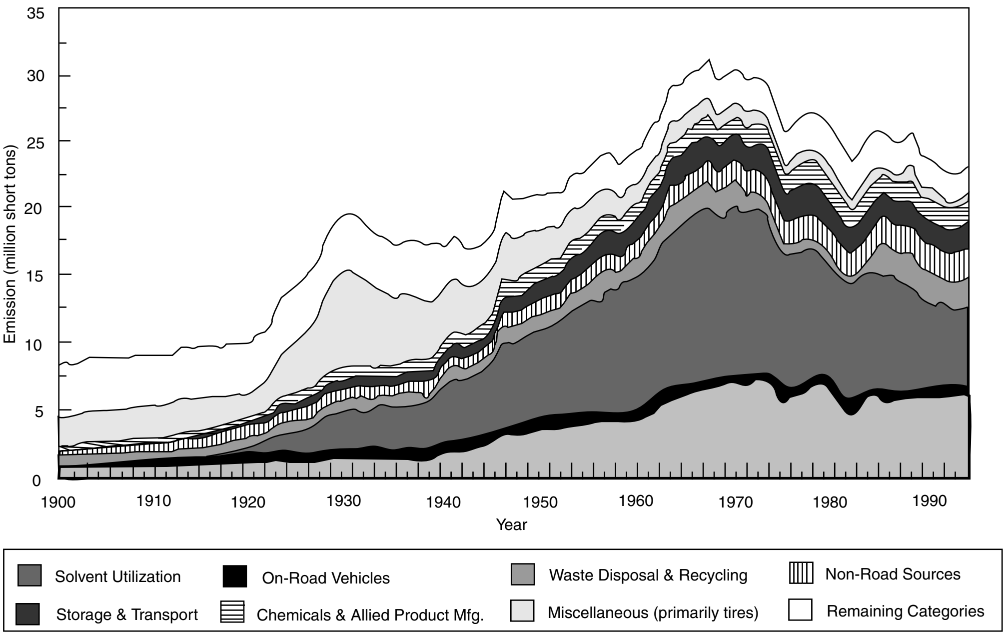


FIGURE 1 Trend in volatile organic compound emissions by seven principal source categories, 1990 to 1994.

TABLE 5
Oxygenates in Exhaust from Simple Hydrocarbon Fuel*

Oxygenate	Concentration range (ppm)
Acetaldyde	0.8–4.9
Acrolein	0.2–5.3
Benzaldehyde	<0.1–13.5
Tolualdehyde	0.1–2.6
Acetone (+ propionaldehyde)	2.3–14.0
Methyl ethyl ketone	0.1–1.0
Methyl vinyl ketone (+ benzene)	0.1–42.6
Acetophenone	<0.1–0.4
Methanol	0.1–0.6
Ethanol	<0.1–0.6
Benzofuran	<0.1–2.8
Methyl formate	<0.1–0.7
Nitromethane	<0.8–5.0

*Reference 3

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
Many of the oxygenated hydrocarbons are themselves toxic, many of them are known human carcinogens and some, especially esters, ketones, and alcohols are known to cause central nervous system disorders (narcosis, etc…)		FROM STATIONARY SOURCES: a) Design equipment to use or consume completely the processed material. b) In the surface coating industry, use a higher percent solids paint to reduce the amount of VOC.

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
2. *Aldehydes*: Aldehydes are one of a group of organic compound with the general formula R-CHO which yield acids when oxidized and alcohols when reduced. They are products of incomplete combustion of hydrocarbons and other organic materials. Formaldehyde and Acrolein-Acetaldehyde cause irritation to the mucous membranes of the eyes, nose, and other portions of the upper respiratory tract. Formaldehyde has also been cited as a potential human carcinogen.	One of the most popular aldehydes used in the chemical process industry is formaldehyde. This is because of its relatively low cost, high purity, and variety of chemical reactions. Among its many uses are as an intermediate in the production of phenolic and amino resins and also in the production of slow release fertilizers. Annual worldwide production capacity now exceeds 12 × 10^6 metrics tons (calculated as 37% solution). In general, aldehydes are produced by the combustion of fuels in motor vehicles, space heating, power generation, and in other combustion activities (such as the incineration of wastes). In addition aldehydes are formed in photochemical reactions between nitrogen oxides and certain hydrocarbons. Natural sources of aldehydes do not appear to be important contributors to air pollution. Some aldehydes are found in fruits and plants.	c) Use materials which have a higher boiling point or are less photochemically active. d) Use control equipment and recycling of organic solvents to reduce emissions. e) Control by absorption, adsorption and condensation. Control methods include more effective combustion as may be obtained in direct flame and the use of catalytic afterburners.
3. *Ethylene:* Ethylene ($H_2C = CH_2$) is the largest volume organic chemical produced today. Ethylene is a colorless hydrocarbon gas of the olefin series, it is generally not toxic to humans or animals, but it is the only hydrocarbon that has adverse effects on vegetation at ambient concentrations of 1 ppm or less. It therefore represents a considerable air pollution problem, for two reasons: 1. it is significantly harmful to plants,	Ethylene may form as a by-product of incomplete combustion of hydrocarbons and other organic substances. Thus, ethylene has been found to be one of the components of automobile and diesel combustion emissions (exhaust and blow by emissions), incinerator effluents, and agricultural waste combustion gases. Ethylene is not normally found in deposits of petroleum or natural gas.	Ethylene poses no peculiar control problem in these emissions and this can be controlled by methods generally used for hydrocarbons. These methods include combustion techniques, absorption techniques, absorption methods, and vapor recovery systems.

TABLE 6

Emissions of Hydrofluorocarbons and Prefluorinated Carbon, 1990 to 1994 Preliminary Estimates (thousand short tons; molecular basis)

Compound	GWP	1990	1991	1992	1993	1994
HFCs						
HFC-23	12,100	6.085	6.206	6.327	2.910	3.075
HFC-125	3,200	0.000	0.000	0.000	0.000	4.211
HFC-134a	1,300	0.551	0.992	1.323	6.526	11.475
HFC-125a	140	0.282	0.292	0.296	1.146	1.687
HFC-227	3,300	0.000	0.000	0.000	0.000	3.946
PFCs						
CF_4	6,300	2.701	2.701	2.701	2.695	2.695
C_2F_6	12,500	0.270	0.270	0.270	0.270	0.270
SF_6	24,900	1.102	1.102	1.102	1.102	1.135

Note(s): Totals presented in this table may not equal the sum of the individual source categories due to rounding.

Source(s): Inventory of U.S. Greenhouse Gas Emissions and Sinks, 1900–1994. Draft Report, U.S. Environmental Protection Agency. September 1995.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
2. it contributes to photochemically produced air pollution. Ethylene is the most abundant (based on mole volume) of the photoreactive hydrocarbons in the lower atmosphere.	In the chemical process industry, virtually all ethylene is consumed as feedstock for a variety of petrochemical products. Ethylene has been known to be used as a ripening agent for fruits and vegetables	
4. *Organic Carcinogens*: These are carbon compounds which cause cancer in experimental animals and are therefore suspected of playing a role in causing human cancer, particularly cancer of the lungs. There is some question as to the carcinogenicity of selected compounds. Polynuclear aromatic hydrocarbons (PAH) in our environment evolve from high-temperature reactions under pyrolytic conditions during incomplete combustion contained in some air pollution source effluents in automobile exhaust fumes, and in moderate concentrations in the air. The major classes of organic carcinogens are as follows: 1. Polynuclear aromatic hydrocarbons (PAH); Benzo-(a)-pyrene (BAP)-substance found in cigarette smoke. Benzo(e)pyrene Benzo(a)anthracene Benzo(e)acetophenthrylene Benzo(b)fluoranthene Chrysene 2. Polynuclear azo-heterocyclic compounds; Dibenz(a,h)acridine Dibenz(a,j)acrydine 3. Polynuclear imino-heterocyclic compounds 4. Polynuclear carbonyl compounds 7H-Benz(de)anthracene-7-one 5. Alkylation agents Aliphatic and alifinic epoxides Peroxide Bactones	The incomplete combustion of matter containing carbon. Heat generation (burning coal, oil and gas) accounts for more than 85%. Sources of heat generation that were tested ranged in size from residential heaters to heavy industrial power plant boilers. Municipal incinerators produce about 5% of emissions. Industrial processes also account for 5%. Organic carcinogens are primarily unwanted by-products of incomplete combustion. However, a few sources of organic carcinogens might be defined as naturally occurring. Bituminous coal contains certain organic carcinogens.	From Motor Vehicle Sources: (Same Controls as Hydrocarbons) From Stationary Sources: 1. Design equipment to use or consume completely the processed material. 2. Use of materials which have a higher boiling point or are less photochemically reactive. 3. Use of control equipment to reduce emissions. 4. Stop open burning of waste by use of multiple-chamber incinerators or disposing of waste in sanitary landfills.
5. *Halogenated Hydrocarbons*: Halogenated hydrocarbons are carbon and hydrogen compounds with one or more of the halide elements of fluorine, chlorine, bromine, or iodine. Of these elements, the most common halogenated hydrocarbons are those containing fluorine and chlorine. Halogenated hydrocarbons were once thought to solve the ozone problem because of their low reactivity. However, many of these compounds are very toxic and thus cause a more immediate threat to human health. Also, there is a great concern of damage caused by these compounds to the stratospheric ozone layer which protects us from the harmful ultraviolet radiation of the sun. These compounds tend to degrade into their elemental components, which include radical alogen, which have a great affinity for ozone.	Halogenated hydrocarbon solvent vapors include those of chloroform ($CHCl_3$), carbon tetrachloride (CCl_4), trichloroethylene (C_2HCl_3), perchloroethylene (C_2Cl_4), etc. From vapors ($CFCl_3$, C_3FCl_3) are very widely used as refrigerants and were once used as propellants. Except for the vicinity of major urban areas, atmospheric halogen concentrations are very low.	The same controls apply for halogenated hydrocarbons as for non-halogenated hydrocarbons. These are adsorption, absorption, etc. However, combustion may be undesirable since free halogen radical combining with water vapor may cause an acid problem. This may damage equipment as well as create a serious environmental problem.
6. *Pesticides*: Pesticides are economic poisons used to control or destroy pests that cause economic losses or adverse human health effects. These chemicals can be grouped as insecticides, herbicides (weed and brush killers, defoliants, and desiccants), fungicides, iscaricides, nematocides, repellants, attractants, and plant growth regulators. In the United States, 300–400 pesticides are registered for use in the production of food. These chemicals	The primary source of pesticides in air is from the application process; a certain amount of drift is unavoidable, even under normal conditions. Pesticides can evaporate into the air from soil, water and treated surfaces. Pesticides contained in dust from the soil can enter the air and be transported for considerable distances before falling back to the earth. Chemical plants manufacturing pesticides also produce pollutant emissions.	Improved application equipment and methods: Improved formulas for pesticides (higher density or use water soluble oils) Wider distribution and use of weather data in area where pesticides are used.

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
have served quite well in the past years in the prevention of famine and disease. However, it must be realized that some pesticides, especially chlorinated hydrocarbons, are metabolized very slowly thus, accumulate in adipose tissue. DDT for example, has been shown to cause tumors in laboratory animals.	Production of pesticides is estimated at 1.1×10^9 lbs.	Control and abatement during production: Venting of solid emissions through bag houses and cyclones Venting of liquid emissions through liquid scrubbers.

TABLE 7
Total National Emissions of Carbon Monoxide, 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
FUEL COMB. -ELEC. UTIL.	4	110	110	237	322	314	322	325
FUEL COMB. -INDUSTRIAL	435	549	661	770	750	677	670	671
FUEL COMB. -OTHER	14,890	10,656	6,250	3,625	6,230	4,072	3.961	3,888
Residential Wood	11,279	7,716	4,743	2,932	5,992	3,781	3,679	3,607
CHEMICAL and ALLIED PRODUCT MFG.	4,190	5,844	3,982	3,397	2,151	1,940	1,998	2,048
Other Chemical Mfg	4,139	5,760	3,775	2,866	1,417	1,522	1,574	1,619
carbon black mfg	4,139	5,760	3,775	2,866	1,417	1,126	1,170	1,207
METALS PROCESSING	2,750	2,910	2,866	3,644	2,246	2,080	2,091	2,166
Ferrous Metals Processing	2,714	2,792	2,540	2,991	1,404	1,394	1,410	1,465
gray iron cupola	1,174	1,551	1,123	1,203	340	262	261	271
PETROLEUM and RELATED INDUSTRIES	221	2,651	3,086	2,179	1,723	435	398	390
Petroleum Refineries and Related Industries	221	2,651	3,086	2,168	1,723	425	388	380
fcc units	210	2,528	2,810	1,820	1,680	389	352	344
OTHER INDUSTRIAL PROCESSES	114	231	342	620	830	717	732	751
Wood, Pulp and Paper and Publishing Products	110	220	331	610	798	657	672	689
SOLVENT UTILIZATION	NA	NA	NA	NA	NA	2	2	2
STORAGE and TRANSPORT	NA	NA	NA	NA	NA	55	56	58
WASTE DISPOSAL and RECYCLING	3,630	4,717	5,597	7,059	2,300	1,686	1,732	1,746
Incineration	2,202	2,711	2,703	2,979	1,246	849	872	879
conical wood burner	1,316	1,613	1,366	1,431	228	18	18	18
Open Burning	1,428	2,006	2,894	4,080	1,054	836	859	867
commercial/institutional	863	1,139	1,509	2,148	47	5	5	5
ON-ROAD VEHICLES	30,121	45,196	64,266	88,034	78,049	62,858	60,202	61,070
Light-Duty Gas Vehicles and Motorcycles	22,237	31,493	47,679	64,031	53,561	40,502	39,163	39,303
Light-Duty Gas Trucks	3,752	6,110	7,791	16,570	16,137	15,084	15,196	15,139
Heavy-Duty Gas Vehicles	4,132	7,537	8,557	6,712	7,189	5,930	4,476	5,244
Diesels	0	54	239	721	1,161	1,342	1,367	1,383
NON-ROAD SOURCES	8,051	11,610	11,575	10,605	12,681	14,642	15,259	15,657
Non-Road Gasoline	3,777	7,331	8,753	9,478	11,004	12,655	13,162	13,452

(*continued*)

TABLE 7 (*continued*)
Total National Emissions of Carbon Monoxide, 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
construction	1,198	2,409	2,262	250	368	395	423	453
industrial	780	1,558	1,379	732	970	1,228	1,285	1,340
lawn and garden	NA	NA	NA	4,679	5,366	6,001	6,212	6,276
farm	1,351	2,716	3,897	46	77	63	70	73
light commercial	NA	NA	NA	2,437	2,680	3,254	3,402	3,519
recreational marine vessels	60	120	518	976	1,102	1,207	1,245	1,256
Non-Road Diesel	32	53	65	543	801	841	903	954
Aircraft	4	934	1,764	506	743	966	1,019	1,063
Railroads	4,083	3,076	332	65	96	122	124	124
MISCELLANEOUS	29,210	18,135	11,010	7,909	8,344	11,173	6,700	9,245
Other Combustion	29,210	18,135	11,010	7,909	8,344	11,173	6,700	9,245
forest wildfires	25,130	11,159	4,487	5,620	5,396	6,079	1,586	4,115
TOTAL ALL SOURCES	93,615	102,609	109,745	128,079	115,625	100,650	94,133	98,017

Note(s): Categories displayed below Tier 1 do not sum to Tier 1 totals because they are intended to show major contributors.
1994 emission estimates are preliminary and will be updated in the next report.
Tier 1 source categories and emissions are shaded.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
B. INORGANIC GASES: The chemistry of the lower atmosphere is controlled by the reactivity of oxygen. In the presence of molecular oxygen (O_2), the stable forms of almost all of the elements are oxides, with the notable exception of nitrogen. Thus, many of the major pollutants are oxides (i.e., CO, SO_2, SO_3, NO, NO_2) and their associated reactive by-products. 1. *Carbon Oxides* Significant amounts of carbon oxides, carbon monoxide (CO) and carbon dioxide (CO_2), are produced by natural and anthropogenic (man made) sources. CO is considered a major atmospheric pollutant because of its significant health effects, whereas, CO_2 is a relatively non-toxic, normal tropospheric (lower atmospheric) constituent and is, therefore, not usually described as a major atmospheric pollutant. However, anthropogenic emissions of CO_2 are of significant concern since large amounts of CO_2 may contribute to global climatic warning. a. *Carbon Monoxide*:	Carbon monoxide (CO) is a colorless, odorless, tasteless gas formed by the incomplete combustion of fossil fuels and other organic matter. During combustion, carbon is oxidized to CO by the following reactions: $$2C + O_2 \longrightarrow 2CO \qquad (1)$$ $$2CO + O_2 \longrightarrow 2CO_2 \qquad (2)$$ CO, formed as an intermediate in the combustion process, is emitted if there is insufficient O_2 present for reaction (2) to proceed. CO is produced naturally by volcanic eruptions, forest fires, lightning and photochemical degradation of various reactive organic compounds. Biologically, CO is formed by certain brown algae, decomposition of chlorophyll in leaves of green plants, various micro-organisms and microbial action in the oceans. Major anthropogenic sources include transportation, industrial processing, solid waste disposal and agricultural burning. it also is present in high concentrations in cigarette smoke. Background concentrations of CO average 0.1 ppm, with peak concentrations in the northern hemisphere during the autumn months due to the decomposition of chlorophyll associated with the color change and fall of leaves. The residence time for CO in the atmosphere is estimated to be 0.1 to 0.3 years. Because CO has a higher affinity (approximately 200 × greater) for blood hemoglobin than oxygen, and also tends to remain more tightly bound, oxygen transport throughout the body	CO can be removed from the atmosphere by the actions of soil micro-organisms which convert it to CO_2. The soil in the U.S. alone is estimated to remove approximately 5×10^8 tons of CO per year, which is far in excess of the anthropogenic emission rate. However, little CO is removed in urban areas since emissions of CO are large and soil is scarce. In automobiles, catalytic convertors are used to reduce CO emissions by combusting the exhaust gases over a catalyst. This catalyst aided reaction combines O_2 with CO to produce CO_2 and water. Similar after-burner processes are used in controlling emissions from stationary sources.

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	of an individual exposed to CO can be greatly reduced. CO is highly toxic at concentrations greater than 1000 ppm. Death results from asphyxiation since body tissues, especially the brain, are deprived of a sufficient supply of oxygen. Because it is colorless, odorless and tasteless, individuals exposed to toxic concentrations are unaware of its presence. However, the concentrations of CO commonly encountered in urban environments are usually only a fraction of those levels which cause asphyxiation. Low-level CO exposure affects the central nervous system with typical behavioral changes including decreased time interval recognition, impairment of brightness, delayed reaction time to visual stimuli, decrease in drying performance and, at concentrations of 100 ppm, dizziness, headache, fatigue and loss of coordinatation. Cigarette smoke contains especially high levels of CO (15,000 to 55,000 ppm) which bind to approximately 3 to 10% of a smoker's hemoglobin. The effects of these high levels would be extremely harmful if it were not for the intermittent nature of the exposure. The inhalation of air between drags greatly reduces the toxic dose. The major effect of CO in cigarette smoke appears to be to increase the risk of angina pectoris patients to myocardial infarcation and sudden death. However, cigarette smoke contains many harmful substances and it is difficult to specifically assess the harmful effects of CO and its exact role in cardiovascular diseases.	
b. *Carbon Dioxide*:	Carbon dioxide (CO_2 is the most commonly emitted air contaminant. It is a product of the complete combustion of carbon in the presence of O_2 as shown in reactions (1) and (2) previously. CO_2 is produced naturally through the decomposition, weathering and combustion of organic matter. Human and animal respiration also contribute CO_2 to the atmosphere. The combustion of coal, oil and natural gas in both stationary and mobile sources is responsible for 90% of anthropogenic CO_2 emissions throughout the world. Solid waste disposal and agricultural burning account for the remaining 10%. Coke ovens and smelters emit significant amounts of CO_2 on a localized basis.	The oceans absorb approximately 50% of anthropogenic CO_2 emissions since CO_2 is highly soluble in water. Green plants also consume large amounts of CO_2 for use in photosynthesis. The use of alternate sources of energy such as nuclear, solar or chemically derived energy is the preferred method to control emissions of CO_2.

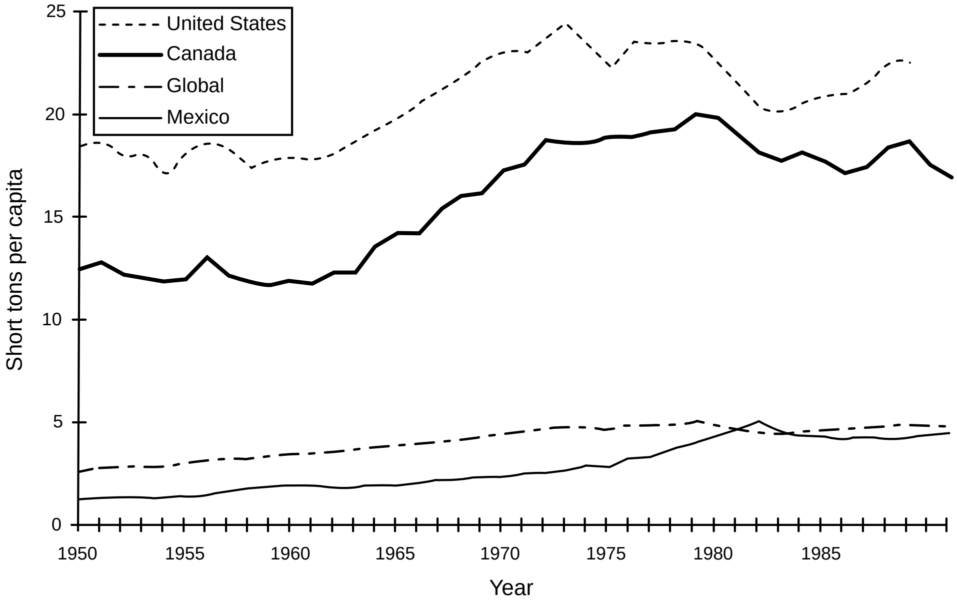


FIGURE 2 Comparison of Per Capita Carbon Dioxide emissions.

Note(s): U.S. per capita emissions data is not presented for 1990 or 1991. See section 10.1 for a discussion of 1990 to 1994 national CO_2 emission estimates.

Sources(s): Marland, G., R.J. Andres, and T.A. Boden 1994. Global, regional and national CO_2 emissions, pp. 9–88. In T.A. Boden, D.P. Kaiser, R.J. Sepanski, and F.W. Stoss (Eds.), *Trends '93: A Compendium of Data on Global Change.* ORNL/CDIAC-65. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn., U.S.A.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	CO_2 is not typically considered a pollutant in air pollution regulations, however, its role in the global heat balance is well recognized. CO_2 can heat up the earth's surface by a phenomenon commonly called the "greenhouse effect." This "greenhouse effect" is caused primarily by water vapor and CO_2, both of which are strong absorbers of infrared radiation. When radiation is absorbed by CO_2 and water, it is reemitted in all directions with the net result being that part of the radiation returns to the earth's surface and raises the temperature. Since 1890, atmospheric CO_2 levels have increased from about 290 to 322 ppm. 25% of this increase has occurred in the past decade. Since 1958, the atmospheric CO_2 levels have increased at a rate of approximately 0.7 ppm per year. If this trend continues, atmospheric CO_2 levels could double by the year 2035A.D. This doubling could result in the warming of surface temperatures by 2.4°C in the midlatitudes, with a greater warming in the polar regions.	
Sulfur Oxides a. *Sulfur Dioxide*:	Sulfur dioxide (SO_2) is a colorless gas whose odor and taste can be detected in the concentration range of 0.3 to 0.1 ppm. Above 3 ppm, it has a pungent, irritating odor. Although SO_2 emissions may occur from volcanic eruptions, most SO_2 (and sulfur trioxide, SO_3) is due to the burning of	In order to reduce the levels of sulfuric acid aerosols in urban air, power plants are often built with tall smokestacks which disperse the SO_2 over a wide area. This reduces the local problem but increases the problem for areas

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	coal and crude oils for electric power and heating. The sulfur content of refined petroleum is usually quite low. At the high temperatures of combustion, the sulfur in these fuels is converted to SO_2 by the reaction: $S + O_2 = SO_2$ (3) Background levels of SO_2 are very low, about 1 ppb. In urban areas maximum concentrations vary from less than 0.1 to over 0.5 ppm. SO_2 itself is a lung irritant and is known to be harmful to people who suffer from respiratory disease. However, it is the sulfuric acid aerosol formed from the oxidation of SO_2 and SO_3 that causes the most damaging health effects in urban areas. The sulfuric acid aerosol is formed by the following reactions which in the atmosphere are photochemically and catalytically accelerated: $2SO_2 + O_2 = 2SO_3$ (4) $SO_3 + H_2O = H_2SO_4$ (5) The sulfuric acid aerosols formed are usually less than 2 microns in diameter and can quite effectively penetrate the innermost passages of the lung, known as the pulmonary region. This is the region where O_2 is exchanged with CO in the blood. Sulfuric acid aerosols irritate the fine vessels of the pulmonary region, causing them to swell and block the vessel passages. Severe breathing impairment may occur. The effect is cumulative, with older people suffering the most severe respiratory problems. SO_2 can also severely damage crops such as spinach, turnip, beets, alfalfa and oats. Trees such as the white pine, white birch and trembling aspen, as well as, ornamental plants such as gladiolus, tulip and sweet pea, can also be damaged.	which are far from the source of the pollutant. The sulfuric acid aerosol is washed out in either rain or snowfall and increases the acidity of local waters downwind from the plant. This condition is known as acid rain. Another approach to SO_2 abatement is to substitute low sulfur coal, sulfur free coals (produced by screening crushed coal) and other sulfur free fuels for high sulfur to low sulfur fuels. This can be seen in urban areas where coal has largely been displaced by petroleum and natural gas. An alternative approach is to remove the SO_2 from the stack gases of the plant by using chemical scrubbers. In the chemical scrubber, the stack gas is passed through a slurry of limestone (calcium carbonate, $CaCO_3$) which removes the SO_2 and produces calcium sulfite which can be collected and disposed of. More commercially valuable abatement processes include catalytic oxidation to produce usable sulfuric acid and reaction with alkalized alumina which allows the recovery of usable sulfur.

TABLE 8

Total National Emissions of Sulfur Dioxide 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
FULE COMB. -ELEC. UTIL.	2,427	4,515	9,264	17,398	17,469	15,898	15,191	14,869
Coal	2,276	4,056	8,883	15,799	16,073	15,227	14,546	14,312
bituminous	1,359	2,427	5,367	9,574	NA	13,365	12,199	11,904
subbituminous	668	1,196	2,642	4,716	NA	1,425	1,796	1,854
anthracite and lignite	249	433	873	1,509	NA	438	551	555
Oil	151	459	380	1,598	1,395	639	612	523
residual	146	453	375	1,578	NA	629	602	512
FULE COMB. -INDUSTRIAL	6,060	5,725	3,864	4,568	2,951	3,106	2,942	3,029
Coal	5,188	4,423	2,703	3,129	1,527	1,843	1,661	1,715
bituminous	3,473	2,945	1,858	2,171	1,058	1,382	1,248	1,289
subbituminous	1,070	907	574	669	326	29	26	26
anthracite and lignite	645	571	272	289	144	81	72	75

(*continued*)

TABLE 8 (*continued*)
Total National Emissions of Sulfur Dioxide 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
Oil	554	972	922	1,229	1,065	823	848	882
residual	397	721	663	956	851	633	662	692
Gas	145	180	189	140	299	352	346	345
FULE COMB. -OTHER	3,642	3,964	2,319	1,490	971	595	599	599
Commercial/Institutional Coal	695	1,212	154	109	110	176	171	169
Commercial/Institutional Oil	407	658	905	883	637	233	241	242
Residential Other	2,517	2,079	1,250	492	211	175	178	177
distillate oil	60	163	295	212	157	137	145	145
bituminous/subbituminous coal	2,267	1,758	868	260	43	30	25	25
CHEMICAL and ALLIED PRODUCT MFG.	215	427	447	591	280	440	450	457
Inorganic Chemical Mfg	215	427	447	591	271	333	341	345
sulfur compounds	215	427	447	591	271	325	332	336
METALS PROCESSING	3,309	3,747	3,986	4,775	1,842	663	667	692
Nonferrous Metals Processing	2,760	3,092	3,322	4,060	1,279	486	488	506
copper	2,292	2,369	2,772	3,507	1,080	300	300	312
lead	80	95	57	77	34	112	114	119
aluminum	4	28	38	80	95	60	60	62
Ferrous Metals Processing	550	655	664	715	562	160	162	168
PETROLEUM and RELATED INDUSTRIES	224	340	676	881	734	440	409	406
OTHER INDUSTRIAL PROCESSES	334	596	671	846	918	401	413	431
Wood, Pulp and Paper, and Publishing Products	0	43	114	169	223	137	141	145
Mineral Products	334	553	557	677	694	257	265	279
cement mfg	318	522	524	618	630	169	176	186
SOLVENT UTILIZATION	NA	NA	NA	NA	NA	1	1	1
STORAGE and TRANSPORT	NA	NA	NA	NA	NA	5	5	5
WASTE DISPOSAL AND RECYCLING	3	3	10	8	33	36	37	37
ON-ROAD VEHICLES	3	103	114	411	521	571	517	295
NON-ROAD SOURCES	3,190	2,392	321	83	175	265	278	283
Marine Vessels	215	215	105	43	117	190	201	206
Railroads	2,975	2,174	215	36	53	68	69	69
MISCELLANEOUS	545	545	554	110	11	14	8	14
Other Combustion	545	545	554	110	11	14	8	14
TOTAL ALL SOURCES	19,953	22,358	22,227	31,161	25,905	22,433	21,517	21,118

Note(s): Categories displayed below Tier 1 do not sum to Tier 1 totals because they are intended to show major contributors.
1994 emission estimates are preliminary and will be updated in the next report.
Tier 1 source categories and emissions are shaded.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
b. *Hydrogen Sulfide*:	Hydrogen sulfide (H_2S) is a colorless gas known by its characteristic rotten egg odor. Natural sources of H_2S include volcanic eruptions, geothermal wells and chemical or bacteriological decomposition of mineral sulfates in springs and lakes. In these natural occurances, other sulfur compounds are nearly always present with the H_2S. Anthropogenic sources include the combustion of coal, natural gas and oil. The refining of petroleum products, coke production, sulfur recovery operations and the kraft process for producing chemical pulp from wood are all major sources of H_2S. The typical rotten egg odor can be detected at very low concentrations, 0.025 to 0.2 ppm, but at these concentrations it has little or no effect upon human health. However, at higher concentrations, H_2S is extremely toxic. Above 150 ppm, the human olfactory apparatus becomes paralyzed, effectively preventing any olfactory warning signal. H_2S is life threatening at 300 ppm since it causes pulmonary edema. At 500 ppm, there is strong stimulation to the nervous system. Above 1000 ppm, there is immediate collapse and respiratory paralysis.	Most removal system for H_2S scrub the gas streams with a suitable absorbent and then remove the absorbed gas from the absorbent for disposal by burning or conversion to usable by-products. Different types of scrubbers can be used such as spray towers, plate towers and venturi scrubbers. Natural removal of H_2S occurs by atmospheric conversion to SO_2 which is subsequently removed from the atmosphere through precipitation and absorption by surfaces and vegetation.
3. *Nitrogen Compounds*: There are five major gaseous forms of nitrogen in the atmosphere: nitrogen gas (N_2), ammonia (NH_3), nitrous oxide (N_3O), nitric oxide (NO), and nitrogen dioxide (NO_2). N_2 is the major gaseous component in the atmosphere and counts for 78% of the atmosphere's mass. NO and NO_2 are important pollutants of the lower atmosphere and because of their interconvertibility in photochemical reactions, are usually collectively grouped as NO_x.	Nitrous oxide (N_2O) is a colorless, slightly sweet, non-toxic gas. It is probably best known as the "laughing gas" which is widely used as an anesthetic in medicine and dentistry. Bacterial action which produces N_2O is the largest single source of any nitrogen oxide on a worldwide basis. It is present in the atmosphere at an average concentration of 0.27 ppm. It is quite inert in the lower atmosphere, but it can react with oxygen atoms that are available in the stratosphere to produce nitric oxide.	
a. *Nitrous Oxide*:		
b. *Nitric Oxide*:	Nitric oxide (NO) is a colorless, odorless, tasteless, relatively non-toxic gas. Natural sources include anaerobic biological processes in soil and water, combustion processes and photochemical destruction of nitrogen compounds in the stratosphere. On a worldwide basis, natural emissions of NO are estimated at approximately 5×10^8 tons per year. Major anthropogenic sources include automobile exhaust, fossil fuel fired electric generating stations, industrial boilers, incinerators, and home space heaters. All of these sources are high temperature combustion processes which follow the reaction: $N_2 + O_2 = 2NO \qquad (6)$ This reaction is endothermic, which means that the equilibrium shifts to the right at high temperatures and to the left at low temperatures. Therefore, as the combustion temperature of a process increases, so will the amount of CO emitted. Background concentrations of NO are approximately 0.5 ppb. In urban areas, one hour average concentrations of NO may reach 1 to 2 ppm. Atmospheric levels of CO are related to the transportation and work cycle, with the highest	

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	concentrations observed during the morning and evening rush hours. Emissions of NO are also greater in the winter months since there is an increase in the use of heating fuels. NO is a relatively non-irritating gas and is considered to pose no health threat at ambient levels. It is rapidly oxidized to nitrogen dioxide, which has a much higher toxicity.	

TABLE 9

Total National Emissions of Nitrogen Oxides, 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
FUEL COMB. -ELEC. UTIL.	660	1,316	2,536	4,900	7,024	7,516	7,773	7,795
Coal	467	1,118	2,038	3,888	6,123	6,698	7,008	7,007
bituminous	255	584	1,154	2,112	3,439	4,600	4,535	4,497
subbituminous	125	288	568	1,041	1,694	1,692	2,054	2,098
Oil	193	198	498	1,012	901	210	169	151
FUEL COMB. -INDUSTRIAL	2,543	3,192	4,075	4,325	3,555	3,256	3,197	3,206
Coal	2,012	1,076	782	771	444	613	550	568
bituminous	1,301	688	533	532	306	445	399	412
Gas	365	1,756	2,954	3,060	2,619	1,656	1,650	1,634
natural	337	1,692	2,846	3,053	2,469	1,436	1,440	1,427
FUEL COMB. -OTHER	529	647	760	836	741	712	726	727
Residential Other	177	227	362	439	356	352	363	364
CHEMICAL and ALLIED PRODUCT MFG.	6	63	110	271	216	276	286	291
METALS PROCESSING	4	110	110	77	65	81	81	84
Ferrous Metals Processing	4	110	110	77	65	53	54	56
PETROLEUM and RELATED INDUSTRIES	105	110	220	240	72	100	95	95
OTHER INDUSTRIAL PROCESSES	107	93	131	187	205	306	315	328
Mineral Products	105	89	123	169	181	216	222	234
cement mfg	32	55	78	97	98	121	124	131
SOLVENT UTILIZATION	NA	NA	NA	NA	NA	2	3	3
STORAGE and TRANSPORT	NA	NA	NA	NA	NA	2	3	3
WASTE DISPOSAL and RECYCLING	110	215	331	440	111	82	84	85
ON-ROAD VEHICLES	1,330	2,143	3,982	7,390	8,621	7,488	7,510	7,530
Light-Duty Gas Vehicles and Motorcycles	970	1,415	2,607	4,158	4,421	3,437	3,680	3,750
light-duty gas vehicles	970	1,415	2,606	4,156	4,416	3,425	3,668	3,737
Light-Duty Gas Trucks	204	339	525	1,278	1,408	1,341	1,420	1,432
light-duty gas trucks 1	132	219	339	725	864	780	828	830
light-duty gas trucks 2	73	120	186	553	544	561	592	603
Heavy-Duty Gas Vehicles	155	296	363	278	300	335	315	333
Diesels	NA	93	487	1,676	2,493	2,375	2,094	2,015
heavy-duty diesel vehicles	NA	93	487	1,676	2,463	2,332	2,047	1,966
NON-ROAD SOURCES	991	1,538	1,443	1,628	2,423	2,843	2,985	3,095
Non-Road Gasoline	122	249	312	81	102	124	122	125
Non-Road Diesel	103	187	247	941	1,374	1,478	1,433	1,494
construction	70	158	157	599	854	944	1,007	1,076

(*continued*)

TABLE 9 (*continued*)
Total National Emissions of Nitrogen Oxides, 1940 through 1994 (thousand short tons)

Source Category	1940	1940	1940	1940	1940	1940	1940	1940
industrial	NA	NA	40	75	99	125	131	136
farm	33	29	50	166	280	230	256	265
airport service	NA	NA	NA	78	113	144	152	159
Aircraft	0	2	4	72	106	139	147	153
Marine Vessels	109	108	108	40	110	173	183	188
Railroads	657	992	772	495	731	929	945	947
MISCELLANEOUS	990	665	441	330	248	373	219	374
TOTAL ALL SOURCES	7,374	10,093	14,140	20,625	23,281	23,038	23,276	23,615

Note(s): Categories displayed below Tier 1 do not sum to Tier 1 totals because they are intended to show major contributors.
1994 emission estimates are preliminary and will be updated in the next report.
Tier 1 source categories and emissions are shaded.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
c. *Nitrogen dioxide*:	Nitrogen dioxide (NO_2) is a colored gas which is a light yellowish orange at low concentrations and reddish brown at high concentrations. It has a pungent, irritating odor. It is relatively toxic and has a rapid oxidation rate which makes it highly corrosive as well. The oxidation of NO to NO_2 follows the reaction: $2NO + O_2 \rightarrow 2NO_2 \quad (7)$ This reaction is slow at low atmospheric levels and accounts for about 25% of all NO conversion. The major NO conversion processes are photochemical, involving hydrocarbons, ozone, aldehydes, carbon monoxide, and other compounds. Background concentrations of NO_2 are approximately 0.5 ppb with one hour average concentrations in urban areas of 0.5 ppm. Peak morning concentrations of NO are followed several hours later by peak levels of NO_2 produced by the chemical and photochemical oxidation of the NO. Since the conversion of NO to NO_2 is related to solar intensity, more NO_2 is produced on warm, sunny days.	In the atmosphere, NO_2 can be photochemically oxidized to nitrates which are subsequently removed by precipitation, dry deposition and surface absorption. In motor vehicles, current methods for controlling NO_x emissions include retardation of spark timing, increasing the air/fuel ratio (i.e., less fuel to air), injecting water into the cylinders, decreasing the compression ratio, and recirculating exhaust gas. All these methods reduce the combustion chamber temperature (which reduces NO_x emissions) without greatly increasing the emissions of hydrocarbons and CO. Catalytic convertors which reduce NO to elemental nitrogen (N_2) can also be used. The use of alternative fuels, such as methyl and ethyl alcohol, which combust at a lower temperature than gasoline can also be used to lower NO_x emissions. For stationary sources, one abatement method is to use a lower NO_x producing fuel; emissions are highest from coal, intermediate with oil and lowest with natural gas. For the numerous methods of control see the article "Nitrogen Oxides" in this Encyclopedia.
4. *Photochemical Oxidants*: Photochemical oxidants are secondary pollutants which result from a complex series of atmospheric actions involving organic pollutants, NO_x, O_2 and sunlight. The main photo-chemical oxidants are ozone, NO_2 (covered in the section on nitrogen compounds) and, to a lesser extent, peroxyacetylnitrate.	Ozone (O_3) is the most important and widely reported of the photochemical oxidants. It is a bluish gas that is 1.6 times heavier than oxygen and is normally found at elevated levels in the stratosphere where it functions to absorb harmful ultraviolet radiation. Ground level ozone is one of the major constituents of photochemical "smog" which is a widespread, urban phenomenon. It is formed when nitrogen dioxide absorbs ultraviolet light energy and dissociates into nitric oxide and an oxygen atom: $NO_2 + h\nu \rightarrow O + NO \quad (8)$	Abatement is achieved through the control of hydrocarbons and nitrogen oxides as discussed in other sections of this chapter.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	These oxygen atoms, for the most part, react with oxygen to form ozone: $O + O_2 \rightarrow O_3 \quad (9)$ In addition, the oxygen atoms can react with certain hydrocarbons to form free radical intermediates and various products such as peroxyacetylnitrate (PAN). Since photochemical oxidants are secondary pollutants formed in the atmosphere as the result of primary pollutants reacting, their concentration in the atmosphere will vary proportionally to the amount of hydrocarbons and NO_2 in the air and the intensity of sunlight. PAN is a very potent eye irritant in addition to being a strong lung irritant like O_3. O_3 is relatively insoluble in respiratory fluids and can be transported into the pulmonary system where it can damage the central airways and terminal pulmonary units such as the respiratory bronchioles and alveolar ducts. Exposure in excess of ambient levels affects lung function causing increased respiratory rates and decreased lung capacity. These effects are more pronounced in smokers and during exercise. Prolonged low-level exposure may result in decreased lung elasticity. Studies on micro-organisms, plants mutagenic, that is, it can cause permanent, inheritable changes in genes. Since mutagens and carcinogens appear to be related, it is possible that O_3 is also carcinogenic.	

(*continued*)

TABLE 10

Summary of U.S. Nitrous Oxide Emissions by Source Category, 1990 to 1994 Preliminary Estimates (thousand short tons)

Source Category	1990	1991	1992	1993	1994
AGRICULTURE					
Crop Waste Burning	4	4	5	4	5
Fertilizers	204	208	210	209	232
Total Agriculture	208	212	215	213	238
MOBILE SOURCE COMBUSTION	108	110	113	115	117
STATIONARY COMBUSTION	39	38	39	39	40
INDUSTRIAL PROCESSES					
Adipic Acid Production	62	65	60	64	68
Nitric Acid Production	44	44	44	45	49
Total Industrial Processes	106	109	104	109	117
TOTAL EMISSIONS	461	465	471	476	512

Note(s): Totals presented in this table may not equal the sum of the individual source categories due to rounding.

Source(s): Inventory of U.S. Greenhouse Gas Emissions and Sinks, 1990–1994. Draft Report, U.S. Environmental Protection Agency. September 1995.

TABLE 11
Ozone Levels Generated in Photoxidation* of various Hydrocarbons with Oxides of Nitrogen

Hydrocarbon	Ozone Level (ppm)	Time (min)
Isobutene	1.00	28
2-Methyl-1,3-butadiene	0.80	45
trans-2-Butene	0.73	35
3-Heptene	0.72	60
2-Ethyl-1-butene	0.72	80
1,3-Pentadiene	0.70	45
Propylene	0.68	75
1,3-Butadiene	0.65	45
2,3-Dimethyl-1,3-butadiene	0.65	45
2,3-Dimethyl-2-butene	0.64	70
1-Pentene	0.62	45
1-Butene	0.58	45
cis-2-Butene	0.55	35
2,4,4-Trimethyl-2-pentene	0.55	50
1,5-Hexadiene	0.52	85
2-Methylpentane	0.50	170
1,5-Cyclooctodiene	0.48	65
Cyclohexene	0.45	35
2-Methylhepatane	0.45	180
2-Methyl-2-butene	0.45	38
2,2,4-Trimethylpentane	0.26	80
3-Methylpentane	0.22	100
1,2-Butadiene	0.20	60
Cyclohexane	0.20	80
Pentane	0.18	100
Methane	0.0	—

* Reference 10.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
Halides a. *Chlorine*:	Chlorine (Cl_2) is a dense, greenish-yellow gas with a distinctive irritating odor. The major anthropogenic sources of chlorine emissions include the chemical decomposition of chlorofluorocarbons (CFCs) used as a refrigerant and propellant in consumer goods, the liquifaction of chlorine cell gas, the loading and cleaning of tank cars, barges and cylinders, dechlorination of spent brine solutions and power or equipment failure. Due to the high reactivity of Cl_2 with many substances, natural emissions of Cl_2 gas are very rare. Volcanic gases contain very small amounts of Cl_2. Low concentrations of Cl_2 may, however, be formed by atmospheric reactions. Since chlorine has strong oxidizing and bleaching properties, it is extremely hazardous to all life forms, as well as corrosive to metals and other materials. Chlorine atoms can destroy ozone	The use of propellants which do not contain CFCs. Industrial emissions can be controlled by the use of scrubbing systems, i.e., water scrubbers, alkali scrubbers and carbon tetrachloride scrubbers.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	molecules and, thus, deplete the earth's protective ozone layer. This stratospheric ozone depletion is a result of the photolytic destruction of CFCs and, subsequent, release of chlorine atoms in the middle stratosphere. Chlorine and ozone react by the reactions: $\mathrm{Cl} + \mathrm{O}_3 \rightarrow \mathrm{ClO} + \mathrm{O}_2 \quad (10)$ $\mathrm{ClO} + \mathrm{O} \rightarrow \mathrm{Cl} + \mathrm{O}_2 \quad (11)$ In these reactions, chlorine acts as a catalyst since it is rapidly regenerated by reaction 11. Estimates have shown that one chlorine atom has the potential to destroy 100,000 ozone molecules before the chlorine atom reacts with hydrogen to form hydrochloric acid and be removed from the cycle.	
b. *Fluorides:*	Fluorine is the 13th element in order of abundance and exists in nature primarily as fluorospar and fluorspatite which contain 49% and 3–4% fluorine, respectively. Fluorospar is the source of nearly all commercially used fluorine. Fluorspatite is also known as phosphate rock and is used in the manufacture of phosphate fertilizers and elemental phosphorous compounds comprising of fluorine. It may occur in extremely low concentrations in the atmosphere as solid particles (sodium and calcium fluoride) or highly irritating and toxic gases (hydrofluoric acid). The processing of fluorospar and fluorspatite are the predominate sources of fluorine air pollutants. Industrial plants manufacturing steel, glass, brick and tile, are among the major emitters. The combustion of coal is another source.	Scrubbers, electrostatic precipitators or baghouses can be used to remove particle emissions while scrubbers can be used to clean gaseous emissions. Most industrial processes require the use of both.

TABLE 12

Total National Emissions of Particulate Matter (PM-10), 1940 through 1994 (thousand short tons)

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
FUEL COMB. -ELEC. UTIL.	962	1,467	2,117	1,775	879	282	268	266
Coal	954	1,439	2,092	1,680	796	269	255	254
bituminous	573	865	1,288	1,041	483	187	184	182
FUEL COMB. -INDUSTRIAL	708	604	331	641	679	240	234	237
FUEL COMB. -OTHER	2,338	1,674	1,113	455	887	553	539	529
Commercial/Institutional Coal	235	224	21	13	8	14	13	13
Residential Wood	1,716	1,128	850	384	818	501	488	478
Residential Other	368	288	194	3	27	18	18	18
CHEMICAL and ALLIED PRODUCT MFG.	330	455	309	235	148	62	63	64
METALS PROCESSING	1,208	1,027	1,026	1,316	622	136	136	141
Nonferrous Metals Processing	588	346	375	593	130	45	45	46
copper	217	105	122	343	32	3	3	3
Ferrous Metals Processing	246	427	214	198	322	86	87	90
Metals Processing NEC	374	254	437	525	170	4	4	5
PETROLEUM and RELATED INDUSTRIES	366	412	689	286	138	28	27	26
Asphalt Manufacturing	364	389	639	217	97	4	4	4

(*continued*)

TABLE 12 (*continued*)
Total National Emissions of Particulate Matter (PM-10), 1940 through 1994 (thousand short tons))

Source Category	1940	1950	1960	1970	1980	1990	1993	1994
OTHER INDUSTRIAL PROCESSES	3,996	6,954	7,211	5,832	1,846	374	377	390
Agriculture, Food, and Kindred Products	784	696	691	485	402	30	31	32
Wood, Pulp and Paper, and Publishing Products	511	798	958	727	183	104	107	111
sulfate (kraft) pulping	470	729	886	668	142	69	71	73
Mineral Products	2,701	5,460	5,563	4,620	1,261	212	211	220
cement mfg	1,363	1,998	2,014	1,731	417	32	33	35
surface mining	62	108	140	134	127	17	17	17
stone quarrying/processing	482	663	1,039	957	421	84	80	83
SOLVENT UTILIZATION	NA	NA	NA	NA	NA	2	2	2
STORAGE and TRANSPORT	NA	NA	NA	NA	NA	57	57	59
WASTE DISPOSAL and RECYCLING	392	505	764	999	273	242	248	250
ON-ROAD VEHICLES	210	314	554	443	397	357	321	311
Diesels	NA	9	15	136	208	250	215	206
NON-ROAD SOURCES	2,480	1,788	201	223	329	372	395	411
Railroads	2,464	1,742	110	25	37	47	48	48
NATURAL SOURCES-wind erosion	NA	NA	NA	NA	NA	4,362	1,978	2,593
MISCELLANEOUS	2,968	1,934	1,244	839	852	36,267	37,905	40,150
Agriculture and Forestry	NA	NA	NA	NA	NA	7,364	7,231	7,121
agricultural crops	NA	NA	NA	NA	NA	6,983	6,837	6,716
agricultural livestock	NA	NA	NA	NA	NA	381	394	405
Other Combustion	2,968	1,934	1,244	839	852	1,178	743	1,017
wildfires	2,179	1,063	428	385	514	590	152	424
managed burning	591	662	606	390	315	529	532	535
Fugitive Dust	NA	NA	NA	NA	NA	27,725	29,930	32,012
unpaved roads	NA	NA	NA	NA	NA	11,338	12,482	12,883
paved roads	NA	NA	NA	NA	NA	5,992	6,095	6,358
other	NA	NA	NA	NA	NA	10,396	11,353	12,771
TOTAL ALL SOURCES	15,956	17,133	15,558	13,044	7,050	43,333	42,548	45,431

Note(s): Categories displayed below Tier 1 do not sum to Tier 1 totals because they are intended to show major contributors.
1994 emission estimates are preliminary and will be updated in the next report.
Tier 1 source categories and emissions are shaded.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
C. PARTICULATES: Particulates are dispersed solid or liquid matter in which the industrial aggregates are larger than single small molecules (about 0.0002 microns in diameter) but smaller than 500 microns. Particulates in the atmosphere range from about 0.1 microns to 10 microns. In general, the smaller particles are quite abundant while the larger particles exist in the atmosphere in very low concentrations. Particulates can remain airborne from a few seconds to several months. Typically, the particulate pollutant category is made up of the products of incomplete fuel combustion, metals, large ions or salts, mists, fumes fugitive dusts and various other solid or liquid particles, for example, acid mist. Small particulates can cause lung irritation and reduce respiratory efficiency by inhibiting the transport of	Sources due to the activities of man include factories such as kraft pulp paper mills, steel mills, and power plants. Mobile sources include the incomplete combustion of fuel in the internal combustion engine, primarily the diesel engine. In many rural areas the wood-burning stove has made a large contribution to airborne particulates. This category includes some compounds which are gaseous while contained, but which condense when they enter into the atmosphere. Included are: aerosols (solids and liquids of microscopic size which are dissolved in gas, forming smoke, fog or mist), large particles and dust, soot (carbon particles impregnated with tar), oil and grease.	Stationary Sources: a) Use of air cleaning techniques and devices by industry and power plants to remove particulate: — Inertial separations or gravitational settling chambers. — Cyclones. — Baghouses and fabric filters. — Electrostatic precipitators. — Scrubbers and venturi scrubbers. b) Control of construction and demolition in the grading of earth, paving roads and parking lots, sand blasting, spray-painting. Techniques include hooding and venting, to air

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
oxygen from the lungs through the circulatory system. Small particulates are also detrimental to health by having adsorbed toxic materials on their surfaces; the particulates are then inhaled into the body. Particulates are also responsible for soiling of materials and reduced visibility. In July 1987, the U.S. Environmental Protection Agency promulgated revised national ambient air quality standard for particulate matter. The new standard placed emphasis on particles less than10 microns in diameter. This revision was based on the finding that fine particulates of less than 10 microns (also known as PM-10) pose a greater hazard to human health than larger particles, because it is these smaller particles that penetrate deep into the lungs. In addition, because of their ability to remain airborne and their refractive properties, the smaller particles also have a greater impact on visibility. In July 1997, based on studies which indicated adverse health effects from the inhalation of very fine particles, the U.S. EPA promulgated a PM-2.5 standard.	Naturally occurring sources of particulates are due to forest fires and windblown dust. Mechanical processes such as wind erosion, grinding, spraying, demolition, industrial activity and salt also contribute to particulate problems. Most of these particulates are in the 1–10 micron range and generally predominate very near the source. Electricity generation, forest product industries, agriculture and its related operations, the crushed stone industry, the cement industry, the iron and steel industry and asbestos mining are other important examples. Surface coating sources emit spray and mist pollutants. These pollutants include organic solvent bases that are used in paints. These volatile organic solvents become airborne during the application of paints to their intended surface.	pollution control equipment and the wetting down of working surfaces with water or oil. c) Disposal of solid waste by sanitary land fill, composting, shredding and grinding rather than incineration. Mobile Sources: The aim is to develop methods of achieving complete combustion. If this is accomplished, particulates (like soot and smoke) would be minimal. To achieve maximum combustion, vehicles in the United States are equipped with catalytic converters which help to completely incinerate unburned fuel. In the U.S. and in many other countries like Canada, Britain and Germany unleaded gasoline is available for use in automobiles. Less lead in the gasoline means less lead particles being emitted into the air. The following are examples of some typical particulate pollutants.

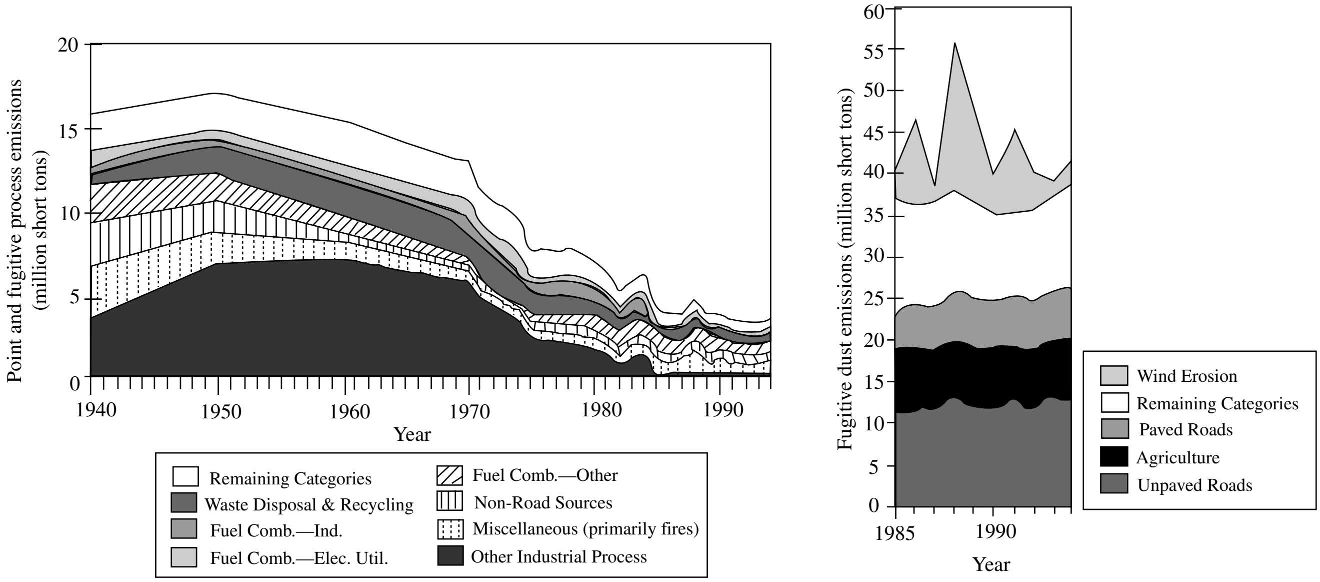


FIGURE 3 Trend in particulate Matter (PM-10) by point and fugitive process sources (1940 to 1994), and by fugitive dust sources (1985 to 1994).

Pollutant types	Sources and abundance	Abatement and control
1. *Aeroallergens*: Aeroallergens (pollens) are airborne materials that elicit a hypersensitivity or allergic response in susceptible individuals. The most common aeroallergens are the pollens of wind-pollinated plants—especially ragweed pollen, which is the main cause of hay fever. In addition to the pollens, aeroallergens include molds, danders, house, cosmetics, and others. It has been estimated that	Most aeroallergens are produced by natural causes, although some may be produced through man-made interferences 1) *Natural sources*. The aeroallergens encompass a wide variety of materials, but pollens are the most important member of this group.	Abatement and control measures for aeroallergens have been directed primarily at the ragweed. Since ragweed grows quickly in areas where the soil has been disturbed, it is not controlled by pulling it up when noticed, since the soil is thus disturbed and the growth may be heavier the following year.

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
between 10 and 15 million people in the United States are affected by seasonal allergic (hay-fever).	a) *Ragweed*—has been found in all 50 states, it produces large quantities of pollen, and the grains are especially adapted for aerial dissemination by virtue of their size (20 μ), shape, and density. It has been estimated that an acre of giant ragweed may produce as much as 50 lbs of pollen during a single season. b) *Fungi*—(molds) usually habitating in soil and dust, can become a menace when airborne. Their concentration in the air is dependent upon the magnitude of the source, their death rate in the air, humidity, temperature and other factors. c) *Danders*—(small particulate organic materials), including feathers of fowl and hair of animals and house dust. 2) *Man-made sources*: a) Flour mills—grain dusts produced in flour-milling plants (have been identified as a cause of asthma). b) Castor bean dust-oil processing plants. Most sources of biological aerosols are natural.	*Herbicide* (*plant killers*)—are sometimes used, but they are not only to ragweed, but to all plants. For eradicating molds, a number of disinfectants have been utilized. Man-made sources are subject to normal particulate control methods as well as good housekeeping practices in plants.
2. *Asbestos*: General name given to a variety of fibrous minerals found in rock masses. The value of asbestos ensues from the indestructible nature of products fabricated from the various grades of mineral fibers. *The major asbestos minerals are*: (Pyroxenes) chrysolite (amphiboles—), crocidolite, amosite, and anthophyllite. Tremolite and actinolite are considerably less important. Over 90% of the asbestos is chrysolite.	*Major sources are*: a) Asbestos mines and factories. b) The wearing of brake linings, roofing insulation and shingles. c) Fireproofing of buildings with sprayed asbestos applications. d) Road surfacing. e) Asbestos cement. f) Asbestos removal.	a) IN MANUFACTURING: Ventilation through fabric sleeve filters carrying out some operations (such as spinning and weaving of asbestos fabrics) as wet processes to eliminate dust. b) IN TRANSPORTATION: Use of plastic-coated bags to transport asbestos. c) IN CONSTRUCTION REMOVAL: Use of insulators to enclose the work area when asbestos fire-proofing is blown onto steel frames. Wetting of asbestos prior to removal.
3. *Non metallic elements*: a. BORON: A non-metallic chemical element which occurs only in combination with other elements as with sodium and other elements (as with sodium and oxygen in borax). Most important pollutants are boron dust and borane fuel. The borones are the most highly toxic of the boron compounds, consists chiefly of pentaborane, decaborane, and diborane.	*Major sources are*: Rocket motor or jet engines which use borane, a compound of boron, for a high energy fuel; combination of petroleum fuels which contain boron as an additive; burning of coal containing boron; manufacturing processes employed to produced boron compounds which are used as wastes softness. *Natural abundance*: Boron is widely distributed in nature, but constitutes only an estimated in 0.001% of the earth's crust. It is present in sea water, and is an essential constituent of a number of rock-forming silicate minerals, such as datolite and tourmaline. Boron occurs naturally only in combined forms, usually as air alkaline earth borate or as boric acid.	1) Prevention of accidental spilling of fuels. 2) Reduction or elimination of boron additives in vehicle fuels.
b. PHOSPHORUS: A solid non-metallic element existing in at least two allotropic forms, one yellow (poisonous, inflammable, and luminous in the dark), the other red (less poisonous, and less inflammable). Elemental phosphorus (yellow) is a protoplasmic poison. Some of its compounds, especially organic phosphates, can also be lethal to man and animal in the case of exposure to high air concentrations.	The compounds known to be emitted in appreciable quantities into the ambient air are phosphorus oxides, phosphoric acid, mostly in agricultural chemicals. Other organic phosphorus compounds are very probably emitted into the ambient air by the chemical industry from processes in which phosphorous products are intermediate or final outputs.	*Major control methods*: Scrubbers cyclones, fiber mist eliminators, high energy wire-mesh contactors and electrostatic precipitators are used in the control of phosphorus emissions.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	Major sources: 1) Oil-fired boilers—0.9% phosphorus in fly ash. 2) Iron and steel industry—phosphorus pentoxide accounts for an average of 0.2% of the total weight of fume from furnances. 3) Transportation sources: Organophosphorus compounds used as fuel additives. *Natural abundance*: Natural phosphates are divided into three classes on the basis of the metal or metals to which it is bound. The three major classes are aluminium (iron) phosphates, calcium-aluminum (iron) phosphates, and calcium phosphates.	
c. SELENIUM: This is a non-metallic element chemically resembling sulfur and tellurium, occurring in several allotropic forms. The soils of the midwestern U.S. are particularly high in selenium content. Selenium has also been found to be an essential nutrient for animals and may be necessary for humans.	*Major sources*: The sources of atmospheric selenium are believed to be terrestrial, such as fuels and ores used by industry (copper refinery), or possibly the burning of trash, particularly paper. *Natural abundance*: In nature, selenium is widely distributed in the earth's crust at a concentration of about 0.09 ppm. Selenium can also be found in coal and igneous rock. Approximately 0.001 $\mu g/m^3$ of selenium has been found in samples of rain, snow, and air.	No study has been made of the methods for control of selenium and its compounds. However, based on the properties and on the methods of recovery and purification of selenium wet scrubbers and high-voltage electrostatic precipitators should be effective.
4. *Heavy metals*. These are the chemically inert electronegative metals with densities of 5 gm/cm^3 and greater. They are chemically inert because their electrons are tightly bound to the metal's nuclei and are unable to readily combine with other elements. However, heavy metals are toxic to the human physiology in that the heavy metals try to bond with enzymatic sulfur atoms in the body. They also attack free amino ($-NH_2$) and organic acid (–COOH) groups found in proteins. A few heavy metals combine with and precipitate some of the body's vital phosphate compounds; other heavy metals catalyze the decomposition of phosphates. Mercury, lead, cadmium, copper combine with all membranes and interfere with the transport of chemicals in and out of the cell. As a result of this heavy metal combination with the tissue, a variety of diseases ranging from cancer to heart disease occurs. The following is a list of metals that are considered to be most detrimental to human health. a. MERCURY: A high density, silver-white metal, is liquid at normal ambient temperatures. Although it is contained in at least 25 minerals, the chief source is cinnabar (HgS).	*Major sources*: The combustion of coal was the largest source of mercury emissions in the United States in 1968. The paint industry was the source of more than 25% of lead emissions in the U.S. in 1968. Marine anti-fouling paints contain mercurial compounds as a toxicant; latex paints use mercurial compounds as a preservative. The third largest source of mercury emissions is the combustion of wastes which accounted for almost 17% of all mercury emissions in 1968. Other sources of mercury include mining and processing of ore. In industrial applications mercury is used in rectifiers, mercury precision lighting, batteries (mercury cell and alkaline energy cell). Laboratory equipment and instruments—such as barometers, thermometers, flow meters, pressure-sensing devices, switches and relays all contain mercury (spillage creates droplets which vaporize). Electrolytic preparation of chlorine. Agricultural use of mercury compounds as pesticides (now declining).	For applications which use mercury at normal temperatures: 1) Proper ventilation in work areas. 2) Cleaning up spilled mercury (sweeping with special vacuum cleaners or chemical treatment). 3) Use of non-porous material for floors, working surfaces and protective clothing. 4) Conventional control of pesticides. For applications which use mercury at high temperatures: Condensing mercury vapors by: Cold-water jacketed condensers, impregnated charcoal. Water scrubbers.
b. LEAD: Lead is a heavy, soft malleable, bluish-gray metallic element. It is the sulfide ore, galena, soil, water, vegetation and animal. It is introduced into the body with the intake of water and in air. Most lead air pollution is in the form of aerosols, fume, and sprays. The largest use of lead 39% of the total in 1968 is in the construction of storage batteries. The second largest use is in the manufacture of gasoline and tetramethyl lead. Approximately 1.30 million tons of lead were consumed in the United States in 1968.	*Major sources*: The *major source* of airborne lead in urban areas is the exhaust from gasoline powered vehicles. Other man-made sources are manufacturing of lead additives for gasoline, processing and manufacture of lead products, the combustion of coal.	1) *From vehicle sources*: Reduction or elimination of lead in fuel; use of particulate traps on vehicle exhausts. 2) *From lead processing and the manufacture of lead products*: Control of operating conditions (temperature and timing): Use of oconventional air cleaning techniques (bag house filters, scrubbers, electrostatic precipitators).

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
		3) *From coal combustion*: use of electrostatic precipitators. 4) *From manufacture of lead additives for gasoline*: Use of water scrubbers and bag house filters. 5) *From transfers and transportation of lead gasoline*: Use of vapor recovery systems; reduction or elimination of lead in gasoline. 6) *From use of pesticide*: Use of pesticides which do not contain lead; improved techniques of pesticide use. 7) *From incineration of refuse*: Use of conventional air cleaning techniques or sanitary land fills instead of incinerators.
c. NICKEL: A grayish white metallic element—hard, rough partially magnetic resistant to oxidation and corrosion. Nickel forms a variety of alloys with other metals. It is very important in making steel alloys and particularly stainless steel. Major pollutants are nickel dust and vapours.	*Major sources*: The processing of nickel to produce various alloys is the major source of emissions: See Table 16: This includes: 1) PLANT producing nickel alloys (including stainless steel) contains anywhere between 3–65% nickel in the alloys. 2) Nickel plating facilities via, electro-plating, electroless plating (chemical plating), electoforming (nickel can on mold) etc.	a) Use of conventional air cleaning devices: 1) Bag filters 2) Precipitators 3) Scrubbers b) Decomposition of gaseous emissions at high temperature forming nickel (which can be removed as a particulate) and carbon monoxide. c) No control methods currently available for vehicle engine exhausts.

TABLE 13

Sources and Health Effects of Some Prominent Heavy Metals[a]

Element	Sources	Health Effects
Mercury	Coal electrical batteries, other industrial	Kidney damage, nerve damage and death
Lead	Auto exhaust, paints	Brain, liver, and kidney damage; convulsions, behavioral disorders, death
Cadmium	Coal, zinc mining, water mains and pipes, tobacco smoke, burning plastics	High blood pressure and cardiovascular disease, interferes with zinc and cooper metabolism
Nickel	Diesel oil, residual oil, coal, tobacco smoke, chemicals and catalysts, steel and nonferrous alloys	Lung cancer
Arsenic	Coal, petroleum, detergents, pesticides, mine tailings	Hazard disputed, may cause cancer
Germanium	Coal	Little innate toxicity
Vanadium	Petroleum (Venezuela, Iran), chemicals and catalysts, steel and nonferrous alloys	Probably no hazard at current levels
Antimony	Industry	Shortened life span in rats

[a] Data from *Chemical & Engineering News 49* (July 19, 1971), 29–33, and other sources.

TABLE 14
National Anthropogenic Mercury Emissions (short tons/year)

Source Category	Mercury	Activity Year
Area Sources		
Electric Lamp Breakage	1.5	1989
Laboratory Use	0.8	1992
Dental Preparations	0.8	1992
Subtotal	3.1	
Combustion Point Sources		
Utility Boilers	54.5	1990
Commercial/Industrial Boilers	29.0	1992
Residential Boilers	3.5	1991
Municipal Waste Combustors	55.0	1991
Medical Waste Incinerators	64.7	1991
Sewage Sludge Incinerators	1.8	1990
Crematories	0.4	1991
Wood-fired Boilers	0.3	1980
Subtotal	209.2	
Manufacturing Sources		
Chlor-alkali Production	6.5	1991
Cement Manufacturing	6.5	1990
Battery Production	0.02	1992
Electrical Apparatus Manufacturing	0.46	1992
Instrument Manufacturing	0.5	1992
Secondary Mercury Production	7.4	1991
Carbon Black Production	0.25	1991
Primary Lead Smelting	9.0	1990
Primary Cooper Smelting*	0.7	1992
Lime Manufacturing	0.7	1992
Fluorescent Lamp Recycling*	0.006	1993
Subtotal	32.0	
TOTAL	244.3	

* Emissions are estimated for only one source, which is scheduled to cease operations by March 31, 1995; nationwide estimates are expected to be higher.

Note(s): Mercury was phased out of paint use in 1991.

Insufficient information was available to estimate emissions for the following source categories:

- Mobile sources;
- Agricultural burning (one study estimates 0.04 tons/year from preharvest burning of sugarcane in Florida everglades area);
- Landfills;
- Hazardous waste incinerators;
- Mercury compounds production;
- By-product coke production; and
- Petroleum refining.

Source(s): Draft Mercury Study Report to Congress, Volume II: Inventory of Anthropogenic Mercury Emissions in the United States. U.S. EPA, Office of Air Quality Planning and Standards, Internal Review Draft, 1995.

TABLE 15
Total National Emissions of Lead, 1970 through 1994 (short tons)

Source Category	1970	1975	1980	1985	1990	1993	1994
FUEL COMB. ELEC. UTIL.	327	230	129	64	64	61	63
Coal	300	189	95	51	46	49	49
FUEL COMB. INDUSTRIAL	237	75	60	30	18	15	15
Coal	218	60	45	22	14	11	11
FUEL COMB. -OTHER	10,052	10,042	4,111	421	418	415	415
Misc. Fuel Comb. (Except Residential)	10,000	10,000	4,080	400	400	400	400
CHEMICAL and ALLIED PRODUCT MFG.	103	120	104	118	136	96	93
Inorganic Chemical Mfg.	103	120	104	118	136	96	93
lead oxide and pigments							
METALS PROCESSING	24,224	9,923	3,026	2,097	2,169	1,887	1,873
Nonferrous Metals Processing	15,869	7,192	1,826	1,376	1,409	1,195	1,171
primary lead production	12,134	5,640	1,075	874	728	604	566
primary copper production	242	171	20	19	19	21	22
primary zinc production	1,019	224	24	16	9	13	14
secondary lead production	1,894	821	481	288	449	353	360
secondary copper production	374	200	116	70	75	70	80
lead battery manufacture	41	49	50	65	78	86	85
lead cable coating	127	55	37	43	50	47	44
Ferrous Metals Processing	7,395	2,196	911	577	576	499	489
coke manufacturing	11	8	6	3	4	3	3
ferroalloy production	219	104	13	7	18	12	13
iron production	266	93	38	21	18	20	19
steel production	3,125	1,082	481	209	138	145	150
gray iron production	3,773	910	373	336	397	319	304
Metals Processing NEC	960	535	289	144	184	193	213
metal mining	353	268	207	141	184	193	212
OTHER INDUSTRIAL PROCESSES	2,028	1,337	808	316	169	54	55
Mineral Products	540	217	93	43	26	27	26
cement manufacturing							
Miscellaneous Industrial Processes	1,488	1,120	715	273	143	28	28
WASTE DISPOSAL and RECYCLING	2,200	1,595	1,210	871	804	829	847
Incineration							
municipal waste	581	396	161	79	67	67	74
other	1,619	1,199	1,049	792	738	762	774
ON-ROAD VEHICLES	171,961	130,206	62,189	15,978	1,690	1,401	1,403
Light-Duty Gas Vehicles and Motorcycles	142,918	106,868	48,501	12,070	1,263	1,046	1,048
Light-Duty Gas Trucks	22,683	19,440	11,996	3,595	400	336	336
Heavy-Duty Gas Vehicles	6,361	3,898	1,692	313	28	19	19
NON-ROAD SOURCES	8,340	5,012	3,320	229	197	179	193
Non-Road Gasoline	8,340	5,012	3,320	229	197	179	193
TOTAL ALL SOURCES	219,471	158,541	74,956	20,124	5,666	4,938	4,956

Note(s): Categories displayed below Tier 1 do not sum to Tier 1 totals because they are intended to show major contributors.
1994 emission estimates are preliminary and will be updated in the next report.
Tier 1 source categories and emissions are shaded.

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
	3) Nickel is used extensively as a catalyst (for i.e. Raney Nickel) used in hydrogenation of organic compounds, dehydrogenation of organic compounds, aging of liquors, etc. 4) Aviation and automobile engines burning fuels containing nickel concentrations range from 1 to 10% nickel. 5) Burning coal and oil-nickel in ash varies from 3 to 10,000 μg/g 6) Incineration of nickel products.	
d. CADMIUM: Is a relatively rare metal which is not found in a free natural state. It is obtained from zinc, lead, copper and other ores that contain zinc minerals. Pollution exists as fumes, and vapors. The major use of cadmium is for electroplating iron and steel. The most common cadmium compounds and their uses are: 1) Electroplating—cadmium cyanide, $Cd(CN_2)$, and cadmium acetate, $Cd(CH_3COO)_3$. 2) Photography and dyeing—cadmium chloride, $CdCl_3$. 3) Manufacture of phosphors, glass in nuclear reactor controls. 4) Manufacture of electrodes for storage batteries—cadmium hydroxide $Cd(OH)_3$. 5) Cadmium iodide, CdI_2, electrode-position of Cd, manufacturing of phosphors. 6) Cadmium oxide—CdO. In phosphores, semi-conductors, manufacture of silver alloys. 7) Cadmium selenide—CdSe. In Photoconductors. 8) Cadmium sulfate—$CdSO_4$. In electrodeposition Cd, Cu, and N. 9) Dimethylcadmium $Cd(H_3)_2$ In organic synthesis.	*Major sources are*: (See Table 17) 1) Mining—Since no ore is mined solely for cadmium recovery, emissions of cadmium dust ore vapors are those that occur during mining and concentration of zinc-bearing ores. 2) Metallurgical processing—most of the atmospheric emissions occur during the roasting and sintering of zinc concentrates as impurities are removed. Cadmium is volatized and condensed to be collected as dust in baghouses or electrostatic precipitators. Lead and copper smelters also process concentrates containing cadmium. 3) Reprocessing—emissions occur during electroplating, manufacturing. 4) Consumptive uses—include use of rubber tires, motor oil, fungicides and fertilizers. 5) Incineration and ether disposal gaseous emissions will occur when scrap metal is melted to make new steel. 6) Cadmium used in plastics and pigments. NATURAL OCCURRENCE: The concentration of cadmium is almost always in direct proportion to that of zinc. The cadmium to zinc ratio varies from about 0.0002 to 0.002. With respect to the cadmium concentration percentage of the earth's crust, it is roughly 0.000055. Small concentrations of cadmium have been estimated in soil and sea water. MAN-MADE OCCURRENCE Mining—2 lbs are emitted per ton of cadmium mined usually as wind loss from tailings. Metallurgical processing.	General control procedures for the prevention of air pollution by dust, fumes, and mists applicable to the metal refinery alloying, and machining industries are considered suitable to these processes in the cadmium industry. —Copper mining and smelting: addition of bag filters and cyclones added to increase the recovery of cadmium. —Use of flue systems to direct the flow of gases to proper receptacles.
e. ARSENIC: a brittle, very poisonous chemical element, found widely distributed over the earth's crust. It is most often found with copper, lead, cobalt, nickel, iron, gold and silver. Arsenic is commonly found as a sulfide, arsenide, arsenite, or arsenate.	*Major sources*: (See Table 18) a) Smelters processing copper, zinc, lead and gold—arsenic is recovered as by product. b) Cotton ginning and the burning of cotton trash. c) Use as a pesticide (DDT). d) Combustion of coal. e) Incineration. Possible sources are: (See Table 18) manufacturing of glass—arsenic pentoxide, As_2O_3, arsenic trisulfide. As_2S_3 manufacturing of ceramics—arsenic trichloride, As_2Cl_3	a) Use of air cleaning devices to remove particulates from smelters and cotton gins. Equipment must operate at temperatures low enough to condense arsenic fumes ($<100°C$) —Electrostatic precipitators —Cooling flues —Bag houses, especially those using wet scrubbing vacuum pumps instead of fabric filters. b) No methods available to control emissions produced by burning cotton trash.

(*continued*)

TABLE 16
Nickel Releases in the U.S. 1979 (metric tons)

Source	
Production and Recovery	
Primary	
Hanna Operations	neg
Mining/Milling	neg
Smelting	
AMAX Operations	
Smelting/Refining	30
Secondary [scrap]:	
Nonferrous Metal	
New scrap: Ni-base	2
Cu-base	2
Al-base	neg
Old scrap: Ni-base	5
Cu-base	5
Al-base	neg
Coproduct and By-product Nickel	
(Copper Industry)	neg
Inadvertant Sources:	
Fossil Fuels	9990
Cement Manufacture	409
Asbestos Manufacture	neg
Tobacco	neg
Use:	
Industrial Processes	
Ferrous Smelting/Refining	52
Nonferrous Smelting/Refining	
Primary	neg
Secondary	neg
Alloys:	
Stainless Steel Heat Resistant	
Steel	340
Other Steel Alloy	95
Super Alloys	15
Nickel-Copper; Copper-Nickel Alloys	10
Permanent Magnet Alloys	1
Other Nickel Alloys	40
Cast Iron	30
Electroplating	neg
Chemicals/Catalysts	neg
Batteries	6
TOTAL	10653

(P. W. McNamara *et al.,* Little (Arthur C.) Inc. Exposure and Risk Assessment for Nickel, U.S. Environmental Protection Agency, EPA 440/4-85/012, December 1981.)

TABLE 17
Cadmium Releases in the U.S. (mt/yr)

Source	
Zn/Pb Mining and Benefication	—
Zn/Cd Smelting	7 (1981)
Electroplating	—
Batteries	1 (1980)
Pigments and Plastics	13 (1980)
Pesticide	—
Other Cd Products	NA
Impurity in Zn Products	NA
Iron and Steel Industry	14 (1981)
Primary Nonferrous/Non-Zinc	218 (1981)
Secondary Nonferrous	2 (1980)
Printing/Photography	—
Other Manufacturing Activity	NA
Coal Mining	—
Coal Combustion	202 (1981)
Oil Combustion	363 (1981)
Gasoline Combustion	13 (1978)
Lubricating Oil	1 (1980)
Tire Wear	5 (1980)
Phosphate Detergent	—
Phosphate Fertilizer	—
Urban Runoff	—
Culturally Hastened Erosion	NA
Natural Weathering	NA
Potable Water Supply	—
POTW Effluent	—
POTW Sludge	14 (1981)
Municipal Refuse	38 (1981)
TOTALS	891

(G. Delos, Cadmium Contamination of the Environment. As Assessment of Nationwide Risks (Final Report), U.S. Environmental Protection Agency, EPA-440/485/023, Feb. 1985.)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
f. VANADIUM: A grayish malleable ductile element found combined in many materials. Vanadium is used primarily to form alloy. Vanadium is also found in coal and oil as an impurity. The consumption of vanadium in 1968 was reported as 5495 tons. Of this total about 80% was used in making various steels. More than 65 vanadium-bearing minerals have been identified. The most important: a) patronite (V_2S_3S) b) Bravoite ($(FeNi)(S_2)$) c) Sulvanite ($3Cu_2S \cdot V_2S_3$)	*Major sources*: Almost all emissions of vanadium in the United States are derived from the combustion of fuel oil and coal both of which contain small amounts of metal. Fuel oil is by far the largest contribution (almost 90% of total emissions). In oil, the concentrations of vanadium pentoxide vary from 0.01% (Continental crude) to 0.06% (Venezuelan crude). The ash from combustion of residual oil varies from 0.002 to 0.3% (by weight). In coal, there is a small contribution of vanadium in the lignite deposit and the ash	*Use of additives*: Use of magnesium oxide in oil-fired burners, resulting in the reduction of fine particulate and amounts of vanadium escaping to the atmosphere. Use of conventional devices to remove particulates. Use of centrifugal collectors to gather ash emissions. Use of efficient fly-ash control equipment such as cyclones, electrostatic precipitators.

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
d) Davidite—titanium ore e) roacoelite ($CaO\cdot 3V_2\cdot S_3\cdot 9H_3O$)	emitted. Vanadium percentages in ash can range anywhere between 0.001 to 0.11%. Other minor sources are the processing of steel, cast iron and nonferrous alloys. Some additional emissions result from the manufacture of glass and ceramics and the use of vanadium as a catalyst.	
g. BERYLLIUM: is a light-weight, grayish metal that has a high strength-to-weight ratio, great stiffness and valuble nuclear properties. A hard metallic element which forms strong, hard alloys with several metals, including copper and nickel. Almost all the presently known beryllium compounds are acknowledged to be toxic in both the soluble and insoluble forms: —beryllium sulfate soluble —beryllium chloride —metallic beryllium insoluble —beryllium oxide In concentrated form, it is found in relatively few minerals, and there are basically compounds of beryllium oxide. The most important such minerals are as follows: Principal ore: Beryl—$3BeO\cdot Al_2O_3\cdot 6SiO_2$ Beryllium is used in nuclear reactors, gas turbines, airplane brakes, optical devices, springs, bellows, diaphragms, electrical contacts especially in high voltage insulation.	*Major sources*: Beryllium is commonly found as an atmospheric pollutant within the confines and in the proximity of industrial plants producing or using beryllium substances. Such plants engage in the extraction, refining, machining and alloying of the metal. b) Combustion of coals and oil containing of on the average 1.9 ppm and 0.08 ppm of beryllium respectively. c) Use of beryllium as additive in rocket fuels. d) During the 1930s, use of beryllium in production of fluorescent lamps was a major source of pollution. NATURAL ABUNDANCE: Beryllium makes up a small portion of the earth's crust (10 ppm) or 0.006%.	1) a) Use of conventional air cleaning devices: scrubbers, venturi scrubbers packed towers, organic wet collectors, wet cyclones. b) For dry processes; conventional bag collectors, reverse-jet bag collectors, electrostatic precipitators, cyclones, unit filters. 2) Discontinuance of the use of beryllium in fluorescent lamp tubes.
h. CHROMIUM: Chromium is a lustrous brittle metallic element usually occurring in compound form with other elements. Most of the chromium ore produced is used in the production of stainless and austenite steels. Chromium (Cr) is commonly known for its use as a decorative finish in chrome plating.	*Major Sources*: Chromium concentrations in urban air average 0.015 $\mu g/m^3$ and range as high as O. 3SO $\mu g/m^3$. Although a complete inventory of sources of ambient chromium has not been made some possible sources are metallurgical industry, chromate-producing industry, chrome plating, the burning of coal, and the use of chromium chemicals as fuel additives, corrosion inhibitors, pigments, tanning agents, etc. *Natural occurrence*: Elemental chromium is not found in nature. The only important commercial chromium mineral is chromite ($FeOCr_2O_3$) which is also never found in the pure form. Most soils and rocks contain small amounts of chromium usually as chromic oxide (Cr_2O_3). The continental crust averages 0.037% by weight, of chromium. In addition, most animal and plant tissues contain small amounts of chromium.	Chromium air pollution usually occurs as particulate emissions, which may be controlled by the usual dust-handling equipment, such as bag filters, precipitators, and scrubbers. *Chrome-plating facilities*: Moisture-extractor vanes in hood-duct systems have been used to break up bubbles in the exhaust gases. *Mist emissions*: Mist emissions from a decorative-chrome plating tank with problems can be substantially eliminated by adding a suitable surface-active agent to the plating solution.

TABLE 18
Arsenic Releases from Production, Use, and Inadvertent Sources (metric tons, 1979)

Source	Air
Production	
ASARCO, Tacoma	210
Use	
Pesticides	1,500
Wood Preservatives	neg
Glass Manufacture	10
Alloys	c
Other	2
Inadvertent Sources	
Fossil Fuel Combustion	2,000
Copper Production, 1° + 2°	1,100
Lead Production, 1° + 2°	230
Zinc Production	280
Iron and Steel	55
Aluminum Production	—
Boron Production	—
Phosphorous Production	—
Manganese Production	10
Antimony Production	—
Cotton Ginning	300
POTW	—
Urban Runoff	—
Inadvertent Releases from Mining and Milling	—
Copper	110
Lead	neg
Zinc	—
Aluminum	neg
Antimony	neg
Coal	—
Iron ore	3
Total	5,813

(Scow *et al.*, Little (Arthur, D.), Risk Assessment for Arsenic (Revised) (Final Report), EPA 440/4-85/005, March 1982.)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
i. IRON: A ductile, malleable silver-white metallic element, scarcely known in a pure condition, but abundantly used in its crude or impure forms containing carbon. Although inhalation of iron oxide is believed to cause a benign pneumoconiosis, there is growing concern about its synergistic effects with sulfur dioxide and carcinogens. Iron particulates may also act to reduce visibility.	*Major sources*: Iron and steel industry, sintering plant, blast furnaces, gray iron cupolas (used to melt gray iron), fuel sources (coal and oil), and incineration. *Natural occurrence*: Iron abounds in nature and is an essential element for both animals and plants. The iron content of the earth's crust has been calculated at 5.6%.	Control of emissions from the iron and steel industry is being accomplished through improvements in steel processing. Dust removal is accomplished by high-efficiency electrostatic precipitators, venturi type scrubbers, or filters.
j. MANGANESE: A hard, brittle grayish-white metallic element whose oxide (MnO_2) is a valuable oxidizing agent, used as alloying agent in steel to give it toughness. Although manganese (Mn) is one	*Major sources*: Air pollution by manganese arises almost entirely from the manganese and steel industries. Fumes from welding rods and organic manganese compounds may also contribute to	Control of manganese from furnaces is accomplished by various types of collectors, including electrostatic

(*continued*)

Part 1. Pollutant Emissions (*continued*)

Pollutant types	Sources and abundance	Abatement and control
of the elements essential to the human body, a high atmospheric concentration may result in poisoning and disease of several types.	air pollution. The organic compounds that have been tested as additives in gasoline, fuel oil, and diesel oil for use in both internal combustion engines and turbine engines may become an increasingly important source of pollution. *Natural occurrence*: Manganese is widely distributed in the combined state, ranking 12th in abundance (100 μg/m^3) among the elements in the earth's crust. Almost all of the manganese in the atmosphere enters as manganese oxides, such as MnO, Mn_3O_3 or Mn_3O_4.	precipitators, high-efficiency scrubbers, and fabric filters.
k. ZINC: A bluish-white metallic element occurring combined as the sulfide, oxide, carbonate, silicate etc. resembling magnesium in its chemical reactions. Although zinc is an essential element of the human and animal body, zinc and its compounds have been found to be toxic under certain conditions.	*Natural occurrence*: Zinc, widely distributed in the earth's crust, occurs in small quantities in almost all igneous rocks. The primary sources of emissions of zinc compounds into the atmosphere are zinc, lead, and copper smelting operations, secondary processing to recover scrap zinc, and possibly the incineration of zinc-bearing materials. Zinc oxide fumes are the zinc compounds most commonly emitted from these sources.	*Zinc gores*—can be collected by electrostatic (rod-curtain-type) precipitates and then further treated in cyclone scrubbers. *Zinc particles*—can be collected by use of electrostatic precipitators, a central cloth-bag collector system, or soil efficient filtering device.
l. BARIUM: A silvery white malleable, active, divalent, metallic element occurring in combination chiefly as barite. Inhalation of barium compounds can cause Baritosis a non-malignant lung disease. Characterized by fibrous hardening.	*Major sources are*: (1) Industrial process involved in mining, refining and production of barium and barium-based chemicals. 2) Use of barium compounds as a fuel additive for the reduction of black smoke emissions from diesel engines. (This is accomplished by the production in vehicle exhaust of micron-sized particles which have minimal effects on visibility.) Concentration of about 0.075 per cent *barium* by weight of additive is most effectively used. *Natural abundance*: Barium frequently appears as gangne in lead and zinc ore deposits. The two main minerals are barite (barium sulfate, $BaSO_4$) and witherite (barium carbonate, $BaCO_3$).	The conventional methods for removal of barium are the same as those for solids, and include bag filters, electrostatic precipitators, and wet scrubbers.

Part 2. Major Air Pollution Sources

Chemicals manufacturing industry	Nature of activity	Type of air pollution problems
ADIPIC ACID	Adipic Acid, COOH × $(CH_2)_4$ × COOH, is a dibasic acid used in the manufacture of synthetic fibers. Adipic acid is produced by the oxidation of cyclohexane by air over a catalyst and then purified by crystallization.	Emissions: The only significant emissions from the manufacture of adipic acid are nitrogen oxides. In oxidizing the cyclohexanol, nitric acid is reduced to nonrecoverable N_2O and potentially recoverable NO and NO_2 emitted into the atmosphere.
AMMONIA	The manufacture of ammonia (NH_3) is accomplished primarily by the catalytic reaction of hydrogen and nitrogen at high temperatures and pressures.	Emissions: Range from CO, HC, to NH_3 gases. Wet scrubbers and water can be utilized to reduce the atmospheric emissions.
CARBON BLACK	Carbon black is produced by reacting a hydrocarbon fuel such as oil and/or gas with a limited supply of air at temperatures of 2500–3000°F. Part of the fuel is burned to CO_2, CO and water, thus generating heat for combustion of fresh feed. The unburnt carbon is collected as a black fluffy particle.	Emissions: A high percentage of the emissions are carbon monoxide and hydrocarbons. The particulate and hydrogen sulfide problem are not as prevalent but do occur at amounts warranting attention. NO_2 emissions are relatively low due to the lack of available oxygen in the reaction.

Part 2. Major Air Pollution Sources (*continued*)

Chemicals manufacturing industry	Nature of activity	Type of air pollution problems
CHARCOAL	Charcoal is generally manufactured by means of pyrolysis, or destructive distillation of wood waste from members of the deciduous hardwood species. Four tons of hardwood are required to produce one ton of charcoal. In the pyrolysis of wood, all the gases, tars, oils, acids, and water are driven off leaving virtually pure carbon.	During pyrolysis of wood, carbon monoxide, hydrocarbons, particulate crude methanol, and acetic acid are emitted into the atmosphere. Some of these gases can be recovered by utilizing a chemical recovery plant.
CHLOR-ALKALI	Chlorine and caustic are produced concurrently by the electrolysis of brine in either the diaphragm or mercury cell.	Emissions from diaphragm and mercury cell chlorine plants include chlorine gas, carbon dioxide, carbon monoxide, and hydrogen. Other emissions include mercury vapor, chlorine, wet scrubbers (alkaline) can be utilized for emission reduction.
EXPLOSIVES	An explosive is a material which, under the influence of thermal or mechanical shock, decomposes rapidly and spontaneously with the evolution of large amounts of heat and gas.	Emissions: Sulfur oxides and nitrogen oxides emissions from processes which produce some of the raw materials for explosives production can be considerable.
HYDROCHLORIC ACID	Hydrochloric acid is manufactured by a number of different chemical processes. Approximately 80% of the hydrochloric acid is produced by the by-product hydrogen chloride process. By-product hydrogen chloride is produced when chloride is added to an organic compound such as benzene, toulene, and vinyl chloride.	The recovery of the hydrogen chloride from the chlorination of an organic compound is the major source of hydrogen chloride emissions. The exit gas from the absorption or scrubbing system is the actual source of the hydrogen chloride emitted.
HYDROFLUORIC ACID	All hydrofluoric acid in the United States is currently produced by reacting acid grade fluorspar with sulfuric acid for 30–60 minutes in externally fired rotary kilns at a temperature of 400–500°F.	The exist gases from the final absorber contain small amounts of HF, silicon tetrafluoride (SiF_4), CO_2, and SO_4 and may be scrubbed with a caustic solution to further reduce emissions. Dust emissions may also result from raw fluorspar grinding and drying operations.
NITRIC ACID	The ammonia oxidation process (AOP) is the principal method of producing commercial nitric acid. It involves high temperature oxidation of ammonia with air over a platinum catalyst from nitric oxide. The nitric oxide air mixture is cooled, and additional air water added to produced nitric acid.	The main source of atmosphere emissions from the manufacture of nitric acid is the tail gas from the absorption tower, which contains unabsorbed nitrogen oxides. These oxides are largely in the form of nitric oxide and nitrogen dioxide.
PAINT AND VARNISH	The manufacture of paint involves the dispersion of a colored oil or pigment in a vehicle, usually an oil or resin, followed by the addition of an organic solvent for viscosity adjustment.	Particulate emissions amount to 0.5 to 1% of the pigment handled; 1 to 2% of the solvent are lost. Hydrocarbons are the pollutant of primary concern.
PHOSPHORIC ACID	Phosphoric acid is produced by two principal methods, the wet process and the thermal process. In the wet process finely-ground phosphate rock is fed into a reactor with sulfuric acid to form phosphoric acid and gypsum. In the thermal process phosphate rock, siliceous flux, and coke are vaporized and placed in contact with water to produce phosphoric acid.	Emissions from the wet process are primarily gas fluorides, consisting mostly of silicon tetrafluoride and hydrogen fluoride. The principal emissions from the thermal process acid are P_2O_2 acid and acid mist. Particulates are also emitted in fairly large quantities.
PHTHALIC ANHYDRIDE	Phthalic anhydride is produced primarily by oxidizing naphthaline vapors with excess air over a catalyst, usually V_2O_5. The phthalic anhydride is then purified by a chemical soak in the sulfuric acid.	The major source of emissions is the excess air from the production system which contains some uncondensed phthalic anhydride, maleic anhydride, quinines, and other organics.
PLASTICS	The manufacture of most resins or plastics begins with the polymerization or linking of the basis compound (monomer) usually a gas or liquid, into high molecular weight noncrystalline solids.	The air contamination from plastics manufacturing are the emissions of raw material or monomer, emissions of solvents or other volatile liquids during the reaction, emissions of sublimed solids such as phthalic anhydride in alkyd production, and emissions of solvents during storage and handling of thinned resins.

(*continued*)

Part 2. Major Air Pollution Sources (*continued*)

Chemicals manufacturing industry	Nature of activity	Type of air pollution problems
PRINTING INK	Printing ink is produced by adding dyes to water and then flushing it with an ink vehicle.	Particulate emissions result from the addition of pigments to the vehicle while gases like terpenses, carbon dioxide, and aldehydes are emitted into the atmosphere, during the preliminary stages of ink production.
SOAP AND DETERGENTS	Soap is manufactured by the catalytic hydrolysis of various fatty acids with sodium or potassium hydroxide to form a glycerol-soap mixture. This mixture is separated by distillation, neutralized and blended to produce soap. In the manufacture of detergents, a fatty alcohol is sulfated, neutralized, and then sprayed dry to obtain the product.	The main atmospheric pollution problem in the manufacture of soap is odor from the spray drying operation, storage of mixing tanks and particulate emissions from the spray drying tower.
CHEMICAL INDUSTRY SODIUM CARBONATE	The Solvay process is used to manufacture over 80% of all soda ash. In this process, the basic raw materials of ammonia, cake, lime-stone (calcium carbonate) and salt (sodium chloride) are purified inabsorbent using ammonia and CO$_2$, to produce sodium bicarbonate as a by-product.	The major emissions from the manufacture of soda ash is ammonia. Small amounts of ammonia are emitted in the vent gases from the brine purification system. Traces of particulate emissions can result from rotary drying, dry solids handling and processing of lime.
SULFURIC ACID	The contact process is responsible for producing 90% of all the sulfuric acid in the United States. In this process sulfuric acid is produced from the contact of SO$_2$ and SO$_3$ with water.	The waste gas contains unreacted sulfur dioxide, unabsorbent sulfur trioxide, as well as sulfuric acid mist and spray. When the waste gas reaches the atmosphere, sulfur trioxide is converted to acid mist.

Food and agricultural industry	Nature of activity	Type of air pollution problems
	This section deals with the manufacture of food and agricultured products and the intermediate steps which present an air pollution problem.	
ALFALFA DEHYDRATING	An alfalfa dehydrating plant produces an animal feed from alfalfa. The dehydrating and grinding of alfalfa constitute the bulk of the manufacturing process of alfalfa meal. It is a very dusty operation most commonly carried out in rural areas.	Sources of dust emissions are the primary cyclone, grinders and air-meal separators. Overall dust loss has been reported as high as 7% by weight of the meal produced. The use of a bag house as a secondary collection system can greatly reduce emissions.
COFFEE ROASTING	Coffee, which is imported in the form of green beans, must be cleaned, blended, roasted and packaged before being sold.	Dust, chaff, coffeebean oils (as mists), smoke, and odors are the principal air contaminants emitted from coffee processing. The major source of particulate emissions and practically the only source of aldehydes, nitrogen oxides and organic acids is the roasting process.
COTTON GINNING	In separating the seed from the lint in raw seed cotton, a large amount of trash is left over. From one ton of cotton approximately one 500 pound bale of cotton can be made, the rest is discarded as trash.	The major sources of particulates from cotton ginning are the unloading fan, the cleaner and the stick and bur machine. When cyclone collectors are used emissions have been reported to be about 90% less.
FEED AND GRAIN MILLS AND ELEVATORS	Grain elevators are primarily transfer and storage units of various sizes. At grain elevator locations the following might occur: recewing, transfer and storages, cleaning, drying and milling or grinding.	Almost all emissions emanating from grain mills are dust particulates (minute grain particulates). The emissions from grain elevator operations are dependent on the type of grain, the moisture content of the grain, amount of foreign material, and the loading and unloading areas.
FERMENTATION	Fermentation occurs when various organisms (as molds, yeast, certain bacteria, etc.) agitate or excite substances into another form. The fermentation industries include the food, beer, whiskey, and wine categories.	Emissions from fermentation process are nearly all gases and primarily consist of carbon dioxide, hydrogen, oxygen, and water vapor, none of which present an air pollution problem. However, particulate emissions can occur in handling of the grain used as raw material, while gaseous hydrocarbons can be emitted during the drying of spent grains.
FISH PROCESSING	The canning, dehydrating, smoking of fish, and the manufacture of fish oil are the important segments of fish processing.	The biggest problem from fish processing is emissions of odors such as hydrogen sulfide and trimethylamine. Some of the methods used to control odors include activated carbon adsorbers, scrubbing with some oxidizing solution and incineration.

Part 2. Major Air Pollution Sources (*continued*)

Food and agricultural industry	Nature of activity	Type of air pollution problems
MEAT SMOKEHOUSES	Smoking is a diffusion process in which food products are exposed to atmosphere of hardwood smoke, causing various organic compounds to be absorbed by the food.	Emissions from smokehouses are generated from the burning hardwood, and included particulates, carbon monoxide, hydrocarbons (CH_4), aldehydes (HCH) and organic acids (acetic).
NITRATE FERTILIZERS	Nitrate fertilizers are the product of the reaction of nitric acid and ammonia to form ammonia nitrate solution or granules.	The main emissions from the manufacture of nitrate fertilizers are the ammonia and nitric oxides lost in the neutralization and drying operation.
PHOSPHATE FERTILIZERS	Nearly all phosphate fertilizers are made from naturally occurring phosphorous-containing minerals such as phosphate rock. The phosphorous content of these minerals is not in a form that is readily available to growing plants so that the minerals must be treated to convert the phosphorous to a plant-available form.	Emissions from manufacturing phosphate fertilizers include vent gases containing particulates ammonia, silicon tetrafluoride, carbon dioxide, steam and sulfur oxides. The sulfur oxides emissions arise from the reaction of phosphate rock and sulfuric acid.
STARCH MANUFACTURING	Starch is obtained through the separation of coarse starch in corn to a fine dry powder form ready for marketing.	The manufacture of starch from corn can result in significant dust emissions from cleaning, grinding, and screening operations.
SUGAR CANE PROCESSING	The processing of sugar cane starts with harvesting crops, then through a series of processes (washing, crushing, milling, diffusing) into the final sugar product.	The largest sources of emissions from sugar cane processing are the open burning in the harvesting of the crop and the burning of bagasse as fuel. Emissions include particulates, CO usually large, HC and nitrogen oxides.

Wood processing industry	Nature of activity	Type of air pollution problems
WOOD PULPING INDUSTRY	Wood pulping involves the production of cellulose from wood by dissolving the lignin that binds the cellulose fiber together. The three major chemical processes for pulp production are the kraft or sulfate process, the sulfite process and the neutral sulfite semi chemical process. The kraft process involves cooking wood chips in sodium sulfide and sodium hydroxide to dissolve the lignin. The excess pulp and impurities are washed away and the remaining clean pulp pressed and dried into the finished product.	Particulate emissions from the kraft process occur primarily from the recovery furnace, the lime kiln and smelt dissolving tank. This characteristic kraft mill odor is principally due to the presence of a variable mixture of hydrogen sulfide and dimethyl disulfide. Some sulfur dioxide emissions result from the oxidation of the sulfur compounds. CO emissions may occur from the recovery furnaces and klins.
PULPBOARD	Pulpboard manufacturing includes the manufacture of fibrous boards from a pulp slurry. After the pulp is washed, it is entered into a board machine and subsequently, dried and ready for fabrication.	Emissions from the paper board machine consist of only water vapor. Little or no particulates are emitted from the dryers.

METALLURGICAL INDUSTRY

The metallurgical industries can be broadly divided into primary and secondary metal production operations. Primary metal industry includes the production of the metal from ore; among these industries are the nonferrous operations involved in aluminum ore reduction, copper smelters, lead smelters, zinc smelters, iron and steel mills, ferro alloys and metallurgical coke manufacture. The secondary metals industry includes the recovery of the metal from scrap and salvage, the production of alloys from ingot, secondary aluminum operations, gray iron foundries, lead smelting, magnesium smelting, steel foundries, and zinc processing.

Metals industry	Nature of activity	Type of air pollution problems
ALUMINUM ORE REDUCTION	Bauxite, a hydrated oxide of aluminum associated with silicon, titanium, and iron, is the base ore for aluminum production. After preliminary purification using the (Boyer) process, the new oxide (Al_2O_3) is reduced in the Hall-Heroult process and pure aluminum is produced. Four tons of bauxite are required to make 1 ton of aluminum.	During the reduction process, the effluent released contains fluorides particulate and gaseous hydrogen fluoride. Particulate matter such as aluminum and carbon from the anodes are also emitted.

(*continued*)

Part 2. Major Air Pollution Sources (*continued*)

Metals industry	Nature of activity	Type of air pollution problems
METALLURGICAL COKE MANUFACTURE	Coking is the process of heating coal in an atmosphere of low oxygen content, i.e., destructive distillation. During the process organic compounds in the coal break down to yield gases and a relatively non-volatile residue.	Visible smoke, hydrocarbons, carbon monoxide, sulfur dioxide, nitrogen oxide and ammonia originate from by-product coking operations.
COPPER SMELTERS	Copper is produced primarily from low-grade sulfide ores, which are concentrated by gravity and subjected to melting and purifying procedures.	The raw waste gases from the process contain significant amounts of dust and sulfur oxides.
FERRO ALLOY PRODUCTION	Ferro alloys is the generic term for alloys consisting of iron and one or more other metals. The major method used to produce ferro alloy for steel making is the electric furnace process. In this process suitable oxides are reduced to the appropriate metals.	Most of the emissions of carbon monoxide and particulates (dust) are a direct result of the electric furnace, which uses carbon as the reducing agent.
BRASS AND BRONZE INGOTS (COPPER ALLOYS)	Obsolete domestic and industrial copper-bearing scrap is the basic raw material of the brass and bronze ingot industry. The ingots are produced from a number of different furnaces through a combination of melting, smelting, refining, and alloying of the process scrap materials.	The exit gas from the furnaces may contain fly ash, soot and smoke and some zinc oxides. Other particulate emissions include the preparation of raw materials and the pouring of ingots.
GRAY IRON FOUNDRY	The major type of furnace used to produce gray iron castings is the cupola, which uses an extremely hot bed of coke to melt the iron.	Emissions from cupola furnaces include CO dust and fumes, smoke, and all vapors.
SECONDARY LEAD SMELTING	Furnaces similar to the ones mentioned above are used to melt impure leaded scraps into desirable products (hard-lead, semi-soft lead, and pure lead).	The primary emissions from lead smelting are particulates, lead oxides, and carbon monoxides.
SECONDARY MAGNESIUM SMELTING	Magnesium smelting is carried out in crucible or pot type furnaces charged with magnesium scraps, melted and poured into perspective molds.	Emissions from magnesium smelting include particulate magnesium (MgO), oxides of nitrogen, sulfur dioxide and chloride gases.
IRON AND STEEL MILLS	To make steel, iron ore is reduced to pig iron, and some of its impurities are removed in a blast furnace. The pig iron is further purified in other processes (open hearth, Bessemer converters, basic oxygen furnaces, or electric furnaces).	Particulates and carbon monoxide are the major pollutant emissions resulting from the various furnace reactions.
LEAD SMELTERS	The ore from primary lead produced contains both lead and zinc. After melting, the metals are concentrated.	Effluent gases from the various concentrating processes include considerable particulate matter and sulfur dioxide.
ZINC SMELTERS	Most domestic zinc comes from zinc and lead ores. Another important source of raw material for zinc metal has been zinc oxide from fuming furnaces, the roasted are electrolytically purified.	Dust, fumes, and sulfur dioxide are evolved from zinc concentrate roasting.
SECONDARY ALUMINUM OPERATIONS	Secondary aluminum operations involve making lightweight metal alloys for industrial castings and ingots. Copper, magnesium, and silicon are the most common alloying constituents.	Emissions from secondary aluminum operations include fine particulate matter and small quantities of gaseous chlorides and fluorides.
STEEL FOUNDRIES	Steel foundries produce steel castings by melting steel metal and pouring it into molds. The basic melting process operations are furnace charging, melting, tapping the furnace into a ladle and pouring the steel into molds.	Particulate emissions from steel foundry operations include iron oxide fumes, sand fires, graphite and metal dust. Gaseous emissions from foundry operations include oxide of nitrogen, oxides of sulfur, and hydrocarbons.
SECONDARY ZINC PROCESSING	Zinc processing includes zinc reclaiming (separation of zinc from the scrap), zinc oxide manufacturing (distilling metallic zinc into dry air stream), and zinc galvanizing (flux cover over zinc).	A potential for particulate emissions, mainly zinc oxide, occur, if the temperature of the furnaces is very high (100°F). Small quantities of ammonia chloride, nitrogen oxides, and carbon monoxides are also emitted into the atmosphere.

Part 2. Major Air Pollution Sources (*continued*)

Mineral products industry	Nature of activity	Type of air pollution problems
	This section involves the processing and production of various minerals. Mineral processing is characterized by particulate emissions in the form of dust. However, most of the emissions from the manufacturing process conventional in this section can be reduced by conventional particulate control equipment such as cyclones, scrubbers, and fabric filters.	
ASPHALT BATCH PLANTS	Hot-mix asphalt paving consists of a combination of aggregates, coarse or fine, uniformly mixed and coated with asphalt cement. The coarse aggregates are usually crushed stone, crushed slag or crushed gravel, while the fine aggregates are usually natural sand and finely crushed stones.	The largest pollutant type is dust, emanating from the rotary dryers and filtering systems, normally used in producing asphalt.
ASPHALT ROOFING	The manufacture of asphalt roofing felts and shingles involves saturating a fiber media with asphalt by means of dipping and/or spraying.	The major pollutants are particulate emissions from asphalt roofing plants during the asphalt blowing operations and the felt saturation. Common methods of control at asphalt saturation plants include complete enclosure of the spray area and saturation followed by good ventilation through one or more collection devices. Some traces of carbon monoxide and hydrocarbons are also present in the emissions from this asphalt process.
BRICKS AND RELATED CLAY PRODUCTS	The manufacture of brick and related products such as clay pipe, pottery and some types of refraction brick involves the grinding, screening, blending of the raw materials, forming, drying or curing, firing and ferial cutting or shaping.	Particulate emissions similar to those obtained in clay processing are emitted from the materials handling process in refractory and brick manufacturing. Gaseous fluorides and nitrogen oxides are also emitted from brick manufacturing operations.
CALCIUM CARBIDE	Calcium carbide is manufactured by heating a mixture of quick-lime (CaO) and carbon in an electric arc furnace when the lime is reduced by the coke to calcium carbide and carbon monoxide. About 1990 pounds of lime and 1300 pounds of coke yield 1 ton of calcium carbide.	Particulates, acetylena, sulfur compounds and some carbon monoxide are emitted from calcium carbide plants.
CASTABLE REFRACTORIES	Castable or fused-cast refraction are manufactured by carefully blending such components as alumina, zirconia, silica, chrome, and magnesium, melting the mixture, pouring into molds, and slowly cooling to the solid state.	Particulate emissions occur from drying, crushing and handling procedures while gaseous fluoride occurs during melting operations.
PORTLAND CEMENT MANUFACTURING	Lime (calcareous), silica (siliceous), alumina (argillaceous) and iron (ferriferous) are the four major components used to manufacture cement. The various substances are crushed in exact proportions, fired in a klin, and then ground in gypsum to be bagged for shipment as cement.	Particulate matter is the primary emission in the manufacture of portland cement and is emitted primarily from crushing operations and rotary kilns. Control systems usually include multicyclones, electrostatic precipitators or combinations of these types of control.
CERAMIC CLAY MANUFACTURE	The manufacture of ceramic clay involves the conditioning of the basic ores, coolinate and mont-morillonite (aluminous-silicate materials), into dry clay products.	Emissions consist primarily of particulates, but some fluorides and acid gases are also emitted in the drying process.
CLAY AND FLY ASH SINTERING	Both the sintering clay and fly ash involve the firing and burning off of residual matter to desirable product. In fly ash, carbon is burned off while in clay, entrained volatile matter is driven off.	Dust is the major pollutant emitted from the screening and sintering process.
COAL CLEANING	Coal cleaning is the process by which undesirable materials are removed from both	Particulates in the form of coal dust constitute the major air pollution problem from coal clearing plants.

(*continued*)

Part 2. Major Air Pollution Sources (*continued*)

Mineral products industry	Nature of activity	Type of air pollution problems
	bituminous and authorite coal. The coal is screened, classified, washed and dried at coal preparation plants.	
CONCRETE BATCHING	Concrete batching involves the proportioning of sand, gravel, cement, and water by means of weight hoppers and conveyors into a mixing receiver.	Particulate emissions consist primarily of cement dust, but some sand and aggregate dust emissions do occur during batching operations.
FIBERGLASS MANUFACTURING	Fiberglass manufactured by melting various raw materials to form glass, drawing the molten glass into fibers, and coating the fibers with an organic material.	The major emissions from fiberglass manufacturing processes are particulates from the glass melting furnace and the product coaling line.
FRIT MANUFACTURING	Raw materials such as borax, feldspar, sodium fluoride and soda ash are melted and then quenched with water to produce shattered small glass particles—called frit. The frit particles are then ground into fine particles used in enameling iron and steel or in glazing porcelain or pottery.	The major emissions from frit-smelting operations are dust and fumes (usually condensed metallic oxide fumes) from the molten charge. A small quantity of hydrogen fluoride also can be detected in the emissions.
GLASS MANUFACTURE	Nearly all glass produced commercially is either soda-lime, lead, fused silica, borasilicate, or 96% silicate. Soda lime glass, being of the largest type, is produced on a massive scale in large, direct fired, continuous melting furnaces in which the blended raw materials are melted at 2700 to form glass.	Emissions from the glass melting operation consist primarily of particulate (only a few microns in diameter) and fluorides, if fluoride-containing fluxes are used in the process.
GYPSUM	Gypsum or hydrated calcium sulfate is a naturally occurring mineral which hardens when in contact with water to form a solid crystalline hydrate. Gypsum is an important building material, and if it loses its water of hydration, becomes plaster of paris.	Gypsum rock dust and partially calcined gypsum dust are emitted into the atmosphere from the grinding and mixing of the gypsum material.
LIME MANUFACTURING	Lime (CaO) is the high temperature product of the calcination of limestone ($CaCO_3$). Lime is manufactured in vertical or rotary kilns fired by coal, oil, or natural gas.	Atmospheric emissions in the lime manufacturing industry include the particulate emissions from the mining, handling, crushing, screening, and calcining of the limestone and the combustion products from the kiln.
MINERAL WOOL	The product mineral wool is made by firing charge material (slag wool and rock wool) in a furnace with silica rock and coke, into long fibrons tails for a “blanket” of wool.	Gases such as sulfur oxides and fluorides are major emissions from cupolas or furnace stacks. Minor particulate emissions are found in existing fumes.
PERLITE MANUFACTURE	Perlite is a glassy, volcanic rock consisting of oxides of silicon and aluminum combined as a natural glass by water of hydration. By a process called exfolication, the material is slowly heated to release water of hydration and thus expand the spherules into low-density particles used primarily as aggregate in plaster and concrete.	A fine dust is emitted from the outlet of the last product collector in a perlite expansion plant. In order to achieve complete control of these particulate emissions a bag-house is needed.
PHOSPHATE ROCK PROCESSING	Phosphate rock preparation involves the benefication to remove impurities, drying to remove moisture, and grinding to improve reactivity.	Emissions in the form of fine rock dust may be expected from drying and grinding operations.
SYNTHETIC FIBERS	Synthetic fibers are classified into two major categories—semi-synthetic, or “True synthetic.” Semi-synthetics, such as viscose rayon and acetate fibers, result when natural polymeric materials such as cellulose are brought into a dissolved or dispersed state and then spun into fine filaments. True synthetic polymers, such as nylon, orlon and dacron result from addition and polymerization reaction to form long chain molecules.	In the manufacture of viscose, rayon, carbon disulfide are the major gaseous emissions. Some examples of minor pollutants emitted from the drying of the finished fiber are hydrocarbons and oil vapor (mist).

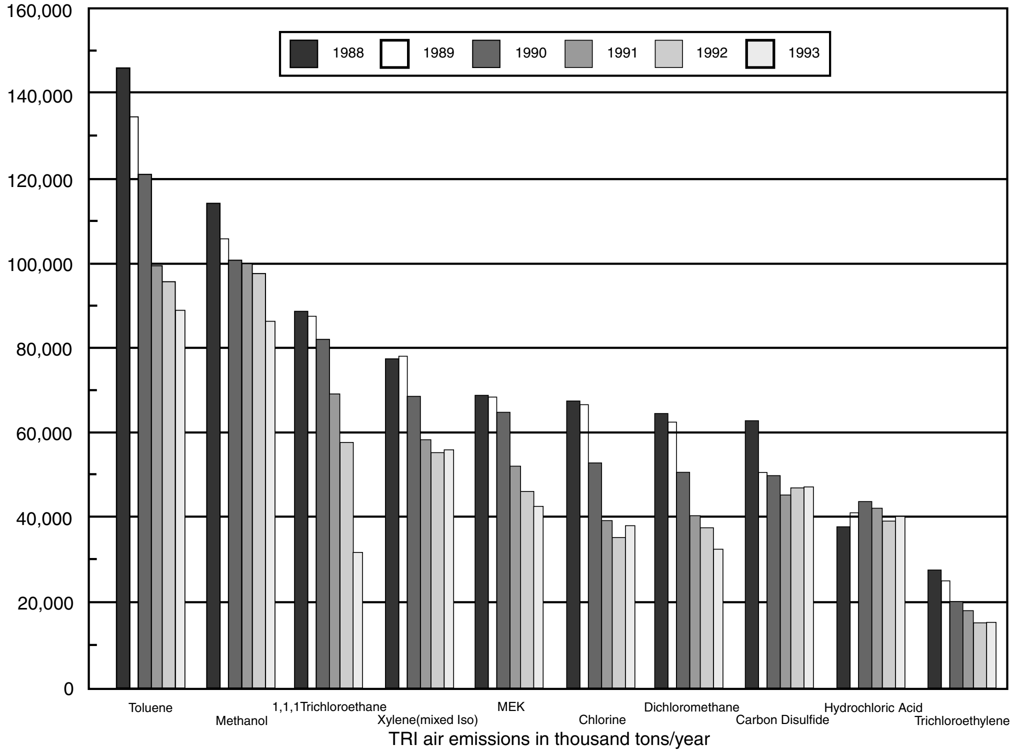


FIGURE 4 Top 10 Hazardous Air Pollutants—1988 Basis.

Part 2. Major Air Pollution Sources (*continued*)

Mineral products industry	Nature of activity	Type of air pollution problems
TEREPHTHALIC ACID	Terephthalic acid is an intermediate in the production of polyethylene terephthalate, which is used in polyester films and other miscellaneous products and by oxidizing paraxylene by nitric acid.	The NO in the off gas from the reactor is the major air contaminant from the manufacture of terephthalic acid.
STONE QUARRYING AND PROCESSING	Rock and gravel products are looosened by drilling and blasting from their deposit beds and removed with the use of heavy equipment. Further processing includes crushing, regrinding, and removal of fines.	Dust emissions occur from many operations in stone quarrying and processing.
Petroleum industry	Nature of activity	Type of air pollution problems
PETROLEUM REFINING	The operations of a petroleum refinery can be divided into four major steps: separation, conversion, treating, and blending. The crude oil is first separated into selected fractions (e.g., gasoline, kerosine, fuel oil, etc.). Some of the less valuable products such as heavy naphtha, are converted to products with a greater sale value such as gasoline. This is done by splitting, uniting, or rearranging the original molecules. The final step is the blending of the refined base stocks with each other and various additives to meet final product specifications.	The major pollutants emitted are sulfur oxides, nitrogen oxides, hydrocarbons, carbon monoxide and malodorons materials. Other emissions of lesser importance include particulates, aldehydes, ammonia, and organic acids. Most of the above mentioned emissions come from boiling process heaters, and catalytic cracking unit regenerators.

Table 19 shows trends of estimated emissions of criteria pollutants from 1970 through 2003. (note: VOCs are precursors to ozone, a criteria pollutant).

Source: *http://www.epa.gov/airtrends/econ-emissons.html*

TABLE 19

National Air Pollutant Emissions Estimates (fires and dust excluded) for Major Pollutants

	Millions of Tons Per Year								
	1970	1975	1980	1985[1]	1990	1995	2001[1]	2002	2003[2]
Carbon Monoxide (CO)	197.3	184.0	177.8	169.6	143.6	120.0	102.4	96.4	93.7
Nitrogen Oxides (NO_x)[3]	26.9	26.4	27.1	25.8	25.1	24.7	22.3	20.8	20.5
Particulate Matter (PM)[4]									
PM10	12.2[1]	7.0	6.2	3.6	3.2	3.1	2.3	2.4	2.3
PM2.5[5]	NA	NA	NA	NA	2.3	2.2	1.8	1.8	1.8
Sulfur Dioxide (SO_2)	31.2	28.0	25.9	23.3	23.1	18.6	16.3	15.3	15.8
Volatile Organic Compounds (VOC)	33.7	30.2	30.1	26.9	23.1	21.6	16.9	15.8	15.4
Lead[6]	0.221	0.16	0.074	0.022	0.005	0.004	0.003	0.003	0.003
Totals[7]	301.5	275.8	267.2	249.2	218.1	188.0	160.2	150.2	147.7

Source: http://www.epa.gov/airtrends/econ-emissions.html

Notes:

1. In 1985 and 1996 EPA refined its methods for estimating emissions. Between 1970 and 1975, EPA revised its methods for estimating particulate matter emissions.
2. The estimates for 2003 are preliminary.
3. NO_x estimates prior to 1990 include emissions from fires. Fires would represent a small percentage of the NO_x emissions.
4. PM estimates do not include condensable PM, or the majority of PM2.5 that is formed in the atmosphere from "precursor" gases such as SO_2 and NO_x.
5. EPA has not estimated PM2.5 emissions prior to 1990.
6. The 1999 estimate for lead is used to represent 2000 and 2003 because lead estimates do not exist for these years.
7. PM2.5 emissions are not added when calculating the total because they are included in the PM10 estimate.

BIBLIOGRAPHY

1. Forster, Christopher F., *Environmental Biotechnology,* Ellis Harwood Limited, p. 445, 1987.
2. Stern, *Air Pollutants Their Transformation and Transport,* Academic Press, p. 65, 1976.
3. Stern, *Air Pollution II and III Sources and Control,* Academic Press, p. 55, 1968.
4. National Air Pollutant Emissions Trends, 1990–1994, Monitoring, and Data Analysis Division U.S. Environmental protection Agency, Research Triangle Park, NC. Publication No. EPA 454/R-95–101, October 1995.
5. Spiro, Thomas G. and William, M. Stigliami, *Environmental Science in Perspective,* State University of New York Press, 1980.
6. Godish, Thad, *Air Quality,* Lewis Publishers Inc., 1985.
7. Altshuller, A.P., Review: Natural volatile organic substances and then effect on air quality in the United States, Atmos. Environ. **17:**2131 (1983).
8. National air quality and emission trends report, 1984, U.S. Environmental Protection Agency, EPA-450/4-86-001, April 1986.
9. Homolya, J.B. and E. Robinson, "Natural and antropogenic emission sources," Chapter A-2 in the *Audio Deposition Phenomena and its Effects: Critical Assessment Review Papers, Vol. 1, Atmospheric Sciences,* A.P. Altshuller, R.A. Linthurst, eds., EPA-600/8-83-016AF, July 1984.
10. Liu, S.C., M. Trainer, F.C. Freshenfeld, D.D. Danish, E.J. Williams, D.W. Fahley, G. Huber, and P.C. Murphy, Ozone production in the rural troposphere and the implications for regional and global ozone distribution, *J. Geophys. Res.* **92:** 4191 (1987).
11. http://www.epa.gov./airlinks/
12. http//www.epa.gov/ttn/atw/eparules.html
13. http://www.epa.gov/airtrends/econ-emissions.html

JEHUDA MENCZEL
U.S. Environmental Protection Agency

AIR POLLUTION SOURCES: see ATMOSPHERIC CHEMISTRY, GREENHOUSE GASES EFFECTS

AQUATIC PRIMARY PRODUCTION

Primary productivity in aquatic systems, like the same process in terrestrial environments, provides the base of the food web upon which all higher levels of an ecosystem depend. Biological productivity is the increase in organic material per unit of area or volume with time. This addition of organic matter is the material from which the various plant and animal communities of an ecosystem are made, and is dependent on the conversion of inorganic matter into organic matter. Conversion is accomplished by plants through the photosynthetic process. Plants are therefore considered to be the *primary producers*, and in an aquatic ecosystem these plants include algae, bacteria, and sometimes higher plants such as water grasses and water lilies. *Primary productivity*, the first level of productivity in a system, can be measured as the rate of photosynthesis, addition of biomass per unit of time (yield), or indirectly by nutrient loss or a measure of respiration of the aquatic community.

METHODS OF STUDY

Standing crop refers to the part of biological production per unit area or per unit volume that is physically present as biomass and that is not lost in respiration. Standing crop measurements over a period of time give an indirect measure of productivity in terms of yield. Plankton, microscopic floating plants and animals, can be collected in a plankton net and may be counted under a microscope or weighed. Aquatic biologists have used standing crop measurements to estimate productivity longer than any other method (e.g. Lohman, 1908). This method is still also used for periphyton (attached algae) or rooted plants.

Only within the past few decades have biologists progressed from merely counting numbers of organisms to calculating biomass, and more recently, to expressing biomass *yield.* Fishery biologists, like farmers, for many years have measured fish productivity in terms of tons produced per acre of water surface per year. Calculating biomass and biomass yield is an important step forward since changes in standing crop reflect the net effect of many biological and physical events and therefore are not directly proportional to productivity. For example, the standing crop of a phytoplankton community may be greatly diminished by predation and water movement, while photosynthetic rates of the survivors may remain high.

The measurement of plant pigments such as chlorophyll *a* is also a standing crop measurement that is frequently used and may now be done through remote sensing by aircraft or satellites.

UPTAKE OF NUTRIENTS

Another early attempt at measuring the rate of production in aquatic ecosystems was to measure the inorganic nutrients taken up in a given system and to calculate the amount of biological production required to absorb this amount. Atkins (1922, 1923) studied the decrease in carbon dioxide and phosphate in measuring production in the North Sea, and Steel (1956), also working in the North Sea, estimated the annual cycle of plant production by considering changes in the inorganic phosphate in relation to vertical mixing of the water mass. Many biologists consider phosphorus to be a difficult element to study in this respect because organisms often store it in excess of their requirements for optimum growth.

Measuring nutrient uptake in an indirect method of determining the rate of productivity in an aquatic ecosystem and is influenced by various other biological activities. Nevertheless, it has been important in the development toward more precise measurements of the dynamic aquatic ecosystem.

MEASUREMENTS OF OXYGEN AND CARBON DIOXIDE

The net rate at which the phytoplankton community of a given ecosystem incorporates carbon dioxide may be estimated in moderately to highly productive aquatic environments by direct measurement of the short-term fluctuations in the dissolved oxygen it produces. The calculations are based on the assumption that a mole of oxygen is released into the environment for each mole of carbon dioxide reduced in photosynthesis. This method precludes the necessity of enclosing the phytoplankton in a bottle. If measurements are made at regular hourly intervals over a 24-hour period, the average hourly decrease in oxygen during times of darkness when no photosynthesis is occurring can be determined. It is assumed that respiration removes this amount of oxygen each hour throughout the day thus giving a measure of the gross rate at which the community incorporates carbon dioxide.

An analogous method exists for recording fluctuations in carbon dioxide.

The pH meter, which measures acidity, has been successfully employed to measure these carbon dioxide changes in the aquatic ecosystem since the removal of carbon dioxide from the water for photosynthesis is accompanied by a proportional rise in pH. This pH shift has been used to estimate both photosynthesis and respiration. The sea and some fresh waters are too buffered against changes in pH to make this method useful in all environments, but it has been employed with success in lakes and for continuously monitoring the growth of cultures. Carbon dioxide may also be directly measured by standard volumetric or gasometric techniques.

Although carbon dioxide and oxygen can be measured with relative precision, the overall precision of productivity measurements made by these techniques is not generally great because of uncertainties in the corrections for diffusion, water movements, or extended enclosure time. Some of the oxygen produced by higher aquatic plants may not be immediately released thus causing a lag period in the evolution of oxygen into the environment. The primary advantage this method has over the more sensitive ^{14}C method is the added benefit of an estimate of community respiration.

Some of the uncertainties of the previous method can be reduced by enclosing phytoplankton samples just long enough in glass bottles for measurable changes in the concentration of oxygen and carbon dioxide to occur, but not long enough for depletion of nutrients or the growth of bacteria on the inside bottle surface. This method is called the *light and dark bottle* method. The name is derived from the fact that identical samples are placed in a transparent "light bottle" and an opaque "dark bottle." Gross and net productivity of the plankton community from which the samples were taken can be estimated by calculating the difference in the oxygen content between the two bottles after a predetermined period of incubation and with that present initially.

Productivity determinations that are dependent on measurements of oxygen are based on some estimated photosynthetic quotient (moles O_2 liberated/moles CO_2 incorporated). For the photosynthesis of carbohydrates the ratio is unity. For the synthesis of an algal cell, however, the expected ratio is higher, and presumably varies with the physiological state of the algae and the nutrients available.

Oxygen methods in general have rather poor sensitivity and are of no use if the gross incorporation of inorganic carbon during the test period is less than about 20 mg of carbon per cubic meter. Several days may be required in many of the less productive aquatic environments for this much photosynthesis to occur and bacteria may develop on the insides of the container during this time, invalidating the results.

Photosynthetic rates can be measured in light and dark bottles also by determining the amount of carbon fixed in particulate form after a short incubation. This can be done by inoculating the bottles with *radioactive carbon* ($Na_2{}^{14}CO_3$). Sensitivities with this method are much greater than the standard method and much shorter periods of incubation are possible. It is possible to obtain easily measurable amounts of ^{14}C in particulate form after only two hours by adjusting the specific activity of the inoculums. However, unlike the oxygen method, the dark bottle results do not provide an estimate of community respiration thus giving the ecologist less information with which to work.

The ^{14}C method has been widely used because it is sensitive and rapid. One outcome of its popularity is that a great deal of scrutiny has been devoted to the method itself. After 18 years of use, however, it is still not clear whether the ^{14}C is measuring gross productivity, net productivity, or something in between. The results probably most closely estimate net productivity, but it may be that this method applies only to a particular set of experimental conditions.

Already mentioned is the evidence that some of the ^{14}C that is fixed during incubation may seep out of the algal cells in the form of water-soluble organic compounds. This material is presumably utilized by bacteria rather than passed on directly to the next higher trophic level as is the remainder of the consumed primary productivity. The amount of primary production liberated extracellularly is large enough to be measured with precision and a number of workers are now routinely including quantitative studies of extracellular products of photosynthesis as part of the measurements of primary productivity.

Calibration of radioactive sources and instruments for measuring radioactivity pose a serious technical problem for the ^{14}C method. In order to calculate productivity in terms of carbon uptake it is necessary to know accurately the amount of ^{14}C added in microcuries and the number of microcuries recovered in particulate form by filtering the sample through a membrane filter.

Further it has been found that phytoplankton cells may become damaged during filtration and calculations based on these conditions will show lower productivity rates than are actually the case.

A point deserving emphasis is that those of us measuring primary productivity are still attempting to determine more precisely what is being measured, and generalizations about the transfer of energy through aquatic food-webs should be made continuously. Neither this nor any other practical technique adequately measures the change in oxidation state of the carbon that is fixed. The subsequent ecological role of newly fixed carbon is even more difficult to measure because of the various ways the photosynthate may be used.

USE OF PRIMARY PRODUCTIVITY MEASUREMENTS IN AQUATIC ECOSYSTEMS

Lindeman (1942) developed a trophic-dynamic model of an aquatic ecosystem and introduced the concept of "energy flow," or the efficiency of energy transfer from one trophic level to the next, to describe its operation. A certain value derived from the measured primary productivity represented the input of energy into the next grazing level, and so forth up the food chain. It was consistent with Lindeman's purpose to express his data as energy units (calories). Subsequent workers have continued to probe the concept of energy flow. However, advances in biochemistry, physiology, and ecology require such a complex model of energy flow that it is difficult to relate it to the natural world. In an imaginary world or model of a system in which the function units are discrete trophic levels, it is not only possible but stimulating to describe the flow of energy through an ecosystem. But when the functional units of the system being investigated are conceived of as macromolecules it is difficult to translate biomass accumulation into energy units.

Besides requiring a portion of their autotrophic production for respiration, phytoplankton communities must also reserve a portion for the maintenance of community structure. In terms of information theory, energy expended for community maintenance is referred to as "*information*." Energy information cost has never been measured directly but there is indirect evidence that it must be paid. For example, when an aquatic ecosystem is altered artificially with the aim of increasing the production of fish, zooplankton and fish may increase in greater proportion than the phytoplankton (McConnell, 1965; Goldman, 1968). Perhaps a large amount of primary production remains with the phytoplankton as information necessary for the maintenance or development of community structure. Grazers then have access only to the production in excess of this threshold level. If the magnitude of the information cost is high relative to primary production, then a small increase in the rate of growth of the primary producers will provide a relatively larger increase in the food supply of grazers and in turn the fish that consume them.

There are difficulties that must be met in the course of fitting measurements of primary productivity to the trophic-dynamic model. A highly variable yet often significant portion of primary production, as measured by ^{14}C light-and-dark bottle experiments, is not retained by the producers but instead moves into the environment in soluble form. It is difficult to measure the absolute magnitude of such excretion by a community of natural plankton because the excreta can rapidly serve as a substrate for bacterial growth and thus find its way back to particulate or inorganic form during the incubation period. Although this excrement is part of the primary productivity and eventually serves as an energy source for organisms at the higher trophic levels, the pathway along which this energy flows does not follow the usual linear sequence modeled for the transfer of energy from phytoplankton to herbivorous zooplankton. There is evidence that the amount of energy involved may sometimes be of the same order of magnitude as that recovered in particulate form in routine ^{14}C productivity studies.

The role of allochthonous material (material brought in from outside the system) in supporting the energy requirements of consumer organisms must also be considered in studies of energy flow. No natural aquatic ecosystem is entirely closed. Potential energy enters in the form of organic solutes and debris. Organic solutes undergo conversion to particulate matter through bacterial action. Sorokin (1965) in Russia found this type of production of particulate matter to be the most important in producing food for crustacean filter-feeders. Particulate and dissolved organic matter may also arise in the aquatic environment through chemosynthesis. This is a form of primary production not usually considered and therefore not usually measured. Although its magnitude may not be great in many systems, Sorokin found it to be very important in the Rybinsk reservoir and in the Black Sea.

PRIMARY PRODUCTION AND EUTROPHICATION

The process of increasing productivity of a body of water is known as eutrophication and in the idealized succession of lakes, a lake would start as oligotrophic (low productivity), becoming mesotrophic (medium productivity) eventually eutrophic (highly productive) and finally dystrophic, a bog stage in which the lake has almost been filled in by weeds and the productivity has been greatly decreased. The concept of eutrophic and oligotrophic lake types is not a new one. It was used by Naumann (1919) to indicate the difference between the more productive lakes of the cultivated lowlands and the less productive mountain lakes. The trophic state of five different aquatic environments will be discussed below.

The general progression from an oligotrophic to an eutrophic and finally to a dystrophic lake (lake succession) is as much a result of the original basin shape, climate, and such edaphic factors as soil, as it is of geologic age. It is unlikely that some shallow lakes ever passed through a stage that could be considered oligotrophic, and it is just as unlikely that the first lake to be considered here, Lake Vanda, will ever become eutrophic. It is also possible that the "progression" may be halted or reversed.

Lake Vanda, located in "dry" Wright Valley near McMurdo Sound in Antarctica, is one of the least productive lakes in the world. The lake is permanently sealed under 3 to 4 meters of very clear ice which transmits 14 to 20% of the incident radiation to the water below. This provides enough light to power the photosynthesis of a sparse phytoplankton population to a depth of 60 meters (Goldman *et al.*, 1967). Lake Vanda can be classified as ultraoligotrophic, since its mean productivity is only about 1 mg $C \cdot m^{-2} \cdot hr^{-1}$.

Lake Tahoe in the Sierra Nevada of California and Nevada is an alpine lake long esteemed for its remarkable clarity. Although it is more productive than Lake Vanda, it is still oligotrophic. The lake is characterized by a deep euphotic (lighted) zone, with photosynthesis occurring in the phytoplankton and attached plants to a depth of about 100 m. Although the production under a unit of surface area is not small, the intensity of productivity per unit of volume is extremely low. Lake Tahoe's low fertility (as inferred from its productivity per unit volume) is the result of a restricted watershed, whose granitic rocks provide a minimum of nutrient salts. This situation is rapidly being altered by human activity in the Tahoe Basin. The cultural eutrophication of the lake is accelerated by sewage disposal in the basin and by the exposure of mineral soils through road building and other construction activities. Since Lake Tahoe's water is saturated with oxygen all the way down the water column, the decomposition of dead plankton sinking slowly towards the bottom is essentially complete. This means that nutrients are returned to the system and because of a water retention time of over 600 years the increase in fertility will be cumulative.

Castle Lake, located at an elevation of 5600 feet in the Klamath Mountains of northern California, shows some of the characteristics of Lake Tahoe as well as those of more productive environments. It, therefore, is best classified as mesotrophic. Although it has a mean productivity of about 70 mg $C{\cdot}m^{-2}{\cdot}hr^{-1}$ during the growing season, it shows a depletion in oxygen in its deep water during summer stratification and also under ice cover during late winter.

Clear lake is an extremely eutrophic shallow lake with periodic blooms of such bluegreen algae as *Aphanizomenon* and *Microcystis* and inorganic turbidity greatly reducing the transparency of the water. The photosynthetic zone is thus limited to the upper four meters with a high intensity of productivity per unit volume yielding an average of about 300 mg $C{\cdot}m^{-2}{\cdot}hr^{-1}$ during the growing season. Because Clear Lake is shallow, it does not stratify for more than a few hours at a time during the summer, and the phytoplankton which sink below the light zone are continuously returned to it by mixing.

Cedar Lake lies near Castle Lake in the Klamath Mountains. Its shallow basin is nearly filled with sediment as it nears the end of its existence as a lake. Numerous scars of similar lakes to be found in the area are prophetic of Cedar Lake's future. Terrestrial plants are already invading the lake, and higher aquatic plants reach the surface in many places. The photosynthesis beneath a unit of surface area amounts to only about 6.0 mg $C{\cdot}m^{-2}{\cdot}hr^{-1}$ during the growing season as the lake is now only about four meters in depth and may be considered a dystrophic lake. Some lakes of this type pass to a bog condition before extinction; in others, their shallow basins may go completely dry during summer and their flora and fauna become those of vernal ponds.

In examining some aspects of the productivity of these five lakes, the variation in both the intensity of photosynthesis and the depth to which it occurs is evident. The great importance of the total available light can scarcely be overemphasized. This was first made apparent to the author during studies of primary productivity and limiting factors in three oligotrophic lakes of the Alaskan Peninsula, where weather conditions imposed severe light limitations on the phytoplankton productivity. The average photosynthesis on both a cloudy and a bright day was within 10% of being proportional to the available light energy.

Nutrient limiting factors have been reviewed by Lund (1965) and examined by the author in a number of lakes. In Brooks Lake, Alaska a sequence of the most limiting factors ranged from magnesium in the spring through nitrogen in the summer to phosphorous in the fall (Goldman, 1960). In Castle Lake potassium, sulfur, and the trace element molybdenum were found to be the most limiting. In Lake Tahoe iron and nitrogen gave greatest photosynthetic response with nitrogen of particular importance. Trace elements, either singly or in combination, have been found to stimulate photosynthesis in quite a variety of lakes. In general, some component of the phytoplankton population will respond positively to almost any nutrient addition, but the community as a whole will tend to share some common deficiencies. Justus von Liebig did not intend to apply his law of the minimum as rigidly as some have interpreted it, and we can best envision nutrient limitation from the standpoint of the balance and interactions of the whole nutrient medium with the community of organisms present at any given time. Much about the nutrient requirements of phytoplankton can be gleaned from the excellent treatise of Hutchinson (1967).

It must be borne in mind that the primary productivity of a given lake may vary greatly from place to place, and measurements made at any one location may not provide a very good estimate for the lake as a whole.

Variability in productivity beneath a unit of surface area is particularly evident in Lake Tahoe, where attached algae are already becoming a nuisance in the shallow water and transparency is often markedly reduced near streams which drain disturbed watersheds. In July, 1962, the productivity of Lake Tahoe showed great increase near areas of high nutrient inflow (Goldman and Carter, 1965). This condition was even more evident in the summer of 1967 when Crystal Bay at the north end of the lake and the southern end of the lake showed different periods of high productivity. This variability in productivity may be influenced by sewage discharge and land disturbance. Were it not for the great volume of the lake (155 km^3), it would already be showing more severe signs of eutrophication.

In the foregoing paper I have attempted to sketch my impressions of aquatic primary productivity treating the subject both as a research task and as a body of information to be interpreted. I believe that biological productivity can no longer be considered a matter of simple academic interest, but of unquestioned importance for survival. The productivity and harvest of most of the world's terrestrial and aquatic environments must be increased if the world population is to have any real hope of having enough to eat. This increase is not possible unless we gain a much better understanding of both aquatic and terrestrial productivity. Only with a more sound understanding of the processes which control productivity at the level of the primary producers can we have any real hope of understanding the intricate pathways that energy moves and biomass accumulates in various links of the food chain. With this information in hand the productivity of aquatic environments can be increased or decreased for the benefit of mankind.

REFERENCES

Atkins, W. R. G. (1922), Hydrogen ion concentration of sea water in its biological relation, *J. Mar. Biol. Assoc. UK*, **12,** 717–771.

Atkins, W. R. G. (1923), Phosphate content of waters in relationship to growth of algal plankton, *J. Mar. Biol. Assoc. UK,* **13,** 119–150.

Fernando, C. H. (1984), Reservoirs and lakes of Southeast Asia, in *Lakes and Reservoirs,* F. B. Taub, Ed., Elsevier, Amsterdam.

Goldman, C. R. (1960), Primary productivity and limiting factors in three lakes of the Alaska Peninsula, *Ecol. Monogr.,* **30,** 207–230.

Goldman, C. R. (1968), Absolute activity of ^{14}C for eliminating serious errors in the measurement of primary productivity, *J. du Conseil,* **32,** 172–179.

Goldman, C. R. and R. C. Carter (1965), An investigation by rapid carbon-14 bioassay of factors affecting the cultural eutrophication of Lake Tahoe, California–Nevada, *J. Water Pollution Control Fed.,* **37,** 1044–1059.

Goldman, C. R., D. T. Mason and J. E. Hobbie (1967), Two Antarctic desert lakes, *Limnol. Oceanogr.*, **12,** 295–310.

Guerrero, R. D. (1983), Talapia farming the Philipines; Practices, problems and prospects. Presented at PCARRD-ICLARM Workshop, Los Baños, Philipines.

Hutchinson, G. E. (1967), *A Treatise on Limnology,* Vol. II. *Introduction to lake biology and the limnoplankton,* John Wiley and Sons, New York.

Junghran, V. G. (1983), *Fish and fisheries of India,* Hindustan Pub. Co.

Kuo, C.-M. (1984), The development of tilapa culture in Taiwan, ICLARM Newsletter, 5(1).

Lindeman, R. L. (1942), The trophic-dynamic aspect of ecology, *Ecology,* **23,** 399–418.

Lohman, H. (1908), Untersuchungen zur Feststellung des vollständigen Gehaltes des Meeres an Plankton, *Wiss. Meeresunters, NF Abt. Kiel,* **10,** 131–370.

Lund, J. W. G. (1965), The ecology of the freshwater phytoplankton, *Biol. Rev.,* **40,** 231–293.

McConnell, W. J. (1965), Relationship of herbivore growth to rate of gross photosynthesis in microcosms, *Limnol. Oceanogr.*, **10,** 539–543.

Naumann, E. (1919), Några synpunkter angående planktons ökologi, Med särskild hänsyn till fytoplankton. *Svensk bot. Tidskr.*, **13,** 129–158.

Petr, J. and J. M. Kapetsky (1990), Tropical reservoir fisheries, *Resource Management and Optimization,* **7,** 3.

Sorokin, Y. A. (1965), On the trophic role of chemosynthesis and bacterial biosynthesis in water bodies, pp. 187–205. *In* C. R. Goldman (ed.), *Primary productivity in aquatic environments,* University of California Press, Berkeley.

Steele, J. H. (1956), Plant production on the Falden Ground. *J. Mar. Biol. Ass. UK,* **35,** 1–33.

CHARLES R. GOLDMAN
University of California, Davis

ATMOSPHERIC: see also **AIR**—various titles

ATMOSPHERIC CHEMISTRY

INTRODUCTION

Atmospheric chemistry is a broadly based area of scientific endeavor. It is directed at determining the quantities of various chemicals in the atmosphere, the origin of these chemicals, and their role in the chemistry of the atmosphere. Many atmospheric chemists are involved in the development of techniques for the measurement of trace quantities of different chemicals in the atmosphere, in emissions, and in depositions. Other atmospheric chemists study the kinetics and mechanisms of chemical reactions occurring in the atmosphere. Still other atmospheric chemists are involved in the development of chemical models of the processes occurring in the atmosphere. Atmospheric chemists work closely with other disciplines: engineers in characterizing anthropogenic emissions; biologists and geologists in characterizing natural emissions and in evaluating the effects of air pollution; physicists in dealing with gas-to-particle conversions; and meteorologists, physicists, computer scientists, and mathematicians in dealing with model development. Atmospheric chemistry plays a key role in maintaining the general well-being of the atmosphere, which is extremely important for maintaining the health of the human race.

In recent years, there has been a growing concern about a number of atmospheric environmental problems, such as the formation of photochemical oxidants, acid deposition, global-scale effects on stratospheric ozone, the sources and fates of toxic chemicals in the atmosphere, and urban and regional haze issues and the presence and effects of fine particulate matter in the atmosphere. These problems are affected by a wide variety of complex chemical and physical processes. Atmospheric chemistry is the broad subject area that describes the interrelationships between these chemical and physical processes.

The principal components of the atmosphere are nitrogen and oxygen. These molecules can absorb a portion of the high-energy solar ultraviolet radiation present in the upper atmosphere and form atoms. These atoms may react with a variety of other species to form many different radicals and compounds. For example, the short-wavelength ultraviolet radiation present in the upper atmosphere can photolyze molecular oxygen to form oxygen atoms. These oxygen atoms may react with molecular oxygen to form ozone. These reactions are only of importance at high altitudes, where the short-wavelength ultraviolet radiation is present. In the lower regions of the atmosphere, only light of wavelengths greater than about 300 nm is present. Table 1 lists the relative concentrations of a number of species present in the atmosphere, near the Earth's surface. The chemistry that is most important at lower altitudes is initiated by a variety of compounds or trace species, which are present in the atmosphere at concentrations of much less than 1 ppm.

One of the most important reasons to understand atmospheric chemistry is related to our need to understand and control air pollution. The air-pollution system, shown in Figure 1, starts with the sources that emit a variety of pollutants into the atmosphere. Those pollutants emitted directly into the atmosphere are called *primary pollutants.* Once these primary pollutants are in the atmosphere, they are subjected to meteorological influences, such as transport and dilution, in addition to chemical and physical transformations to secondary pollutants. *Secondary pollutants* are those formed by reactions in the air. The pollutants in the air may be removed by a variety of processes, such as wet and dry deposition. An ambient-air-monitoring program is used to provide detailed information about the compounds present in the atmosphere.

TABLE 1
Relative composition of the atmosphere near the Earth's surface

Species	Concentration (ppm)
N_2	780,840
O_2	209,460
H_2O	<35,000
Ar	9,340
CO_2	335
Ne	18
He	5.2
CH_4	1.7
Kr	1.14
H_2	0.53
N_2O	0.30
CO	<0.2
Xe	0.087
O_3	0.025

Source: Adapted from J. Heicklen (1976), *Atmospheric Chemistry,* Academic Press, New York; and R.P. Wayne (1985), *Chemistry of Atmospheres,* Clarendon Press, Oxford.

One of the principal goals of air-pollution research is to obtain and use our detailed knowledge of emissions, topography, meteorology, and chemistry to develop a mathematical model that is capable of predicting concentrations of primary and secondary pollutants as a function of time at various locations throughout the modeling domain. These model results would be validated by comparison with ambient-air-monitoring data. Model refinement continues until there is acceptable agreement between the observed and predicted concentrations. This type of air-quality model, on an urban scale, is called an *airshed model.* Airshed models treat the effects of a set of stationary and mobile sources scattered throughout a relatively small geographical area (~100 km^2). These models are intended to calculate concentrations of pollutants within this geographical area and immediately downwind.

It is also necessary to develop a detailed knowledge of the impacts of pollutants on the various important receptors, such as humans, plants, and materials. This impact information is used to identify the pollutants that need to be controlled. An airshed model can be used to predict the effectiveness of various proposed control strategies. This information can be passed on to legislative authorities, who can evaluate the costs and benefits of the various strategies and legislate the best control measures.

Unfortunately, there are significant gaps in our knowledge at every step throughout this idealized air-pollution system.

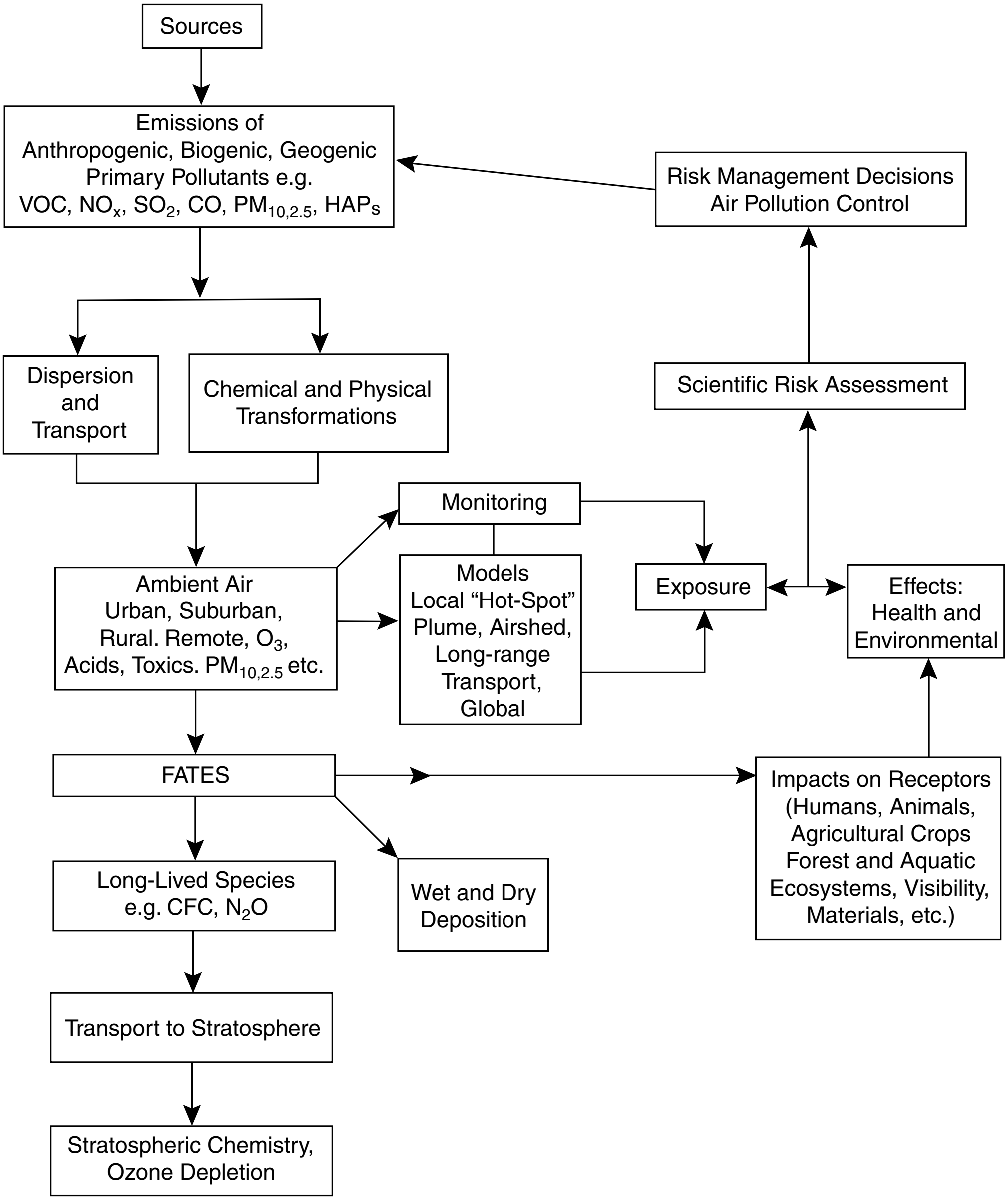


FIGURE 1 The atmospheric air-pollution system. From Finlayson-Pitts and Pitts (2000). (HAPs—hazardous air pollutants). With permission.

Hence, there is considerable room for continued research. Atmospheric chemistry is involved in several steps through the air-pollution system. First is chemically characterizing and quantifying the emissions of primary pollutants. Second is understanding the chemical and physical transformations that these primary pollutants undergo. Third is measuring the quantities of the various pollutants in the ambient air. Fourth is quantifying the deposition processes for the various pollutants. Finally, a mathematical formulation of the sources, chemical and physical transformations, and removal processes must be incorporated into the atmospheric model.

The chemistry of the formation of secondary pollutants is extremely complex. It requires the identification of all of the important reactions contributing to the chemical system. There must be a thorough investigation of each specific reaction, which can be achieved only when the reaction-rate constant has been carefully determined for each elementary reaction involved in the properly specified reaction mechanism. Because of the large number of important reactions that take place in the atmosphere, the rapid rates of many of them, and the low concentrations of most of the reactants, the experimental investigations of these atmospheric chemical kinetics is an enormously large and complex task.

In the United States, a set of National Ambient Air Quality Standards (NAAQS) have been established, as shown in Table 2. The primary standards are designed to protect the public health of the most susceptible groups in the population. Secondary NAAQS have also been set to protect the public welfare, including damage to plants and materials and aesthetic effects, such as visibility reduction. The only secondary standard that currently exists that is different from the primary standard is for SO_2, as shown in the table. For comparison purposes, Table 3 shows recommended limits for air pollutants set by the World Health Organization and various individual countries.

To illustrate the importance and complexity of atmospheric chemistry, a few examples will be presented and discussed: (1) urban photochemical-oxidant problems, (2) secondary organic aerosols, (3) chemistry of acid formation, and (4) stratospheric ozone changes in polar regions. These examples also illustrate the differences in the spatial scales that may be important for different types of air-pollution problems. Considering urban problems involves dealing with spatial distances of 50 to 100 km and heights up to a few kilometers, an urban scale or mesoscale. The chemistry related to acid formation occurs over a much larger, regional scale, extending to distances on the order of 1000 km and altitudes of up to about 10 km. For the stratospheric ozone-depletion problem, the chemistry of importance occurs over a global scale and to altitudes of up to 50 km. Secondary organic aerosol formation can be an urban to regional scale issue.

TABLE 2

U.S. National Ambient Air Quality Standards

Pollutant	Primary	Averaging Times	Secondary
Carbon monoxide	9 ppm	8-hour[1]	None
	35 ppm	1-hour[1]	None
Lead	1.5 $\mu g/m^3$	Quarterly average	Same as primary
Nitrogen dioxide	0.053 ppm	Annual (arith. mean)	Same as primary
Particulate matter (PM_{10})	50 $\mu g/m^3$	Annual[2] (arith. mean)	Same as primary
	150 $\mu g/m^3$	24-hour[1]	
Particulate matter ($PM_{2.5}$)	15 $\mu g/m^3$	Annual[3] (arith. mean)	Same as primary
	65 $\mu g/m^3$	24-hour[4]	
Ozone	0.08 ppm	8-hour[5]	Same as primary
	0.12 ppm	1-hour[6]	Same as primary
Sulfur oxides	0.03 ppm	Annual (arith. mean)	—
	0.14 ppm	24-hour[1]	—
	—	3-hour[1]	0.5 ppm

[1] Not to be exceeded more than once per year.

[2] To attain this standard, the expected annual arithmetic mean PM_{10} concentration at each monitor within an area must not exceed 50 $\mu g/m^3$.

[3] To attain this standard, the 3-year average of the annual arithmetic mean $PM_{2.5}$ concentrations from single or multiple community-oriented monitors must not exceed 15 $\mu g/m^3$.

[4] To attain this standard, the 3-year average of the 98th percentile of 24-hour concentrations at each population-oriented monitor within an area must not exceed 65 $\mu g/m^3$.

[5] To attain this standard, the 3-year average of the fourth-highest daily maximum 8-hour average ozone concentrations measured at each monitor within an area over each year must not exceed 0.08 ppm.

[6] *(a)* The standard is attained when the expected number of days per calendar year with maximum hourly average concentrations above 0.12 ppm is ≤ 1.

(b) The 1-hour NAAQS will no longer apply to an area one year after the effective data of the designation of that area for the 8-hour ozone NAAQS.

Source: Data is from the U.S. EPA Web site: http://www.epa.gov/air/criteria.html.

TABLE 3
Recommended ambient air-quality limits for selected gases throughout the world.

Country	CO (ppm)	SO$_2$ (ppm)	O$_3$ (ppm)	NO$_2$ (ppm)	PM$_{10}$ (µg/m^3)
WHO	26 (1 hr)	0.048 (24 hr)	0.061 (8 hr)	0.105 (1 hr)	
	8.7 (8 hr)	0.019 (annual)		0.021 (annual)	
EU	8.7 (8 hr)	0.132 (1 hr, <24x)	0.061 (8 hr)	0.105 (1 hr, <18x)	50 (24 hr, <35x)
		0.047 (24 hr, <3x)	(<25x/yr, 3 yr avg.)	0.021 (annual)	40 (annual)
		0.008 (annual)			
UK	10 (8 hr)	0.132 (1 hr, <24x)	0.050 (8 hr)	0.105 (1 hr, <18x)	50 (24 hr, <35x)
		0.047 (24 hr, <3x)		0.021 (annual)	40 (annual)
		0.008 (annual)			
Russia	4.4 (24 hr)	0.02 (24 hr)		0.045 (24 hr)	
Australia	9 (8 hr)	0.20 (1 hr)	0.10 (1 hr)	0.12 (1 hr)	50 (24 hr, <5x)
		0.08 (24 hr)	0.08 (4hr)	0.03 (annual)	
		0.02 (annual)			
New Zealand	9 (8 hr, <9x)	0.132 (1 hr, <9x)	0.08 (1 hr)	0.105 (1 hr, <9x)	50 (24 hr, <5x)
China	9 (1 hr)	0.19 (1 hr)	0.10 (1 hr)	0.13 (1 hr)	150 (24 hr)
	3.5 (24 hr)	0.06 (24 hr)		0.06 (24 hr)	100 (annual)
		0.02 (annual)		0.04 (annual)	
Japan	20 (8 hr)	0.10 (1 hr)	0.06 (1 hr)	0.04–0.06 (24 hr)	200 (1 hr)
	10 (24 hr)	0.04 (annual)			100 (24 hr)
Hong Kong	26 (1 hr, <3x)	0.30 (1 hr, <3x)	0.12 (1 hr, <3x)	0.16 (1 hr, <3x)	180 (24 hr)
	9 (8 hr)	0.13 (24 hr)		0.08 (24 hr)	55 (annual)
		0.03 (annual)		0.04 (annual)	
Thailand	30 (1 hr)	0.30 (1 hr)	0.10 (1 hr)	0.17 (1 hr)	120 (24 hr)
	9 (8 hr)	0.12 (24 hr)			50 (annual)
		0.04 (annual)			
Philippines	30 (1 hr)	0.06 (24 hr)		0.08 (24 hr)	150 (24 hr)
	9 (8 hr)	0.023 (annual)			60 (annual)
Nepal	9 (8 hr)	0.027 (24 hr)		0.042 (24 hr)	120 (24 hr)
		0.02 (annual)		0.021 (annual)	
Bangladesh		0.03 (annual)		0.04 (annual)	200 (annual)
India	3.5 (1 hr)	0.03 (24 hr)		0.04 (24 hr)	100 (24 hr)
(Residential)	1.7 (8 hr)	0.023 (annual)		0.03 (annual)	60 (annual)
Saudi Arabia	35 (1 hr, 2x/30)	0.28 (1 hr, 2x/30)	0.15 (1 hr, 2x/30)	0.35 (1 hr, 2x/30)	340 (PM15 24 hr)
	9 (8 hr, 2x/30)	0.14 (24hr)		0.05 (annual)	80 (PM15 annual)
		0.03 (annual)			
Egypt	26 (1 hr)	0.13 (1 hr)	0.10 (1 hr)	0.20 (1 hr)	70 (24 hr)
	9 (8 hr)	0.06 (24 hr)	0.06 (8 hr)	0.08 (24 hr)	
		0.02 (annual)			
South Africa		0.30 (1 hr)	0.12 (1 hr)	0.20 (1 hr)	180 (24 hr)
		0.10 (24 hr)		0.10 (24 hr)	60 (annual)
		0.03 (annual)		0.05 (annual)	
Canada			0.065 (8 hr)		30 (PM2.5 24 hr)
Mexico	11 (8 hr)	0.13 (24 hr)	0.11 (1 hr)	0.21 (1 hr)	150 (24 hr)
		0.03 (annual)			50 (annual)
Brazil	35 (1 hr)	0.14 (24 hr)		0.17 (1 hr)	150 (24 hr)
	9 (8 hr)	0.03 (annual)	0.08 (1 hr)	0.05 (annual)	50 (annual)

Source: Data was collected from Web sites from the individual countries and organizations.

Note: Numbers in parentheses represent the averaging time period and number of exceedances allowed.

URBAN PHOTOCHEMICAL OXIDANTS

The photochemical-oxidant problems exist in a number of urban areas, but the Los Angeles area is the classic example of such problems. Even more severe air-pollution problems are occurring in Mexico City. The most commonly studied oxidant is ozone (O_3), for which an air-quality standard exists. Ozone is formed from the interaction of organic compounds, nitrogen oxides, and sunlight. Since sunlight is an important factor in photochemical pollution, ozone is more commonly a summertime problem. Most of the ozone formed in the troposphere (the lowest 10 to 15 km of the atmosphere) is formed by the following reactions:

$$NO_2 + h\nu\ (\lambda \leq 430\text{ nm}) \rightarrow NO + O(^3P) \qquad (1)$$

$$O(^3P) + O_2 + M \rightarrow O_3 + M \qquad (2)$$

Nitrogen dioxide (NO_2) is photolyzed, producing nitric oxide (NO) and a ground-state oxygen atom (designated as $O(^3P)$). This oxygen atom will then react almost exclusively with molecular oxygen to form ozone. The M in reaction (2) simply indicates that the role of this reaction depends on the pressure of the reaction system. NO can also react rapidly with ozone, reforming NO_2:

$$NO + O_3 \rightarrow NO_2 + O_2 \qquad (3)$$

These three reactions allow one to derive the photostationary state or Leighton relationship

$$[O_3]\,[NO]/[NO_2] = k_1/k_3 \text{ or } [O_3] = k_1[NO_2]/k_3[NO]$$

This relationship shows that the O_3 concentration depends on the product of the photolysis rate constant for NO_2 (k_1) times the concentration of NO_2 divided by the product of the rate constant for the NO reaction with O_3 (k_3) times the NO concentration. This photolysis rate constant (k_1) will depend on the solar zenith angle, and hence will vary during the day, peaking at solar noon. This relationship shows that the concentration of ozone can only rise for a fixed photolysis rate as the $[NO_2]/[NO]$ concentration ratio increases. Deviations from this photostationary state relationship exist, because as we will see shortly, peroxy radicals can also react with NO to make NO_2.

Large concentrations of O_3 and NO cannot coexist, due to reaction (3). Figure 2 shows the diurnal variation of NO, NO_2, and oxidant measured in Pasadena, California. Several features are commonly observed in plots of this type. Beginning in the early morning, NO, which is emitted by motor vehicles, rises, peaking at about the time of maximum automobile traffic. NO_2 begins rising toward a maximum value as the NO disappears. Then the O_3 begins growing, reaching its maximum value after the NO has disappeared and after the NO_2 has reached its maximum value. The time of the O_3 maximum varies depending on where one is monitoring relative to the urban center. Near the urban center, O_3 will peak near noon, while further downwind of the urban center, it may peak in the late afternoon or even early evening.

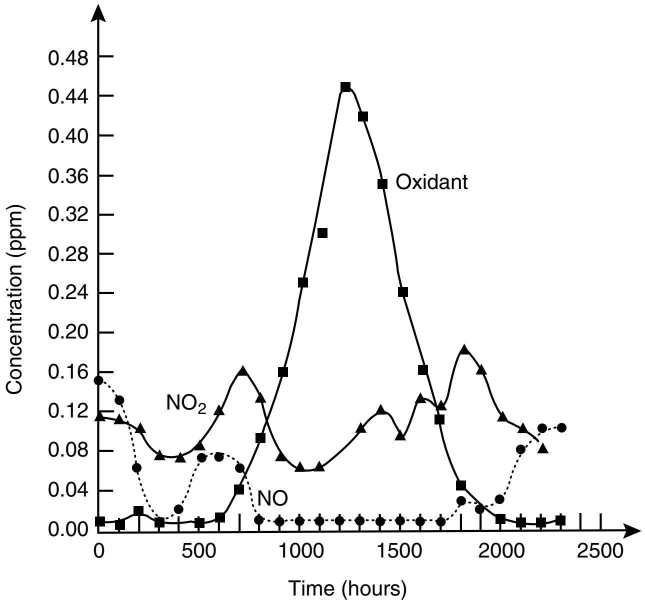


FIGURE 2 Diurnal variation of NO, NO_2, and total oxidant in Pasadena, California, on July 25, 1973. From Finlayson-Pitts and Pitts (2000). With permission.

Hydrocarbon Photooxidation

The chemistry of O_3 formation described thus far is overly simplistic. How is NO, the primary pollutant, converted to NO_2, which can be photolyzed? A clue to answering this question comes from smog-chamber studies. A smog chamber is a relatively large photochemical-reaction vessel, in which one can simulate the chemistry occurring in the urban environment. Figure 3 shows a plot of the experimentally observed loss rate for propene (a low-molecular-weight, reactive hydrocarbon commonly found in the atmosphere) in a reaction system initially containing propene, NO, and a small amount of NO_2. The observed propene-loss rate in this typical chamber run was considerably larger than that calculated due to the known reactions of propene with oxygen atoms and ozone. Hence, there must be another important hydrocarbon-loss process.

Hydroxyl radicals (OH) react rapidly with organics. Radicals, or free radicals, are reactive intermediates, such as an atom or a fragment of a molecule with an unpaired electron. Let's look at a specific sequence of reactions involving propene.

The hydroxyl radical reacts rapidly with propene:

$$OH + CH_3CH{=}CH_2 \rightarrow CH_3CHCH_2OH \qquad (4a)$$

$$OH + CH_3CH{=}CH_2 \rightarrow CH_3CHOHCH_2 \qquad (4b)$$

These reactions form radicals with an unpaired electron on the central carbon in (4a) and on the terminal carbon in (4b). These alkyl types of radicals react with O_2 to form alkylperoxy types of radicals.

$$CH_3CHCH_2OH + O_2 \rightarrow CH_3CH(O_2)CH_2OH \qquad (5a)$$

$$CH_3CHOHCH_2 + O_2 \rightarrow CH_3CHOHCH_2(O_2) \qquad (5b)$$

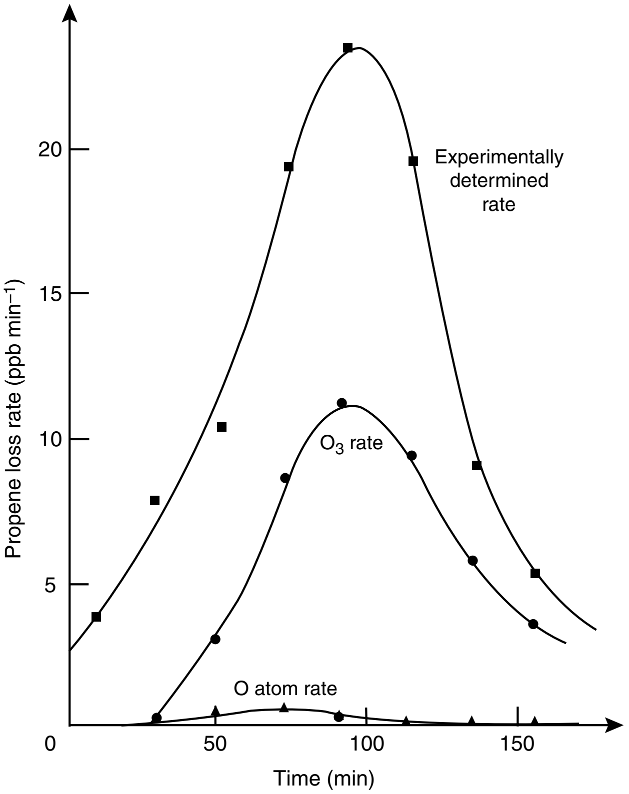


FIGURE 3 Experimentally observed rates of propene loss and calculated loss rates due to its reaction with O_3 and O atoms. From Finlayson-Pitts and Pitts (1986).

In both cases the unpaired electron is on the end oxygen in the peroxy group (in parentheses). These peroxy radicals react like all other alkylperoxy or hydroperoxy radicals under atmospheric conditions, to oxidize NO to NO_2:

$$\begin{aligned}&CH_3CH(O_2)CH_2OH + NO \rightarrow\\ &\qquad CH_3CH(O)CH_2OH + NO_2 &(6a)\\ &CH_3CHOHCH_2(O_2) + NO \rightarrow\\ &\qquad CH_3CHOHCH_2(O) + NO_2 &(6b)\end{aligned}$$

The resulting oxy radicals are then expected to dissociate to

$$CH_3CH(O)CH_2OH \rightarrow CH_3CHO + CH_2OH \quad (7a)$$
$$CH_3CHOHCH_2(O) \rightarrow CH_3CHOH + CH_2O \quad (7b)$$

Forming CH_3CHO (acetaldehyde or ethanal) and a new, one-carbon radical (7a) and HCHO (formaldehyde or methanal) and a new, two-carbon radical (7b). These new radicals are expected to react with O_2 to form the appropriate aldehyde and a hydroperoxy radical, which can oxidize NO to NO_2.

$$CH_2OH + O_2 \rightarrow HCHO + HO_2 \quad (8a)$$
$$CH_3CHOH + O_2 \rightarrow CH_3CHO + HO_2 \quad (8b)$$
$$HO_2 + NO \rightarrow OH + NO_2 \quad (9)$$

So far in this hydrocarbon oxidation process, two NO molecules have been oxidized to two NO_2 molecules, a formaldehyde and an acetaldehyde molecule have been formed, and the hydroxyl radical that initiated the reaction sequence has been re-formed. This mechanism shows the importance of the hydroxyl radical in explaining the excess removal rate of propene observed in smog-chamber studies. In addition, it provides a clue about how NO is converted to NO_2 in the atmosphere.

Hydroxyl radicals are present in the atmosphere at very low concentrations. Since the hydroxyl radical is reformed in the atmospheric photooxidation of hydrocarbons, it effectively acts as a catalyst for the oxidation of hydrocarbons. Figure 4 illustrates the role of the hydroxyl radical in initiating a chain of reactions that oxidize hydrocarbons, forming peroxy radicals that can oxidize NO to NO_2 and re-form hydroxyl radicals. The NO_2 can photolyze, leading to the formation of ozone.

PAN Formation

Acetaldehyde may react with hydroxyl radicals, forming the peroxyacetyl radical ($CH_3C(O)O_2$) under atmospheric conditions:

$$CH_3CHO + OH \rightarrow CH_3CO + H_2O \quad (10)$$
$$CH_3CO + O_2 \rightarrow CH_3C(O)O_2 \quad (11)$$

The peroxyacetyl radical may react with NO:

$$CH_3C(O)O_2 + NO \rightarrow CH_3C(O)O + NO_2 \quad (12)$$
$$CH_3C(O)O + O_2 \rightarrow CH_3O_2 + CO_2 \quad (13)$$

oxidizing NO to NO_2 and producing a methylperoxy radical. The methylperoxy radical can oxidize another NO to NO_2, forming a HO_2 (hydroperoxy) radical and a molecule of formaldehyde:

$$CH_3O_2 + NO \rightarrow CH_3O + NO_2 \quad (14)$$
$$CH_3O + O_2 \rightarrow HCHO + HO_2 \quad (15)$$

Alternatively, the peroxyacetyl radical may react with NO_2 to form peroxyacetyl nitrate ($CH_3C(O)O_2NO_2$, or PAN):

$$CH_3C(O)O_2 + NO_2 \leftrightarrow CH_3C(O)O_2NO_2 \quad (16)$$

Which reaction occurs with the peroxyacetyl radical depends on the relative concentrations of NO and NO_2 present.

PAN, like ozone, is a member of the class of compounds known as photochemical oxidants. PAN is responsible for much of the plant damage associated with photochemical-oxidant problems, and it is an eye irritant. More recent measurements of PAN throughout the troposphere have shown that PAN is ubiquitous. The only significant removal process for PAN in the lower troposphere is, as a result of its thermal decomposition, the reverse of reaction (16). This thermal decomposition of PAN is both temperature- and pressure-dependent. The lifetime for PAN ranges from about 30 minutes at 298 K to several months under conditions of the upper troposphere (Seinfeld and Pandis, 1998). In the upper troposphere, PAN is relatively stable and acts as an important reservoir for NO_x. Singh *et al.* (1994) have found that PAN is the single most abundant reactive nitrogen-containing compound

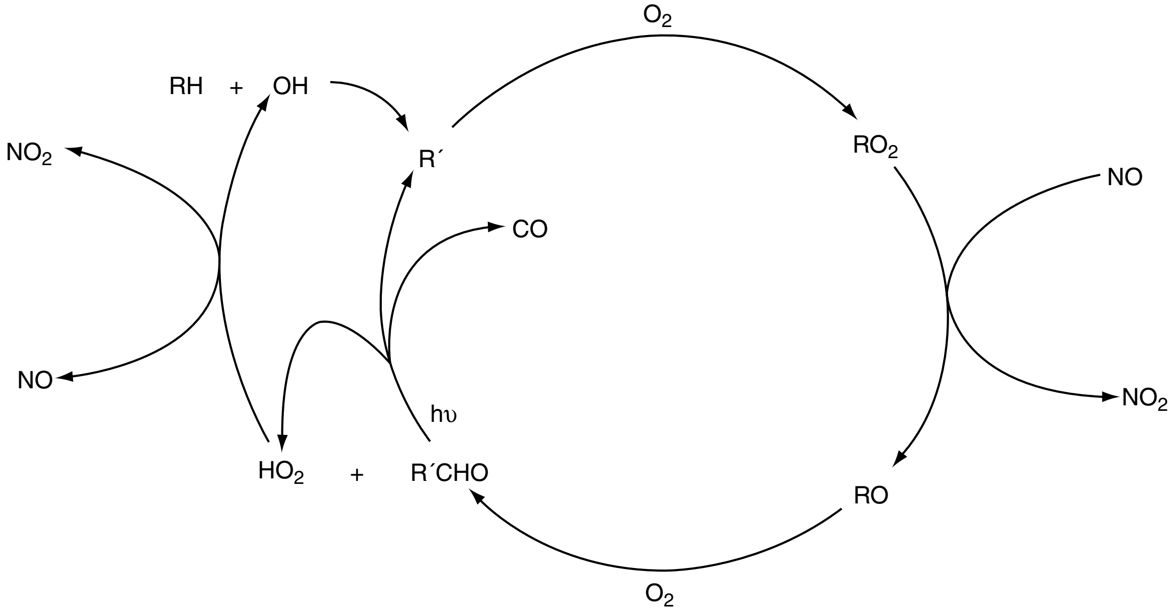


FIGURE 4 Schematic diagram illustrating the role of the hydroxyl-radical-initiated oxidation of hydrocarbons in the conversion of NO to NO_2.

in the free troposphere. Talukdar *et al.* (1995) have found that photolysis of PAN can compete with thermal decomposition for the destruction of PAN at altitudes above about 5 km. The reaction of the hydroxyl radical with PAN is less important than thermal decomposition and photolysis throughout the troposphere.

The oxidation of hydrocarbons does not stop with the formation of aldehydes or even the formation of CO. It can proceed all the way to CO_2 and H_2O. CO can also react with hydroxyl radicals to form CO_2:

$$OH + CO \rightarrow H + CO_2 \qquad (17)$$
$$H + O_2 + M \rightarrow HO_2 + M \qquad (18)$$

The chain of reactions can proceed, oxidizing hydrocarbons, converting NO to NO_2, and re-forming hydroxyl radicals until some chain-terminating reaction occurs. The following are the more important chain-terminating reactions:

$$HO_2 + HO_2 \rightarrow H_2O_2 + O_2 \qquad (19)$$
$$RO_2 + HO_2 \rightarrow ROOH + O_2 \qquad (20)$$
$$OH + NO_2 + M \rightarrow HNO_3 + M \qquad (21)$$

These reactions remove the chain-carrying hydroxyl or peroxy radicals, forming relatively stable products. Thus, the chain oxidation of the hydrocarbons and conversion of NO to NO_2 are slowed.

Radical Sources

This sequence of hydrocarbon oxidation reactions describes processes that can lead to the rapid conversion of NO to NO_2. The NO_2 thus formed can react by (1) and (2) to form O_3. In order for these processes to occur, an initial source of hydroxyl radicals is required. An important source of OH in the nonurban atmosphere is the photolysis of O_3 to produce an electronically excited oxygen atom (designated as $O(^1D)$):

$$O_3 + h\nu\ (\lambda \leq 320\ \text{nm}) \rightarrow O(^1D) + O_2 \qquad (22)$$

The excited oxygen atom can either be quenched to form a ground-state oxygen atom or react with water vapor (or any other hydrogen-containing compound) to form hydroxyl radicals:

$$O(^1D) + H_2O \rightarrow 2OH \qquad (23)$$

Other possible sources of hydroxyl radicals include the photolysis of nitrous acid (HONO), hydrogen peroxide (H_2O_2), and organic peroxides (ROOH):

$$HONO + h\nu\ (\lambda \leq 390\ \text{nm}) \rightarrow OH + NO \qquad (24)$$
$$H_2O_2 + h\nu\ (\lambda \leq 360\ \text{nm}) \rightarrow 2OH \qquad (25)$$

The atmospheric concentration of HONO is sufficiently low and photolysis sufficiently fast that HONO photolysis can only act as a radical source, in the very early morning, from HONO that builds up overnight. The photolysis of H_2O_2 and ROOH can be significant contributors to radical production, depending on the quantities of these species present in the atmosphere. Another source of radicals that can form OH radicals includes the photolysis of aldehydes, such as formaldehyde (HCHO):

$$HCOC + h\nu\ (\lambda \leq 340\ \text{nm}) \rightarrow H + HCO \qquad (26)$$
$$HCO + O_2 \rightarrow HO_2 + CO \qquad (27)$$

forming HO_2 radicals in (27) and from H atoms by reaction (18). These HO_2 radicals can react with NO by reaction (9) to form OH. The relative importance of these different sources for OH and HO_2 radicals depends on the concentrations of the different species present, the location (urban or rural), and the time of day.

Organic Reactivity

Atmospheric organic compounds have a wide range of reactivities. Table 4 lists calculated tropospheric lifetimes for selected volatile organic compounds (VOCs) due to photolysis and reaction with OH and NO_3 radicals and ozone (Seinfeld and Pandis, 1998). All of the processes identified in the table lead to the formation of organic peroxy radicals that oxidize NO to NO_2, and hence lead to ozone production. But we can see that in general the reaction of the organic molecule with the hydroxyl radical is the most important loss process.

The most important chain-terminating process in the urban atmosphere is the reaction of OH with NO_2. Hence, comparing the relative rates of the OH reaction with VOCs to that of OH with NO_2 is important for assessing the production of ozone. Seinfeld (1995) found that the rate of the OH reaction with NO_2 is about 5.5 times that for the OH reactions with a typical urban mix of VOCs, where NO_2 concentrations are in units of ppm and VOC concentrations are in units of ppm C (ppm of carbon in the VOC). If the VOC-to-NO_2 ratio is less than 5.5:1, the reaction of OH with NO_2 would be expected to predominate over the reaction of OH with VOCs. This reduces the OH involved in the oxidation of VOCs, hence inhibiting the production of O_3. On the other hand, when the ratio exceeds 5.5:1, OH preferentially reacts with VOCs, accelerating the production of radicals and hence O_3. Different urban areas are expected to have a different mix of hydrocarbons, and hence different reactivities, so this ratio is expected to change for different urban areas.

TABLE 4

Estimated tropospheric lifetimes for selected VOCs due to photolysis and reaction with OH and NO_3 radicals and ozone

	Lifetime Due to Reaction with			
	OH[a]	O_3[b]	NO_3[c]	hν
n-Butane	5.7 days	—	2.8 yr	
Propene	6.6 h	1.6 days	4.9 days	
Benzene	12 days	—	—	
Toluene	2.4 days	—	1.9 yr	
m-Xylene	7.4 h	—	200 days	
Formaldehyde	1.5 days	—	80 days	4 h
Acetaldehyde	11 h	—	17 days	5 days
Acetone	66 days	—	—	38 days
Isoprene	1.7 h	1.3 days	0.8 h	
α-Pinene	3.4 h	4.6 h	2.0 h	
β-Pinene	2.3 h	1.1 days	4.9 h	
Camphene	3.5 h	18 days	1.5 days	
2-Carene	2.3 h	1.7 h	36 min	
3-Carene	2.1 h	10 h	1.1 h	
d-Limonene	1.1 h	1.9 h	53 min	
Terpinolene	49 min	17 min	7 min	

Source: From Seinfeld and Pandis (1998). With permission.

[a] 12-hour daytime OH concentration of 1.5×10^6 molecules cm^{-3} (0.06 ppt).

[b] 24-hour average O_3 concentration of 7×10^{11} molecules cm^{-3} (30 ppb).

[c] 12-hour average NO_3 concentration of 2.4×10^7 molecules cm^{-3} (1 ppt).

Carter and Atkinson (1987) have estimated the effect of changes in the VOC composition on ozone production by use of an "incremental reactivity." This provides a measure of the change in ozone production when a small amount of VOC is added to or subtracted from the base VOC mixture at the fixed initial NO_x concentration. The incremental reactivity depends not only on the reactivity of the added VOC with OH and other oxidants, but also on the photooxidation mechanism, the base VOC mixture, and the NO_x level. Table 5 presents a table of maximum incremental reactivities (MIR) for several VOCs. The concept of MIR is useful in evaluating the effect of changing VOC components in a mixture of pollutants.

TABLE 5

Maximum incremental reactivities (MIR) for some VOCs

VOC	MIR[a] (grams of O_3 formed per gram of VOC added)
Carbon monoxide	0.054
Methane	0.015
Ethane	0.25
Propane	0.48
n-Butane	1.02
Ethene	7.4
Propene	9.4
1-Butene	8.9
2-Methylpropene (isobutene)	5.3
1,3-Butadiene	10.9
2-Methyl-1,3-butadiene (isoprene)	9.1
α-Pinene	3.3
β-Pinene	4.4
Ethyne (acetylene)	0.50
Benzene	0.42
Toluene	2.7
m-Xylene	8.2
1,3,5-Trimethylbenzene	10.1
Methanol	0.56
Ethanol	1.34
Formaldehyde	7.2
Acetaldehyde	5.5
Benzaldehyde	−0.57
Methyl *tert*-butyl ether	0.62
Ethyl *tert*-butyl ether	2.0
Acetone	0.56
C_4 ketones	1.18
Methyl nitrite	9.5

Source: From Finlayson-Pitts and Pitts (2000). With permission.

[a] From Carter (1994).

This concept of changing the VOC mixture is the basis for the use of reformulated or alternative fuels for the reduction of ozone production. Oxygenated fuel components, such as methanol, ethanol, and methyl t-butyl ether (MTBE), generally have smaller incremental reactivities than those of the larger alkanes, such as n-octane, which are more characteristic of the fuels used in automobiles. The use of these fuels would be expected to reduce the reactivity of the evaporative fuel losses from the automobiles, but the more important question is how they will change the reactivity of the exhaust emissions of VOCs. The data that are currently available suggests that there should also be a reduction in the reactivity of the exhaust emissions as well.

Ozone Isopleths

Ozone production depends on the initial amounts of VOC and NO_x in an air mass. Ozone isopleths, such as those shown in Figure 5, are contour diagrams that provide a convenient means of illustrating the way in which the maximum ozone concentration reached over a fixed irradiation period depends on the initial concentrations of NO_x and the initial concentration of VOCs. The ozone isopleths shown in Figure 5 represent model results for Atlanta, using the Carbon Bond 4 chemical mechanism (Seinfeld, 1995). The point on the contour plot represents the initial conditions containing 600 ppbC of anthropogenic controllable VOCs, 38 ppbC of background uncontrollable VOCs, and 100 ppb of NO_x. These conditions represent morning center-city conditions. The calculations are run for a 14-hour period, as chemistry proceeds and the air mass moves to the suburbs, with associated changes in mixing height and dilution. The air above the mixing layer is assumed to have 20 ppbC VOC and 40 ppb of O_3. The peak ozone concentration reached in the calculation is about 145 ppb, as indicated at the point. The isopleths arise from systematically repeating these calculations, varying the initial VOC and initial NO_x with all other conditions the same.

The base case corresponds to the point, and the horizontal line represents a constant initial NO_x concentration. At a fixed initial NO_x, as one goes from the point to a lower initial VOC, the maximum O_3 decreases, while increasing the initial VOC leads to an increase in the maximum O_3 concentration until the ridge line is reached. The ridge line represents the VOC-to-NO_x ratio that leads to the maximum ozone production at the lowest concentrations of both VOC and NO_x. The region of the isopleth diagram below the ridge line is referred to as the NO_x-limited region; it has a higher VOC:NO_x ratio. The region of the diagram above the ridge line is referred to as the VOC-limited region; it has a lower VOC:NO_x ratio. In

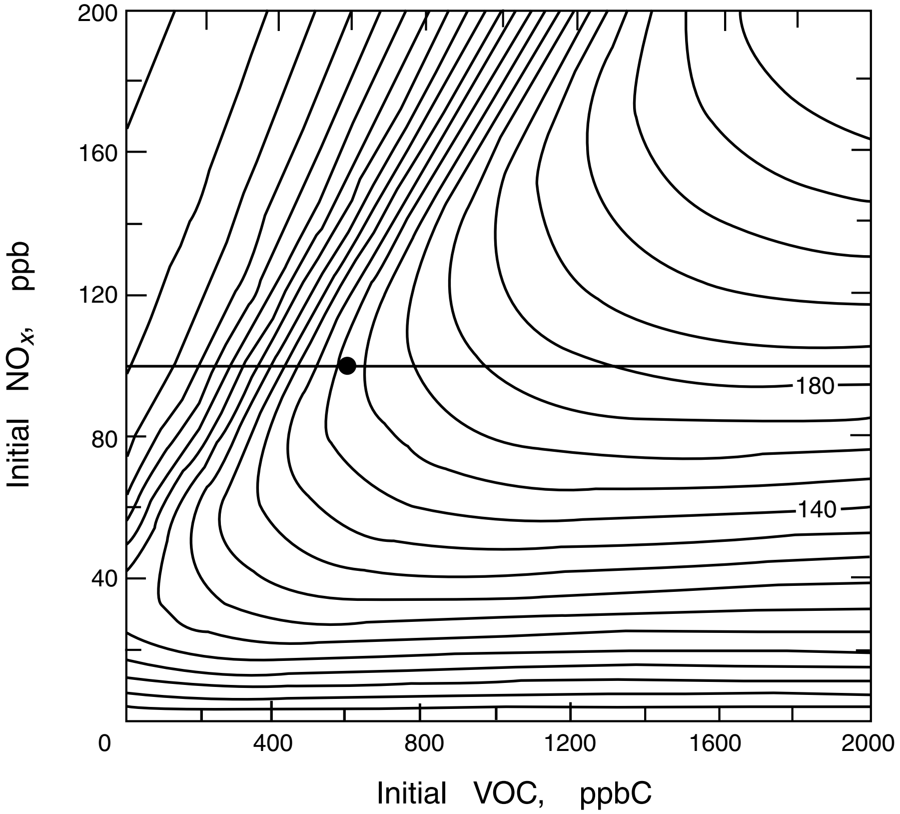


FIGURE 5 Ozone isopleth diagram for Atlanta, Georgia. Adjacent ozone isopleth lines are 10 ppb different. The point on the constant NO_x line represents the base case. From Seinfeld (1995). With permission.

the NO_x-limited region, there is inadequate NO_x present to be oxidized by all of the peroxy radicals that are being produced in the oxidation of the VOCs. Adding more NO_x in this region increases ozone production. The base-case point in Figure 5 is located in the VOC-limited region of the diagram. Increasing NO_x from the base-case point actually leads to a decrease in the maximum ozone that can be produced.

Nighttime Chemistry

At night, the urban atmospheric chemistry is quite different than during the day. The ozone present at night may react with organics, but no new ozone is formed. These ozone reactions with organics are generally slow. Ozone can react with alkanes, producing hydroxyl radicals. This hydroxyl-radical production is more important for somewhat larger alkenes. The significance of this hydroxyl-radical production is limited by the available ozone. Besides reacting with organics, ozone can react with NO_2:

$$O_3 + NO_2 \rightarrow O_2 + NO_3 \tag{28}$$

forming the nitrate radical (NO_3). NO_3 radicals can further react with NO_2 to form dinitrogen pentoxide (N_2O_5), which can dissociate to reform NO_3 and NO_2:

$$NO_3 + NO_2 + M \rightarrow N_2O_5 + M \tag{29}$$

$$N_2O_5 \rightarrow NO_3 + NO_2 \tag{30}$$

establishing an equilibrium between NO_3 and N_2O_5. Under typical urban conditions, the nighttime N_2O_5 will be 1 to 100 times the NO_3 concentration. These reactions are only of importance at night, since NO_3 can be photolyzed quite efficiently during the day.

NO_3 can also react quickly with some organics. A generic reaction, which represents reactions with alkanes and aldehydes, would be

$$NO_3 + RH \rightarrow HNO_3 + R \tag{31}$$

The reactions of NO_3 with alkenes and aromatics proceed by a different route, such as adding to the double bond. NO_3 reacts quite rapidly with natural hydrocarbons, such as isoprene and α-pinene (Table 4), and cresols (Finlayson-Pitts and Pitts, 2000). Not much is known about the chemistry of N_2O_5, other than it is likely to hydrolyze, forming nitric acid:

$$N_2O_5 + H_2O \rightarrow 2HNO_3 \tag{32}$$

Summary

The discussion of urban atmospheric chemistry presented above is greatly simplified. Many more hydrocarbon types are present in the urban atmosphere, but the examples presented should provide an idea of the types of reactions that may be of importance. In summary, urban atmospheric ozone is formed as a result of the photolysis of NO_2. NO_2 is formed by the oxidation of the primary pollutant NO, which accompanies the hydroxyl-radical-initiated chain oxidation of organics. Hydroxyl radicals can be produced by the photolysis of various compounds. Ozone formation is clearly a daytime phenomenon, as is the hydroxyl-radical attack of organics.

SECONDARY ORGANIC AEROSOLS

With the implementation of air-quality standards for fine (or respirable) particulate matter in the atmosphere, there has been increasing interest in the composition and sources of this fine particulate matter. It has long been recognized that particles in the atmosphere have both primary (direct emission) and secondary (formed in the atmosphere) sources. Among the secondary particulate matter in the atmosphere are salts of the inorganic acids (mostly nitric and sulfuric acids) formed in the atmosphere. It has been found that there is a significant contribution of carbonaceous particulate matter, consisting of elemental and organic carbon. Elemental carbon (EC), also known as black carbon or graphitic carbon, is emitted directly into the atmosphere during combustion processes. Organic carbon (OC) is both emitted directly to the atmosphere (primary OC), or formed in the atmosphere by the condensation of low-volatility products of the photooxidation of hydrocarbons (secondary OC). The organic component of ambient particles is a complex mixture of hundreds of organic compounds, including: n-alkanes, n-alkanoic acids, n-alkanals, aliphatic dicarboxylic acids, diterpenoid acids and retene, aromatic polycarboxylic acids, polycyclic aromatic hydrocarbons, polycyclic aromatic ketones and quinines, steroids, N-containing compounds, regular steranes, pentacyclic triterpanes, and iso- and anteiso-alkanes (Seinfeld and Pandis, 1998).

Secondary organic aerosols (SOAs) are formed by the condensation of low-vapor-pressure oxidation products of organic gases. The first step in organic-aerosol production is the formation of the low-vapor-pressure compound in the gas phase as a result of atmospheric oxidation. The second step involves the organic compound partitioning between the gas and particulate phases. The first step is controlled by the gas-phase chemical kinetics for the oxidation of the original organic compound. The partitioning is a physicochemical process that may involve interactions among the various compounds present in both phases. This partitioning process is discussed extensively by Seinfeld and Pandis (1998).

Figure 6 (Seinfeld, 2002) illustrates a generalized mechanism for the photooxidation of an n-alkane. The compounds shown in boxes are relatively stable oxidation products that might have the potential to partition into the particulate phase. Previous studies of SOA formation have found that the aerosol products are often di- or poly-functionally substituted products, including carbonyl groups, carboxylic acid groups, hydroxyl groups, and nitrate groups.

A large number of laboratory studies have been done investigating the formation of SOAs. Kleindienst *et al.* (2002)

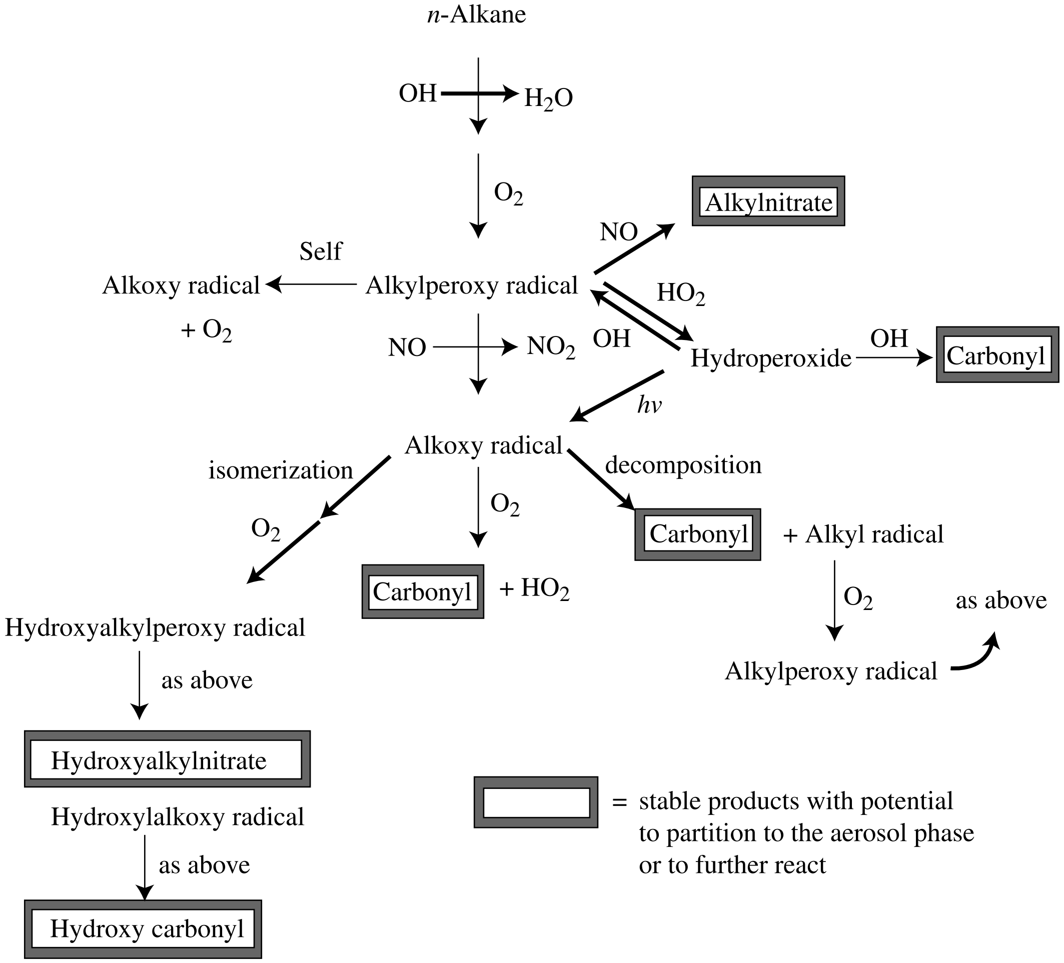


FIGURE 6 Generalized mechanism for the photooxidation of an n-alkane. The products shown in boxes are expected to be relatively stable organic products that might be able to partition into the particulate phase, if they have sufficiently low vapor pressures. From Seinfeld (2002). With permission.

have shown significant SOA formation from the irradiation of simulated auto exhaust. Griffin *et al.* (1999) have shown that the oxidation of biogenic hydrocarbons can also be important contributors to SOAs. This work also investigated the role of individual oxidation pathways, by ozone, nitrate radicals, and hydroxyl radicals. It was found that each of these oxidants can be quite important depending on the biogenic hydrocarbon with which they are reacting. Figure 7 (Seinfeld, 2002) shows an example of the partitioning of products of the ozone reaction with α-pinene between the gas and particulate phases. From this figure it is clear that the partitioning can change a lot between the various poly-functional products of the oxidation of α-pinene.

Jang *et al.* (2002) suggested that acidic aerosol surfaces may catalyze heterogeneous reactions that could lead to the formation of additional SOAs. As we will see in the next section, there is considerable potential for having acidic aerosols present in the atmosphere. The authors present data that suggests larger secondary-aerosol yields in the presence of an acid seed aerosol than occurs in the presence of a non-acid seed aerosol. The suggestion is that the acid is capable of catalyzing the formation of lower-volatility organic products, maybe through polymerization.

Pandis *et al.* (1991) have found no significant SOA formation from the photooxidation of isoprene, due to its small size and the high volatility of its oxidation products. Significant SOAs are formed from biogenic hydrocarbons larger than isoprene. Claeys *et al.* (2004) suggest that the yield of SOAs from the photooxidation of isoprene in the Amazonian rain forest, where NO_x is low (<100 ppt), is about 0.4% on a mass basis. Even with its low particulate yield, since the global annual isoprene emissions are about 500 Tg per year, the SOAs from isoprene photooxidation alone could account for about 2 Tg/yr. This is a significant fraction of the Intergovernmental Panel on Climate Change (Houghton *et al.*, 2001) estimate of between 8 and 40 Tg/yr of SOAs from biogenic sources. The oxidation of the other biogenic hydrocarbons are expected to have much higher SOA yields.

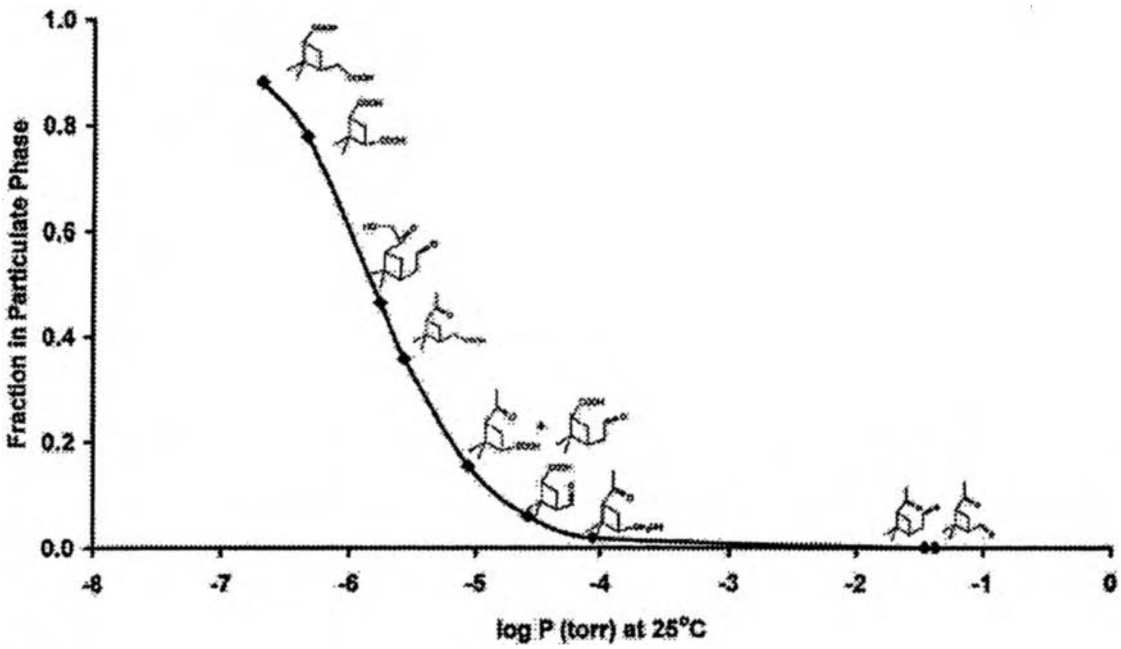


FIGURE 7 Partitioning of the products of the ozone reaction with α-pinene between the gas and particulate phases, assuming a total organic aerosol loading of 50 μg/m^3. From Seinfeld (2002). With permission.

CHEMISTRY OF ATMOSPHERIC ACID FORMATION

Acid deposition has long been recognized to be a serious problem in Scandinavian countries, and throughout Europe, much of the United States, and Canada. Most of the concerns about acid deposition are related to the presence of strong inorganic acids, nitric acid (HNO_3) and sulfuric acid (H_2SO_4), in the atmosphere. Sulfur dioxide (SO_2) and nitrogen oxides (NO_x) are emitted from numerous stationary and mobile combustion sources scattered throughout the industrialized nations of the world. As this polluted air is transported over large distances, 500 km and more, the sulfur and nitrogen oxides can be further oxidized, ultimately to the corresponding acids. The 1990 Clean Air Act Amendments require significant reductions in SO_2 from power plants in the eastern portion of the United States. Less significant reductions of NO_x emissions are also required.

As was suggested earlier, one of the primary goals of air-pollution research is to take information about emissions, topography, meteorology, and chemistry and develop a mathematical model to predict acid deposition in the model area. The type of model used to do this is known as a long-range transport (LRT) model, where the dimensions are on the order of 1000 km or more. The acid deposition that is observed is produced by the chemical processes occurring in the atmosphere during the transport. Prediction of the effects of any reduction in emissions of sulfur and nitrogen oxides requires a detailed understanding of the atmospheric reactions involved in the oxidations.

Pollutant emissions are transported by the winds for hundreds of kilometers within the boundary or “mixing” layer of the atmosphere. This layer is approximately 1000 m deep and well mixed, allowing pollutants to be dispersed both horizontally and vertically throughout this layer. In the boundary layer, a variety of chemical and physical processes affect the concentrations of the pollutants. To form the acids, the sulfur and nitrogen oxides must react with some oxidants present in the atmosphere. The most important gas-phase oxidants were discussed above. These oxidation processes may occur in the gas phase, or they may occur as aqueous phase reactions in clouds. The gas-phase oxidations of sulfur and nitrogen oxides are better quantified than are the aqueous-phase oxidations.

Gas-Phase Processes

There are three potentially important gas-phase oxidation processes for producing nitric acid. These processes were identified earlier: the reaction of hydroxyl radicals with NO_2 (21), hydrogen abstraction reactions from organics by NO_3 (31), and the reaction of N_2O_5 with water (32). During the day, the dominant process leading to the formation of HNO_3 is reaction (21). At night, the N_2O_5 reaction with water vapor (32) is important. The hydrogen atom abstraction reaction of NO_3 with organics is expected to be of relatively minor importance. The 24-hour averaged rate of NO_2 conversion to HNO_3 during the summer at 50% relative humidity is expected to be between 15%/hour and 20%/hour.

Calvert and Stockwell (1983) have shown that the gas-phase oxidation of sulfur dioxide is primarily by the reaction of the hydroxyl radical with SO_2:

$$HO + SO_2 + M \rightarrow HOSO_2 + M \quad (33)$$

$$HOSO_2 + O_2 \rightarrow HO_2 + SO_3 \quad (34)$$

$$SO_3 + H_2O \rightarrow H_2SO_4 \quad (35)$$

In this sequence of reactions, the OH radical initiates the oxidation of SO_2. The bisulfite radical ($HOSO_2$) product reacts rapidly with oxygen to form sulfur trioxide (SO_3) and HO_2. The HO_2 radical can be converted back to OH by reaction (9), and the SO_3 can react with water to form sulfuric acid. The details of the kinetics of these processes have been presented by Anderson *et al.* (1989). This sequence of reactions can be simplified for modeling purposes to the reaction

$$OH + SO_2 \ (+ \ O_2, H_2O) \rightarrow H_2SO_4 + HO_2 \quad (36)$$

The modeling suggests that for moderately polluted and mildly polluted cases described above, the maximum SO_2 oxidation rates were 3.4%/hour and 5.4%/hour. These maximum conversions occurred near noon, when the OH concentration was a maximum. The conversion of SO_2 to H_2SO_4 for a clear summertime 24-hour period was 16% and 24% for the moderately and mildly polluted conditions. The gas-phase oxidation of both NO_2 and SO_2 vary considerably, depending on the concentrations of other species in the atmosphere. But the gas-phase oxidation of SO_2 is always going to be much slower than that for NO_2.

Aqueous-Phase Chemistry

Aqueous-phase oxidations of nitrogen oxides are not believed to be very important in the atmosphere. On the other hand, the aqueous-phase oxidations of sulfur dioxide appear to be quite important. Sulfur dioxide may dissolve in atmospheric water droplets, to form mainly the bisulfite ion (HSO_3^-):

$$SO_2 + \text{Cloud} \rightarrow SO_2{\cdot}H_2O \rightarrow HSO_3^- + H^+ \quad (37)$$

The concentration of the bisulfite ion in the droplet is dependent on the Henry's law constant (H), which determines the solubility of SO_2 in water, the equilibrium constant (K) for the first dissociation of the hydrated SO_2, the gas-phase SO_2 concentration, and the acidity of the solution.

$$[HSO_3^-] = KH\,[SO_2]_{gas}/[H^+]$$

$SO_2{\cdot}H_2O$, HSO_3^-, and SO_3^{2-} are all forms of sulfur (IV) (S(IV)). At normal pH levels, the bisulfite ion is the predominate form of sulfur (IV) in aqueous systems, and the form that needs to be oxidized to the sulfate ion (SO_4^{2-}), sulfur (VI). HSO_3^- can be oxidized by oxygen, but this process is very slow. The reaction may be catalyzed by transition metal ions, such as manganese (Mn^{2+}) and iron (Fe^{3+}). The importance of these metal-catalyzed oxidations depends strongly on the concentration of metal ions present. This may be enhanced by passing through heavily industrialized areas, where there might be sources of these metals for the atmosphere.

Ozone and hydrogen peroxide are likely to be more important catalysts for the oxidation of S(IV). The rate of ozone-catalyzed oxidation of S(IV) decreases as the pH of the solution decreases (or as the solution becomes more acidic). Since the HSO_3^- concentration depends inversely on $[H^+]$, the rate of oxidation of S(IV) slows down considerably as the pH decreases ($[H^+]$ increases). This reaction is likely to be of importance at pH $\geqslant$ 4.5.

Hydrogen peroxide, on the other hand, is much more soluble than ozone. Hence, even though the gas-phase concentrations are much lower than ozone, the aqueous concentrations can be high. The rate constant for the hydrogen-peroxide-catalyzed reaction increases as the pH decreases, down to a pH of about 2.0. At a pH of 4.5, the oxidation catalyzed by 1 ppb of gaseous H_2O_2 in equilibrium with the aqueous phase is about 100 times faster than the ozone-catalyzed oxidation by 50 ppb of gaseous O_3 in equilibrium with the aqueous phase. Figure 8 shows a comparison of aqueous-phase

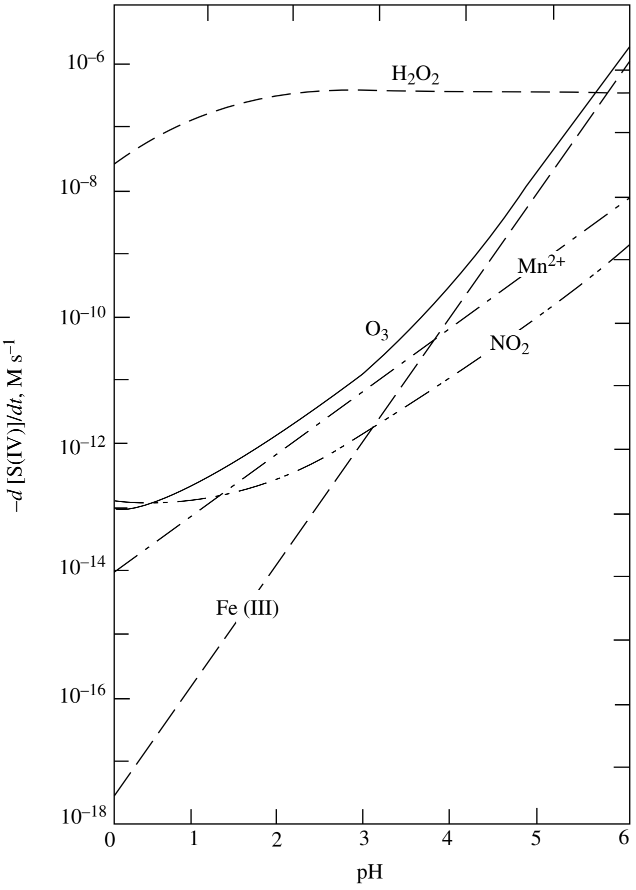


FIGURE 8 Comparison of aqueous-phase oxidation paths; the rate of conversion of S(IV) to S(VI) as a function of pH. Conditions assumed are: $[SO_2(g)] = 5$ ppb; $[NO_2(g)] = 1$ ppb; $[H_2O_2(g)] = 1$ ppb; $[O_3(g)] = 50$ ppb; $[Fe(III)] = 0.3$ µM; and $[Mn(II)] = 0.03$ µM. From Seinfeld and Pandis (1998). With permission.

catalyzed SO_2 oxidation paths as a function of pH. In the case of the H_2O_2-catalyzed oxidation of S(IV), the rate of oxidation will be limited by the H_2O_2 present in the cloud or available to the cloud. This leads to the rate of S(IV) conversion to S(VI) being limited by the rate at which gaseous H_2O_2 is incorporated into the aqueous phase of the clouds by updrafts and entrainment.

Natural Sources of Acids and Organic Acids

There are a variety of potential natural sources of acids in the atmosphere. Dimethyl sulfide (DMS) is one of the most important natural sulfur compounds emitted from the oceans (Cocks and Kallend, 1988). Hydroxyl radicals may react with DMS by either of two possible routes:

$$OH + CH_3SCH_3 \rightarrow CH_3S(OH)CH_3 \qquad (38)$$

$$OH + CH_3SCH_3 \rightarrow CH_3SCH_2 + H_2O \qquad (39)$$

addition to the sulfur or abstraction of a hydrogen atom from one of the methyl groups. For the first case, the product is proposed to react with oxygen:

$$CH_3S(OH)CH_3 + 2O_2 \rightarrow CH_3SO_3H + CH_3O_2 \qquad (40)$$

eventually forming methane sulfonic acid (CH_3SO_3H, or MSA). Many organic S(IV) compounds are easily hydrolyzed to inorganic S(IV), which can be oxidized to S(VI). For the second path, the alkyl-type radical is expected to react with molecular oxygen to form a peroxy-type radical, followed by the oxidation of NO to NO_2:

$$CH_3SCH_2 + O_2 \rightarrow CH_3SCH_2O_2 \qquad (41)$$

$$CH_3SCH_2O_2 + NO + 2O_2 \rightarrow NO_2 + HCHO + SO_2 + CH_3O_2 \qquad (42)$$

The details of this mechanism are not well established, but the suggestion is that DMS, which is produced by biogenic processes, can be partially oxidized to SO_2, hence contributing to the SO_2 observed in the atmosphere. This SO_2 would be oxidized by the same routes as the anthropogenic SO_2. Several of the papers included in the volume by Saltzman and Cooper (1989) have presented a much more complete discussion of the role of biogenic sulfur in the atmosphere.

In recent years, it has become increasingly obvious that there are substantial contributions of organic acids (carboxylic acids) in the atmosphere (Chebbi and Carlier, 1996). It has been found that formic acid (HCOOH) and acetic acid (CH_3COOH) are the most important gas-phase carboxylic acids identified in the atmosphere. Concentrations in excess of 10 ppb of these compounds have been observed in polluted urban areas. Concentrations of these acids have been observed in excess of 1 ppb, in the Amazon forest, particularly during the dry season. A very wide range of mono- and dicarboxylic acids have been observed in the aqueous phase, rain, snow, cloud water, and fog water. Dicarboxylic acids are much more important in aerosol particles, since they have much lower vapor pressures than do monocarboxylic acids. Carboxylic acids have been observed as direct emissions from biomass burning, in motor-vehicle exhaust, and in direct biogenic emissions. Carboxylic acids are also produced in the atmosphere. The most important gas-phase reactions for the production of carboxylic acids are as a product of the ozone oxidation of alkenes. Aqueous-phase oxidation of formaldehyde is believed to be a major source of formic acid, maybe more important than the gas-phase production. Carboxylic acids are, in general, relatively unreactive; their primary loss processes from the atmosphere are believed to be wet and dry deposition.

Summary

Much of the atmospheric acidity results from the oxidation of nitrogen oxides and sulfur oxides. In the case of nitrogen oxides, this oxidation is primarily due to the gas-phase reaction of OH with NO_2. In the case of sulfur oxides, the comparable reaction of OH with SO_2 is much slower, but is likely to be the dominant oxidation process in the absence of clouds. When clouds are present, the aqueous-phase oxidation of SO_2 is expected to be more important. At higher pH, the more important aqueous oxidation of SO_2 is likely to be catalyzed by ozone, while at lower pH, the H_2O_2-catalyzed oxidation is likely to be more important. Organic acids also contribute significantly to the acidity observed in the atmosphere.

POLAR STRATOSPHERIC OZONE

In 1974, Molina and Rowland proposed that chlorofluorocarbons (CFCs) were sufficiently long-lived in the troposphere to be able to diffuse to the stratosphere, where effects on ozone would be possible. They shared in the 1995 Nobel Prize in chemistry for this work.

More recently an ozone "hole" has been observed in the stratosphere over Antarctica, which becomes particularly intense during the Southern Hemispheric spring, in October. This led attention to be shifted to the polar regions, where effects of CFCs on stratospheric ozone content have been observed. Before dealing with this more recent discovery, it is necessary to provide some of the background information about the stratosphere and its chemistry.

The stratosphere is the region of the atmosphere lying above the troposphere. In the troposphere, the temperature of the atmosphere decreases with increasing altitude from about 290 K near the surface to about 200 K at the tropopause. The tropopause is the boundary between the troposphere and the stratosphere, where the temperature reaches a minimum. The altitude of the tropopause varies with season and latitude between altitudes of 10 and 17 km. Above the tropopause, in the stratosphere, the temperature increases with altitude up to about 270 K near an altitude of 50 km. In the troposphere, the warmer air is below the cooler air. Since warmer air is less dense, it tends to rise; hence there is relatively good vertical mixing in the troposphere. On the other hand, in the stratosphere the warmer air is on top, which leads to poor vertical mixing and a relatively stable atmosphere.

Stratospheric Ozone Balance

In the stratosphere, there is sufficient high-energy ultraviolet radiation to photolyze molecular oxygen:

$$O_2 + h\nu\ (\lambda \leq 240\ \text{nm}) \rightarrow 2O(^3P) \tag{43}$$

This will be followed by the oxygen-atom reaction with O_2 (2) forming ozone. These processes describe the ozone production in the stratosphere. They are also the processes responsible for the heating in the upper stratosphere. This ozone production must be balanced by ozone-destruction processes. If we consider only oxygen chemistry, ozone destruction is initiated by ozone photolysis (22), forming an oxygen atom. The oxygen atom can also react with ozone, re-forming molecular oxygen:

$$O(^3P) + O_3 \rightarrow 2O_2 \tag{44}$$

Reactions (43), (2), (22), and (44) describe the formation and destruction of stratospheric ozone with oxygen-only chemistry. This is commonly known as the Chapman mechanism.

Other chemical schemes also contribute to the chemistry in the natural (unpolluted) stratosphere. Water can be photolyzed, forming hydrogen atoms and hydroxyl radicals:

$$H_2O + h\nu\ (\lambda \leq 240\ \text{nm}) \rightarrow H + OH \tag{45}$$

The OH radical may react with ozone to form HO_2, which may in turn react with an O atom to reform OH. The net effect is the destruction of odd oxygen (O and/or O_3).

$$OH + O_3 \rightarrow HO_2 + O_2 \tag{46}$$
$$HO_2 + O \rightarrow OH + O_2 \tag{47}$$
$$O + O_3 \rightarrow 2O_2 \tag{Net}$$

These reactions form a catalytic cycle that leads to the destruction of ozone. An alternative cycle is

$$H + O_3 \rightarrow OH + O_2 \tag{48}$$
$$OH + O \rightarrow H + O_2 \tag{49}$$
$$O + O_3 \rightarrow 2O_2 \tag{Net}$$

Other catalytic cycles involving HO_x species (H, OH, and HO_2) are possible. Analogous reactions may also occur involving NO_x species (NO and NO_2),

$$NO + O_3 \rightarrow NO_2 + O_2 \tag{3}$$
$$NO_2 + O \rightarrow NO + O_2 \tag{50}$$
$$O + O_3 \rightarrow 2O_2 \tag{Net}$$

and ClO_x species (Cl and ClO),

$$Cl + O_3 \rightarrow ClO + O_2 \tag{51}$$
$$ClO + O \rightarrow Cl + O_2 \tag{52}$$
$$O + O_3 \rightarrow 2O_2 \tag{Net}$$

These processes are some of the ozone-destruction processes of importance in the stratosphere. These types of processes contribute to the delicate balance between the stratospheric ozone production and destruction, which provide the natural control of stratospheric ozone, when the stratospheric HO_x, NO_x, and ClO_x species are of natural origin.

Ozone plays an extremely important role in the stratosphere. It absorbs virtually all of the solar ultraviolet radiation between 240 and 290 nm. This radiation is lethal to single-cell organisms, and to the surface cells of higher plants and animals. Stratospheric ozone also reduces the solar ultraviolet radiation up to 320 nm, wavelengths that are also biologically active. Prolonged exposure of the skin to this radiation in susceptible individuals may lead to skin cancer. In addition, stratospheric ozone is the major heat source for the stratosphere, through the absorption of ultraviolet, visible, and infrared radiation from the sun. Hence, changes in the stratospheric ozone content could lead to significant climatic effects.

Stratospheric Pollution

Over the past 30 years, there has been considerable interest in understanding the ways in which man's activities might be depleting stratospheric ozone. Major concerns first arose from considerations of flying a large fleet of supersonic aircraft in the lower stratosphere. These aircraft were expected to be a significant additional source of NO_x in the stratosphere. This added NO_x could destroy stratospheric O_3 by the sequence of reactions (3) and (50) and other similar catalytic cycles. The environmental concerns, along with economic factors, were sufficient to limit the development of such a fleet of aircraft.

More recently, environmental concern has turned to the effects of chlorofluorocarbons on the stratospheric ozone. These compounds were used extensively as aerosol propellants and foam-blowing agents and in refrigeration systems. The two most commonly used compounds were $CFCl_3$ (CFC-11) and CF_2Cl_2 (CFC-12). These compounds are very stable, which allows them to remain in the atmosphere sufficiently long that they may eventually diffuse to the stratosphere. There they may be photolyzed by the high-energy ultraviolet radiation:

$$CFCL_3 + h\nu\ (\lambda \leq 190\ \text{nm}) \rightarrow CFCl_2 + Cl \tag{53}$$

This reaction, and similar reactions for other chlorinated compounds, leads to a source of chlorine atoms in the stratosphere. These chlorine atoms may initiate the catalytic destruction of ozone by a sequence of reactions, such as reactions (51) and (52). Numerous other catalytic destruction cycles have been proposed, including cycles involving combinations of ClO_x, HO_x, and NO_x species.

In recent years, our ability to model stratospheric chemistry has increased considerably, which allows good comparisons between model results and stratospheric measurements. Based upon our improved understanding of the stratosphere and the continuing concern with CFCs, about 45 nations met during the fall of 1987 to consider limitations on the production and consumption of CFCs. This led to an agreement to freeze consumption of CFCs at 1986 levels, effective in September 1988, and requirements to reduce consumption by 20% by 1992 and by an additional 30% by 1999. In November 1992, the Montreal Protocol on Substances That Deplete the Ozone Layer revised the phase-out schedule for CFCs to a complete ban on production by January 1, 1996. In November 1995, additional amendments were adopted to freeze the use of hydrogen-containing CFCs (HCFCs) and methyl bromide (CH_3Br) and eliminate their use by 2020 and 2010, respectively. These agreements were very important steps to addressing the problem of CFCs in the atmosphere. This has also led to major efforts to find environmentally safe alternatives to these compounds for use in various applications.

Antarctic Ozone

Farman *et al.* (1985) observed a very significant downward trend in the total ozone column measured over Halley Bay, Antarctica (Figure 9). Solomon (1988) has reviewed this and other data from Antarctica, and has concluded that there has been a real decrease in the ozone column abundance in the South Polar region. Other data suggest that the bulk of the effect on ozone abundance is at lower altitudes in the stratosphere, between about 12 and 22 km, where the stratospheric ozone concentrations decrease quickly and return to near normal levels as the springtime warms the stratosphere. The subsequent discussion will outline some of the chemical explanations for these observations. Some atmospheric dynamical explanations of the ozone hole have been proposed, but these are not believed to provide an adequate explanation of the observations.

Figure 10 shows plots of results from flights in the Antarctic region during August and September 1987 (Anderson *et al.*, 1991). Ozone- and ClO-measurement instrumentation was flown into the polar stratosphere on a NASA ER-2 aircraft (a modified U-2). This figure shows a sharp increase in ClO concentration as one goes toward the pole and a similar sharp decrease in stratospheric ozone. On the September 16th flight, the ClO concentration rose from about 100 to 1200 ppt while the ozone concentration dropped from about 2500 to 1000 ppb. This strong anticorrelation is consistent with the catalytic ozone-destruction cycle, reactions (51) and (52).

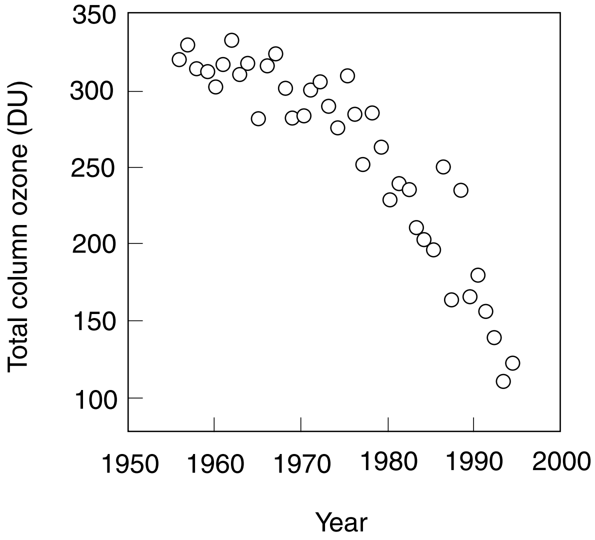


FIGURE 9 Average total column ozone measured in October at Halley Bay, Antarctica, from 1957 to 1994. Ten additional years of data are shown in this plot beyond the period presented by Farman *et al.* (1985). From Finlayson-Pitts and Pitts (2000). With permission.

Solomon (1988) has suggested that polar stratospheric clouds (PSCs) play an important role in the explanation of the Antarctic ozone hole. PSCs tend to form when the temperature drops below about 195 K and are generally observed in the height range from 10 to 25 km. The stratosphere is sufficiently dry that cloud formation does not occur with water-forming ice crystals alone. At 195 K, nitric acid-trihydrate will freeze to form cloud particles, and there is inadequate water alone to form ice, until one goes to an even lower temperature. Significant quantities of nitric acid are in the cloud particles below 195 K, while they would be in the gas phase at higher temperatures. PSCs are most intense in the Antarctic winter and decline in intensity and altitude in the spring, as the upper regions of the stratosphere begin warming. It was proposed that HCl(a) ((a)—aerosol phase), absorbed on the surfaces of PSC particles, and gaseous chlorine nitrate, $ClONO_2(g)$, react to release Cl_2 to the gas phase:

$$ClONO_2(g) + HCl(a) \rightarrow Cl_2(g) + HNO_3(a) \qquad (54)$$

Subsequent research identified several other gas-surface reactions on PSCs that also play an important role in polar stratospheric ozone depletion

$$ClONO_2(g) + H_2O(a) \rightarrow HOCl(g) + HNO_3(a,g) \qquad (55)$$
$$HOCl(g) + HCl(a) \rightarrow Cl_2(g) + H_2O(a) \qquad (56)$$
$$N_2O_5(g) + H_2O(a) \rightarrow 2HNO_3(a,g) \qquad (57)$$

Reactions (55) and (56) have the same net effect as reaction (54), while reaction (57) removes reactive nitrogen oxides from the gas phase, reducing the rate of ClO deactivation by

$$ClO + NO_2 \rightarrow ClONO_2 \qquad (58)$$

Webster *et al.* (1993) made the first *in situ* measurement of HCl from the ER-2 aircraft. These results suggested that HCl is not the dominant form of chlorine in the midlatitude lower stratosphere, as had been believed. These results suggested that HCl constituted only about 30% of the inorganic chlorine. This has led to the belief that $ClONO_2$ may be present at concentrations that exceed that of HCl.

Figure 11 shows a chronology of the polar ozone-depletion process. As one enters the polar night, $ClONO_2$ is the dominant inorganic chlorine-containing species, followed by HCl and ClO. Due to the lack of sunlight, the temperature decreases and polar stratospheric clouds form, permitting reactions (54), (55), and (56) to proceed, producing gaseous Cl_2. Both HCl and $ClONO_2$ decrease. As the sun rises, the Cl_2 is photolyzed, producing Cl atoms that react

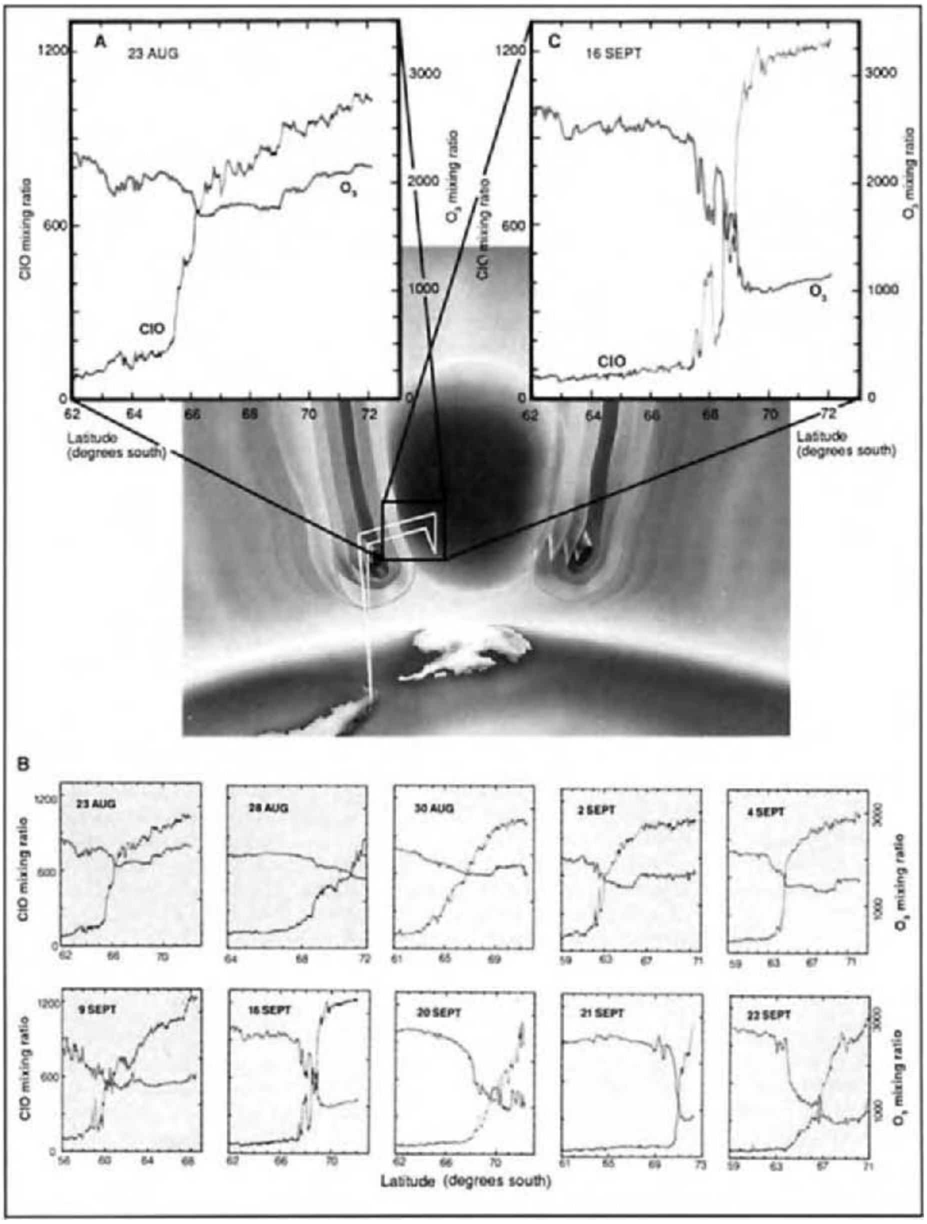


FIGURE 10 Rendering of the containment provided by the circumpolar jet that isolates the region of highly enhanced ClO (shown in green) over the Antarctic continent. Evolution of the anticorrelation between ClO and O_2 across the vortex transition is traced from: (A) the initial condition observed on 23 August 1987 on the south-bound log of the flight; (B) summary of the sequence over the ten-flight series; (C) imprint on O_3 resulting from 3 weeks of exposure to elevated levels of ClO. Data panels do not include dive segment of trajectory; ClO mixing ratios are in parts per trillion by volume; O_3 mixing ratios are in parts per billion by volume. From Anderson *et al*. (1991). With permission.

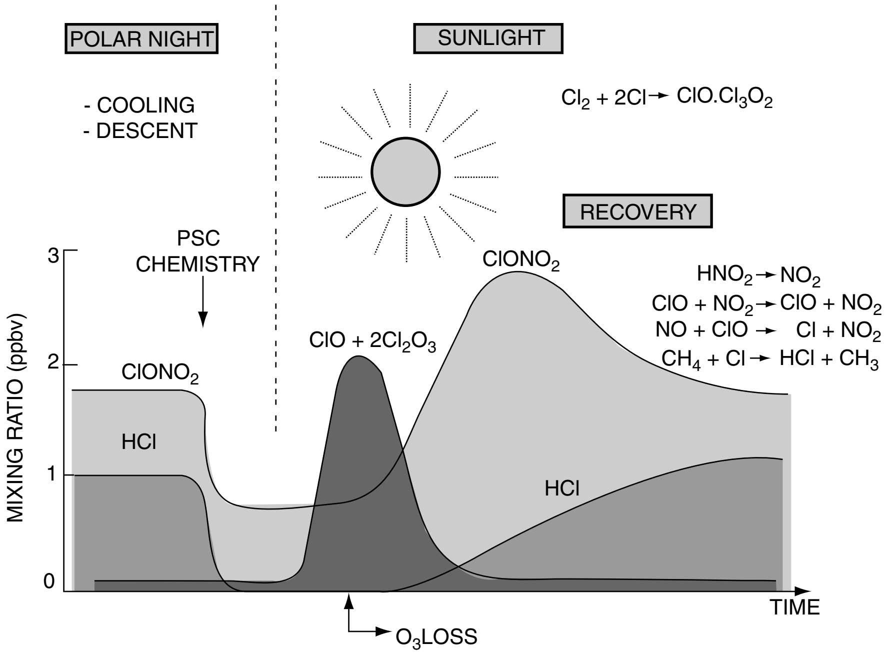


FIGURE 11 Schematic of the time evolution of the chlorine chemistry, illustrating the importance of the initial HCl/$ClONO_2$ ratio, the sudden formation of ClO with returning sunlight, the way in which $ClONO_2$ levels can build up to mixing ratios in excess of its initial values, and the slow recovery of HCl levels. From Webster *et al.* (1993). With permission.

with ozone to form ClO. This ClO may react with itself to form the dimer, $(ClO)_2$:

$$ClO + ClO + M \rightarrow (ClO)_2 + M \qquad (59)$$

Under high-ClO-concentration conditions, the following catalytic cycle could be responsible for the destruction of ozone:

$$\begin{aligned} 2 \times (Cl + O_3 &\rightarrow ClO + O_2) && (51)\\ ClO + ClO + M &\rightarrow (ClO)_2 + M && (59)\\ (ClO)_2 + h\nu &\rightarrow Cl + ClOO && (60)\\ ClOO + M &\rightarrow Cl + O_2 && (61)\\ 2O_3 &\rightarrow 3O_2 && (\text{Net}) \end{aligned}$$

This ClO-driven catalytic cycle can effectively destroy O_3, but it requires the presence of sunlight to photolyze Cl_2 and $(ClO)_2$. The presence of sunlight will lead to an increase in temperature that releases HNO_3 back to the gas phase. The photolysis of HNO_3 can release NO_2, which can react with ClO by reaction (58) to re-form $ClONO_2$. This can terminate the unusual chlorine-catalyzed destruction of ozone that occurs in polar regions.

Anderson (1995) suggests that the same processes occur in both the Arctic and Antarctic polar regions. The main distinction is that it does not get as cold in the Arctic, and the polar stratospheric clouds do not persist as long after the polar sunrise. As the temperature rises above 195 K, nitric acid is released back into the gas phase only shortly after Cl_2 photolysis begins. As nitric acid is photolyzed, forming NO_2, the ClO reacts with NO_2 to form $ClONO_2$ and terminate the chlorine-catalyzed destruction of ozone. Anderson (1995) suggests that the temperatures warmed in late January 1992, and ozone loss was only 20 to 30% at the altitudes of peak ClO. The temperatures remained below 195 K until late February 1993, and significantly more ozone will be lost. The delay between the arrival of sunlight and the rise of temperatures above 195 K are crucial to the degree of ozone loss in the Arctic.

Summary

The observations made in the polar regions provided the key link between chlorine-containing compounds in the stratosphere and destruction of stratospheric ozone. These experimental results led to the Montreal Protocol agreements and their subsequent revisions to accelerate the phase-out of the use of CFCs. A tremendous amount of scientific effort over many years has led to our current understanding of the effects of Cl-containing species on the stratosphere.

CLOSING REMARKS

Our knowledge and understanding has improved considerably in recent years. Much of the reason for this improved knowledge is the result of trying to understand how we are affecting our environment. From the foregoing discussion, it is clear that atmospheric chemistry is quite complex. It has been through the diligent research of numerous individuals, that we have been able to collect pertinent pieces of information that can be pulled together to construct a more complete description of the chemistry of the atmosphere.

REFERENCES

Anderson, J.G. (1995), Polar processes in ozone depletion, in *Progress and Problems in Atmospheric Chemistry,* World Scientific Publishers, Singapore, pp. 744–770.

Anderson, J.G., D.W. Toohey, and W.H. Brune (1991), Free radicals within the Antarctic vortex: The role of CFCs in Antarctic ozone loss, *Science,* **251,** 39–46.

Anderson, L.G., P.M. Gates, and C.R. Nold (1989), Mechanism of the atmospheric oxidation of sulfur dioxide by hydroxyl radicals, in *Biogenic Sulfur in the Environment,* E.S. Saltzman and W.J. Cooper, eds., American Chemical Society, Washington, D.C., pp. 437–449.

Calvert, J.G., and W.R. Stockwell (1983), Acid generation in the troposphere by gas-phase chemistry, *Environ. Sci. Technol.,* **17,** 428A–443A.

Carter, W.P.L. (1994), Development of ozone reactivity scales for volatile organic compounds, *J. Air & Waste Manage. Assoc.,* **44,** 881–899.

Carter, W.P.L., and R. Atkinson (1987), An experimental study of incremental hydrocarbon reactivity, *Environ. Sci. Technol.,* **21,** 670–679.

Chebbi, A., and P. Carlier (1996), Carboxylic acids in the troposphere, occurrence, sources, and sinks: A review, *Atmos. Environ.,* **30,** 4233–4249.

Claeys, M., B. Graham, G. Vas, W. Wang, R. Vermeylen, V. Pashynska, J. Cafmeyer, P. Guyon, M.O. Andreae, P. Artaxo, and W. Maenhaut (2004), Formation of secondary organic aerosols through photooxidation of isoprene, *Science,* **303,** 1173–1176.

Cocks, A. and T. Kallend (1988), The chemistry of atmospheric pollution, *Chem. Britain,* **24,** 884–888.

Farman, J.C., B.G. Gardiner, and J.D. Shanklin (1985), Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, *Nature,* **315,** 207–210.

Finlayson-Pitts, B.J., and J.N. Pitts, Jr. (1986), *Atmospheric Chemistry: Fundamentals and Experimental Techniques,* Wiley & Sons, New York.

Finlayson-Pitts, B.J., and J.N. Pitts, Jr. (2000), *Chemistry of the Upper and Lower Atmosphere,* Academic Press, San Diego, CA.

Griffin, R.J., D.R. Cocker III, R.C. Flagan, and J.H. Seinfeld (1999), Organic aerosol formation from the oxidation of biogenic hydrocarbons, *J. Geophys. Res.,* **104D,** 3555–3567.

Heicklen, J. (1976), *Atmospheric Chemistry,* Academic Press, New York.

Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson, eds. (2001), *Climate Change 2001: The Scientific Basis,* published for the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge. http://www.grida.no/climate/ipcc_tar/wg1/index.htm

Jang, M., N.M. Czoschke, S. Lee, and R.M. Kamens (2002), Heterogeneous atmospheric aerosol production by acid-catalyzed particle-phase reactions, *Science,* **298,** 814–817.

Kleindienst, T.E., E.W. Corse, W. Li, C.D. McIver, T.S. Conver, E.O. Edney, D.J. Driscoll, R.E. Speer, W.S. Weathers, and S.B. Tejada (2002), Secondary organic aerosol formation from the irradiation of simulated automobile exhaust, *J. Air & Waste Manage. Assoc.,* **52,** 259–272.

Molina, M.J. and F.S. Rowland (1974), Stratospheric sink for chlorofluoromethanes: Chlorine atom-catalysed destruction of ozone, *Nature,* **249,** 810–812.

Pandis, S.N., S.E. Paulson, J.H. Seinfeld, and R.C. Flagan (1991), Aerosol formation in the photooxidation of isoprene and β-pinene, *Atmos. Environ.,* **25,** 997–1008.

Saltzman, E.S. and W.J. Cooper, eds. (1989), *Biogenic Sulfur in the Environment,* American Chemical Society, Washington, D.C.

Seinfeld, J.H. (1995), Chemistry of ozone in the urban and regional atmosphere, in *Progress and Problems in Atmospheric Chemistry,* J.R. Barker, ed. World Scientific Publishers, Singapore, pp. 34–57.

Seinfeld, J.H. (2002), Aerosol formation from atmospheric organics, presented at DOE Atmospheric Sciences Program Annual Meeting, Albuquerque, NM, March 19–21. http://www.atmos.anl.gov/ACP/2002presentations/Seinfeld02.pdf

Seinfeld, J.H. and S.N. Pandis (1998), *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change,* Wiley & Sons, New York.

Singh, H.B., D. Herlth, D. O'Hara, K. Zahnle, J.D. Bradshaw, S.T. Sandholm, R. Talbot, G.L. Gregory, G.W. Sachse, D.R. Blake, and S.C. Wofsy (1994), Summertime distribution of PAN and other reactive nitrogen species in the northern high-latitude atmosphere of eastern Canada, *J. Geophys. Res.,* **99D,** 1821–1836.

Solomon, S. (1988), The mystery of the Antarctic ozone "hole," *Rev. Geophys.,* **26,** 131–148.

Talukdar, R.S., J.B. Burkholder, A.M. Schmoltner, J.M. Roberts, R.R. Wilson, and A.R. Ravishankara (1995), Investigation of the loss processes for peroxyacetyl nitrate in the atmosphere: UV photolysis and reaction with OH, *J. Geophys. Res.,* **100,** 14163–14173.

Wayne, R.P. (1985), *Chemistry of Atmospheres,* Clarendon Press, Oxford.

Webster, C.R., R.D. May, D.W. Toohey, L.M. Avallone, J.G. Anderson, P. Newman, L. Lait, M. Schoeberl, J.W. Elkins, and K.R. Chan (1993), Chlorine chemistry on polar stratospheric cloud particles in the Arctic winter, *Science,* **261,** 1130–1134.

LARRY G. ANDERSON
Joint Graduate School of Energy and Environment at
King Mongkut's University of
Technology Thonbury—Bangkok
While on leave from
University of Colorado at Denver

B

BIOLOGICAL TREATMENT OF WASTEWATER

1. INTRODUCTION

Biological treatment is the most widely used method for removal, as well as partial or complete stabilization of biologically degradable substances present in waste-waters. Most often, the degradable substances are organic in nature and may be present as suspended, colloidal or dissolved matter. The fraction of each form depends on the nature of wastewater. In the operation of biological treatment facilities, the characteristics of wastewater are measured in terms of its chemical oxygen demand, COD, biochemical oxygen demand, BOD, total organic carbon, TOC, and volatile suspended solids, VSS; concepts of which have been discussed elsewhere.[1]

Most of the conventional biological wastewater treatment processes are based on naturally occurring biological phenomena, but are carried out at accelerated rates. These processes employ bacteria as the primary organisms; however, certain other microorganisms may also play an important role. Gates and Ghosh[2] have presented the biological component system existing in the BOD process and it is shown in Figure 1. The degradation and stabilization of organic matter is accomplished by their use as food by bacteria and other microorganisms to produce protoplasm for new cells during the growth process. When a small number of microorganisms are inoculated into a bacteriological culture medium, growth of bacteria with time follows a definite pattern as depicted in Figure 2 by plotting viable count and mass of bacteria against time.[3] The population dynamics of bacteria in biological treatment processes depends upon various environmental factors including pH, temperature, type and concentration of substrate, hydrogen acceptor, availability and concentration of essential nutrients like nitrogen, phosphorous, sulfur, etc., and essential minerals, osmotic pressure, toxicity of media or by-products, and degree of mixing.[4] In recent years, cultures have been developed for biological treatment of many hard-to-degrade organic wastes.

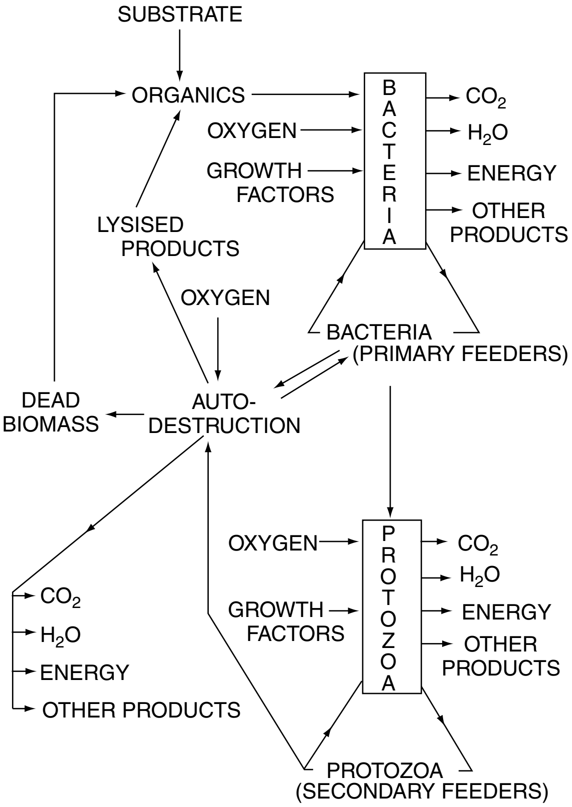


FIGURE 1 Biological component system existing in BOD process.

2. METABOLIC REACTIONS

The metabolic reactions occurring within a biological treatment reactor can be divided into three phases: oxidation, synthesis and endogenous respiration. Oxidation–reduction may proceed either in the presence of free oxygen, aerobically, or in its absence, anaerobically. While the overall reactions carried out may be quite different under aerobic and anaerobic conditions, the processes of microbial growth and energy utilization are similar. Typical reactions in these three phases are formulated below:

- Organic Matter Oxidation (Respiration)
  $C_xH_yO_z + O_2 \rightarrow CO_2 + H_2O + \text{Energy}$
- Inorganic Matter Oxidation (Respiration)
  $NH_4^+ 2O_2 \rightarrow NO_3^- + H_2O + 2H^+ + \text{Energy}$
- Protoplasm (Cell Material) Synthesis
  $C_xH_yO_z + NH_3 + O_2 + \text{Energy} \rightarrow C_5H_7NO_2 + H_2$
  $C_xH_yO_z + H^+ + NO_3^- + \text{Energy} \rightarrow C_5H_7NO_2 + CO_2 + H_2O$
- Protoplasm (Cell Material) Oxidation
  $C_5H_7NO_2 + 5O_2 \rightarrow 5CO_2 + 2H_2O + NH_3 + \text{Energy}$

Therefore, bacterial respiration in living protoplasm is a biochemical process whereby energy is made available for endothermic life processes. Being dissimilative in nature, respiration is an important process in wastewater treatment practices. On the other hand, endogenous respiration is the internal process in microorganisms that results in auto-digestion or self-destruction of cellular material.[3] Actually, bacteria require a small amount of energy to maintain normal functions such as motion and enzyme activation and this basal-energy requirement of the bacteria has been designated as endogenous respiration. Even when nutrients are available, endogenous metabolism proceeds with the breakdown of protoplasm.[5] According to Bertalanffy's hypothesis,[6] the microbial growth is the result of competition between two opposing processes: Aufban—assimilation, and Abban—endogenous metabolism. The rate of assimilation is proportional to the mass of protoplasm in the cell and the surface area of the cell, whereas the endogenous metabolism is dependent primarily on environmental conditions.

In the presence of enzymes produced by the living microorganisms, about 1/3 of the organic matter removed is oxidized into carbon dioxide and water in order to provide energy for synthesis of the remaining 2/3 of the organic matter into the cell material. Metabolism and process reactions occurring in typical biological wastewater treatment processes are explained schematically by Stewart[7] as shown in Figure 3. Thus, the basic equations for biological metabolisms are:

Organic matter metabolized
= Protoplasm synthesized + Energy for synthesis

and

Net protoplasm accumulation
= Protoplasm synthesized − Endogenous respiration.

"Growth Kinetics"

Irvine and Schaezler[8] have developed the following expression for non-rate limited growth of microorganisms in logarithmic phase:

$$\frac{dN}{dt} = k_0 N \qquad (1)$$

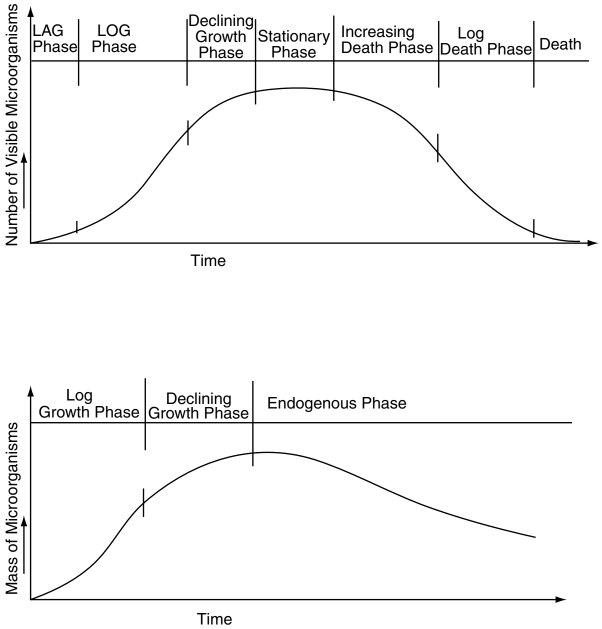
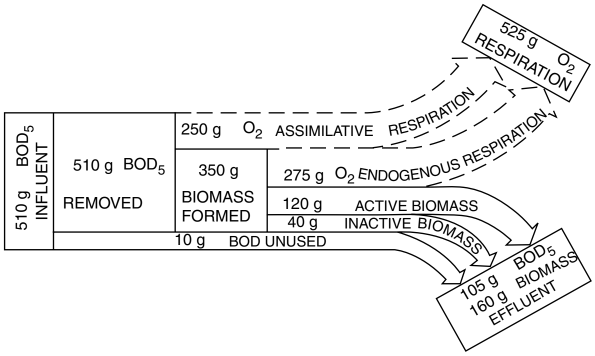


FIGURE 2 Growth pattern of microorganisms.



SYSTEM METABOLISM FOR SOLUBLE WASTES

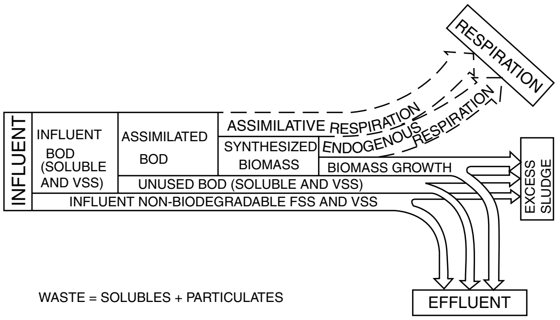


FIGURE 3 Metabolism and process reactions.

or

$$N_t = N_o e^{k_o t}$$

where:

N_0 = Number of viable microorganisms per unit volume at time $t = 0$
$N_t = N$ = Number of viable microorganisms per unit volume at time t

and

k = Logarithmic growth rate constant, time^{-1}.

In wastewater treatment practices, the growth pattern based on mass of microorganisms has received more attention than the number of viable microorganisms. If each microorganism is assumed to have an average constant mass, then N in Eq. 1 can be replaced with X, the mass of active microorganisms present per unit volume to obtain:

$$\frac{dX}{dt} = k_0 X. \tag{2}$$

The growth of bacterial population may become limited either due to exhaustion of available nutrients or by the accumulation of toxic substances. The growth rate of bacteria starts slowing down, and Eq. 1 changes to the form:

$$\frac{dN}{dt} = k_t N \tag{3}$$

where growth rate factor k_t, varies with time and becomes a function of temperature, T, pH, substrate concentration, S, and concentration of various nutrients, C_{n1}, C_{n2}, etc., i.e.:

$$k_t = V_1\,(T, \text{pH}, C_s, C_{n1}, C_{n2}, \ldots).$$

Figure 4 shows variation in growth rate k_t with change in nutrient concentrations, assuming that T and pH are held constant and substrate concentration, S, is greater than the critical substrate concentration, S^*, above which k_t, is independent of S. Several interesting observations are made from these curves.[8] First, the maximum value of k_t is essentially constant. Second, the shape of the curve and the limiting concentration is different for each nutrient. Third, k_t is shown to be zero when any of the nutrients is missing. Fourth, as the biological reaction proceeds, all nutrients are consumed. Thus, even if all nutrients are initially in excess, the growth may eventually become limited. Finally, as the concentration drops to zero, a stationary phase is reached, i.e., dN/dt becomes zero.

In case of a substrate limited system, rate of growth is given by:

$$\frac{dN}{dt} = \mu N \tag{4}$$

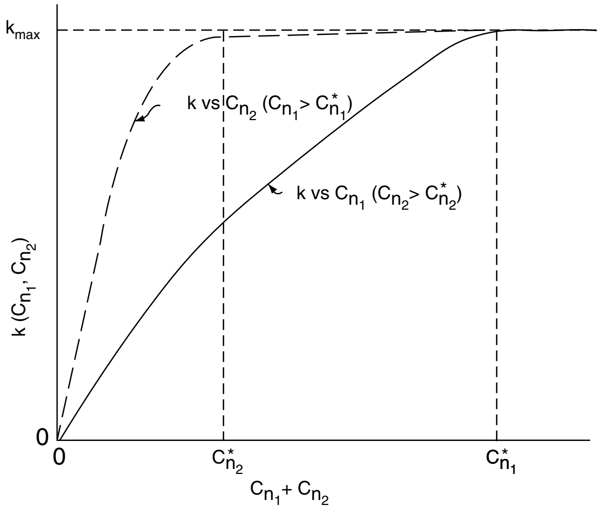


FIGURE 4 k vs nutrient concentration.

or

$$\frac{dX}{dt} = \mu X.$$

The following simple relationship between specific growth rate of microorganisms, μ, and substrate concentration, S, was developed by Monod[9] and has been widely accepted:

$$\mu = \frac{dN}{N dt} = \frac{dX}{X dt} = \mu_{\max} \frac{S}{K + S} \tag{5}$$

where K is a constant called half velocity coefficient and $\mu_{\max}$ is maximum specific growth rate.

It is postulated that the same amount of substrate is incorporated in each cell formed. Therefore, the rate of increase in number or mass of microorganisms in logarithmic growth phase, dN/dt, or dX/dt, is proportional to the rate of substrate consumption, dS/dt, or dL/dt, if the substrate concentration is measured in terms of its BOD, L, and the following relationship can be stated:

$$\frac{dX}{dt} = Y \frac{dS}{dt} \tag{6}$$

or

$$\Delta X = Y \Delta S$$

where Y is called the growth yield coefficient, ΔX is the cell mass synthesized in a given time, and ΔS is substrate removed in the same time. The substrate utilization rate, q, per unit biomass has been defined as:

$$q = \frac{dS}{X dt} \tag{7}$$

Combining Eqs. 4, 6 and 7 yields:

$$q = \frac{\mu}{Y} \tag{8}$$

and

$$q = q_{\max} \frac{S}{K + S}. \tag{9}$$

Under conditions of rate limited growth, i.e., nutrient exhaustion or auto-oxidation, Eq. 6 becomes:

$$\frac{dX}{dt} = Y \frac{dS}{dt} - bX \tag{10}$$

where b is the auto-oxidation rate or the microbial decay rate. In absence of substrate, this equation is reduced to:

$$\frac{dX}{dt} = -bX. \tag{11}$$

Several kinetic equations have been suggested for analysis and design of biological wastewater treatment systems and the following have been applied frequently:[10–13]

$$\frac{dS}{dt} = \frac{q_{\max} SX}{(K + S)} \tag{12}$$

$$\frac{dS}{dt} = qSX \tag{13}$$

$$\frac{dS}{dt} = qX \frac{S^2}{S_0} \tag{14}$$

where S_0 is the initial substrate concentration. Combining Eqs. 10 and 12 gives the net specific growth rate:

$$\mu = \frac{dX}{X dt} = \frac{q_{\max} YS}{K + S} - b \tag{15}$$

A similar kinetic relationship can be obtained by combining Eq. 10 with Eqs. 13 and 14.

Effect of Temperature

One of the significant parameters influencing biological reaction rates is the temperature. In most of the biological treatment processes, temperature affects more than one reaction rate and the overall influence of temperature on the process becomes important. The applicable equation for the effect of temperature on rate construct is given by:

$$k_T = k_{20} \theta^{T-20} \tag{16}$$

where θ is the temperature coefficient. This equation shows that reaction rates increase with increase in temperature.

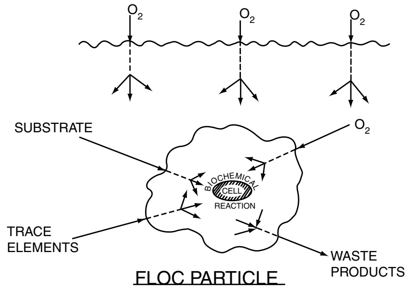
Methods of BOD Removal

In wastewater treatment processes, the microorganisms are not present as isolated cells, but are a collection of microorganisms such as bacteria, yeast, molds, protozoa, rotifers, worms and insect larvae in a gelatinous mass.[13] These microorganisms tend to collect in a biological floc, called biomass, which is expected to possess good settling characteristics. The biological oxidation or stabilization of organic matter by the microorganisms present in the floc is assumed to proceed in the following sequence:[13,14]

(a) An initial high rate of BOD removal from wastewater on coming in contact with active biomass by adsorption and absorption. The extent of this removal depends upon the loading rate, the type of waste, and the ecological condition of the biomass.
(b) Utilization of decomposable organic matter in direct proportion to biological cell growth. Substances concentrating on the surface of biomass are decomposed by the enzymes of living cells, new cells are synthesized and end products of decomposition are washed into the water or escape to the atmosphere.
(c) Oxidation of biological cell material through endogenous respiration whenever the food supply becomes limited.
(d) Conversion of the biomass into settleable or otherwise removable solids.

The rates of reactions in the above mechanisms depend upon the transport rates of substrate, nutrients, and oxygen in case of aerobic treatment, first into the liquid and then into the biological cells, as shown in Figure 5.[15] Any one or more of these rates of transport can become the controlling factors in obtaining the maximum efficiency for the process. However, most often the interfacial transfer or adsorption is the rate determining step.[14]

In wastewater treatment, the biochemical oxygen demand is exerted in two phases: carbonaceous oxygen demand to oxidize organic matter and nitrogenous oxygen demand to oxidize ammonia and nitrites into nitrates. The nitrogenous oxygen demand starts when most of the carbonaceous oxygen demand has been satisfied.[15] The typical progression of carbonaceous BOD removal by biomass with time, during biological purification in a batch operation, was first shown by Ruchhoft[16] as reproduced in Figure 6. The corresponding metabolic reactions in terms of microorganisms to food ratio, M/F, are shown in Figure 7. This figure shows that the food to microorganisms ratio maintained in a biological reactor is of considerable importance in the operation of the process. At a low M/F ratio, microorganisms are in the log-growth phase, characterized by excess food and maximum rate of metabolism. However, under these conditions, the settling characteristic of biomass is poor because of their dispersed



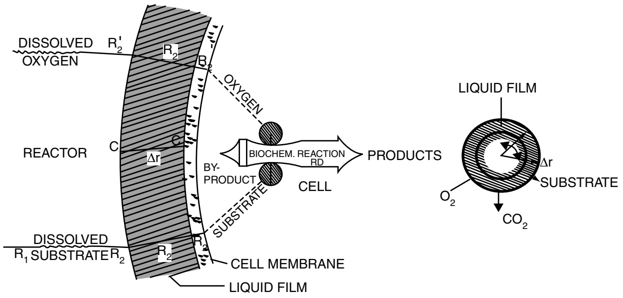


FIGURE 5 Mass transfer in biofloc.

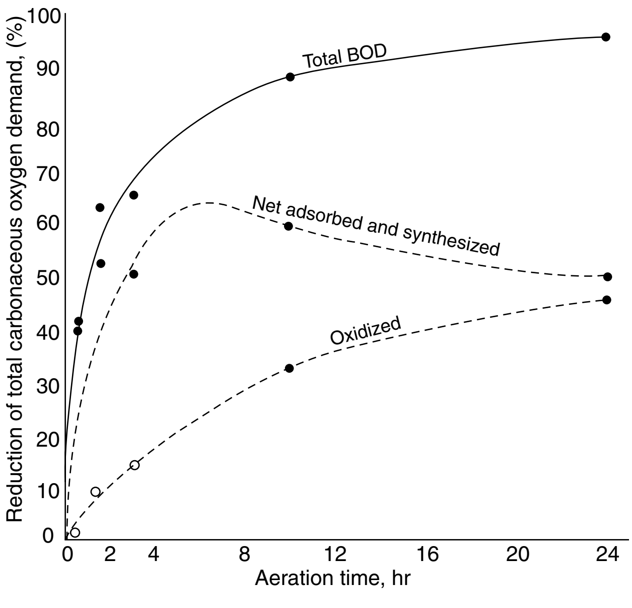


FIGURE 6 Removal of organic inbalance by biomass in a batch operation.

growth; also, the BOD removal efficiency is poor as the excess unused organic matter in solution escapes with the effluent. On the other hand, high M/F ratio means the operation is in the endogenous phase. Competition for a small amount of food available to a large mass of microorganisms results in starvation conditions within a short duration and continued aertion under these conditions results in auto-oxidation of biomass. Although the rate of metabolism is relatively low at high M/F ratio, settling characteristics of biomass are good and BOD removal efficiency is high.

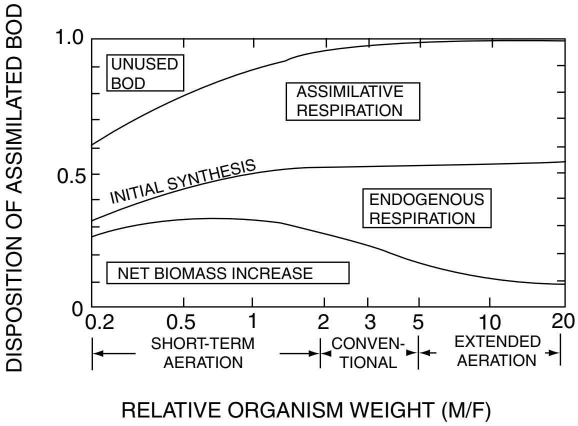


FIGURE 7 Metabolic reactions for the complete spectrum.

Goodman and Englande[17] have suggested that the total mass concentration of solids, X_T, in a biological reactor is composed of an inert fraction, X_i, and a volatile fraction, X_v, which can be further broken down into an active fraction, X, and non-biodegradable residue fraction, X_n, resulting from endogenous respiration, i.e.:

$$X_T = X_i + X_v = X_i + X + X_n. \qquad (17)$$

The total mass concentration of solids in wastewater treatment is called suspended solids, whereas its volatile fraction is called volatile suspended solids, X. In a biological reactor, volatile suspended solids, X, is assumed to represent the mass of active microorganisms present per unit volume.

3. TOXICITY

Toxicity has been defined as the property of reaction of a substance, or a combination of substances reacting with each other, to deter or inhibit the metabolic process of cells without completely altering or destroying a particular species, under a given set of physical and biological environmental conditions for a specified concentration and time of exposure.[18] Thus, the toxicity is a function of the nature of the substance, its concentration, time of exposure and environmental conditions.

Many substances exert a toxic effect on biological oxidation processes and partial or complete inhibition may occur depending on their nature and concentration. Inhibition may result from interference with the osmotic balance or with the enzyme system. In some cases, the microorganisms become more tolerant and are considered to have acclimatized or adapted to an inhibitory concentration level of a toxic substance. This adaptive response or acclimation may result from a neutralization of the toxic material produced by the biological activity of the microorganisms or a selective growth of the culture unaffected by the toxic substance. In some cases, such as cyanide and phenol, the toxic substances may be used as substrate. Rates of acclimation to lethal factors vary greatly. Thus, the toxicity to microorganisms may result due to excess concentrations of substrate itself, the presence of inhibiting substances or factors in the environment and/or the production of toxic by-products.[19–23]

The influence of a toxicant on microorganisms depends not only on its concentration in water, but also on its rate of absorption, its distribution, binding or localization in the cell, inactivation through biotransformation and ultimate excretion. The biotransformations may be synthetic or nonsynthetic. The nonsynthetic transformations involve oxidation, reduction or hydrolysis. The synthetic transformation involve the coupling of a toxicant or its metabolite with a carbohydrate, an amino acid, or a derivative of one of these. According to Warren[19], the additive interaction of two toxic substances of equal toxicity, mixed in different proportions, may show combined toxicity as shown in Figure 8. The combined effects may be supra-additive, infra-additive, no interaction or antagonism. The relative toxicity of the mixture is measured as the reciprocal of median tolerance limit.

Many wastewater constituents are toxic to microorganisms. A fundamental axiom of toxicity states that all compounds are toxic if given to a text organism at a sufficiently high dose. By definition, the compounds that exert a deleterious influence on the living microorganisms in a biological treatment unit are said to be toxic to those microorganisms. At high concentrations, these substances kill the microbes whereas at sublethal concentrations, the activity of microbes is reduced. The toxic substances may be present in the influent stream or may be produced due to antagonistic interactions.

Biological treatment is fast becoming a preferred option for treating toxic organic and inorganic wastes in any form;

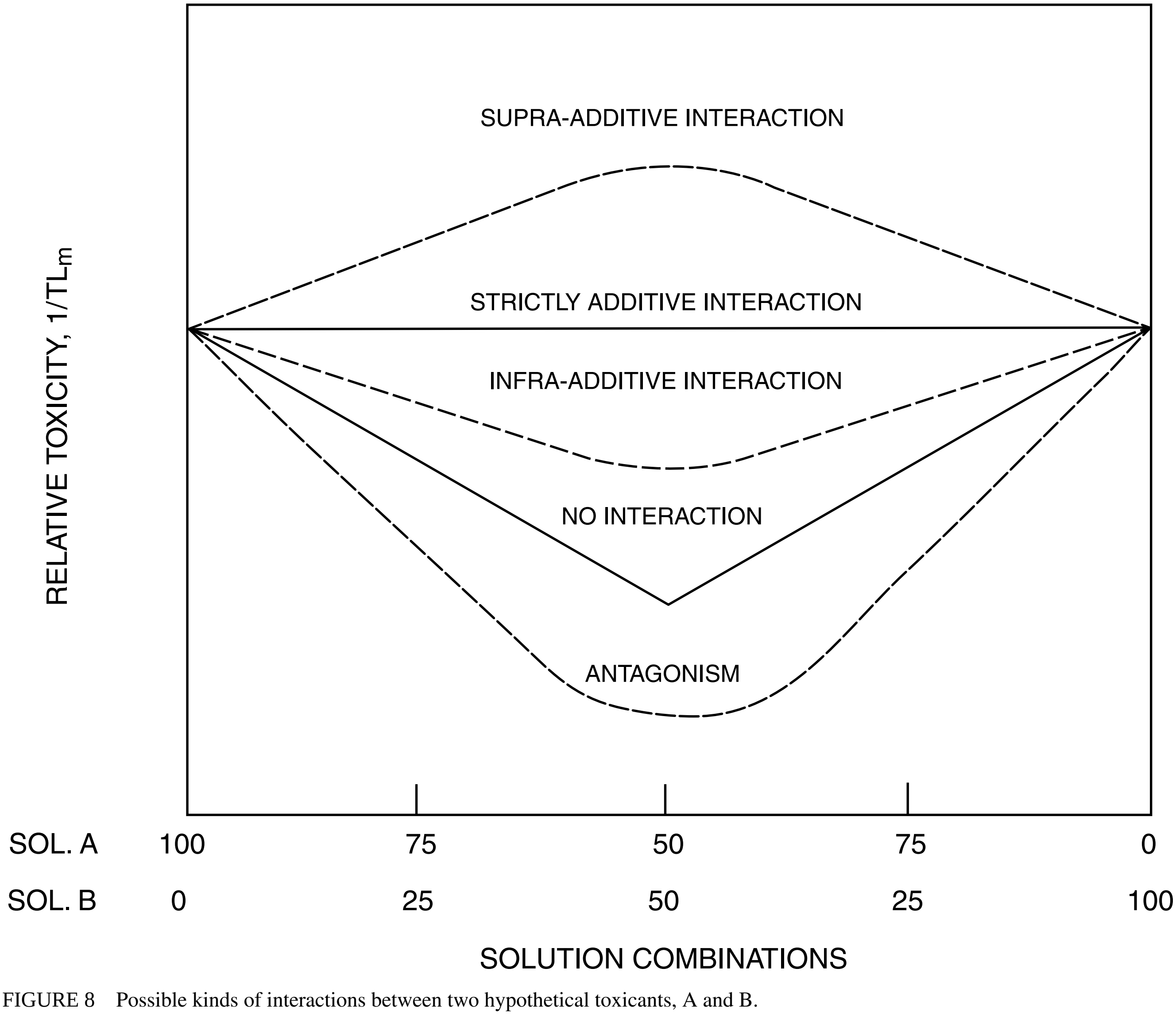


FIGURE 8 Possible kinds of interactions between two hypothetical toxicants, A and B.

solid, liquid or gaseous. The application of biological processes in degradation of toxic organic substances is becoming popular because (i) these have an economical advantage over other treatment methods; (ii) toxic substances have started appearing even in municipal wastewater treatment plants normally designed for treating nontoxic substrates; and (iii) biological treatment systems have shown a resiliency and diversity which makes them capable of degrading many of the toxic organic compounds produced by the industries.[24] Grady believes that most biological treatment systems are remarkably robust and have a large capacity for degrading toxic and hazardous materials.[25] The bacteria and fungi have been used primarily in treating petroleum-derived wastes, solvents, wood preserving chemicals and coal tar wastes. The capability of any biological treatment system is strongly influenced by its physical configuration.

As mentioned previously, the Michelis–Menten or Monond equation, Eq. 5, has been used successfully to model the substrate degradation and microbial growth in biological wastewater treatment process. However, in the presence of a toxic substance, which may act as an inhibitor to the normal biological activity, this equation has to be modified. The Haldane equation is generally accepted to be quite valid to describe inhibitory substrate reactions during the nitrification processes, anaerobic digestion, and treatment of phenolic wastewaters.[24,26,27]

$$\text{Haldane Equation} \quad \mu = \frac{\mu_{\max} S}{S + K + S^2/K_i} \qquad (18)$$

where K_i is the inhibition constant.

In the above equation, a smaller value for K_i indicates a greater inhibition. The difference between the two kinetic equations, Monod and Haldane, is shown in Figure 9, in which the specific growth rate, μ, is plotted for various substrate concentrations, *S*. The values for $\mu_{\max}$, K_s and K_i are assumed to be 0.5 h^{-1}, 50 mg/L and 100 mg/L, respectively.

Behavior of Biological Processes

The behavior of a biological treatment process, when subjected to a toxic substance, can be evaluated in three parts:

1. Is the pollutant concentration inhibitory or toxic to the process? How does it affect the biodegradation rate of other pollutants?
2. Is the pollutant concentration in process effluent reduced to acceptable level? Is there a production of toxic by-products?
3. Is there an accumulation of toxic substances in the sludge?

The above information should be collected on biological systems that have been acclimated to the concerned toxic substances. Pitter[28] and Adam *et al.*[29] have described the acclimation procedures.

Generally, biological processes are most cost-effective methods to treat wastes containing organic contaminants. However, if toxic substances are present in influents, certain pretreatment may be used to lower the levels of these contaminants to threshold concentrations tolerated by acclimated microorganisms present in these processes. Equalization of toxic load is an important way to maintain a uniform influent and reduce the shock load to the process. Also, various physical/chemical methods are available to dilute, neutralize and detoxicate these chemicals.

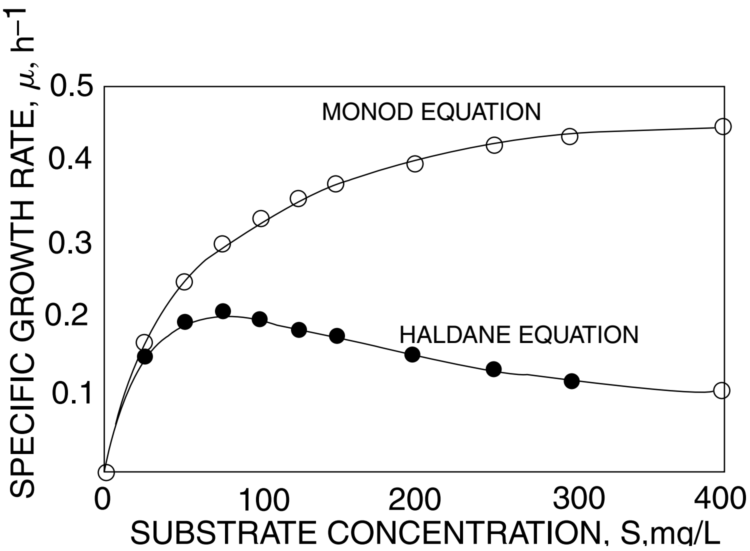


FIGURE 9 Change of specific growth rate with substrate concentration (inhibited and uninhibited).

Genetically Engineered Microorganisms

One of the promising approaches in biodegradation of toxic organics is the development of genetically engineered microorganisms. Knowledge of the physiology and biochemistry of microorganisms and development of appropriate process engineering are required for a successful system to become a reality. The areas of future research that can benefit from this system include stabilization of plasmids, enhanced activities, increased spectrum of activities and development of environmentally safe microbial systems.[30]

4. TYPES OF REACTORS

Three types of reactors have been idealized for use in biological wastewater treatment processes:

(a) Batch Reactors in which all reactants are added at one time and composition changes with time;
(b) Plug Flow or Non-Mix Flow Reactors in which no element of flowing fluid overtakes another element; and
(c) Completely Mixed or Back-Mix Reactors in which the contents are well stirred and are uniform in composition throughout.

Most of the flow reactors in the biological treatment are not ideal, but with negligible error, some of these can be considered ideal plug flow or back-mix flow. Others have considerable deviations due to channeling of fluid through the vessel, by the recycling of fluid through the vessel or by the existence of stagnant regions of pockets of fluid.[31] The non-ideal flow conditions can be studied by tagging and following each and every molecule as it passes through the vessel, but it is almost impossible. Instead, it is possible to measure the distribution of ages of molecules in the exit stream.

The mean retention time, $\bar{t}$ for a reactor of volume V and having a volumetric feed rate of Q is given by $\bar{t} = V/Q$. In non-ideal reactors, every molecule entering the tank has a different retention time scattered around $\bar{t}$. Since all biological reactions are time dependent, knowledge on age distribution of all the molecules becomes important. The distribution of ages of molecules in the exit streams of both ideal and non-ideal reactors in which a tracer is added instantaneously in the inlet stream is shown in Figure 10. The spread of concentration curve around the plug flow conditions depends upon the vessel or reactor dispersion number, *Deul,* where D is longitudinal or axial dispersion coefficient, u is the mean displacement velocity along the tank length and l is the length dimension.[32] In the case of plug flow, the dispersion number is zero, whereas it becomes infinity for completely mixed tanks.

Treatment Models

Lawrence and McCarty[11] have proposed and analyzed the following three models for existing continuous flow aerobic or anaerobic biological wastewater treatment configurations:

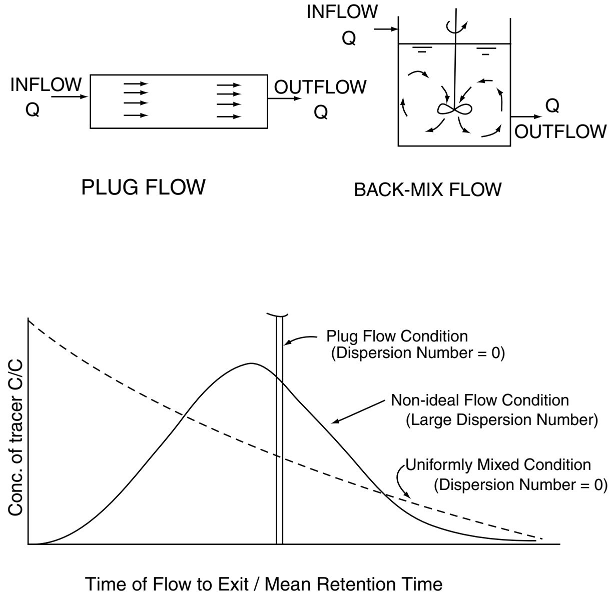


FIGURE 10 Hydraulic characteristics of basins.

(a) a completely mixed reactor without biological solids recycle,
(b) a completely mixed reactor with biological solids recycle, and
(c) a plug flow reactor with biological solids recycle.

These configurations are shown schematically in Figure 11. In all these treatment models, the following equations can be applied in order to evaluate kinetic constants,[33] where Δ indicates the mass or quantity of material:

- *Solid Balance Equation*

$$\begin{bmatrix}\Delta\text{Cells}\\ \text{Reactor}\end{bmatrix} = \begin{bmatrix}\Delta\text{Cells}\\ \text{Growth}\end{bmatrix} - \begin{bmatrix}\Delta\text{Cells}\\ \text{Decay}\end{bmatrix} - \begin{bmatrix}\Delta\text{Cells}\\ \text{Effluent Loss}\end{bmatrix} \tag{19}$$

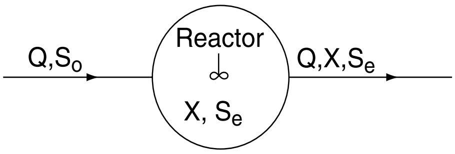
- *Substrate Balance Equation*

$$\begin{aligned}\begin{bmatrix}\Delta\text{Substrate}\\ \text{Reactor}\end{bmatrix} &= \begin{bmatrix}\Delta\text{Substrate}\\ \text{Influent}\end{bmatrix} - \begin{bmatrix}\Delta\text{Substrate}\\ \text{Growth}\end{bmatrix}\\ &\quad - \begin{bmatrix}\Delta\text{Substrate}\\ \text{Effluent Loss}\end{bmatrix}\end{aligned} \tag{20}$$

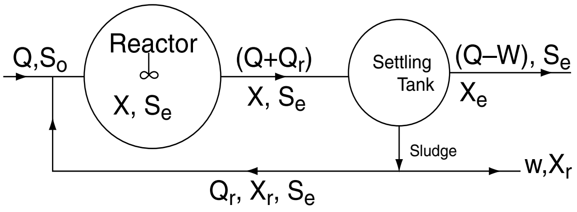
Parameters for Design and Operation

Various parameters have been developed and used in the design and operation of biological wastewater treatment processes and the most significant parameters are:

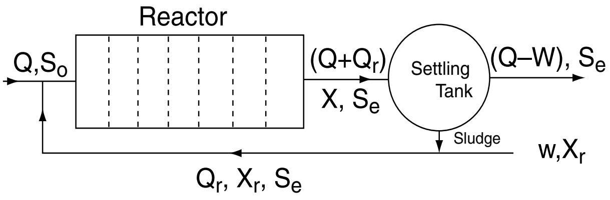
θ_x– Biological Solids Retention Time, or Sludge Age, or Mean Cell Retention Time, is defined



I- Completely Mixed-No biological solids recycle



II- Completely Mixed-Biological solids recycle



III- Plug Flow-Biological solids recycle

FIGURE 11 Treatment models.

as the ratio between total active microbial mass in treatment system, X_T, and total quantity of active microbial mass withdrawn daily, including solids wasted purposely as well as those lost in the effluent, $\Delta X_T/\Delta t$. Regardless of the fraction of active mass, in a well-mixed system the proportion of active mass wasted is equal to the proportion of total sludge wasted, making sludge age equal for both total mass and active mass.

U– Process Loading Factor, or Substrate Removal Velocity, or Food to Microorganisms Ratio, or Specific Utilization, is defined as the ratio between the mass of substrate utilized over a period of one day, $\Delta S/\Delta t$, and the mass of active microorganisms in the reactor, X_T.

$\bar{t}$– Hydraulic Retention Time or Detention Time, or Mean Holding Time, is defined as the ratio between the volume of Reactor, V, and the volumetric feed rate, Q.

B_V– Volumetric Loading Rate or Hydraulic Loading Rate is defined as the ratio between the mass of substrate applied over a period of one day, $S_T/\Delta t$ and the volume of the reactor, V.

E– Process Treatment Efficiency or Process Performance is defined as percentage ratio between the substrate removed, $(S_0 - S_e)$, and influent substrate concentration, S_0.

A desired treatment efficiency can be obtained by control of one or more of these parameters separately or in combination.

5. BIOLOGICAL TREATMENT SYSTEMS

The existing biological treatment systems can be divided into the following three groups:

(a) Aerobic Stationary-Contact or Fixed-Film Systems: Irrigation beds, irrigation sand filters, rotating biological contactors, fluidized bed reactors, and trickling filters fall in this group. In these treatment processes, the biomass remains stationary in contact with the solid supporting-media like sand, rocks or plastic and the wastewater flows around it.

(b) Aerobic Suspended-Contact Systems: Activated sludge process and its various modifications, aerobic lagoons and aerobic digestion of sludges are included in this group. In these treatment processes, both the biomass and the substrate are in suspension or in motion.

(c) Anaerobic Stationary-Contact and Suspended Contact Systems: Anaerobic digestion of sludges and anaerobic decomposition of wastewater in anaerobic lagoons fall in this category.

A typical layout of a wastewater treatment plant incorporating biological treatment is shown in Figure 12. Primary sedimentation separates settleable solids and the aerobic biological treatment is designed to remove the soluble BOD. The solids collected in primary sedimentation tanks and the excess sludge produced in secondary treatment are mixed together and may be digested anaerobically in digesters. Trickling filter and activated sludge processes are most common secondary treatment processes for aerobic treatment and are discussed in detail. Discussion of sludge digestion by anaerobic process and use of biological nutrient removal as a tertiary treatment have also been included.

In addition to conventional pollutants present in municipal and industrial wastewaters, significant concentrations of toxic substances such as synthetic organics, metals, acids, bases, etc., may be present due to direct discharges into the sewers, accidental spills, infiltration and formation during chlorination of wastewaters. It is import to have a knowledge of both the scope of applying biological treatment and the relevant engineering systems required to achieve this capability. Thus, the kinetic description of the process and the deriving reactor engineering equations and strategies for treatment of conventional and toxic pollutants are essential for proper design and operation of biological waste treatment systems.[24]

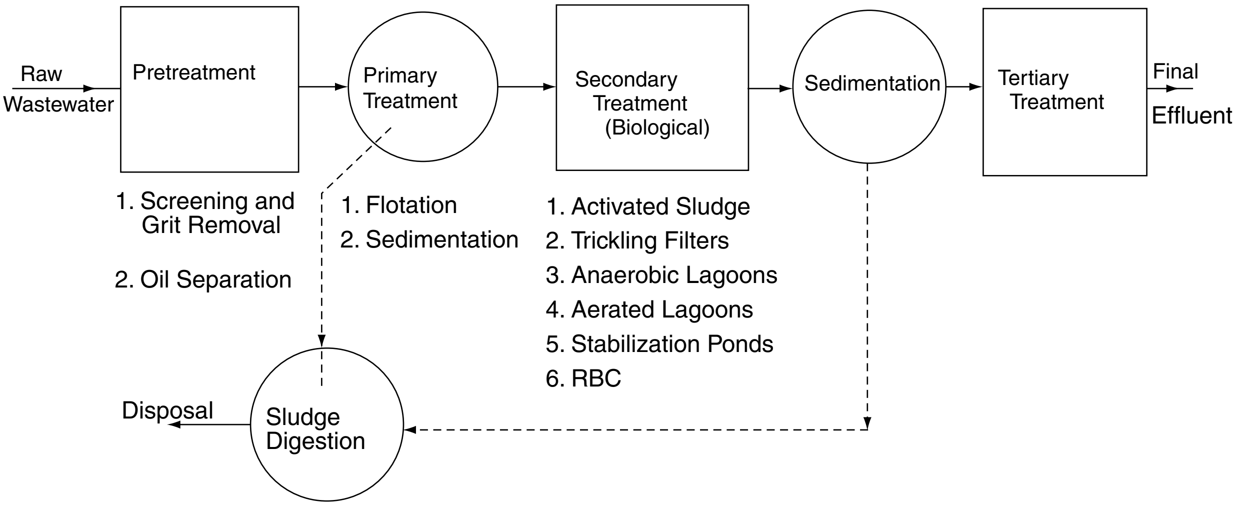


FIGURE 12 Typical wastewater treatment sequence.

The available information strongly indicates that immobilized biological systems are less sensitive to toxicity and have a higher efficiency in degrading toxic and hazardous materials.[34] Fixed-film wastewater treatment processes are regarded to be more stable than suspended growth processes because of the higher biomass concentration and greater mass transfer resistance from bulk solution into the biofilm in fixed-films.[35] The mass transfer limitation effectively shields the microorganisms from higher concentrations of toxins or inhibitors during short-term shock loads because the concentrations in biofilms change more slowly than in the bulk solution. Also, since the microorganisms are physically retained in the reactor, washout is prevented if the growth rate of microorganisms is reduced.[34,35] The biofilm systems are especially well suited for the treatment of slowly biodegradable compounds due to their high biomass concentration and their ability to immobilize compounds by adsorption for subsequent biodegradation and detoxification.[34]

Trickling Filters

Wastewater is applied intermittently or continuously to a fixed bed of stones or other natural synthetic media resulting in a growth of microbial slime or biomass on the surface of this media. Wastewater is sprayed or otherwise distributed so that it slowly trickles through while in contact with the air. For maximum efficiency, food should be supplied continuously by recirculating, if necessary, the treated wastewater or settled sludge or both. Oxygen is provided by the dissolved oxygen in influent wastewater, recirculated water from the air circulating through the interstices between the media to maintain aerobic conditions.

Active microbial film, biomass, consisting primarily of bacteria, protozoa, and fungi, coats the surface of filter media. The activity in biological film is aerobic, with movement of oxygen, food and end-products in and out of it as shown in Figure 13. However, as the thickness of the film increases, the zone next to the filter medium becomes anaerobic. Increased anaerobic activity near the surface may liquify the film in contact with the medium, resulting in sloughing or falling down of the old film and growth of a new film. The sloughed solids are separated in a secondary settling tank and a part of these may be recirculated in the system. Two types of trickling filters are recognized, primarily on the basis of their loading rates and method of operation, as shown in Table 1. In low-rate trickling filter, the wastewater passes through only once and the effluent is then settled prior to disposal. In high-rate trickling filter, wastewater applied

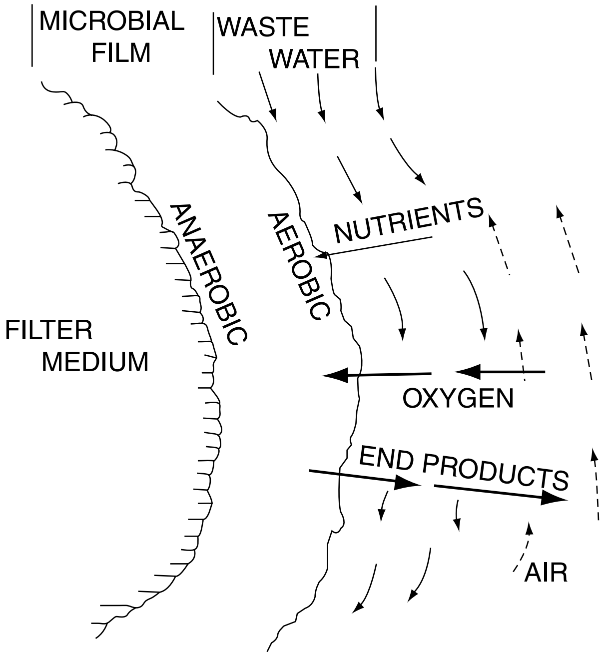


FIGURE 13 Process of BOD removal in trickling filters.

TABLE 1

Comparison of low-rate and high-rate filters

Parameters	Low-Rate Filters	High-Rate Filters
Hydraulic Loading		
US gallons per day per square foot	25 to 100	200 to 1000
Million US gallons per day per acre	1.1 to 4.4	8.7 to 44
Cubic metre per day per square metre	1.0 to 4.1	8.1 to 40.7
Organic Loading (BOD)		
Pounds of BOD per day per 1000 cubic feet	5 to 25	25 to 300
Pounds of BOD per day per acre-foot	220 to 1100	1100 to 13000
g of BOD per day per cubic metre	80 to 400	400 to 4800
Recirculation	Generally absent	Always provided R = 0.5 to 3
Effluent Quality	High nitrified, lower BOD	Not fully nitrified, higher BOD

to filters is diluted with recirculated flow of treated effluent, settled effluent, settled sludge, or their mixture, so that it is passed through the filter more than once. Several recirculation patterns used in high-rate filter systems are shown in ASCE Manual.[36] Sometimes two filter beds are placed in series and these are called Two-Stage Filters.

The advantages and disadvantages of recirculation are listed below:

Advantages of Recirculation

(a) Part of organic matter in influent wastewater is brought into contact with growth on filter media more than once.

(b) Recirculated liquid contains active microorganisms not found in sufficient quantity in raw wastewater, thus providing seed continually. This continuous seeding with active microorganisms and enzymes stimulates the hydrolysis and oxidation and increases the rate of biochemical stabilization.

(c) Diurnal organic load is distributed more uniformly. Thus, when plant flow is low, operation is not shut off. Also, stale wastewater is freshened.

(d) Increased flow improves uniformity of distribution, increases sloughing and reduces clogging tendencies.

(e) Higher velocities and continual scouring make conditions less favourable for growth of filter flies.

(f) Provides for more flexibility of operation.

Disadvantages

(a) There is increased operating cost because of pumping. Larger settling tanks in some designs may increase capital cost.

(b) Temperature is reduced as a result of number of passes of liquid. In cold weather, this results in decreased biochemical activity.

(c) Amount of sludge solids to digesters may be increased.

The *ACE Manual*[36] lists the following factors affecting the design and operation of filters:

(a) composition and characteristics of the wastewater after pretreatment,
(b) hydraulic loading applied to the filter,
(c) organic loading applied to the filter,
(d) recirculation, system, ratio and arrangement,
(e) filter beds, their volume, depth and air ventilation,
(f) size and characteristics of media, and
(g) temperature of wastewater.

Assuming that the flow through the packed column could be approximated as plug flow, and if BOD removal rate occurs by first order reaction, Eq. 13, then the formula to use in trickling filters will become:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = qSX = k_f S$$

or

$$\frac{S_e}{S_0} = \mathrm{e}^{-k_f t}. \tag{21}$$

Another equation suggested for application in trickling filters[13] is:

$$\frac{S_e}{S_0} = \frac{1}{1+qXt} = \frac{1}{1+k_f}t \tag{22}$$

where trickling filter rate coefficient, k_f, is a function of active film mass per unit volume and remains constant for a given specific area and uniform slime layer. Contact time, t, is related to filter depth, H, volumetric rate of flow per unit area, Q_a, and specific surface area of filter media, A_v. Sinkoff, Porges, and McDermott[37] have proposed the following relationship based on their experiments:

$$t = c_1 H \left[\frac{A_v}{Q_a} \right]^n \tag{23}$$

c_1 is assumed to be a constant and exponent n ranges between 0.53 and 0.83 depending upon the type of filter medium and the hydraulic characteristics of the system. Substitution of this value of t in Eq. 21 gives:

$$\frac{S_e}{S_0} = \exp - k_f \left[\frac{A_v}{Q_a} \right]^n Hc_1 = \mathrm{e}^{-k'_f H/Q_a^n}. \tag{24}$$

Eckenfelder[13] suggests that the amount of active surface film covering the filter medium decreases with depth H; therefore, combining Eqs. 22 and 23 and substituting $c_1 \propto 1/H^m$, gives:

$$\frac{S_e}{S_0} = \frac{1}{1 + k_f A_v^n H(1-m)/Q_a^n} = \frac{1}{1 + k''_f H(1-m)/Q_a^n}. \tag{25}$$

For treatment of domestic wastewater on rock filters, Eckenfelder has obtained the values of $n = 0.5$, $m = 0.33$ and $k''_f = 2.5$ with H in ft and q in MGD/acre. Several empirical relationships for process efficiency in trickling filters have been proposed and successfully applied. Most significant of these are the National Research Council Formula and Rankin's Formula which have been described in detail in ASCE Manual.[36] Eckenfelder and O'Connor[13] have reported a value of 1.035 for overall temperature coefficient, θ, in Eq. 16. An adjustment in process efficiency due to variation in temperature should be provided.

Activated Sludge Process

It is a biological treatment process in which biologically active mass, called activated sludge, is continuously mixed with the biodegradable matter in an aeration basin in the presence of oxygen. The combination of wastewater and activated sludge is called the mixed liquor. The oxygen is supplied to the mixed liquor either by diffusing compressed air or pure oxygen into the liquid or by mechanical aeration. The activated sludge is subsequently separated from the mixed liquor by sedimentation in a clarifier and a part of this sludge is recirculated to the aeration basin. The rest of this sludge, indicating net excess production of biological cell material, is disposed of. Activated sludge treatment plants vary in performance due to variation in unit arrangements, methods of introducing air and wastewater into the aeration basin, aeration time, concentration of active biomass, aerator volume, degree of mixing, etc. Some important types of activated sludge processes are discussed below and their operating parameters are summarized in Table 2.

TABLE 2
Activated sludge process parameters

Parameters	Conventional	Step Aeration	Short Term	Biosorption	Pure Oxygen	Complete Mixing	Extended Aeration	Aerated Lagoons
Organic Loading Rate—B_v 1b BOD_5 per day per 1000 cubic feet	30–40	50–150	100–400	30–70	150–250	125–180	10–20	5
g BOD_5 per day per cubic metre	480–640	800–2400	1600–6400	480–1120	2400–3200	2000–2880	160–320	80
Process Loading Factor, U 1b BOD_5 per day per 1b 1b MLVSS								
or								
kg BOD_5 per day per kg MLVSS	0.2–0.5	0.2–0.5	2–5	0.2–0.5	0.4–1.0	0.6–1.0	0.05–0.2	0.2
Sludge Age, days, θ_x	3–4	3–4	0.2–0.5	3–4	0.8–2.3		14–∞	3–5
Aeration Time, hours, $\bar{t}$	6–7.5	6–7.5	2–4	0.5–1.5 (aeration)	1–3	3–5	20–30	70–120
BOD_5 removal, %, E	90–95	90–95	60–85	85–90	88–95	85–90	85–90	85–90
$\frac{\text{Normal Return Sludge}}{\text{Average Resign Flow}} \times 100$	30 (15–75)*	50 (20–75)*	20 (10–50)*	100 (50–150)*	25 (20–50)*	100 (50–150)*	100 (50–200)*	0
Primary Settling Required	Yes	Yes	No	Optional	Yes	Optional	No	No

*Provision in design should be made for these maximum and minimum values.

Kinetic Rate: Depending upon the design and operating conditions, one or more of the kinetic rate Eqs. 10, 12, 13 and 14 for BOD removal can be applied to different types of the activated sludge processes.

Oxygen Requirement: Oxygen is used to provide energy for synthesis of biological cells and for endogenous respiration of the biological mass. The total oxygen requirement, ΔO_2, can be expressed with the following equation;

$$\Delta O_2 = a'\Delta S + b'X_T \tag{26}$$

where a' is the fraction of BOD removed that is oxidized for energy and b' is the oxygen used for endogenous respiration of the biological mass, per day. In conventional aeration basins, an hourly oxygen demand of 50 to 80 mg/L per 1000 mg/L of VSS is exerted near the beginning of the tank and is reduced to 20 mg/L per 1000 mg/L of VSS in the course of 4 to 6 hours.[14]

Excess Sludge Yield: By applying material balance for volatile suspended solids in activated sludge system, and using the concept shown in Figure 3:

Excess solids in activated sludge system = Non-biodegradable suspended solids in influent + Biomass Synthesized during BOD removal – Biomass broken down by endogenous respiration
or

$$\Delta X = fX_0 + a\Delta S - bX_T \tag{27}$$

where:

ΔX = Net accumulation of volatile suspended solids, g/day

f = Fraction of volatile suspended solids present in the influent which are non-degradable

X_0 = Influent volatile suspended solids, g/day

Temperature Effect: According to Eckenfelder and O'Connor,[13] the value of temperature coefficient in Eq. 12 varies between 1.0 for low loading rates to 1.04 for high loading rates. Friedman and Schroeder[38] have studied in detail the effect of temperature on growth and the maximum cell yield occurred at 20°C.

Elements of a conventional activated sludge system are shown in Figure 14. In this system, the settled waste is mixed with the return sludge at the inlet end of the aeration tank. The microorganisms receive the full impact of any shock load and respond accordingly with sudden increase in oxygen demand during growth. By the time microorganisms leave the aeration tank, the organic matter has been stabilized and the microorganism population starts dying off. Thus, the microbial population undergoes a continual shifting and never reaches a relatively constant equilibrium.[7]

A mass of activated sludge of three to four times the mass of the daily BOD load must be kept in the system in order to consume all the new food and also acquire good settling properties. These types of plants have been used for treating domestic wastewaters of low biochemical oxygen demands. In conventional activated sludge plants

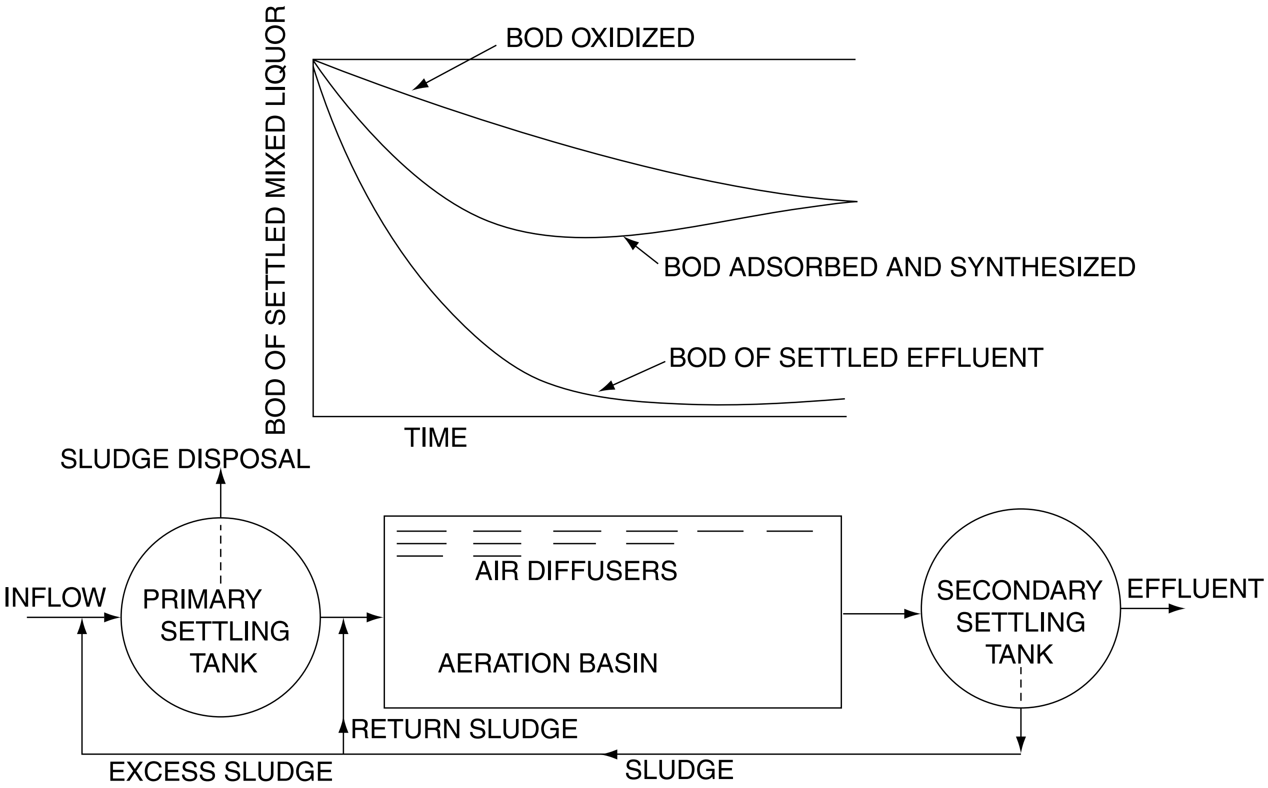


FIGURE 14 Conventional activated sludge.

that have plug flow design, high BOD in influent causes higher oxygen demand at that point in the mixed liquor and this oxygen demand diminishes as the flow passes down the aeration tank. Most of the plants designed these days are provided with tapered aeration, with highest air supply near the inlet end and lowest near the outlet end of the aeration tank.

Modifications of the Conventional Activated Sludge Process

A. Step Aeration Activated Sludge

Step aeration process, developed by Gould[39] at New York City, offers more flexibility than the conventional activated sludge process. In this process, wastewater is introduced at four or more points along the aeration tank in order to maintain a uniformly distributed loading. In addition to evening out the oxygen demand, this also keeps sludge reaerated in the presence of substrate. This process remains biologically more active instead of reaching the endogenous phase near the end of the conventional aeration tank. Step aeration system layout and fluctuations in BOD in aeration tank are shown in Figure 15. This method has been successfully employed in the treatment of domestic wastewaters and industrial wastewaters of similar nature.

B. Short Term Aeration or High Rate or Modified Activated Sludge

These systems have very high loading rates, both in terms of organic and volumetric loading, and low mixed liquor volatile suspended solids, thus requiring small aeration tank capacities and reduced air requirements. Because of shorter aeration time and lower mass of organisms, this process provides an intermediate degree of treatment. Organic matter is removed largely by synthesis, thus exerting a high rate of oxygen demand and producing a relatively large volume of sludge per unit mass of BOD removed. Since the sludge still contains certain unstabilized organic matter, the settled sludge in secondary settling tanks should be removed rapidly in order to avoid its anaerobic decomposition and floatation. The flow diagram is similar to the conventional system as shown in Figure 14.

C. Contact Stabilization or Biosorption

The elements of this type of plant are shown in Figure 16. This system is ideally suited to the treatment of wastewaters in which a large portion of BOD is

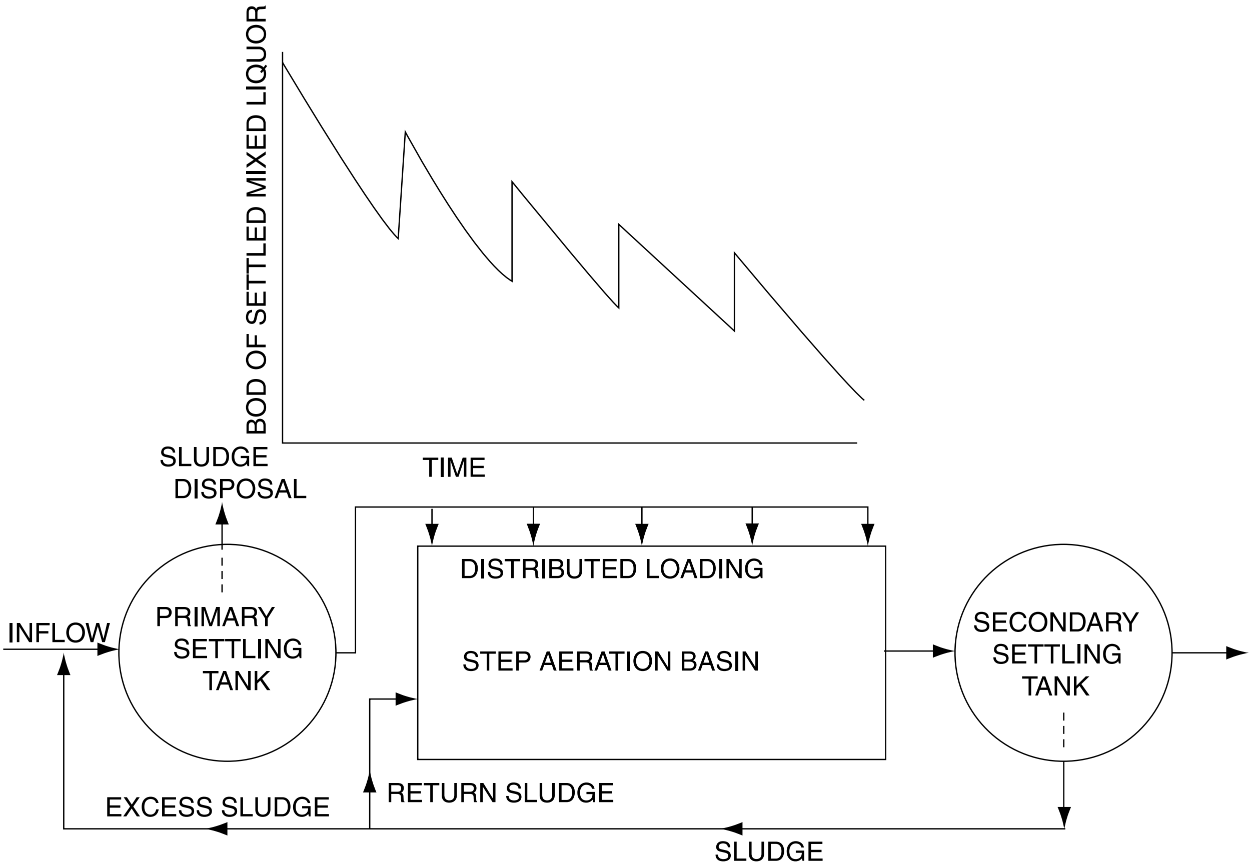


FIGURE 15 Step aeration activated sludge.

present in suspended or colloidal form. The suspended BOD is rapidly absorbed in a short period, ½ to 1½ hours, by the well-activated organisms and a part of soluble BOD is metabolized. In the activation tank, the sludge is reaerated for bio-oxidation and stabilization of adsorbed food; and when returned to the aeration tank, it is activated for higher BOD removal as compared to the conventional plant where sludge has become lean and hungry in the absence of a food supply. The additional advantage of this process is the reduced overall tank volume required as compared to the conventional system. However, the operation of such plants is more complex and less flexible than conventional ones.

D. Completely Mixed Activated Sludge

"Complex mix" approach is with respect to combining the return sludge and wastewater in order to maintain the entire contents of the aeration chamber in essentially a homogenous state. Wastewater introduced into the aeration basin is dispersed rapidly throughout the mass and is subjected to immediate attack by fully developed organisms throughout the aeration basin. Biological stability and efficiency of the aeration basin is enhanced by this design. Layout of a completely-mixed activated sludge plant and variation in BOD are shown in Figure 17.

In this mathematical analysis, McKinney[5] considered the complete mixing activated sludge process as the one in which the untreated wastes are instantaneously mixed throughout the aeration tank. In effect, the organic load on the aeration tank is uniform from one end to the other end and consequently a uniform oxygen demand and a uniform biological growth are produced. It is assumed to reduce the effect of variations in organic loads that produce shock loads on conventional units, retain a more biological population and hence, produce a more uniform effluent, and be able to treat organic wastes of any concentration and produce an effluent of any desired concentration.[5] Using Treatment Model II, Figure 11, as an example of a completely mixed system, Lawrence and McCarty[11] have shown analytically that although the complete-mixing will reduce the shock loads due to variations in organic loads, plug flow type conventional units, Treatment Model III, are more efficient.

Assuming that Eq. 13 is applicable for BOD removal rate, and since the BOD in a completely mixed aerator, S, is equal to the effluent BOD, S_e, therefore under steady state conditions:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{S_0 - S_e}{t} = qXS_e$$

or

$$\frac{S_e}{S_0} = \frac{1}{1 + qXt}. \tag{28}$$

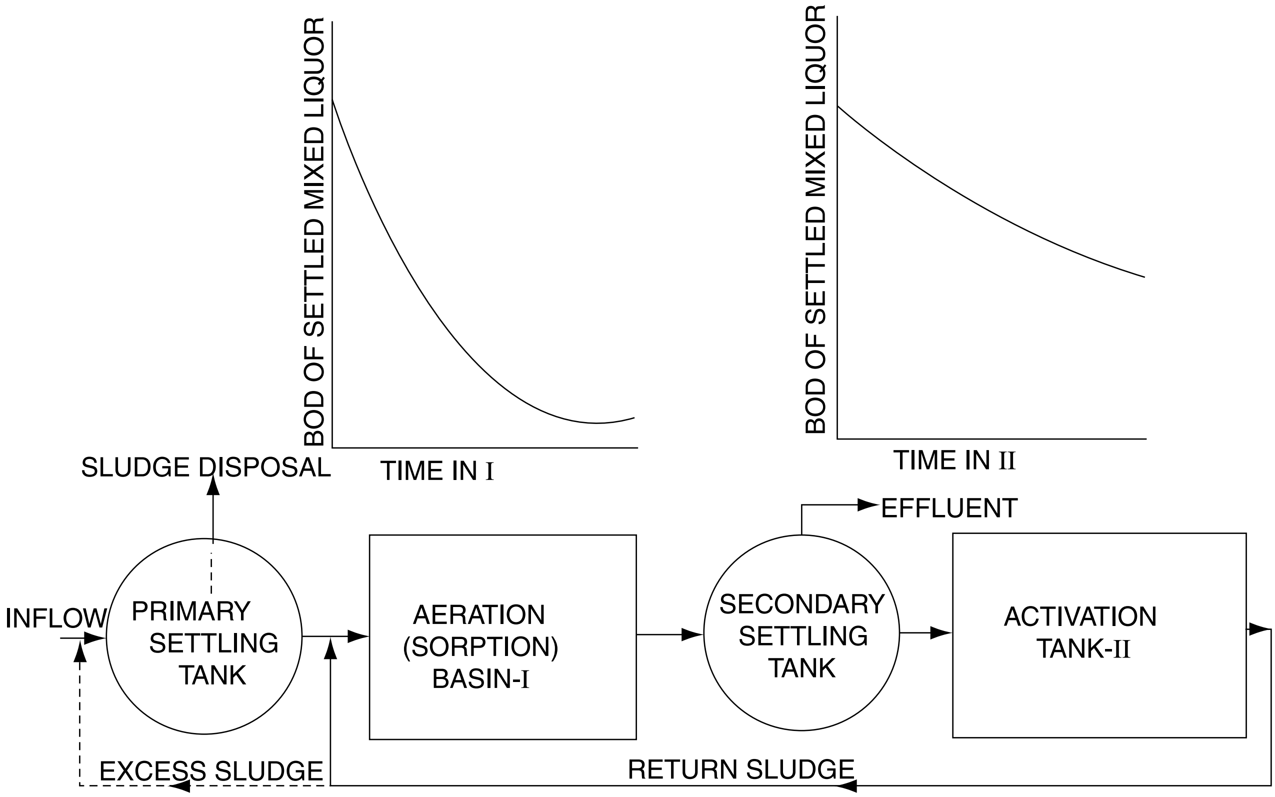


FIGURE 16 Biosorption (contact stabilization) activated sludge.

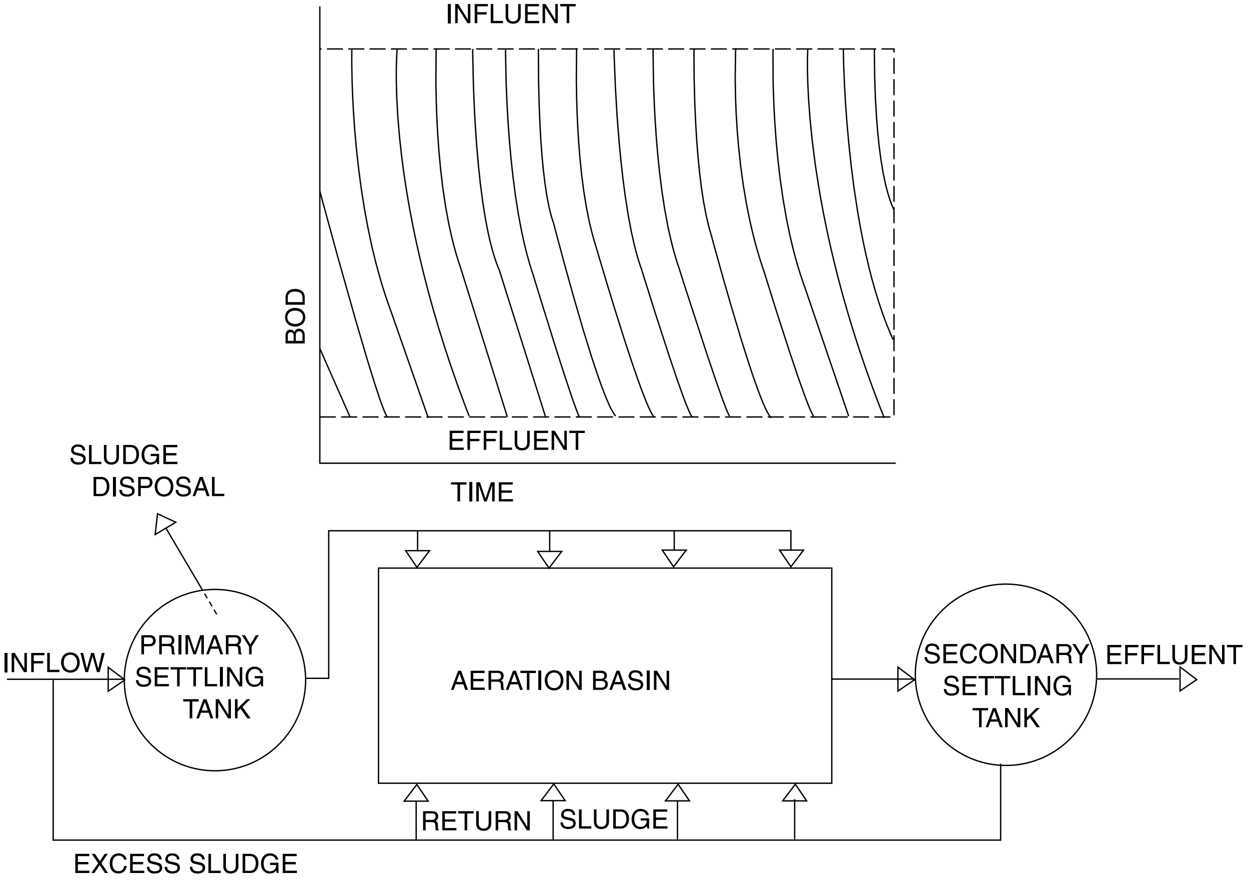


FIGURE 17 Complete mixing activated sludge.

E. In recent years, several wastewater treatment plants have been designed to operate with pure oxygen instead of conventional use of air in activated sludge treatment process. The obvious advantage of pure oxygen aeration is the higher oxygen concentration gradient maintained within the liquid phase, and this condition permits higher concentration of biomass in the aeration tank. This process has been shown to be more economical due to less energy requirements and in some cases has produced a better quality effluent. Significant increase in volumetric loading rate, reduction in sludge production, elimination of foaming problems and decrease in treatment costs are claimed to be advantages.[40]

A pure oxygen activated sludge system developed by Union Carbide Corporation is shown in Figure 18. This process is operated at MLSS values between 3000–10000 mg/L and the settling rate of sludge is considerably improved.

F. Extended Aeration

Extended aeration plant is the one where the net growth rate is made to approach zero, i.e., rate of growth becomes approximately equal to rate of decay. This is achieved by increasing the aeration time in order to keep the sludge in the endogenous growth phase for a considerable time. In practice, it is impossible to operate an extended-aeration system without sludge accumulation, because certain volatile solids, mainly polysaccharides in nature and inert organisms in activated sludge process, accumulate in the plant. Excess sludge is not generally wasted continuously from an extended aeration, but instead, the mixed liquor is allowed to increase in suspended solids concentration and a large volume of the aeration tank content or return sludge is periodically pumped to disposal. Oxidation ditch plants are designed and operated on this principle. Layout of a typical extended-aeration plant and variation in BOD in aeration tank are shown in Figure 19.

G. Aerated Lagoons

These are similar to the activated sludge system but without recirculation of sludge. Mechanical or diffused aeration devices are used for supplying oxygen and also providing sufficient mixing. All suspended solids may or may not be kept in suspension, depending upon the degree of mixing. Deposited solids may undergo anaerobic decomposition. Mathematically, the BOD removal rate in aerated lagoons is given by Eq. 13 and assuming the aerated lagoon to be a completely mixed system, without recycle and maintaining sufficient turbulence,

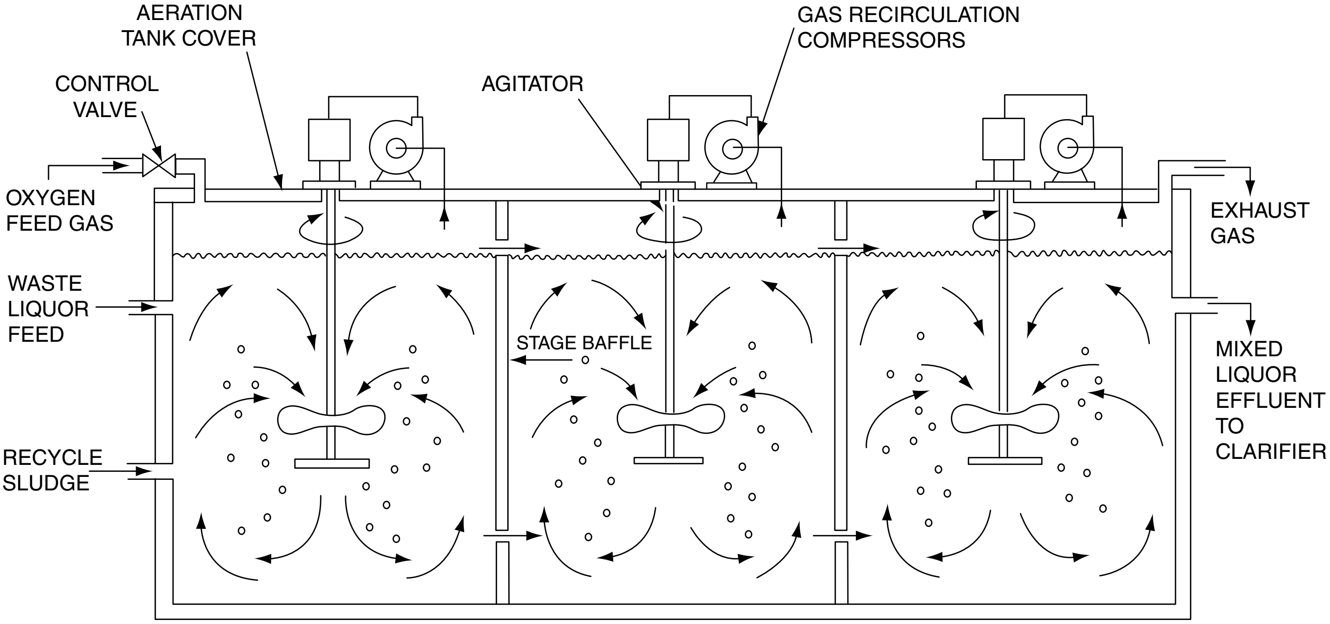


FIGURE 18 Schematic diagram of "unox" system with rotating sparger.

this equation becomes similar to Eq. 28. In practice, this equation has proven to represent a generalized response function for design of most aerated lagoons.[33]

The exact solid level in an aerated lagoon can be approximated by applying a material balance around the lagoon, under equilibrium conditions:

Solids In + Net Synthesis In Basin = Solids Out

or

$$X_0 + (Y\Delta S - b\,X_e t) = X_e$$

or

$$X_e = \frac{X_0 + Y\Delta S}{1 + bt} \tag{29}$$

Because of a very low solid concentration, the detention time in aeration basins is very high and a large volume of aeration basins is required. Therefore, the temperature variation exerts a profound effect on the rate of BOD removal. Eckenfelder and Ford[10] have given a relationship for estimating the lagoon temperature at both extreme conditions. Once this temperature is established, a corrected k_T value should be obtained from Eq. 16, using θ equal to 1.035 and then adopted in the kinetic Eq. 28.

Several other modifications in the activated sludge process have been discussed elswhere;[41] but most of these modifications are similar in concepts to one or more of the types discussed above. For example, in Hatfield and Kraus systems the supernatant from digestion tanks or even digested sludge are added to the reaeration tank to provide nutrients. Similarly, an Activated Aeration Plant is a combination of a conventional activated sludge process and the short-term aeration process.

Rotating Biological Contactors

As mentioned earlier, the traditional aerobic biological wastewater treatment processes have been divided into two groups: fixed film or stationary contact systems like trickling filters and suspended contact systems like activated sludge process. Rotating biological contactors, RBC, are more like trickling filters in operation, but adopt certain characteristics of suspended growth systems. In this process, large lightweight plastic disks of 2–4 m diameter are half submerged in the wastewater flowing continuously through cylindrical bottomed tanks. The disks are rotated slowly at a speed of 1–2 rpm. The biomass grows on the plastic disks and the substrate is absorbed by this biomass while it is submerged in the wastewater. The oxygen absorption occurs when the biomass is in direct contact with air, generally at a rate higher than that obtained in trickling filters.

These units have been operated successfully at extreme temperature conditions both for municipal and industrial wastewaters having very high BOD values. Antoine and Hynek[42] have concluded that RBC are stable, versatile and competitive with the activated sludge process.

In Canada, an important parameter regulating the pulp and paper wastewater treatment is toxicity reduction, measured by rainbow trout standard bioassay tests. The results of bioassay tests conducted by Antoine[43] showed RBC was effective in treating the toxic paper mill wastewater, when

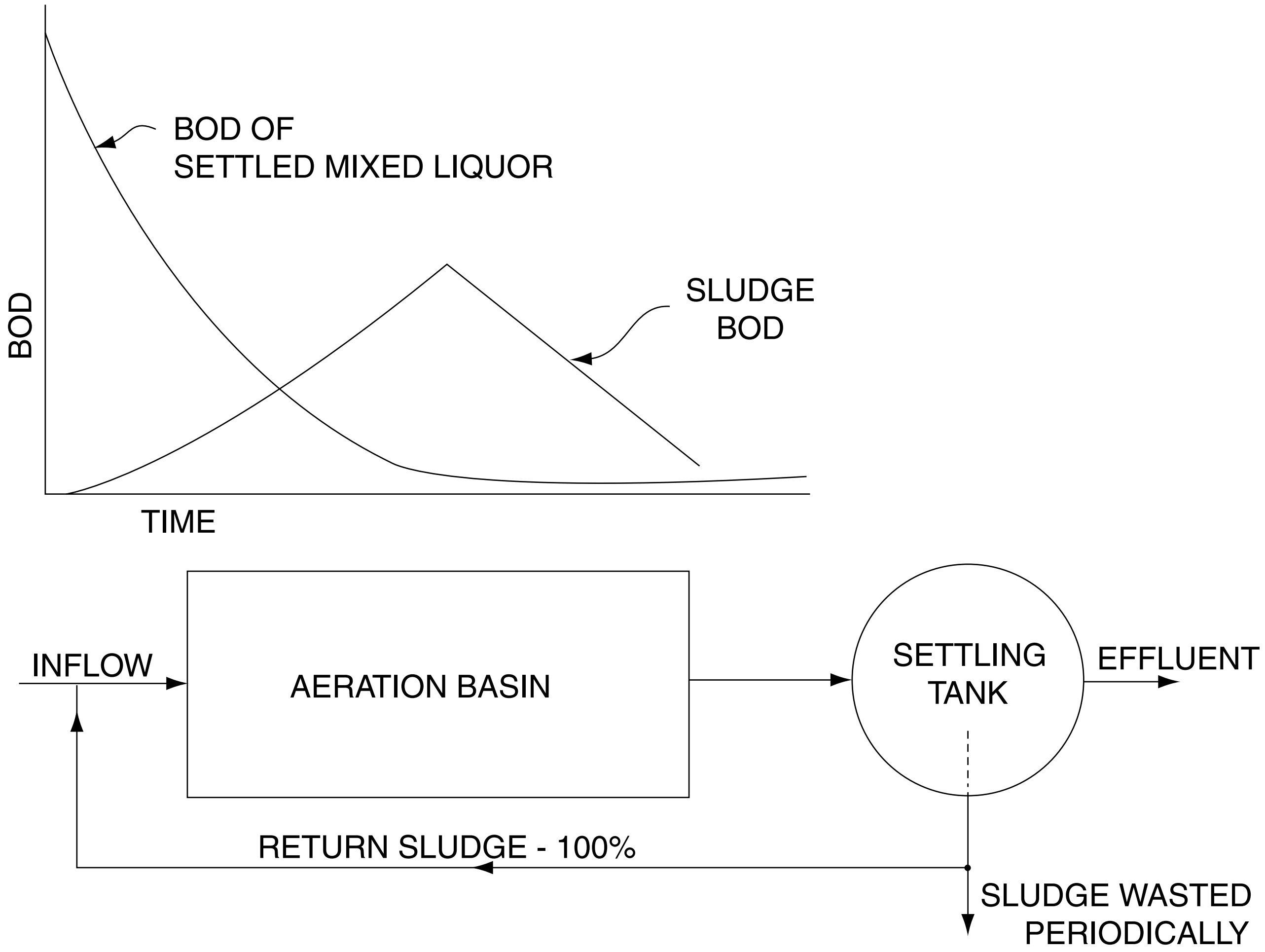


FIGURE 19 Extended aeration activated sludge.

it was operated at disk speeds of 13 and 17 rpm and flow rates of 1.9 to 2.5 LPM (0.5 to 0.65 USGPM). Similarly, Antoine observed that the RBCs were able to produce acceptable effluents for boardmill, kraft and sulfite wastewaters. For sulfite wastewater, the loading rate had to be reduced to increase the detention time. On the other hand, the suspended growth treatment of pulp and paper wastes has not consistently produced effluents of an acceptable level. B.C. Research had conducted tests on the use of the rotating biological contactor process for refinery waste containing phenols and observed it to be an effective method with proper control on operation.[43]

Anaerobic Treatment

In this process, anaerobic bacteria stabilize the organic matter in absence of free oxygen. Anaerobic treatment has been used widely for stabilization of sludges collected from primary and secondary settling tanks and recently is being adopted for treatment of soluble wastes in anaerobic lagoons, anaerobic filters, etc. One of the important advantages of anaerobic processes over aerobic processes is a high percentage conversion of organic matter to gases and liquid and a low percentage conversion to biological cells. McCarty[44] has mentioned that efficient anaerobic treatment of soluble wastes with BOD concentration as low as 500 mg/L is now feasible. Wastes with lower BOD can also be treated anaerobically, although the waste treatment efficiency will not be of the same magnitude as expected from aerobic treatment.

Anaerobic treatment of wastewaters takes place in two stages as shown in Figure 20. In the first stage, complex organic materials like protein, fats, carbohydrates, are converted into simple organic acids by acid forming bacteria, but with little change in BOD or COD value. In the second stage, these fatty acids are converted to carbon dioxide and methane, thereby stabilizing the BOD or COD.

In a conventional anaerobic treatment process, the substrate is fed into the digester continuously or intermittently. In most of the existing digesters, the contents are mixed, mechanically or with compressed gas collected from digesters. There is no recirculation of digested sludge and the system is a typical flow through system. The hydraulic detention time, $\bar{t}$ in

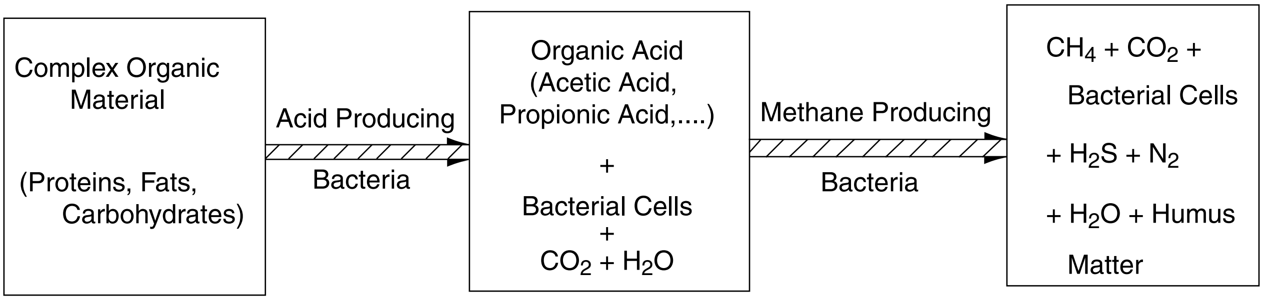


FIGURE 20 Sequential mechanism of anaerobic waste treatment.

the conventional process becomes equal to the solid retention time, θ_x. Recently, several modifications have been made in the conventional anaerobic treatment process. McCarty[44] has grouped the basic anaerobic process designs into Conventional Process, Anaerobic Activated Sludge Process, and Anaerobic Filter Process. Operating conditions of these process designs are shown in Figure 21. It is suggested that the conventional process be used for concentrated wastes like sludges where economical treatment can be obtained by keeping hydraulic detention time, $\bar{t}$ equal to the desired solid retention time, θ_x. The economic treatment of diluted wastes, however, requires hydraulic detention time, $\bar{t}$, much below the desired solid retention time, θ_x, and thus, anaerobic contact processes become more applicable.[44]

Anaerobic treatment processes are more sensitive to operating parameters and their environments as compared to aerobic processes. The best parameter for controlling the operation of anaerobic treatment is the biological retention time or solid retention time, SRT. A minimum SRT exists below which the critical methane producing bacteria are removed from the system faster than they can reproduce themselves. In practice, SRT values of two to ten times this minimum value are used. Thus, the hydraulic detention time and solid retention time maintained in anaerobic treatment processes are very high and the net growth of biological solids becomes very low due to significant decay as given by Eq. 12.

Mixing of the digester content is becoming a common practice. The advantages of mixing are better contact between food and microorganisms, uniform temperature, reduction in scum formation, accelerated digestion and distribution of metabolic inhibitors.

Certain cations, such as sodium, potassium, calcium, or magnesium show a toxic or inhibitory effect on anaerobic treatment when present in high concentrations, as shown in Table 3.[45]

Soluble sulfides exhibit toxicity because only they are available to the cells. If the concentration of soluble sulfides exceeds 200 mg/L, then the metabolic activity of methanogenic population will be strongly inhibited leading to the process failure.[21] Concentrations up to 100 mg/L can be tolerated without acclimation and sulfide concentrations between 100 and 200 mg/L can be tolerated after acclimation.

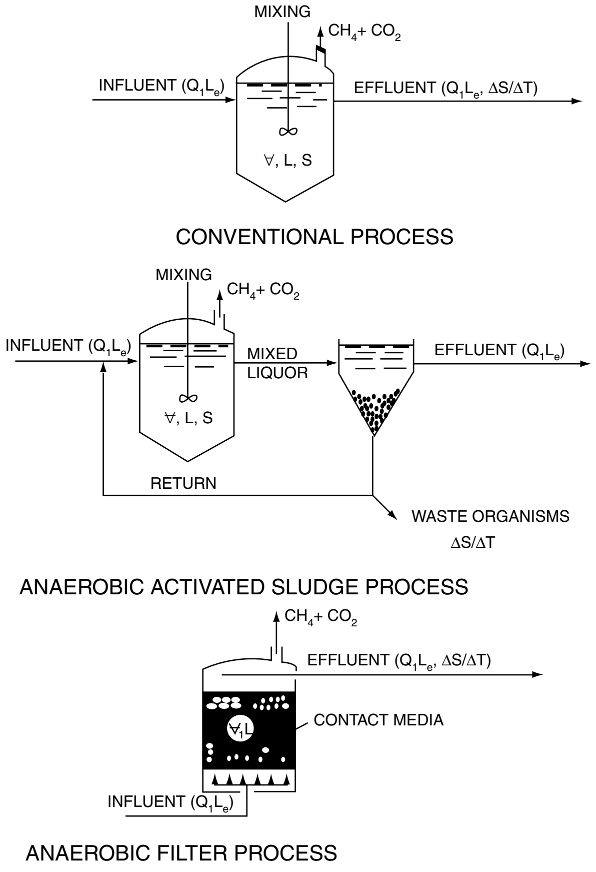


FIGURE 21 Basic anaerobic process designs.

TABLE 3
Stimulatory and inhibitory concentrations of light metal cations to anaerobic processes

Cation	Stimulatory Con., mg/L	Strong Inhibitory Con., mg/L
Sodium	100–200	8000
Potassium	200–40	12000
Calcium	100–200	8000
Magnesium	75–150	3000

Depending on pH, ammonia can be toxic to anaerobic bacteria and free ammonia is more toxic. If concentration of free ammonia exceeds 150 mg/L, severe toxicity will result, whereas the concentration of ammonium ions must be greater than 3000 mg/L to have the same effect. At a concentration of 1600 mg/L as *N*, ammonia can upset the process.[20] The volatile acids cause little inhibition in anaerobic reactors at neutral pH.[21] Operating parameters of conventional anaerobic digesters are shown in Table 4.

6. NUTRIENT REMOVAL

Biological nitrification and denitrification is one of the common methods for nitrogen removal from wastewaters. In warmer climates, nitrification may occur to a considerable degree in conventional aerobic biological treatment processes, followed by serious adverse effects of denitrification in settling tanks and/or the receiving bodies of water. In northern cold climates, below 18°C, a three-stage biological system as shown in Figure 22 is considered necessary for nutrient removal.[46] In the first stage, carbonaceous BOD is reduced to a level below 50 mg/L. In the second stage, the ammonia, present in effluent from the first stage, is oxidized to nitrites and nitrates by nitrosomonas and nitrobacters, respectively, as shown below:

$$2NH_4^+ 3O_2 \xrightarrow{\text{Nitrosomonas}} 2NO_2^- + 2H_2O + 4H^+$$
$$2NO_2^- + O_2 \xrightarrow{\text{Nitrobacter}} 2NO_3^-$$

The third stage accomplished denitrification–conversion of nitrites and nitrates to atmospheric nitrogen under anaerobic conditions:

$$3NO_3^- + CH_3OH \rightarrow 3NO_2^- + CO_2 + 2H_2O$$
$$2NO_2^- + 2CH_3OH \rightarrow N_2 + CO_2 + H_2O + 2OH^-$$

TABLE 4
Operating parameters of conventional anaerobic digesters

	Parameters	Unmixed	Mixed
B_v	– Loading Rate,	0.02–0.05	0.1–0.3
	1b VSS/day/cubic ft	0.32–0.80	1.6–3.2
	kg VSS/day/cubic metre		
$\bar{t}$	– Detention time, days	30–90	10–15
E	– Volatile Solids	50–70	50
	Reduction percent		
Mixing		Absent	Present
pH		6.8–7.4	6.8–7.4
Temperature, °C		30–35	30–35

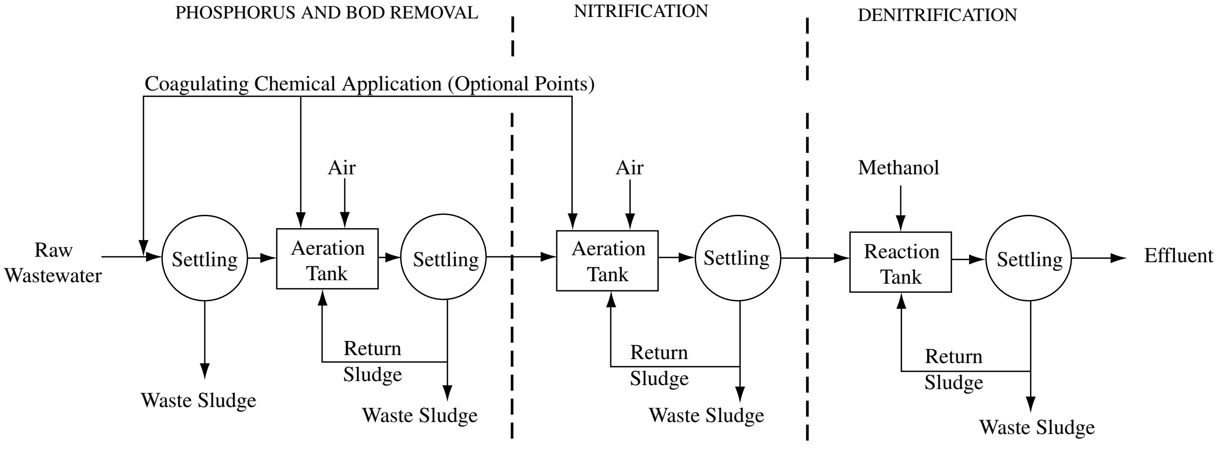


FIGURE 22 Typical three-stage treatment process for nutrient removal.

A supplemental source of carbonaceous BOD must be added in this stage to reduce the nitrates to nitrogen gas in a reasonable period of time. This has been accomplished either by adding a cheap organic substrate like methanol or by bypassing a part of the wastewater containing carbonaceous BOD in the first stage. In some cases, the carbonaceous and nitrogeneous oxidation steps are combined in a one-stage aerobic biological system. Another system uses fixed-film reactors, such as gravel beds, separately for nitrification and denitrification stages. Effluent nitrogen concentrations of 2 mg/L have been proposed as the upper limit in a biological process. Many full scale biological nitrogen removal facilities are now in operation. Nitrifying bacteria are subject to inhibition by various organic compounds, as well as by inorganic compounds such as ammonia. Free ammonia concentrations of 0.1 to 1.0 mg/L and free nitrous acid concentrations of 0.22 to 2.8 mg/L, start inhibiting *Nitrobacters* in the process.[20]

The majority of phosphorus compounds in wastewaters are soluble and only a very small fraction is removed by plain sedimentation. The conventional biological treatment methods typically remove 20 to 40 percent of phosphorus by using it during cell synthesis. A considerably higher phosphorus removal has been achieved by modifying the processes to create "luxury phosphorus uptake." Factors required for this increased phoshorus removal are plug-flow reactor, slightly alkaline pH, presence of adequate dissolved oxygen, low carbon dioxide concentration and no active nitrification.[46] However, the most effective method of phosphate removal is the addition of alum or ferric salts to conventional activated sludge processes.

Nomenclature

A_v = Specific surface area of filter media, Length^{-1}

B_v = Volumetric loading rate; mass per unit volume per unit time

D = Longitudinal dispersion coefficient, (Length)^2 per unit time

E = Process treatment efficiency, ratio

H = Filter depth, length

K = Half velocity coefficient = substrate concentration at which rate of its utilization is half the maximum rate, mass per unit volume

K_i = Inhibition constant, mass per unit volume

L = Substrate concentration around microorganisms in reactor, measured in terms of BOD, mass per unit volume

N_0 = Number of microorganisms per unit volume at time $t = 0$

N_t = N = Number of microorganisms per unit volume at time t

ΔO_2 = Amount of oxygen requirement, mass per unit time

Q = Volumetric rate of flow, volume per unit time

Q_a = Volumetric rate of flow per unit area, Length per unit time

Q_r = Volumetric rate of return flow, volume per unit time

R = Recycle ratio

S = Substrate concentration, mass per unit volume

ΔS = Substrate removed, mass per unit time

S_e = Effluent BOD or final substrate concentration, mass per unit volume

S_0 = Influent BOD or in the initial substrate concentration, mass per unit volume

T = Temperature, °C

U = Process loading factor, time^{-1}

V = Volume of the reactor, volume

X = Mass of active microorganisms present per unit volume

ΔX = Cell mass synthesized, mass per unit time

X_e = Effluent volatile suspended solids, mass per unit volume

X_0 = Influent volatile suspended solids, mass per unit volume

X_r = Volatile suspended solids in return sludge, mass per unit volume

X_T = Total mass of microorganisms in the reactor, mass

Y = Growth yield coefficient, dimensionless

a' = Fraction of BOD removed that is oxidized for energy

b = Microorganisms decay coefficient, time^{-1}

b' = Oxygen used for endogenous respiration of biological mass, time^{-1}

c_1 = Constant

f = Fraction of volatile suspended solids present in the influent which are non-degradable

k_f, k'_f, k''_f = Rate coefficient in filters, time^{-1}

k_0 = Logarithmic growth rate constant, time^{-1}

k_t = Growth rate factor, time^{-1}

k' = Growth rate factor, (time)$^{-1}$ (mass per unit volume)$^{-1}$

l = Length dimension in reactor, Length

m = Constant

n = Trickling filter exponent

q = dS/Xdt = Substrate utilization rate per unit biomass

$q_{\max}$ = Maximum substrate utilization rate per unit biomass

t = Contact time in filter or any other reactor, time

$\bar{t}$ = V/Q = Mean retention time, time

u = Mean displacement velocity in reactor along length, length per unit time

w = Volumetric rate of flow of waste sludge, volume per unit time

θ = Temperature coefficient for microbial activity

θ_x = Mean cell retention time, time

μ = dx/Xdt = Specific growth rate of microorganisms, time^{-1}

$\mu_{\max}$ = Maximum specific growth rate of microorganisms, time^{-1}

D/ul = Reactor dispersion number, dimensionless

M/F = Microorganisms to food ratio in a reactor

$\mathrm{d}L/\mathrm{d}t$ = Rate of waste utilization measured in terms of BOD, mass per unit volume per unit time

$\mathrm{d}N/\mathrm{d}t$ = Rate of growth in number of microorganisms, Number per unit volume per unit time

$\mathrm{d}S/\mathrm{d}t$ = Rate of substrate consumption, mass per unit volume per unit time

$\Delta S/\Delta t$ = Mass of substrate utilized over one day, mass per unit time

$S_T/\Delta t$ = Total mass of substrate applied over a period of one day, mass per unit time

$\mathrm{d}X/\mathrm{d}t$ = Rate of growth of mass of active microorganisms, mass per unit volume per unit time

$\Delta X_T/\Delta t$ = Total quantity of active biomass withdrawn daily, mass per unit time

REFERENCES

1. MacInnis, C., Municipal Wastewater, *Encyclopedia of Environmental Science and Engineering,* Vol. 1, edited by J.R. Pfafflin and E.N. Ziegler, Gordon and Breach, New York.
2. Gates, W.E. and S. Ghosh, Biokinetic Evaluation of BOD Concepts of Data, *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers,* Vol. 97, no. SA3, June 1971, pp. 287–309.
3. McKinney, R.E., *Microbiology for Sanitary Engineers,* McGraw-Hill Book Company, Inc., New York, 1962.
4. Stanier, R.Y., M. Doudoroff and E.A. Adelberg, *The Microbial World,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1957.
5. McKinney, R.E., Mathematics of Complete Mixing Activated Sludge, *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers*, **88**, SA3, May 1962, pp. 87–113.
6. Tsuchiya, H.M., A.G. Frederickson and R. Avis, Dynamics of Microbial Cell Populations, *Advances in Chemical Engineering,* Vol. 6, edited by T.B. Drew, J.W. Hoopes, Jr. and T. Vermeulen, Academic Press, New York, 1966.
7. Stewart, M.J., Activated Sludge System Variations, Specific Applications, *Proceedings of the Fifteenth Ontario Industrial Waste Conference,* June 1968, pp. 93–115.
8. Irvine, R.L. and D.J. Schaezler, Kinetic Analysis of Date from Biological Systems, *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers,* Vol. 97, No. SA4, August 1971, pp. 409–424.
9. Monod, J., The Growth of Bacterial Cultures, *Annual Review of Microbiology,* 3, 371, 1949.
10. Eckenfelder, W.W. and D.L. Ford, *Water Pollution Control,* Jenkins Publishing Company, Austin, Texas, 1970.
11. Lawrence, A.W. and P.L. McCarty, Unified Basis for Biological Treatment Design and Operation, *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers,* **96**, SA3, June 1970, pp. 757–778.
12. Pearson, E.A., Kinetics of Biological Treatment, *Advances in Water Quality Improvement,* edited by E.F. Gloyne and W.W. Eckenfelder, Jr., University of Texas Press, Austin, 1970.
13. Eckenfelder, W.W. and D.J. O'Connor, *Biological Waste Treatment,* Pergamon Press, New York, 1961.
14. Fair, G.M., J.C. Geyer and D.A. Okun, *Water and Wastewater Engineering,* Vol. 2, John Wiley and Sons, Inc., 1968.
15. Bewtra, J.K., Droste, R.L. and Ali, H.I., The Significance of Power Input in the Testing and Biological Treatment of Industrial Wastewater, *Treatment and Disposal of Liquid and Solid Industrial Wastes,* edited by K. Curi, Pergamon Press, New York, 1980, pp. 23–47.
16. Ruchhoft, C.C., Studies of Sewage Purification-IX, *Public Health Reports,* **54**, 468, 1939.
17. Goodman, B.L. and A.J. Englande, Jr., A Unified Model of the Activated Sludge Process, *Journal of the Water Pollution Control Federation,* 46, February 1974, p. 312.
18. Parker, H.W., *Wastewater Systems Engineering,* Prentice-Hall Inc., Englewood Cliffs, 1975.
19. Warren, C.E., *Biology and Water Pollution Control,* W.B. Saunders Company, Toronto, 1971.
20. Eckenfelder, W.W., Jr., *Principles of Water Quality Management,* CBI Publishing Company, Inc., Boston, 1980.
21. Grady, C.P., Jr. and H.C. Lim, *Biological Wastewater Treatment—Theory and Applications,* Marcel Dekker, Inc., New York, 1980.
22. Bewtra, J.K., Biological Treatment of Wastewater, *Encyclopedia of Environmental Science and Technology,* Vol. I, edited by E. Ziegler and J. Pfafflin, Gordon and Breach Science Publishers Inc., New York, 1982, pp. 81–102.
23. Bewtra, J.K., Toxocity Effects on Biological Processes in Waste Treatment, *New Directions and Research in Waste Treatment and Residual Management,* Vol. 2, Proceedings of International Conference held at the University of British Columbia, Vancouver, B.C., June 1985, pp. 807–827.
24. Gaudy, A.F., Jr., W. Lowe, A. Rozich and R. Colvin, Practical Methodology for Predicting Critical Operating Range of Biological Systems Treating Inhibitory Substrates, *Water Pollution Control Federation Journal,* Vol. 60, No. 1, 1988, pp. 77–85.
25. Grady, C.P.L., Jr., Biodegradation of Hazardous Wastes by Conventional Biological Treatment, *Hazardous Wastes and Hazardous Materials,* **3**, 1986, pp. 333–365.
26. Gaudy, A.F., Jr., A.F. Rozick and E.T. Gaudy, Activated Sludge Process Models for Treatment of Toxic and Nontoxic Wastes, *Water Science and Technology,* Vol. 18, 1986, pp. 123–137.
27. Godrej, A.N. and J.H. Sherrard, Kinetics and Stoichiometry of Activated Sludge Treatment of a Toxic Organic Wastewater, *Water Pollution Control Federation Journal,* Vol. 60, No. 2, 1988, pp. 221–226.
28. Pitter, P., Determination of Biological Degradability of Organic Substances, *Water Research,* **10**, 1976, pp. 231.
29. Adam, C.E., D.L. Ford and W.W. Eckenfelder, Jr., *Development of Design and Operational Criteria for Wastewater Treatment,* Enviro Press, Inc., Nashville, 1981.
30. Pierce, G.E., Potential Role of Genetically Engineered Microorganisms to Degrade Toxic Chlorinated Hydrocarbons, *Detoxication of Hazardous Wastes,* edited by J.H. Exner, Ann Arbor Science Publishers, Ann Arbor, 1982, pp. 315–322.
31. Levenspiel, O., *Chemical Reaction Engineering,* John Wiley and Sons, Inc., New York, 1967.
32. Murphy, K.L. and B.I. Boyko, Longitudinal Mixing in Spiral Flow Aeration Tanks, *Journal of the Sanitary Engineering Division, Proceedings of the American Society of Civil Engineers,* **96**, SA2, April 1970, pp. 211–221.
33. Parker, C.E., *Anaerobic–Aerobic Lagoon Treatment for Vegetable Tanning Wastes,* Report prepared for the Federal Water Quality Administration Environmental Protection Agency, U.S. Government Printing Office, Washington, D.C., December 1970.
34. Stevens, D.K., Interaction of Mass Transfer and Inhibition in Biofilms, *Journal of Environmental Engineering,* Vol. 114, No. 6, 1988, pp. 1352–1358.
35. Toda, K. and H. Ohtake, Comparative Study on Performance of Biofilm Reactors for Waste Treatment, *Journal of General Applied Microbiology,* Vol. 31, No. 2, 1985, pp. 177–186.
36. *Wastewater Treatment Plant Design,* American Chemical Society of Civil Engineers Manual of Engineering Practice, No. 36, New York, NY, 1977.
37. Sinkoff, M.D., R. Porges and J.H. McDermott, Mean Residence Time of a Liquid in a Trickling Filter, *Journal of the Sanitary Engineering*

Division, Proceedings of the American Society of Civil Engineers, **85**, SA6, 1959.
38. Friedman, A.A. and E.D. Schroeder, Temperature Effects on Growth and Yield for Activated Sludge, presented at 26th Purdue Industrial Waste Conference, Lafayette, Indiana, May 4–6, 1971.
39. Gould, R.H., Tallmans Island Works Opens for World's Fair, *Municipal Sanitation,* Vol. 10, No. 4, April 1939, p. 185.
40. McWhirter, J.R., Oxygen and the Activated Sludge Process, Chapter 3 in *The Use of High Purity Oxygen in the Activated Sludge Process,* Vol. 1 edited by J.R. McWhirter, CRC Press Inc., West Palm Beach, 1978.
41. Srinda, R.T. and R.F. Ward, Activated Sludge Processes: Conventional Processes and Modifications-Applications, presented at Short Course in Water Quality Control, Department of Civil Engineering, University of Massachusetts, Amherst, Mass., March 1970.
42. Antoine, R.L. and R.J. Hynek, Operating Experience with Bio Surf Process Treatment of Food Processing Wastes, *Proceedings of 28th Industrial Wastes Conference,* Purdue University, Lafayette, Indiana, May 1973.
43. Antoine, R.L. Fixed Biological Surfaces—Wastewater Treatment, CRC Press, Cleveland, Ohio, 1976, pp. 93–122.
44. McCarty, P.L., Anaerobic Treatment of Soluble Wastes, *Advances in Water Quality Improvement,* edited by E.F. Gloyne and W.W. Eckenfelder, Jr., University of Texas Press, Austin, 1970, pp. 336–352.
45. McCarty, P.L., Anaerobic Waste Treatment Fundamentals, *Public Works,* Vol. 95, No. 9–12, 1964, pp. 95–126.
46. Bouck, D.W., Nutrient Removal in Three-Stage Processing, Chapter 5 in *Advances in Water and Wastewater Treatment—Biological Nutrient Removal,* edited by M.P. Wanielista and W.W. Eckenfelder, Jr., Ann Arbor Science, Ann Arbor, MI, 1978, pp. 65–78

J.K. BEWTRA
N. BISWAS
University of Windsor

BROWNFIELDS

The American Society for Testing and Materials (ASTM) defines *brownfields* as “abandoned, idled, or underutilized properties where expansion or redevelopment is complicated by the potential or confirmed existence of chemical(s) of concern.” The U.S. Environmental Protection Agency (EPA) Web page states, “Brownfields are abandoned, idled, or under-used industrial and commercial facilities where expansion or redevelopment is complicated by real or perceived environmental contamination.”

In the mid-twentieth century, *brownfield* was a planners’ term for urban blight. Brownfields had existed for decades, perhaps centuries, but a strong focus on cleaning up these properties did not happen until the 1970s (see Table 1). Concurrent with the federal use of *brownfields,* several local and state governments adopted the term for their efforts to bring about economic revitalization. Most governments have adopted specific legal definitions of brownfields. These definitions reflect differing environmental and economic conditions but have strong similarities to the federal definition.

In everyday language, a brownfield is an area that is contaminated or perceived to be contaminated. Most brownfields can be redeveloped, revitalized, and reused after assessment and cleanup. The EPA’s brownfields program helps communities work together to create jobs and put abandoned properties back into productive use. The EPA, together with other federal, state, and local agencies, provides funds, coordination, and advice for the cleanup of brownfields.

Politically, brownfields have been contrasted to “greenfields.” Greenfields are rural areas that are in danger of being converted to industrial areas. The goals of many brownfield programs include saving farmland and open spaces in addition to putting brownfields back into industrial use. Development of greenfields can be economically and environmentally problematic, because it means building shipping and utility infrastructures that are essential for most industrial development. Development of brownfields can be economically and environmentally more desirable because they often have utility connections like water, sewer, and electricity as well as train access with sidings. Further, cities and counties can regain or enhance their tax base by cleaning and redeveloping brownfields.

This article discusses the history of brownfields, lists some common and legal definitions, discusses the associations with social justice and banking issues, and gives a case study of a showcase community that demonstrates creative successes. It does not discuss the associated new urbanism movement within architecture and urban planning.

HISTORY

The history of brownfields is intertwined with the history of hazardous-waste cleanups and the EPA. In the 1960s the United States grappled with the challenge of many unused and contaminated facilities. These properties were across the United States, from the shuttered steel mills in Pennsylvania and Cleveland to mining operations in Montana and Arizona to closed timber mills in Washington and Oregon. The facilities represented many industries, including closed smelters, metal-plating factories, machine shops, and chemical plants. Many facilities had complied with the few environmental regulations of the early 1900s.

In response to a fire on the Cuyahoga River, President Richard Nixon created the EPA in 1969 by presidential directive. The new EPA was faced with such media disasters as Love Canal, the Valley of the Drums, and Bridgeport. The EPA began regulatory efforts with the Clean Air Act and Clean Water Act, closely followed by regulations to control hazardous substances. In 1976, the Resource Conservation and Recovery Act (RCRA) and Toxic Substance Control Act (TSCA) initiated cleanup regulations. In 1980, the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA, or Superfund) ushered in two decades of federal Superfund cleanups. Many states created analogous legislation and regulations. From 1984 to 1995, Superfund cleanups were financed by a tax on the industries that synthesized or manufactured chemicals.

In response to the widespread economic-development obstacles posed by urban brownfields, the EPA announced its original Brownfields Action Agenda in January 1995. Brownfields were an adaptation from the EPA cleanup-enforcement-driven pattern over to economic and environmental local collaboration with support from the EPA. The impetus to bring about this change came from several large Midwest and East Coast cities that led the movement to revitalize their abandoned industrial areas with funding from federal and private sources. Initially, both the EPA and the John D. and Catherine T. MacArthur Foundations funded a series of brownfield forums in Chicago in the early 1990s. These forums developed a set of brownfield redevelopment principles that have been adopted, adapted, and standardized.

TABLE 1

Brownfields time line

Brownfields Timeline	1976	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	2000	01	02	03
Resources Conservation & Recovery Act (RCRA)—1976, 1984																												
Toxic Substance Control Act (TSCA)—1976																												
Waste-treatment facility fire, Bridgeport, New Jersey—1977																												
Children hurt at waste dump, Love Canal, New York—1978																												
Discovery of Valley of the Drums site, Kentucky—1979																												
Comprehensive Environmental Response, Compensation, & Liability Act (CERCLA or Superfund)—1980																												
$1.6 billion tax-funded trust fund authorized—1980																												
Environmental Justice Movement, Warren County, North Carolina—1982																												
Superfund Amendments & Reauthorization Act (SARA)—1986																												
Superfund trust fund increased to $8.5 billion—1986																												
Brownfields Initiatives announced—1993																												
Small Business Liability Relief and Brownfields Revitalization Act—1993																												
EPA Brownfield Action Agenda—1995																												
CERCLA/SARA tax on chemical & petroleum industry sunsetted—1995																												
Brownfield National Partnership—1997																												
Superfund Redevelopment Initiative—1999																												
Brownfield Initiative/Harvard’s Innovation in Government Award—2000																												
Expanded Brownfield Cleanup Revolving Loan Fund—2000																												
Relaim Pennysylvannia—cleanup of mining grayfields—2002																												
Nationwide movement to clean up abandoned grayfields—2003																												

Some of these principles have become engineering practice in the "Standard for Process of Sustainable Brownfields Redevelopment" from the ASTM.

In association with the forums, the Chicago Brownfield Initiative began with a pilot cleanup and redevelopment program in 1993. The Chicago Department of Environmental Protection, in partnership with the mayor's office and the Chicago Departments of Planning and Development, Buildings, and Law, coordinated the brownfields pilot program. The pilot program involved the cleanup up of five abandoned polluted industrial sites and initiated redevelopment. The five pilots resulted in new construction activity and the creation of jobs. The city's experience with these sites became a national model for continued innovation at large-scale cleanups.

Chicago shared its experiences by hosting another brownfield forum to discuss the legal, financial, and ethical issues related to urban disinvestments. The forum, which included business leaders, industrialists, environmentalists, bankers, regulators, and city officials, generated a list of recommended actions to facilitate brownfield cleanups and redevelopments. Cities across the United States began to use the successful Chicago-recommended actions. Chicago revisited its forum recommendation in late 1997 to assess local and national progress. This conference increased national attention and validated the work since the first conference. More urban areas took the model and made it theirs.

One city that took the model and made it theirs is St. Louis, Missouri. St. Louis, like many older cities, had deteriorated commercial districts that imposed a blighting effect on surrounding residential neighborhoods. St. Louis began one of the earliest brownfields programs in the mid-1990s. By 2000, St. Louis had cleaned up many sites using the brownfields approach. Mayor Freeman Bosley detailed the experiences at several sites in congressional testimony. In one targeted area, the city paid to assemble, clear, and clean a corner site critical to the shopping district's viability. According to the mayor, the owners of this area had not been able to command sufficient rent to maintain their property. When cleanup was accomplished, a private company invested in what is now a thriving commercial business district that provides employment, generates sales, and helps to attract patrons to other retail and eating establishments in the area.

On May 13, 1997, Vice President Al Gore announced the Brownfields National Partnership Action Agenda (National Partnership), outlining the EPA's activities and future plans to help states and communities implement and realize the benefits of the brownfields initiative. The National Partnership brings together federal agencies to address brownfield cleanups and redevelopments in a more coordinated approach.

AGENCIES, CITIES, AND UNIVERSITIES INVOLVED IN BROWNFIELDS

The other seven federal agencies involved are: the Department of Housing and Urban Development (HUD), the Department of Transportation (DOT), the General Services Administration (GSA), the National Oceanic and Atmospheric Administration (NOAA), the Department of Health and Human Services (DHHS), the Department of Labor (DOL), and the Department of Energy (DOE). HUD administers the Brownfields Economic Development Initiative (BEDI) as the key competitive grant program to stimulate and promote economic- and community-development activities under Section 108(q) of the Housing and Community Development Act of 1974. Through BEDI, HUD administers these grants to stimulate local governments and private-sector parties to redevelop or continue phased redevelopment efforts on brownfield sites where environmental conditions are known and redevelopment plans exist. The DOT has multiple approaches to support transportation-related brownfields by funding cleanups as part of its infrastructure development, work with other agencies on brownfields for transportation-related uses, encourage consideration of transportation access in redevelopment planning, and identify policies that discourage transportation-related brownfields redevelopment. With thousands of federal properties located throughout the country, the GSA is partnering with communities to ensure that underutilized federal properties are an active component in the redevelopment of our nation's urban centers. NOAA has a signed agreement with the EPA to lay the groundwork for revitalizing aging port-city waterfronts. The DHHS specifies essential services to be provided by its health-related agencies and the larger public-health community that must be applied to each brownfields project to assure public-health protection. The DOL, through its Office of Environmental Management, Office of Intergovernmental and Public Accountability, has developed an electronic access (Internet-based) system to provide technical assistance and increase community members' capacity to understand and resolve environmental issues related to brownfields. The DOE provides technical assistance in brownfield efforts from its Headquarters Program Offices and the National Laboratories and Technology Centers.

Many major urban areas, through both cities and counties, have associated with the federal brownfields, and some have continued their own brownfields efforts. Pittsburgh, Pennsylvania, is a city that develops brownfields innovations in association with Carnegie Mellon University while it continues to work with the EPA. Another city with a strong university affiliation is Cincinnati, Ohio, where collaboration with the University of Cincinnati provides training and environmental-justice support and broadens community affiliations.

COMMON AND LEGAL DEFINITIONS

The EPA and other environmental- and health-protection agencies base their regulations and implementation on science. Most often they adapt technical definitions that are measurable and science-based into regulations. The terms surrounding brownfields do not follow this pattern. Brownfields definitions bring a community-based sensibility. The complexity and plasticity of brownfields begins in the definitions and continues through implementation.

Legally, the EPA uses the definition of brownfield in Public Law 107-118 (HR 28869), the Small Business Liability Relief and Brownfields Revitalization Act, signed into law January 11, 2002. This definition says that "within certain legal exclusions, the term 'brownfield site' means real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant." Following the definition are pages of exclusions that primarily detail sections of other laws with priority. The primary statutory authorities citied are: the Clean Water Act, as amended in 1977; CERCLA of 1980, commonly called Superfund; the RCRA, as amended in 1984; the Superfund Amendments and Reauthorization Act (SARA) of 1986; the Medical Waste Tracking Act of 1988; the Great Lake Critical Programs Act of 1990; the Clean Air Act of 1990; the Clean Water Act of 1990; and the Housing and Community Development Act of 1992.

On many publications and Web pages on brownfields, the EPA discusses them as "abandoned, idled or underused industrial and commercial facilities where expansion or redevelopment is complicated by real or perceived environmental contamination." This definition facilitates the EPA's Brownfields Economic Redevelopment Initiative in empowering states, communities, and other stakeholders involved in brownfield revitalization to work together on redevelopment.

The definitions of brownfields by states are varied, as are the patterns of implementation. They vary from Missouri, with one of the oldest and most defined brownfields programs; to Washington, with an operating program without a definition; to Alaska, with no definition or program.

Missouri's complex definition of brownfields comes from its 1995 brownfields legislation in Chapter 447 of the Revised Statutes of Missouri (commonly known as the Brownfields Redevelopment Program). The Missouri Department of Natural Resources (MDNR) and the Missouri Department of Economic Development jointly have the following definition. To be a brownfield in Missouri, a project must meet two criteria:

1. All projects must enter and be accepted into the MDNR Voluntary Cleanup Program which provides property owners with oversight of and concurrence with all cleanup activities.
2. A project will be considered eligible if it meets the following criteria:

   (a) The property must have been abandoned for at least three years or underutilized. Real property is underutilized if less than 35% of its commercially usable space is used for its most commercially profitable and economically productive use.

   (b) The property is owned by a local, state, or federal governmental agency, or by a private party who is to the potential responsible party, and the project is endorsed by the local municipality;

   (c) The property is contaminated, or perceived to be contaminated, by a hazardous substance; and,

   (d) The property is planned to undergo redevelopment for a commercial, mixed-use, or industrial use that is expected to create at least 10 jobs or retain at least 25 jobs, or some combination thereof.

The state of Washington discusses brownfields as "the shorthand term for historically contaminated and underutilized or vacant industrial property" on its Web page. In some literature it defines brownfields as "properties that are abandoned or underused because of environmental contamination for past industrial or commercial practices." However, there is not a definition in any state statute or regulation. If the public thinks a site is a brownfield—it is. Nevertheless, Washington maintains an active brownfields program, with a showcase project in Seattle and King County as its model. That project is discussed in the case study below.

ASSOCIATED ISSUES: ENVIRONMENTAL JUSTICE AND BANKING

Many contaminated properties are located in areas such as older urban centers, where a high proportion of the residents are minorities, have low incomes, or do not have English as their first language. These common problems reflect the economic limitations faced by disadvantaged individuals. Therefore, disadvantaged communities must overcome special barriers to effectively advocate for their community interests during the review and permitting of projects with potential environmental impacts. This created environmental injustice. In 1982, Warren County was the poster child for environmental injustice and documented racism. That year, citizens banded together and made the Warren County PCB landfill protest a seminal event for the environmental-justice movement. The North Carolina Environmental Justice Network (NCEJN) was formed, and it became a catalyst that galvanized people of color in the fight for environmental justice. The struggle in Warren County was the spark that lit that national environmental-justice movement in the 1980s.

In its most basic interpretation, environmental justice (EJ) is the principle that all people have the right to be protected from environmental pollution and to receive a fair share of environmental benefits. It combines environmental protection with considerations of nondiscrimination and civil rights. Many organizations have been formed on the model of the NCEJN to support these principles locally. Additionally, governments have provided support through agencies such as the Oregon Governors Environmental Justice Advisory Board.

EJ policies seek to level the playing field by providing disadvantaged communities with technical and organizational support, by providing special scrutiny for proposed projects in EJ communities that might result in significant environmental impact, and by offering incentives for certain desirable types of development, including brownfields. EJ policies go beyond brownfields. However, brownfields are an effective means for advancing EJ principles.

When HUD administers brownfields grants, it has EJ requirements. HUD works with community organizations, the private sector, local and state governments, and other federal agencies to provide equitable reinvestment in communities with fair employment opportunities. Other agencies, such as the Agency for Toxic Substance and Disease Registry's Office of Urban Affairs, are actively involved in issues such as public-health issues that are linked to EJ.

Bankers make lending decisions that affect brownfields. Initially, bankers chose to lend on greenfields, rather than brownfields, because brownfields bring unpredictable expense and liability—this despite the fact that greenfields may be more expensive because of the infrastructure that needs to be built. The unpredictable expense and liability of brownfields came from the wide variability in cleanup and associated legal costs. Without some predictability for cleanup and liability costs, banks were biased toward the more predictable greenfields without infrastructures, like rail connections, sewer, water, electricity, and nearby communities to provide. The EPA's brownfield program brought predictability through its multiagency collaborative approach. This approach, combined with funding, often overcame the reluctance of bankers to fund the development of brownfields. Additionally, bankers rely on technical standards developed by other fields to make decisions. With the engineering standards that have been developed in the 1990s such as the "Standard for Process of Sustainable Brownfields Redevelopment" from the ASTM, banks have reliable technical standards. Finally, banks are members of the community and are positively influenced by brownfields because of the community support and process.

From the success of the brownfield cleanups, an associated movement to clean up grayfields has developed. "Grayfields" are defined as blighted or obsolete buildings sitting on land that is not necessarily contaminated. Grayfields range from aging shopping malls in the suburbs to mining reclamation across the Pennsylvania countryside. Many regions hope to have grayfield successes using some of the partnerships and methods developed by brownfield programs.

SEATTLE AND KING COUNTY CASE STUDY

Case studies are written discussions of a topic containing an applied example of the topic. Case studies are used in legal, business, and environmental studies. There are many case studies to select from, because from 1993 to 2000, the EPA has provided over $250 million in brownfields funding in the form of grants and loans. More that 50 brownfield-related job-training and redevelopment demonstration projects have been funded. Projects have ranged from innovative test pilots for heavily contaminated areas in large cities to small communities with a large brownfields.

The case study below discusses a brownfield showcase community initiative in Seattle, Washington, that uses differing levels of technology and different levels of private–public cooperation at several sites. The Seattle and King County Brownfields Initiative was one of the 11 initially funded under the EPA Brownfields Showcase Communities Initiative. The funding comes through the King County and Seattle Offices of Economic Development and has been renewed because of a track record of successes.

This initiative has two tracks. First, several small businesses have received assistance from the brownfields program that has enabled them to return contaminated industrial properties to productive businesses. Second, area-wide projects have made cleanups more attainable for all businesses under their umbrellas.

One of the businesses receiving funding was an auto-wrecking yard, All City Wrecking, that has been cleaned up and redeveloped as a neighborhood store and gas station. This 2-acre site supported a family-owned auto wrecking yard for 30 years. As the owners neared retirement, they ceased operating their business with the hopes of selling their property. The presence of contamination posed challenges to that sale. The site was contaminated with oil, petroleum products, and heavy metals. The Environmental Extension Service (EES), a contractor under the grant, was able to help this business overcome the difficulties of addressing the contamination.

The EES provided free assistance at every stage of the project. The EES helped the owners properly dispose of liquid wastes on the site, and obtained a local matching grant to defray disposal costs. The EES then assisted in selecting and hiring consultants to perform both the assessment and cleanup on the property, reviewed and interpreted consultant reports for the owners, and made recommendations for how to proceed with assessment and cleanup. Within approximately 8 months, the All City Wrecking site underwent environmental testing, cleanup, and compliance monitoring. This process ended with a "No Further Action" designation by the Washington Department of Ecology and has enabled this property to be sold, redeveloped, and recycled for a new productive use as a neighborhood store and gas station. There were many such cleanups that were facilitated by the umbrella projects described below.

Two wide-ranging projects facilitated the cleanup of all properties in their respective ranges. The largest, Washington's newly established risk-based cleanup standards for total petroleum hydrocarbons (TPH), was statewide. The other project was the localized Duwamish Corridor Groundwater Study. This study characterized the groundwater in a heavily industrial area that has been created with material dredged from the river and washed from the hills and documented that the groundwater was not a drinking-water aquifer. Both of these government efforts had the effect of streamlining projects and reducing the cleanup costs. The more flexible TPH cleanup standards enabled this project to clean the soil up to a commercial, rather than a residential, cleanup level. The groundwater study, funded by King County from state and federal grants, helped to streamline the evaluation and regulatory process for each site. Both reduced the time needed to collect background information on sites, thereby lowering the costs of site evaluation.

This and other brownfield cleanups in Seattle and King County were facilitated by:

1. A 5-year project to improve the science for characterizing and guiding the cleanup of petroleum-contaminated sites statewide. The changes to state law recommended by this project were ecological as part of the revisions to Washington's Model Toxics Control Act Regulation.
2. An interagency project that provided the Duwamish Corridor Groundwater Study of an industrial area by a river that is important to shipping. The area included parts of south Seattle and adjacent King County.
3. The creation of a technical-assistance center (the EES) run by the nonprofit Environmental Coalition of South Seattle (ECOSS), which provides direct, door-to-door assistance to manufacturing and industrial businesses in environmental cleanup and pollution-prevention practices.
4. A revolving loan fund for environmental cleanup for which a partnership among King County, the city of Seattle, the city of Tacoma, and the state of Washington manage the EPA grant money.

CONCLUSION

Brownfield programs are a highly successful phase of environmental cleanups in the United States. The first phase was science-based and regulation-driven cleanups. That phase began in 1976 and continues to this day. Occasionally, these cleanups involve economically viable properties that go right back into use. More often, the cleaned-up sites involve abandoned, idled, or underutilized properties. In those cases, the expansion or redevelopment is complicated because of the potential or confirmed contamination. Therefore, the brownfield approach was added in 1993 as a phase that ideally works with the cleanup and then continues through redevelopment. The two approaches continued concurrently. When the federal tax to fund Superfund cleanups was sunsetted in 1995, the number of cleanups began to decline. As Superfund monies run out, brownfield funding will become more important. Brownfield programs coordinate agency and private-sector interests to work together to create jobs and put abandoned properties back into productive use. Problems may arise when the brownfield cleanups are underfunded, the local economy is weak, or cooperation is not achieved. Despite these obstacles, brownfield pilots and projects have been documenting success stories for over a decade. Brownfields have sparked social economic movements such as EJ and economic revitalization of grayfields.

The next phase of environmental cleanups has not yet arrived. Currently, brownfield programs are active across the United States. Their goal is to have all contaminated sites cleaned cooperatively and put back into use. If cleanup and brownfield sites remain clean and no further sites are created, cleanup programs may work themselves into obsolescence. Related movements like EJ and grayfields begun from brownfields will separate as their goals differ. However, brownfields are likely to remain at a smaller and increasingly more sophisticated level for decades.

REFERENCES

ASTM, Standard for Process of Sustainable Brownfields Redevelopment, E-1984–98, November 10 (1998), published January 1999.
http://discover.npr.org/rundowns/segment.jhtml?wfld=1760130
http://dnr.metrokc.gov/swd/brownfields/demonstration.shtml
http://environment.fhwa.dot.gov/guidebook/vol1/doc7c.pdf
http://state.nj.us/dep/srp/brownfields/bda/
http://stlcin.missouri.org/cerp/brownfields/stlouis.cfm
http://www.atsdr.cdc.gov/OUA/RRCMH/borwnf.htm
http://www.brownfields2003.org/
http://www.ci.chi.il.us/Environment/Brownfields/History.htm
http://www.cpeo.org/lists/brownfields/1997/00000118.htm
http://www.dep.state.pa.us/dep/local_gov/envirodir/toolkit_g.htm
http://www.epa.gov/brownfields/
http://www.epa.gov/brownfields/glossary.htm
http://www.epa.gov/brownfields/html-doc/97aa_fs.htm
http://www.epa.gov/R5Brownfields/
http://www.ecy.wa.gov/biblio/97608.html
http://www.ecy.wa.gov/ecyhome.html
http://www.environews.com/Features/env_justice.htm
http://www.gsa.gov/Portal/gsa/ep/contentView.do?contenteId=10033&contentType=GSA_OVERVIEW
http://www.hmdc.state.ng.us/brownfields/history.html
http://www.hud.gov/offices/cpd/economicdevelopment/programs/bedi/index.cfm
http://www.hud.gov/offices/cpd/economicdevelopment/programs/bedi/index.cfm
http://www.metrokc.gov/exec/news/2000/120500.htm
http://www.nemw.org/brown_stateimpacts.pdf
http://www.nemw.org/cmclean1.htm
http://www.noaanews.noaa.gov/oct1702.html
http://www.planersweb.com/w226.html
International City/County Management Association, Brownfields Blueprints, A Study of the Showcase Communities Initiative, 2000.
United States Environmental Protection Agency, Brownfields, Office of Solid Waste and Emergency Response (5102G), EPA 542-B-97-002.

LEE DORIGAN
King County Department of Natural Resources

C

CHEMICAL EFFECTS: see EFFECTS OF CHEMICALS; AIR POLLUTANT EFFECTS; POLLUTION EFFECTS ON FISH

CHEMICAL TREATMENT: see PHYSICAL AND CHEMICAL TREATMENT OF WASTEWATERS

COAL GASIFICATION PROCESSES

In spite of temporary oil "gluts," elements of a new coal-based synthetic fuels industry are slowly emerging in oil importing nations. This coal conversion activity involves the commercial integration of process and power systems. Overcoming both process engineering and environmental problems will be crucial factors in the development of coal liquefaction and gasification plants. Depending upon project size and complexity, the associated expenditures for the total compliance effort could require multimillion dollar budgeting.

The concept of gasification of coal is not a new one. John Clayton proved conclusively that gas could be obtained from coal in the early 1680s. His initial experiments were observations of the products formed upon heating coal. In the presence of air, heat will invariably be generated by burning a portion of the coal. In order to increase the yield of secondary fuels with higher hydrogen to carbon (H/C) ratio than that of coal, it is required to gasify the coal in the presence of steam and an oxygen containing gas. The products formed during high yield gasification are typically hydrogen, carbon monoxide, and variable amounts of light hydrocarbons, especially methane. Carbon dioxide may be scrubbed from the product. The coal, steam, air mixtures are contacted at temperatures above 700°C in fluidized, entrained flow or moving bed configurations.

Liquefaction of coal may be accomplished by reacting with heavy oil derivative hydrocarbons at temperatures of 400 to 500°C. Contaminants are typically hydrogenated to gases which may be absorbed (sulfur to H_2S, nitrogen to ammonia and oxygen to water).

According to Quig and Granger (1983), a coal conversion facility impacts the environment through the handling of large amounts of coal, and discharges from the conversion process and associated facilities. Also, there will be impacts related to the construction and operation of any large industrial complex. The major health concerns for both occupational and offsite populations include potential exposure to particulates, sulfur compounds, trace elements, aromatic amines, and other nitrogenous compounds and radioactive nuclides.

Considerations of these issues and concerns for this facility will begin with the coal handling facilities. Fugitive dust, consisting mainly of coal fines, is generated by the disturbance of the coal in the unloading, transfer and storage facilities. Particulates can remain airborne and be transported from the site under certain meteorological conditions and therefore must be evaluated in terms of their potential impacts and control mechanisms. Coal pile runoff and coal wetting wastewater contain varying amounts of coal fines and dissolved constituents depending on variables such as rainfall intensity and duration, contact time, coal storage configuration and coal pile sealing techniques. Values of over 2000 mg/l total suspended solids and 10,000 mg/l total dissolved solids have been reported by EPA and TVA for runoff from coal piles. The magnetic separation of metallic materials from the coal during preliminary coal cleaning operations will generate a variable quantity of pyretic solid waste which must be addressed. The coal processing facilities, that is coal grinding and slurry preparation, include controls which minimize the discharges from these operations.

Some of the more important coal gasification processes include those of Texaco, Shell, Dow & British Lurgi. These are carried out at high temperature 600 to 3000°F and high pressure 25 to 80 atmospheres. The most developed process is Cool Water integrated gasification/combined cycle (IGCC) described by Holt (1988) and Spencer *et al.* (1986) which uses a Texaco gasifier. Makansi (1987) compares the performance of various systems. Important emissions data for IGCC projects are presented at the end of the current review.

Additional information is presented below on the status of coal gasification environmental effects. A comparison of the impacts on water streams of various processes is given in Table 1.

Pruschek *et al.* (1995) discusses the removal of pollutants from a coal gasification plant in a more efficient and economical manner than in previous designs by conserving energy in the cleaning sections of the plant. A zinc titanate catalyst is being tested for hot (1000°F) gas cleanup potential at Tampa Electric's 260 MW coal gasification power plant in Lakeland, Fla.

Waste gas emissions are reduced by scrubbing the raw gases leaving the gasifier in an acid gas removal system and converting the H_2S (via a modified Claus process) to sulfur. Sulfur dioxide is thus drastically reduced in the final stack emissions. NO_x levels are reduced by saturating the gas prior to gas turbine combustion (see Spencer 1986) or Makansi (1987). Advances in process efficiency are possible, through the use of a combined cycle configuration and by reducing gasifier energy losses. Figure 1 illustrates the Shell Coal gasification process. The product gas would typically be fired in a combustion turbine followed by an HRSG and a steam turbine (i.e., combined cycle) to complete the IGCC. Heitz (1985) presented data on end uses of various gasifier process streams (see Table 2). The analysis of a typical product gas stream appears in Table 3.

From an economic point of view it is desirable to construct an IGCC in phases, Le *et al.* (1986). In the typical scenario the first phase would be installation of simple cycle gas turbines for peaking power. As of 1989 the maximum single gas turbine output is about 150 MW. In the second phase a heat recovery boiler is used to generate steam for either cogeneration or to power a steam turbine (i.e., ordinary combined cycle). Zaininger Engineering (Lewis, 1988) indicate that there is an optimum time at which the gasifier plant could be added as fuel cost/availability would dictate. Normal combined cycle efficiency can be approximately 50% (LHV) whereas IGCC values range from 37 to 42%. However, new hot gas cleanup processes (such as limestone throwaway or metal oxide catalyst) are being developed which may increase IGCC efficiency to about 48%.

TABLE 1
Coal gasification wastewater concentrations (mg/l, unless noted otherwise). (Adapted from Epstein, 1987)

Component	KILnGAS (Illinois No. 6) Moving Bed	Lurgi Dry Ash (Montana Rosebud) Moving Bed	Lurgi Dry Ash (High-Sulfur Eastern Coal at Sasol) Moving Bed	Lurgi Dry Ash (Lignite at Kosovo) Moving Bed	British Gas-Lurgi Slagger (Pittsburgh No. 8) Moving Bed	Grand Forks Slagger (Lignite) Moving Bed	HYGAS (Illinois No. 6) Fluidized Bed	Texaco (Illinois No. 6) Entrained Flow
Chemical oxygen demand (COD)	4100–6100	21,000–23,000	12,000	20,000	20,000	25,400	4050	1100
Total organic carbon (TOC)	810–1610	—	3500	6000	—	—	—	—
Total phenols	260–660	4200–4400	3800	3000	3000	5100	710	<1
Cyanides and thiocynates	130–300	8–19	<3	80	1150	150	30	50
Total nitrogen	1200–2300	—	—	4300	—	5200	3700	—
Ammonia	840–1700	4000–14,000	7000	3700	3000	—	—	2100
Total sulfur	430–1030	—	950	—	700	—	—	—
Chloride	450–710	40–45	670	—	1500	—	—	3500
Total suspended solids (TSS)	150–700	—	—	—	—	—	—	—
Total dissolved solids (TDS)	1070–2100	1700–4000	—	2000	7000	—	—	2220
Oil and grease	25–340	—	50	900	—	300	—	6
pH	8.3–8.8	8–10	9	9	9	—	—	—
Temp. °F, 100s	—	9–12	9–12	9–12	9–12	—	16–18	24–28

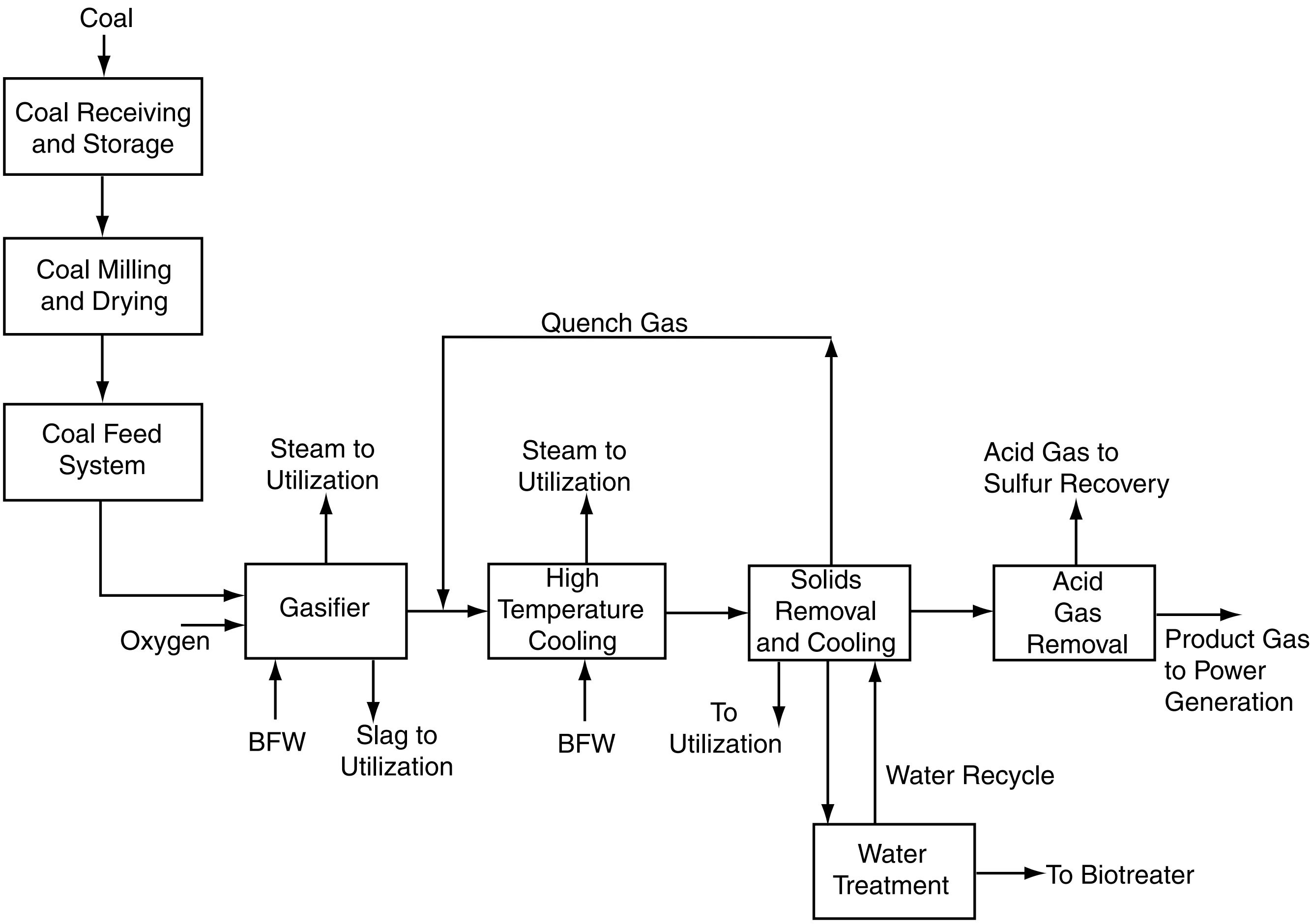


FIGURE 1 Shell coal gasification process block flow diagram for SCGP-1.

The stack emission values presented in Table 4, as per Holt's (1988) review of Cool Water data, are impressively low for a coal-based plant. Sulfur dioxide in the gases leaving the tail gas incinerator is about one fifth the quantity leaving in the turbine exhaust stream. The turbine exhaust did not contain sulfur compounds other than SO_2.

A typical gasification process will generate both solid and liquid wastes. The solid waste retains the natural impurities inherent in the coal such as heavy metals, chlorides as well as organic compounds formed by the combustion of coal in the reactor. Testing, as defined in rules and regulations under Section 3001 of the 1976 Resource Conservation and Recovery Act RCRA, must be conducted to determine the nature of the solid waste generated by the gasification of the particular coal with the specified process. The wastewater stream exiting the carbon-ash scrubbers may contain dissolved and suspended fly ash and unconverted carbon trapped in the scrubber. The treatment of this process wastewater for discharge will generate sludge which will have to be controlled and assessed.

A gasification process may not have any atmospheric emissions during normal operating conditions. Gases are processed and recirculated so that the desired discharge from the system is product gas alone. During upset conditions, however, venting will occur causing releases of gaseous wastes.

TABLE 2
Uses of products, co-products and effluents (from Heitz, 1985)

Product gas	Power generation
Steam	General steam system
Acid gas	Sulfur plant/sulfur sales
Slag and ash	Roads, fill, other products
Runoff and stripped	Biotreater

Sources of non-gasifier liquid wastes which must be evaluated include blowdown from clarifiers and cooling towers, drainage from equipment and floor areas, demineralizer wastes and sanitary wastes. Contaminants in these streams are generally a function of the intake water quality and any chemical additives.

TABLE 3
Typical treated product gas composition. (Adapted from Heitz, 1985 for SCGP-1)

Component	% Volume (N_2 Free)
Hydrogen	30
Carbon monoxide	69
Carbon dioxide	<0.1
Hydrogen sulfide	<0.01
Carbonyl sulfide	<0.01
Hydrogen cyanide	<0.001
Ammonia	<0.001
Hydrogen chloride	<0.001
Hydrogen fluoride	<0.001
Methane	0.03
Water	0.6
Argon	0.2

TABLE 4
Cool water stack emissions. (adapted from Holt 1988)

	HRSG Emissions (Lbs/Million Btu)			
Pollutant	U.S. EPA NSPS	Cool Water Permits*	Cool Water Test Results**	Incinerator combustion products
NO_x	0.60	0.13	0.056	140 ppmv
SO_2	0.24	0.034	0.017	116 ppmv
Particulates	0.03	0.01	0.008	0.18 g/m^3

* Cool Water Permit for Low Sulfur Coal Operation (Low sulfur coal is defined in the permit as coal containing less than 0.7 wt.% sulfur).

** 1987 EPA Performance Test Results for SUFCo Coal, Holt (1988).

The main problems found with sulfur sorbents involve mechanical property degradation and/or loss of sulfur capacity over many sulfidation-regeneration cycles. The sorbents receiving the most attention are all zinc based. Building on the Cool Water technology, various zinc titanate formulations and proprietary materials were developed by the U.S. Department of Energy (DOE)/Morgantown Energy Technology Center (METC) and used at Tampa Electric—Swisher *et al.* (1995). A hot gas pilot scale desulfurization is currently operating at METC. It uses a simulated coal gas mixture at volumetric flow rates of up to 120,000 standard cubic feet per hour and 400 psia and up to 1200°F. Fluidized bed technology is ideal for reactors that continuously circulate reactive and regenerated adsorbent. For further information and updates the reader is referred to the following publication and the DOE websites:

http://www.netl.doe.gov/publications/proceedings/96/96ps/ps_pdf/96ps1_5.pdf, and http://www.netl.doe.gov/publications/proceedings/02/GasCleaning/1.05paper/pdf

One of the world's largest coal gasification plants for electricity generation, a 253 MWe power plant based on the Shell IGCC process, was built and started-up in Buggenum, in the Netherlands, in 1994. Since October 2001, the installation has been owned and operated by NUON Power as a fully commercial electric generating unit. Pulverized coal is gasified at about 400 psi and 2700°F using pure oxygen rather than air. Fly ash bearing raw gas exiting the gasifier is passed through cyclones to remove the larger particles. The remaining fines are collected in a hot gas ceramic candle filter. The filter treats about 1 million cubic feet per hour of syngas at about 500°F and 380 psia. The filter contains tube modules with elements made from a structure of silicon carbide supporting a porous Mullite grain membrane with a pore size of about 4×10^{-4} inch. Future plans include co-gasifying waste materials such as biomass chicken litter, wood products and sewage sludge.

Further cleanup of the syngas involves removal of acid gases such as COS, H_2S and chlorides as well as a process to recover sulfur (see Fig. 1). Also removal of ammonia compounds is included in the water treatment process.

Sources of non-gasifier solid wastes include sludge from raw water treatment and spent bauxite catalyst from the Claus unit, spent cobalt/molybdenum catalyst from the tail gas treatment, silica gel from the air separation plant, and various resins from the water treatment plant.

Sources of non-gasifier atmospheric emissions are cooling tower evaporation and drift, and the gas turbine. The cooling tower releases will cause salt deposition in the surrounding area. The nature and extent must be determined and evaluated using intake water quality data, design information and mathematical modeling. The on-site gas turbine will emit low levels of NO_x, SO_2 and particulates which will have to be considered.

Other impacts associated with any large scale project are those related to the construction of the plant. There will be an increase in traffic by the construction work force on the existing roads. There are potential construction work force impacts related to workers immigration, shortage of certain skilled labor categories, etc. which must be assessed. In addition, there is a potential for noise impacts during the construction or operation of the facility.

The increase in traffic on the roads, rails and rivers near a proposed site due to coal deliveries will create additional demands on the system. The construction of a facility will have unavoidable impacts on the land use at the site.

Construction and operation may impact on the surface water bodies and the potable aquifers in the area but due to the current state of waste water treatment technology, minimal impact on water quality is expected after control measures are applied.

The authors wish to acknowledge the work of Quig and Granger (1983) in an earlier version of this article, some of which appears here.

REFERENCES

Epstein, M., Elec. Pwr. Res. Inst. Jnl. pp. 39–42 Advanced Power System, April/May (1987).

Heitz, W.L., Status of the Shell Coal Gasification Process (SCGP) presented at 5th Annual EPRI Contr. Conference, Palo Alto, CA, Oct. 30 (1985).

Holt, N.A., EPRI Report AP-5931, Oct. (1988).

Le, T.T., J.T. Smith and M.T. Sander, Elec. Pwr. Res. Inst. EPRI Report AP 4395 Jan. (1986).

Lewis, A., EPRI Report AP-5467, Feb. (1988).

Makansi, J., Power, pp. 75–80, Apr. (1987).

Pruschek, R., G. Oeljeklaus and V. Brand, "Combined cycle power plant with integrated coal gasification, CO shift and CO_2 washing." Energy Conversion and Management, June/September 1995, pp. 797–800.

Quig, R.H., *Chem. Eng. Prog.,* **76,** 47–54, March (1980).

Quig, R.H. and T. Granger, *Encycl. of Envir. Sci. and Eng.* Vol. 1, 103–113, (1983).

Spencer, D.F., S.B. Alpert and H.H. Gilman, *Science* **232,** 609–612, May (1986).

Swisher, J.H., J. Yang, and R.P. Gupta, *Ind. & Eng Chem. Research,* Vol. 34, 4463–4471 (1995).

ROBERT J. FARRELL
ExxonMobil

EDWARD N. ZIEGLER
Polytechnic University

COMMUNITY HEALTH

GENERAL

Health is the ability to cope with activities of daily living, ever a norm yet always relative.[1] Most people enjoy and expect good health, boast of their abilities to perform but resent being unhealthy. Health is valued most highly after it has been impaired.

Health and sickness are two contrasting aspects of our world's life. Sickness is a deviation from normal healthy functioning, much like any other system breakdown. Each attracts a legion of helpers ready to aid ailing persons return to their diverse pre-illness states. So it is with a community's health.[2]

Evaluations of community health must consider many factors. In addition to specialized and general medical care, variables include housing availability and ownership, transportation adequacy for work or pleasure, entertainment and recreation, severity of usual weather conditions, educational opportunities, and social and religious factors, among others. Whereas a plethora of medical and surgical options contribute to what are sometimes claimed as miraculous recoveries from a personal illness, environmental engineering quietly strives to create basic healthy living conditions and protect communities from potential physical dangers, widely ranging in severity and in size from small to large.

Environmental health of a community is rated by the degrees of satisfaction that result from engineering works and organized community efforts that improve the physical freedom, comfort, and efficiency of residents. Cities and towns are judged on the numbers and quality of existing recreation opportunities, schools and colleges, and transportation routes, and fiscal services. Yet, while the processes that caused those amenities to be developed are important, the systems and organizations that currently maintain growth or stability must also be carefully noted and weighed. A community's health is not static and must be cared for continually, much like the health of any resident.

For millennia, communities have depended on engineering skills to install, maintain and improve necessities and amenities of life that contribute to the well being of citizens, animals, and plant life. Ancient Crete used conical terra cotta pipes to keep sewage flowing, an early example of the venturi principle, and for centuries massive aqueducts provided ample supplies of fresh water to Roman communities. Excellent roads and transport brought foodstuffs from farms to imperial cities and helped speed commerce and communications between businesses and governments. Cities around the world before and after Rome have turned to engineers to solve urban problems of drainage, flooding, and crowding while at the same time pyramids and temples, cathedrals and castles, universities and markets were being built, all designed to maintain good health and integrity of communities.

Today, efficient land, sea, and air rapid transport networks help deliver farm products promptly to processors, relying heavily on refrigeration to maintain freshness and nutritional quality. Water supplies, a vital element in human health as well as a basic unit of industry, no longer depend on insanitary streams or cisterns but are collected behind large dams, delivered over great distances through well-engineered conduits to filtration plants and there purified, chlorinated and fluoridated for safe consumption. Epidemic diseases like typhoid fever and cholera, amoebiasis, malaria and yellow fever no longer threaten communities in developed nations, thanks to engineering that provides potable water, free from harmful parasites and available for human waste disposal.

Evaluation of any community's health is both quantitative and qualitative. Planners need to know how many hospitals exist within the city limits, where they are in relation to centers of population, and whether transportation for patients, staffs, and visitors is adequate. Numbers of primary and secondary schools, technical training centers and universities, each with details of the students being served are important data in judging community ambience. Local governments want to know how many of the people who work in a city actually reside there, how many residents rent apartments or own their own homes and in what direction these numbers are changing, measures of migration in and out of the jurisdiction.

Businesses and factories that are seeking new locations look closely at pertinent employment rates and skills of available workers that can support general manufacturing, or contribute to growth or modification of an enterprise. Other statistics reflect the fire safety of a community, like the numbers of residential or business fires, annual dollar losses due to conflagrations, or the average response times of fire apparatus to alarms. Law enforcement is rated by numbers and categories of crimes, numbers of crimes solved and the convictions that result. Social health indices would include the numbers of persons on public welfare or assistance, and the numbers and rates of out-of-wedlock school-age pregnancies.

When a community's ability to provide effective sickness care is weighed, the ratios of total population to physicians and other health professions, the number, size, accessibility, and types of hospitals, with attention to emergency services and rehabilitation care are important. Slightly different studies for evaluation of care must be given to rural areas where hospital accessibility, highway networks, and estimated transit times for ambulance services, become critical elements in area's health study.

Residents and visitors, businesses and news media, continually assess the health of a community by studying reports that describe adherence to or departures from desired norms of community living. Governmental publications provide estimates and statistical analyses of regularly collected data about commerce, industry, and banking. Visitors fairly accurately sense or rate a community's ambience by noting the frequency and types of cultural activities for residents or visitors, the numbers and quality of public and private schools, and available recreational facilities. Competent urban planning encourages and supports neighborhoods that have identity and local pride, perhaps with islands of green park land, all served by excellent roads and public mass transit that make for easy access to cultural and recreational areas. Community health depends upon diverse activities, many that result from excellent engineering and some directly related to personal health or sickness care of citizens.

The central official health authority for a community is its public health department, a major health agency under direct control of the mayor or county executive. The health department is established by statute, and its chief health officer is a legal guardian of community health with an authority that includes subpoena power to enforce applicable health codes and regulations. Sizes and complexity of official health agencies range from two or three persons in a town or village to major departments of states and territories, or the national U.S. Department of Health and Human Services (DHHS) and the international World Health Organization (WHO).

Health agencies or departments have a tradition of professional knowledge and compassionate service, two characteristics that usually provide a modest freedom from administrative interventions. This varies somewhat in proportion with budget allocations but it allows health departments a certain flexibility to design their services as needed to carry out community health tasks. Child health care in an affluent suburb will be patterned differently from clinics for expectant mothers and infants and family planning in low income or poverty areas. Centers for treatment of drug abuse and addictions or patients recently discharged from state mental health hospitals need to be close to where those conditions are frequently encountered.

Basic services of most health departments include the epidemiology of infectious diseases; treatment clinics with emphases on sexually transmitted diseases and tuberculosis; family planning services, often a euphemism for contraception but they may also include advice on abortions in the first trimester of a pregnancy; child health care in well-baby clinics where newborn infants and preschoolers are followed closely with childhood immunizations and school health services; pediatric dentistry; services for drug and alcohol abusers who want relief from these personal burdens; and special clinics or services when needed by special communities, such as victims of AIDS or chronic psychiatric ailments. Health education, once a major activity in a local health department, now is carried out by schools and numerous single-focus volunteer organizations, like those for heart disease or cancer. Health promotion, too, has received much attention in the public press and from an increase in general awareness of the advantages of stopping smoking tobacco, weight control, cholesterol levels, and balanced diets.

Environmental control services, a major division of any health department, include restaurant inspection and food services control, with added concerns for area-wide sanitation, housing, industrial hygiene, and animal control. Control and abatement of environmental hazards vary as an emphasis shifts from human health effects to a general community salubriousness, pushed by legislative actions that reflect the relative effectiveness of lobbyists for human health adherents or environmental activists.

Concentrations of population and their needs vary among cities, towns, or counties whose geographies, ethnic compositions, and industrial bases differ sufficiently so that each health unit needs flexibility in designing services. Many health departments have citizen or legislative oversight through various boards of health, and functional tasks vary with the problems that confront the individual jurisdictions. Malaria control, for example, is a prime concern for tropical and some developing countries while the energies of urban health agencies may be heavily invested in tuberculosis detection and treatment. Each is governed by regulations or distinct health codes that have been written to implement laws enacted by respective legislatures. Although occasional conflicts can arise between neighboring communities when confusion or contradiction exist in regulations or information flows, relatively consistent scientific and engineering standards developed by professional societies or governmental experts are available to make health codes more uniform, despite restrictions of jurisdictional boundaries.

Public health responsibilities were once synonymous with local and urban health departments and a few public health schools that were based in major universities. Since about 1960, however, many other community agencies have assumed much of traditional public health's former activities and have blurred the previously clear image of official health departments.

Federal funds to carry out public health programs once were limited to health departments but have been made available as grants to groups such as voluntary associations, hospitals, medical schools, neighborhood associations, and Indian tribes or ethnic groups. Interested community health care provider groups now provide needed auxiliary services such as health education and promotion, medical services to disadvantaged groups, and family planning. They also carry out limited data collection and disease specific epidemiology that support their ongoing research activities. Coordination of numerous separate and isolated studies and consolidation of focused findings can be difficult for local health departments who may not have had a role in the investigations. William Foege, a former director of the Centers for Disease Control, has described the environment for public health practice as changing beyond any predictions and urged support of government for health monitoring systems and interventions.[3]

Private health care has gradually become part of the greater public health system, largely due to change in funding and demands from the public and its elected officers for better quality control of health services. These interests have led to closer supervision of medical services delivery. Licensing boards and related quality assurance organizations have increased their efforts to assure that the professionals who provide health care are fully qualified to practice the disciplines for which they were trained. Terminology like licenses, certificates, or permits tend to be specific for each profession even though a general public may use the names interchangeably. Commonly, physicians and dentists are licensed, nurses are registered, midwives are certified, and physician assistants receive permits from specific state boards.

Each board evaluates the training that its applicants have received and administers a licensing examination that must be passed before permission to practice is awarded. Licenses or certificates can be withdrawn when the specific board has, always by careful legal action, determined that the holder of that license has breached its standards for quality performance. Sanctions of licensees or permit holders vary and may include a written admonition or reprimand, suspension for brief periods, requirement that the professional undergo special training or medical treatment, or even complete withdrawal of the permit to practice. Failure to perform in accordance with accepted standards may result in disciplinary action by hospitals as well, and the results of such decisions are shared with all agencies that have official public responsibility for quality care, consistent with existing standards of confidentiality. Categories of other health care practitioners who must possess licenses vary slightly from state to state but often include dental hygienists, psychologists, social workers, physician assistants, chiropractic, physical therapists, and acupuncturists.

Personal health care services, once known as the "private practice" of physicians and nurses, are gradually shifting from solo practitioners to corporate or group practices. Remuneration has also shifted from personal payments by the patient or a guardian to public payments from sources like Medicare, Medicaid, and insurance programs. Some of the new groups of health providers are independent practice organizations (IPAs) in which physicians are separate providers but relate to a central management group; preferred provider organizations (PPOs) where physicians retain independence but engage in contractual services; and health maintenance organizations (HMOs) in which physicians and other professionals may be contractual employees of a public or private entrepreneurial organization, rather than independent professionals.

Free-standing medical centers provide initial care for relatively minor illnesses or prompt referral of more serious conditions, sometimes known as "urgicenters," have arisen to care for persons who have no regular medical attendant or whose physicians are not immediately available. Free-standing non-governmental medical clinics are often located at vacation resorts or in busy shopping malls and are part of a community's medical resources. These are also known as walk-in clinics or colloquially as "doc in a box" services of various complexity and usually meant to suffice only for short-term care.

Mergers of smaller health care groups create progressively larger corporations which control wide-ranging delivery sites with well-equipped outpatient clinics for initial care and specialty consultations of medical and surgical illnesses, ambulatory surgical centers for one-day surgery, as well as more complex invasive and diagnostic procedures, similar to those provided in hospitals. Pharmacies and laboratories and rehabilitation centers may be part of major health care groups. Special medical units have been created to provide general care services for entire families at university medical centers with their own professional schools, or by major hospital groups or industries whose employees might otherwise lack quality medical care.

Urban and rural US communities have always had cadres of health care providers or "healers" who were considered by traditional licensed physician groups as being less well trained and to whom medical licensure was denied. Some of these care givers in remote or isolated communities were folk medicine healers who had received instruction from older practitioners or even been self-taught. These indigenous healers and a large number of other allied health care providers have been roughly designated as complementary or alternative health care.

Complementary and alternative are general terms applied loosely to more than a hundred or so unrelated healing methods whose adherents believe each is effective. Complementary health care refers to those skills or systems that will support customary medical or surgical care directed by a licensed physician. Acupuncture is a complementary system, an insertion of needles in parts of the body in accord with a complex system to relieve pain or cure ailments.[4] The practice of chiropractic, also, is increasingly accepted as complementary to physician directed care. In contrast, alternative health care is a term to denote a therapy or system that is meant to supplant regular medical care of a physician. Iridology, the diagnosis of human disorders by examining the ocular iris, and naturopathy are two of many alternative care systems.

Homeopathy and chiropractic practitioners, for example, are recognized as healers to be licensed by their own boards in many states, thus subjecting them to a degree of oversight and legal responsibility. Several popular journals now exist in which the various supportive health care modalities are featured. Meanwhile, any of the other alternative care practitioners may be found in local jurisdictions with little or no official supervision or control, unless medical licensure laws are violated or use of a specific method causes harm to a subject under care.[5] Quiet controversy or open disagreement exists about the value of any one system in these two fields, but a federal office has been created in the Department of Health and Human Services (DHHS) to study and evaluate efficacy and safety in complementary or alternative health care practice.

One school of fully qualified healers, the osteopathic physicians, were once considered as having a discipline that was distinctly different from allopathic physicians. Today, osteopathic physicians graduate from osteopathic medical schools that have full governmental certification and provide training of equivalent quality to other U.S. medical schools. Osteopathic physicians are licensed in all states and osteopathic specialty certifications are acceptable for full practice privileges and positions on hospital medical staffs.

Modern socialized health care that began in Germany in 1876 has spread in one form or another to almost all of the world's developed nations, notably excepting the United States of America. Socialized health systems have varied from care that is nearly totally government controlled to that which is delivered by independent practitioners whose patients have freedom to select their caregivers. Payment systems also differ, with mixtures of public pay from government directed services to private pay for personal services to individuals.

In the U.S., mixtures of private or public insured care and public entitled care exists. Health care insurance is available to many workers as part of their employment obtained through union efforts or purchased by employers. Federal Medicare is available to persons over 65 years of age and those who receive Social Service disability payments. Public entitlements include military personnel, certain categories of persons in need, such as pregnant mothers, children, prisoners, and Medicaid for persons receiving governmental subsistence aid.

Federal grants were given to medical schools during WW I to speed the training of medical students and produce the physicians needed by an expanding military establishment. Funding for medical education and research increased regularly during World War II and continued to escalate after hostilities ceased. Medical school enrollments grew and new schools were opened to accommodate the soldiers who were returning to seek new careers. Premedical courses were expanded to provide basic instruction to prepare medical school candidates, and hospital residency programs expanded rapidly to provide training after medical school in the burgeoning medical specialties.

Concurrently, efforts were under way to make sure that patients and entire communities that once suffered tragically from diseases like acute poliomyelitis and chronic tuberculosis would no longer be threatened by these scourges. Protective oral and injectable vaccines have nearly eliminated polio, and effective antituberculosis drugs have rendered sanitoria for very long-term care no longer necessary. The severe late effects of sexually transmitted diseases [STDs], like cardiac complications of syphilis and its severe central nervous system damage, or late complications of gonorrhea that resulted in sterility, pelvic abscesses, or purulent arthritis, are now forestalled by early treatment with antibiotics. Scarlet fever, child-bed fever or puerperal sepsis, rheumatic fever, and erysipelas, streptococcal infections are nearly medical rarities. However, drug-resistant forms of all common bacteria are viewed as major therapeutic dangers, and new diseases like those of the Ebola and Hanta viruses begin to challenge infectious disease specialists, leading to a surging interest in global infectious diseases.

Major advances occurred in basic health care technology and pharmaceutical products manufacturing but also in the engineering skills and sciences that produced computed tomography, magnetic resonance imaging, molecular biology, laparoscopic surgery, and rapid, convenient laboratory methods. A current view of these fields has been described by N.P. Alazraki.[6]

Medical divisions of all major universities and large private hospitals compete vigorously for federal funding for basic and advanced research in health sciences, and insurance reimbursements and private bequests or donations of money to enable the expansion of existing facilities and structures needed for new health technologies and the related technical staffs.

Much controversy exists over rising costs of medical care, specialized equipment, laboratory procedures, and growing numbers of professional personnel. Efforts to remedy the large U.S. medical care system and reduce expenditures met much opposition from organized caregivers and thwarted the efforts of the Clinton administration to carry out sweeping changes in the way that health care would be delivered. Perhaps paradoxically, much attention is now being given to curbing excessive or elaborate care through better organization, shorter hospital stays, living wills and durable powers of attorney that help patients to terminate useless care under predetermined conditions, and even a medical concept of "futility care" that attempts to deal with hopeless cases. Despite many condemnations of modern medical practice that has grown rich, the balance of curative efforts has been positive. Abel Wolman, the famed water and sanitation engineer, once summed up the progress that federal and other monies have supported by saying, "Money is the root of all … good!"

HEALTH AGENCIES

The health department of any political subdivision is the governmental unit that enforces the health code and the sanitary ordinances and regulations that follow. Its director is a governmental officer for health sworn to uphold the laws that relate to health and sanitation and to recommend new measures to counter any recognized dangers to the public's health. No other licensed physician group or hospital or voluntary health agency has this key legal authority to maintain health in a political jurisdiction. Each health department is responsible for guarding the health of its entire assigned geographic and political subdivision, a legal commitment to abate any health hazard that is as broad as the political units to which they belong. Contrasted to the border-to-border obligations of a health officer, the wall-to-wall scope of hospitals is limited to medical events within its buildings or in contractual obligations of outreach programs. Physicians and other healthcare providers are more or less limited in their concerns to person-to-person duties in the care of individual patients.

Official public health services provide some sickness care but are legally charged to assure that all citizens have safe and healthy environments in which to live and work. Health departments must respond to complaints from anyone within their jurisdiction, and their officers are expected to act promptly to abate health dangers. These agencies have a duty to intervene and a right to enter whenever a hazard exists that threatens the community's health. Although actual physical entry can legally be obtained only by permission or by warrant, health departments have long been considered to have a right to enter any dispute that involves community health. In this respect, official health departments are truly safety agencies equal to fire or police departments.

In addition to this unique legal status, the health officer is a member of the central staff or cabinet of the chief elected official and provides advice on health matters. The mayor, county executive, or governor can direct that health officer to begin appropriate corrective actions whenever the public health is endangered. To help in fulfilling statutory and directed functions, health departments have four essential community functions: Epidemiology, disease prevention, needed services, rescue and protection.

EPIDEMIOLOGY

The importance of epidemiology, or the knowledge of a disease that exists in a community, was clearly stated as early as 1879 in a Baltimore City health ordinance that began with this statement: "It shall be the duty of the Commissioner of Health, from time to time to make a circuit of observation to the several parts of the city and its environs" to detect dangers to health and promptly order their correction. This legal charge codified the duty of the health officer personally or through delegated agents to gather information about the healthiness of the community or its inhabitants. Often known as "shoe leather epidemiology," this direction is similar to a military axiom that tells commanders to "go to the sound of the guns," to go where troubles exists. Further, the Baltimore ordinance directed the health commissioner to take immediate steps to correct noxious conditions and to report to the mayor with advice on further actions needed to maintain a healthy city.

All of these steps are a part of epidemiology, or wisdom "among a people." Health departments collect information about communicable diseases, analyze reports of births and deaths, follow trends in accidental injuries and childhood development, and provide annual summaries of the health of a community. Possibly the earliest example of modern measurement of health statistics occurred in 1882 when vital records of Baltimore, MD, and Washington, DC, were analyzed by sensing punch cards. John Shaw Billings, who would later design the Johns Hopkins Hospital and the New York Public Library, asked an electrical engineer, Herman, Hollerith, to devise a means of electrically sensing and counting the holes in a series of punch cards that contained census information about the respective populations. This venture was successful and the information obtained is listed in the annual reports of the 1880 U.S. census for the two cities. Gathering health facts and analyzing them is such a distinctive health department function that a well-known national health officer once urged that they be named "Departments of Epidemiology" to emphasize this importance.[7]

The discipline of epidemiology determines and analyzes the distribution and dynamics, or changing characteristics, of diseases in human populations. Physical health in a community is measured inversely by the incidence of a disease or injury over time or its prevalence at any given moment or time period.[8] The lower the incidence or prevalence, the better the community's epidemiologic health. Health officers utilize epidemiology to determine where, how and why disease exists in their communities. Other measures of the impact of diseases on community health are found in statistics of morbidity and mortality, as well as a disease intensity in a neighborhood, where a disease is spreading, and any deaths as noted by gender, age, and race. Epidemiology is a basic tool for allocation of departmental protective and preventive resources and services.

Health departments gather and analyze data supplied in reports to central agencies received from physicians and health care workers who are required by law to so report on official forms each occurrence of certain human diseases, at least 52 of which must be reported to the U.S. Centers for Disease control. Reporting enables health officials to deploy public health workers and resources who can act to limit the spread of infection to susceptible inhabitants. A health officer can require that patients with certain communicable diseases can be isolated in a single household, or order a community-wide investigation of food poisoning or a measles outbreak in local schools that is preliminary to focused prevention. When spread of a disease can be minimized by vaccination, localized or mass specific inoculation programs are offered to all persons at risk, some of whom may also require specific antibiotic therapy as prophylactic against infection.

DISEASE PREVENTION

The first case of an infectious disease, the "index case," is the well-spring of epidemiology (Sartwell). Certain diseases spread rapidly in a community and cause severe economic loss if workers are incapacitated or children are disabled by severe complications. Infections like measles and rubella (German measles), are no longer accepted as unavoidable risks to children but have been markedly reduced by protective vaccines. Poliomyelitis has nearly been eliminated in the Americas and small pox has in fact been eliminated from the world by mass immunizations.[9] Epidemic cholera appeared in Peru in the last decade and spread through neighboring countries, and Lyme disease was identified several decades ago in New England; recent findings show that it also has a world-wide distribution. Diseases are spread easily by tourists and business people who return to homes by air, unwittingly carrying early infections of exotic diseases. These may include malaria, typhus, hemorrhagic dengue fever, and others, many of which require prompt public health attention to minimize or eliminate transmission to U.S. residents.

An initial report of a case of tuberculosis, for example, will require that all persons who have contacted the patient be skin tested (tuberculin test) several times for evidence of existing tuberculosis; those persons who test positive must be further examined to determine if active or progressive disease is present. Prophylactic treatment with isoniazid can be safely offered to persons with positive skin tests if they are under the age of 35 or 40 years, but full therapeutic courses of antibiotic drugs are recommended for those with clearly evident pulmonary or extra-pulmonary infection. Some strains of tuberculosis germs (mycobacteria) have arisen that are resistant to several of the four customary therapeutic drugs. The specific type of mycobacterium responsible for a patient's disease must be determined so that the initial drug regime will be effective from the beginning, not after months of partially effective therapy have elapsed and the infection has not been controlled. To prevent the spread of drug resistant mycobacteria, initial courses of multiple antituberculosis drugs are individually administered by nurses who see to it at each scheduled dosing that an infected patient ingests the total amount of drugs that have been prescribed. This mode of anti-tuberculosis therapy is "Directly Observed Treatment" (DOT) and, while costly in staff time, it has resulted in effective control of a dangerous disease. In many political jurisdictions, when a patient with tuberculosis is judged to be a public danger because of failure to cooperate in treatment and control of the spread of infection, public health officers can request courts to incarcerate that patient so that treatment can be carried out under close supervision to protect the public.

Reported tuberculosis has increased as a complication of the human immunodeficiency virus (HIV), the causative agent of the AIDS epidemic. Because of this, the Centers for Disease Control (CDC) in Atlanta have advised state and local health departments to tuberculin skin test all persons with HIV infection.[10] Conversely, persons with tuberculosis are urged to be tested for HIV, and the registries of patients with these two diseases should be matched at least annually.[11]

When many cases of enteric diseases are reported, either localized as endemics or widespread as epidemics, health departments mount intensive searches to determine the nature and source of infections. Control measures to prevent further spread can include mass vaccination, provision of safe food supplies and potable water, official quarantine or isolation of sick persons and their excreta, evacuation of critically ill persons for definitive therapy, and vigorous public information programs to help uninfected persons at risk take appropriate steps to avoid the disease.[12]

Infectious disease outbreaks can result in heavy loads of sick or injured persons that may overwhelm community resources and require help from allied governmental or voluntary health providers for the short term as in disasters or damage to water supplies, or over the long term to remedy the ill effects of continued poverty or devastation.

Rats and mice, the animals chiefly regarded as signs of environmental deterioration, easily find homes and food in the debris customarily associated with poverty housing. Rat urine spreads leptospirosis, an infection that results in jaundice and occasionally renal impairment or aseptic meningitis, but small children who must live in rat infested neighborhoods are also at risk of rat bites. Major control projects strive to teach communities to deny rats an access to food in accumulated garbage, and eliminate easy harborage in piles of casually discarded trash or poorly maintained, dilapidated residential structures. Rats have proved to be courageous adversaries and survive despite major attempts to eliminate them with poisons, traps, gases, high frequency sound, and education of human residents. Yet, the dangers from rat bites are intolerable and health departments continue to war on rat populations.

ENVIRONMENTAL HAZARDS

Huge plumes of dark smoke billowing from industrial stacks was once a sign of a community's prosperity, but no longer. Oil and electricity heat have replaced coal as fuels for residences, industry, locomotives, and ships. Municipalities have banned or severely curbed the operations of home and apartment incinerators, and those that remain must have stack scrubbers and other equipment to drastically reduce emissions of particulates into the atmosphere. Half a century ago, the dark, gray lungs of city dwellers that resulted from inhalation of coal dusts and other particulates were easily distinguished at autopsies from the pinkish gray ones of rural citizens. Funded by federal grants, states are now well equipped to measure gaseous atmospheric pollutants and gather the data needed to support limiting toxic gaseous emissions by motor vehicles and major industries, and human pulmonary health has markedly improved.

Control of hazardous materials (HAZMATs) may be divided between health and other agencies yet emphasize safe storage and transport to protect neighborhoods and safe usages for workers. Local fire departments and state fire marshals or departments of environment are charged with the responsibility for containing spills and subsequent decontamination. Fire officials, for example, at all times carry handbooks and catalogs of toxic substances in their vehicles, and have been trained in the appropriate responses to hazardous spills. Frequent reconnaissance inspections of known locations where hazardous materials are stored helps to insure safe management of toxic materials. Departments of public works, health, and police may also be called upon to assist in management of spills.

Official health agencies now participate as consultants for community health in air pollution management, or advisors on the health aspects of building construction when toxic substances are used incorrectly. When lines of authority are not clear, duplication of municipal or county services and conflicting regulations may result in failure to respond promptly to early warning signs of environmental hazards. Health officials may be called after a chemical spill has been abated to counsel communities about possible delayed hazardous effects on humans.

Health sanitarians and technical inspectors of other agencies investigate complaints made by workers, nearby residents, or the general public about unsafe workplace environments. Officials inspect sites where alleged infractions of regulations have occurred, interview complainants and workers, and take samples or measurements of gases, dusts, or other possible toxic agents. Corrections may be made on the scene, or the respective departments will initiate official regulatory actions or seek legal sanctions. Airborne particles of asbestos, silica, metallic dusts, chemical gases or fumes, and organic fibers can pose dangers to workers or the general public and must be promptly abated.

Once widely used in fireproofing, building construction, and insulation, asbestos, if inhaled, results in fibrosis or chronic scarring of lungs with moderate to severe respiratory disability. Late effects include cancer (mesothelioma) of the lining membranes of the pleura (chest) or the peritoneum (abdominal wall) and is uniformly fatal. Abatement of existing asbestos-containing material from limited building areas that do not involve major razing must be sealed off to prevent dusts from spreading to unaffected areas; workers engaged in asbestos removal must be carefully garbed in protective clothing and equipped with approved protective respiratory equipment. All asbestos demolition activities, and the debris containing asbestos must be handled with strict attention to federal and local laws and regulations.

Inhaled coal dusts result in coal miner pneumoconiosis or black lung disease, a disabling condition more often seen in mines producing anthracite or hard coal. Unprotected inhalations of dusts from the floors of factories that produce pottery or electrical insulators has resulted in silicosis, a severe pulmonary fibrosis that may be associated with recurrences of old pulmonary tuberculosis.

Control of known toxic volatile inhalants is also spread among several agencies at all governmental levels, occasionally with no clear delineation of regulatory boundaries or with control and enforcement distributed among several departments. Urged by insurance companies and pressured by legal actions to improve worker and community safety, industries strive to provide healthy workplace environments and still achieve satisfactory manufacturing profits.

Federal funding remains for lead paint poisoning prevention, rodent control in cities, water fluoridation, air pollution control, protection against hazardous materials, and staff education or training. The blood lead level that is considered to be toxic in children and adults continues to decrease, but the problems of plumbism (lead poisoning) associated with chipped and worn lead paint in homes persist. Although paints containing lead were banned in Baltimore city by an ordinance more than half a century ago, walls and woodwork of old homes retain lead paints under successive covers of new lead-free paint. Nevertheless, children continue to suffer from plumbism, albeit lower levels, because they inhale lead dusts in homes or eat lead paint chips that fall from old painted surfaces. Adults who are unaware of the toxic effects of lead paint dust may suffer from a more acute form of plumbism when they fail to wear protective masks of industrial-quality in de-leading homes.

To correct environmental hazards, officials benefit from epidemiologic data to locate where toxic wastes have been dumped and to measure the effects of toxic substances on community health. After chemically damaged land has been identified, reclamation or redevelopment as prime commercial and residential sites is often difficult without costly decontamination. States have established official registries of toxic substances that include listings of associated cancer cases. Toxic substance registries are valuable in the long run but high initial costs, operational difficulties, and challenges from local industries have hampered their development.

Air pollutants may merely annoy residents in a neighborhood because of unpleasant odors but, when residents suspect that pollution can be hazardous, high levels of general community anxiety can result. Persons who work or dwell in a polluted area cannot avoid breathing the air that may cause acute or chronic respiratory difficulties. When pollutants are known to have serious short-term or long-term effects, citizens or workers demand that regulatory action be taken to reduce or eliminate the offending pollutants. However, if threats of job losses occur should the cost of controls cause major industries to move away, low level pollution may be tolerated. Detection, accurate measurement and clearly defined pathologic effects of fumes and dusts on human and animal health have been recorded, as well as toxic damage to fabrics or delicate machinery and other personal or industrial equipment.

Air pollution from industrial processes or solvents used in manufacturing is closely monitored by most industries but small, non-union or inexperienced producers may be unaware that hazards exist in a workplace. In some cases, managers will ignore or defy governmental regulatory control. Health department inspection staffs may be too small to carry out regular inspections of the thousands of small industries in their jurisdictions and complaints from citizens or workers may be needed to give the earliest warnings of serious small-factory air pollutions. Health inspectors may be called upon to test and measure for toxicity of fumes in workplaces from chemical processes, spills or misuse of dangerous mixtures, or emanations from treated fabrics, compressed wood products, and other finished products. Local health departments, however, are no longer responsible for water supply, sewage treatment, and trash and garbage disposal.

The emphases of governmental air pollution control efforts range from entire geographic areas, such as area-wide ambient levels of industrial stack pollutants and motor vehicle emissions, to residences that may be contaminated with radon emissions from soils or rocks. Although state or federal governments are responsible for monitoring large area contaminations, toxic emissions control in local industries is often a community task.

Water pollution results when surface contaminants from industry spills or runoffs from soil fertilizers and pest control agents enter streams and reservoirs or filter into an area's aquifer and wells. Water filtration plants carefully monitor bacterial and chemical contaminants of inflows and outflows to assure that pleasantly tasting potable water and low mineral content is delivered to a community. This is a major task when a downstream community depends on river water to supply its needs while upstream another community dumps its sewerage and industrial wastes. Water potability and safety must meet national standards but water qualities for drinking, food preparation, and industrial uses may vary widely. Adequate water supplies are vital to community health, recreation and cleanliness, but also to industry and related employment. Carefully controlled use of low mineral content non-potable water can be used carefully for garden plots and residential lawns and general use in irrigation technologies for agriculture.

National dental societies have urged that minimum levels of fluoride be added to residential water supplies because of the protective action of this chemical against tooth decay. However, opposition has arisen in some communities because high fluoride levels may cause annoying dental discoloration. Effects of chlorination, also, has been reported to be remotely associated with malignant conditions in humans but the net value of better control of intestinal diseases is believed to far outweigh any possible small toxic effect of chlorinated water. Water with high mineral content, known as “hard water,” even though potable, may require water softening equipment in homes to make it satisfactory for laundry and food preparation or manufacturing.

Anti-smoking campaigns have emphasized documented tragic ill effects from tobacco smoke produced by burning tobacco products, both for the smoker and for persons who are breathing “second hand smoke,” a mixture of exhaled smokes and that which rises from the tips of lit cigarettes. Increased mortality from lung cancer, emphysema, and cardiovascular disease has been thoroughly documented and other cancers of the gastrointestinal and genitourinary tracts are believed to result in part from the swallowing of saliva that contains nicotine and other constituents of tobacco smoke, and the urinary excretion of tobacco-associated chemicals through those channels. Bans on indoor use of smoking tobacco have been promulgated by many local and state governments to protect non-smokers from immediate or long-term effects of noxious chemicals in tobacco smoke. Existing efforts to restrict the purchase of tobacco products to adults and prevent sales to adolescents have been increased. So-called smokeless tobacco (snuff) has been shown to cause precancerous lesions of oral structures and ultimately to lead to localized cancer, even in young adults.

Carbon monoxide (CO) emissions are highly toxic because the CO radical bonds tightly to hemoglobin in circulating blood cells, markedly reducing their oxygen-carrying capacity and resulting in serious impairment of cerebral or cardiovascular function and death. Poisonings have occurred from CO in motor vehicle exhausts in home garages or poorly maintained underground parking garages with deficient ventilation, in homes that use fossil fuels for heating when flues have been obstructed, and when unvented space heaters are used in small, closed spaces and room oxygen has been markedly diminished or exhausted. Poorly maintained mufflers or exhaust systems may leak CO into automobile or passenger compartments at any time, even when driving through traffic. Slow and quiet development of this highly toxic gas can result in unrecognized but fatal levels of CO in vacation cabins, recreational boating, and private airplanes.

Radon is a radioactive gaseous chemical element formed as a first product in the atomic disintegration of radium. Rather frequently encountered in homes of what is known as the “Reading Prong,” a geographic area around the city in Pennsylvania, radon levels vary from home to home. This radioactive gas may be an environmental hazard when high levels exist in confined human habitations that are surrounded by radioactive soils or rocks. Ambient atmospheric levels can usually be reduced to safe levels by adequate ventilation.

Additional disease prevention activities of community health departments include sanitary control of food production and distribution, large-scale childhood vaccination programs, strict quarantine or relative isolation of persons with infectious diseases, and animal control to minimize animal bites or rabies and other zoönoses (diseases spread by animals to humans). In Baltimore, for example, when animal control was transferred to the health department and the collection of all stray dogs was emphasized, reported animal bites of humans decreased from over 8000 to under 1500 in a period of four years. Feral animals are dangers to all humans, to each other, as well as their excreta being a major source of area insanitation. Increasingly, local laws provide improved community protection by requiring that animals be leashed and that owners collect and dispose of animal feces in a sanitary fashion.

NEEDED SERVICES

Governments, large and small, provide health care services in many bureaus other than the official health department, and often with little or no coordination. For example, a county may operate jail medical care, a personnel system for employees, fire fighters and police, school health programs, health units in a department of social service, care for indigent elderly, ambulance services, a public general hospital, and more. The totality of community health services operated by a government is generally poorly comprehended, frequently underestimated and understated.

Rescue and protection require health departments to work with police, fire and other rescue agencies where human health is endangered. When local sickness care units refuse to care for ill or injured persons, for whatever reasons, health departments are expected to supply medical care. Because treatment of homeless or abandoned indigent persons can be costly for diseases such as tuberculosis or complex conditions such as AIDS (acquired immunodeficiency syndrome), these tasks may fall to health departments when other health providers fail to meet their obligations or are overwhelmed by patient loads of epidemics.

Official health departments in cities and counties operate personal health clinics for special populations. Although these services vary with local needs, they have been categorized by one local health officer as traditionally treating “unwashed patients with dirty diseases, who live in hard to reach places and can’t pay.”[13] These groups of patients constitute indigent or needy citizens with ailments like tuberculosis, leprosy, sexually transmitted diseases, or AIDS.

Major health problems that confront community health departments include teen-age pregnancies and inadequate child health care, homicides that are highest among young males, substance abuse (drugs and alcohol) that lead to severe personal and community deterioration and high crime rates in any community, and AIDS with its costly and complex terminal care. The total number of cases since first reporting in 1981 of AIDS worldwide in 1995 was 436,000, with 295,493 deaths. In one state alone, namely Maryland, the total number of cases was 13,082 since 1981, with 7,507 deaths. This epidemic of a terrible new disease has been devastating. The disease is transmitted via body fluids, in heterosexual as well as homosexual contacts, by inadequately screened blood transfusions, and by the use of inadequately sterilized parenteral equipment such as shared needles in illicit drug use.

An office of the state medical examiner, often incorrectly identified as the coroner, is a public health agency that is closely allied with the judiciary. Directed by a physician who is a specialist in forensic pathology, this department operates a morgue to which is taken a person who has been found dead under suspicious circumstances or who died without medical attention. A medical examiner performs an autopsy to determine the cause of death, and may be called upon to testify to this fact in criminal cases. A medical examiner should have immediately available a complete forensic or criminal laboratory to examine human tissues, clothing stains, and body fluids that might relate to or explain a crime. The skills of these pathologists are often called upon to identify bodies from comparison of oral structures to dental records, or even by sending specimens or entire parts of a body to federal crime laboratories for analysis. Although often a grisly business, a prime function of medical examiners is to assure that justice is served and diverse community concerns are assuaged.

Public general hospitals (PGHs) are owned by the political subdivision and have major fiscal support from any of several governments. The PGH reports directly to the chief executive officer (CEO) of the respective jurisdiction or the hospital is supervised by an appointed or elected board. The future of this once important resource of a city or county medical system is in jeopardy, threatened by growth of voluntary, not-for-profit, and profit-making hospitals. Funding equivalent to that of private hospitals has not readily reached public city and county hospitals perhaps for many reasons. It may have been that inexperienced or inattentive elected officials, medically unskilled appointed governing boards, serious budgetary limitations or other critical but unmet needs have all been responsible for the decrease in public hospital care efforts. Whatever the reasons for poor support of these public hospitals that were often the major care centers for immediate and long-term care of needy citizens, many of them have closed and some have been sold to voluntary groups or medical schools. Some have simply been closed, leaving communities with inadequate sickness care.

Others remained in central city locations, inadequately funded, surrounded by blighted areas and required to serve large populations with major health problems. These hospitals are still unable to accomplish easily their missions of helping critically and chronically ill needy residents. Often located in population centers where per capita income and insurance payments are low, public general hospitals serve patients who, for whatever reasons can find no other medical resource. Their populations also present with severe illnesses, chronic conditions, and with multiple complications of various addictions, poverty lifestyles, and AIDS. Obsolescent or convoluted governmental policies set by local, state and federal statutes can hamper smooth operations of these municipal or county hospitals. Employment practices, budgeting, purchasing and plant maintenance may be insensitive to the needs of hospitals. Medical care demands rapid response by all persons involved if services must adapt to changing professional practices.

For better patient care, larger PGHs may be linked with nearby professional schools to help train graduate physicians, medical students and other health care workers. In this exchange, the teaching and research staffs of medical schools bring with them a bevy of talented professionals who supervise patient care and prepare physicians for a life of learning and service. It remains to be seen whether the fresher and better funded investor-owned and corporately managed HMOs and hospitals will find it possible before the end of the century to assume a sizable portion of the heavy load of poverty patients with complex problems, and still make the profits demanded by their stockholders.

Hospitals in the U.S. struggle to maintain fiscal stability in a health care system that is undergoing rapid change. Some may not easily or willingly accept indigent patients suffering from complex medical and social problems. Maryland and several other states in the USA, however, have enacted legal requirements to ease fiscal burdens of hospitals by mandating that private and public patient care costs be shared by all hospitals and all payers in an equitable fashion.

Fiscal support for any general hospital is derived in part from local taxes, federal grants for disease management, and private philanthropy for specified tasks. Considerable income is also generated by fees for service charged to and paid by Medicare (elderly persons), Medicaid (indigent patients), payments from Blue Cross and other insurance companies, and from direct self-payments.

Any community that operates a jail or prison faces a growing need to provide quality medical care to its prisoners. Penal systems, once condemned for their inattention to the medical needs of inmates, have instituted policies to expand services while shortening sick-call lines, improve staff and prisoner morale, and reduce risky transport of prisoners to hospitals for medical or surgical consultations. Large jail populations, also regarded as a class of "regulated communities," include inmates who suffer from drug abuse, alcoholism, mental illness and behavioral disorders. Some jails have created obstetrical and infant care facilities to serve female populations, or special care units for geriatric prisoners.

Designers of new jails must also consider special handling of two major illnesses among prisoners when health units are being planned—pulmonary tuberculosis and AIDS (acquired immunodeficiency syndrome). Patients with tuberculosis infections may ascribe a chronic cough to smoking, or be nearly free of symptoms yet spread the infection to other inmates via sputum droplets that contain live tuberculosis bacteria. The customary elements (volume, rate of change, direction, etc.) in ventilation and air flow in prison cells, hospital suites, recreation and dining areas, need careful attention to minimize the spread of tuberculosis and other respiratory infections from patients with unrecognized disease to other inmates. Health suites should provide for temporary isolation of tuberculosis patients in the early phase of therapy. Longer term isolation may be required rarely but can be needed if the disease is advanced. Moreover, some tuberculosis infections are being increasingly recognized as due to "multiple drug resistant" (MDR) bacilli and the usual anti-tuberculosis drug regimes are not effective. New and expensive drugs, carefully administered, are needed to achieve successful treatment of stubborn MDR infections.

Although AIDS is believed transmissible only through body fluids like blood or semen, isolation may be needed to provide special care for persons who are critically ill with AIDS. These patients are at high risk of severe illnesses from "opportunistic" infections due to organisms that opportunistically take advantage of AIDS patients whose resistance to infections is compromised. These organisms may be bacteria, viruses, or even fungi and other parasites that are normally part of a prison environment. Management of patients with advanced HIV disease can be difficult and requires that the health staff be aware of the continuing advances in therapy that may be available. Medical or administrative protective isolation may be needed to remove AIDS patients from cell blocks where they might be exposed to physical danger from other inmates.

Health services for governmental employees are similar to those of industry and vary in size and complexity with the jurisdiction served. Pre-employment examinations and on-the-job injury care follow customary standards, and worker's compensation claims are processed in accord with state or federal regulations. Medical advisory programs (MAPs) for employees whose on-the-job performances are thought to be due to emotional disorder or substance abuse helps guide affected workers to appropriate care and permit continued employment. If existing buildings are to be altered for use as employee health clinics to include a wide range of health services, special attention must be given to shielding staff workers from x-radiation. A surgical suite should be available for initial care and follow-up of injured workers, and adequate soundproofing provided for rooms where psychiatric counseling is to be delivered.

RESCUE AND PROTECTION

Ambulance services have become an important part of any medical care system, with marked expansion after the mid 1970s. Although turn-of-the century horse-drawn ambulances had long been replaced by hospital-based motor vehicles, only major cities provided this vital emergency transport. The city of Baltimore, for example, had provided a crude transport that used canvas litters in police patrol wagons (paddy wagons) until 1928. At that time several limousine-type ambulances were purchased to be based in and operated by the Fire Department. Fire fighters who volunteered for this duty received basic first aid training and were assigned permanently to this service for better patient care of citizens who could not afford a private ambulance. Commercial ambulance services, many being operated by funeral directors who used converted hearses and personnel with little or no training in patient care, were not supervised or licensed until federal funding was assured with the passage of Medicare in 1965. Local governments developed regulations to assure that patients being transported would have emergency care immediately available and that they would be secure. Regulations specified the training that should prepare ambulance attendants, as well as the medications and patient care equipment to be carried in ambulance vehicles.

Today, municipal or county ambulance services are usually based in fire departments, to take advantage of a communication network that is already in place. Fire houses provide ambulance bases that are open around the clock, staffed by fire fighters who are a uniformed and disciplined corps, and who have a tradition of rapid response to emergencies. Air ambulance services using helicopters are operated by state police departments, in the absence of functioning fire departments. Some local governments may contract with private ambulance companies to provide both emergency and routine medical transport, and hospitals may also be used both as bases and to provide quick and ready access to medical critical care specialists for urgent consultations.

An ambulance is a sick-room on wheels for urgent care. Whereas once these emergency vehicles and staff operated simply in a "scoop and run" format, providing rapid transit for critically ill or injured persons, today's ambulances have state-of-the-art design and equipment, and their crews have expert training in the provision of immediate and effective responses to life-threatening conditions. Communities have quickly recognized the value of highly trained attendants who can stabilize critically ill or injured persons prior to careful transport to hospital centers for definitive care. Radio two-way communications for voice and electronic equipment, like electrocardiographs, enable attendants to administer initial medical care beyond usual first aid, and provide hospitals with information that helps them have ready any special equipment or staff for prompt attention on arrival of the ambulance.

Ambulance attendants, or emergency medical technicians, ambulance (EMT-As), are the first-responders in any emergency transport system. To be designated an EMT-A, each applicant must undergo special training, pass a stringent examination, and be certified by an official body of the local or state jurisdiction before being permitted to care for patients with medically emergent conditions. Ambulances dedicated to critical emergency care are stocked with approved medications for use by EMT-As in the field. Technical equipment is readily available to reestablish damaged airways or stop dangerously irregular heart action by electrical defibrillation,[14] immobilize fractured bones, or provide intravenous fluids. All ambulance equipment and EMT-A training is approved by official boards of physicians and other health care professionals; it is used in accord with tested protocols and the support of hospital physicians via radio communication when needed.

Helicopter emergency transport of patients, first used by several municipal ambulance systems in the early 1960s, was proven of value in military use. Helicopters and emergency technicians provide rapid patient evacuation over difficult terrain or locations that wheeled vehicles can reach only with difficulty. Helicopters are dispatched from a central location and radio communication provides directions to the helicopter pilot and continuous medical consultation for the medical attendants. As in any other ambulance, expert patient care is provided both initially on the ground and then in the air en route to a designated hospital center, thus avoiding rough highway carries. Vital data for on-going patient care is supplied by electrocardiograms (ECGs) and other clinical information is transmitted directly to a base hospital trauma center where a skilled physician is in direct voice radio contact and guides the prompt resuscitation of patients. Additional electronic support for advanced mobile sick care, including image transmission and limited or special laboratory services can be expected to be developed in this decade.

Regional hospital-based trauma centers receive casualties according to protocols that have been developed by regional emergency medical service committees. Each is staffed and equipped to provide complete skilled care for critically injured or severely ill patients. Trauma centers are regulated by a state agency and based at hospitals after a careful selection process that considers need, existing transport networks, and available staffing. Each must agree to maintain superior medical and surgical staff capability, and be ready to receive patients with urgent medical problems or extensive injuries around the clock. Critical care support must be extended in hospitals allied with the trauma center to provide patients needed long-term care through expected periods of recovery and initial rehabilitation. A trauma center concept best fits hospitals that are closely allied with medical schools, and it relieves small community hospitals of legal and economic burdens of severe trauma. Trauma centers may also accept urgent transfers from efficient community hospitals when severe injuries have been stabilized sufficiently for transport.

A community hospital need not operate an emergency care department but, if it does, all persons seeking medical, surgical, or psychiatric help must be accepted. Each patient must receive prompt and appropriate treatment and disposition, even if continued care is accomplished elsewhere. Although residents of a community may view a nearby emergency department as a walk-in clinic for the management of minor ailments, each of these special departments must be fully equipped and adequately staffed for critical care conditions. Physicians should be qualified for advanced life support care (ALS) and the nurses for basic life support (BLS), as a minimum. Emergency physicians preferably are certified by an appropriate emergency specialty board. The high costs of hospital emergency care is offset by assured payments for a wide range of costly services to patients with full insurance coverage who are admitted.

COMPUTERS AND TELEMEDICINE

Computer storage and transmission of information in clinical practice has lagged behind industrial usage, largely due to the confidential nature of patient data. A relative inadequate computer literacy of older, experienced physicians who must learn to enter data according to prescribed formats may account for some continued reluctance in the acceptance of electronic clinical data systems. Despite more general availability of more compact personal computers and electronic bookkeeping systems now installed in the offices of most physicians and clinics, maintaining patient records in computerized systems has advanced slowly.

In 1991, the Institute of Medicine (IOM) released a report that urged adoption of a computer-based patient record as standard medical practice in the United States.[15] In 1995, Beverly Woodward discussed many aspects of computer use in medical care.[16] Attempts to design algorithms that replicate complex thought processes a trained physician uses to probe the history of a patient's illness or solve a clinical problem have not been uniformly commercially successful. Generation, storage, and easy retrieval of clinical records expose the data to improper use by persons who are not entitled to confidential information. Data that include treatment for AIDS or psychiatric illnesses, for example, are particularly sensitive because of their possible misuse by employers, insurance companies, or legal systems.

Computers have proven useful in hospitals to store and have readily available cumulative information that records hourly and daily patient care data and provides easy review of a patient's progress. But start-up costs are high and training of hospital staffs includes nursing personnel, administrators, and other care givers, such as visiting personal physicians and consultants. Lindberg and Humphreys of the National Library of Medicine (NLM) have discussed other uses for computers in medicine that are now being explored.[17]

Boards in each state that license physicians and national professional specialty societies regularly use computers to store data and track careers of individual physicians and other regulated health care professionals. State licensing boards that need to research a physician's application for medical licensure can access information in the National Practitioner Data Bank that briefly describes pertinent disciplinary experiences in any of the 50 states.[18] National specialty societies have computerized the credits earned in cumulative continuing medical education (CME), critical information that each member can access when applying for license or specialty certification renewals.

Telemedicine and related modem technology provide access through e-mail or the World Wide Web to specialists and the National Library of Medicine in Bethesda, MD for instant consultations, making it possible for distant and isolated communities quickly to obtain needed expert medical and surgical advice and guidance.

Telemedicine is the transmission of laboratory information, radiologic interpretations, and clinical information, including actual examination of patients, by audio-visual communication via fiber optic telephone line or satellite transmission between isolated areas and medical specialists in urban teaching hospitals. Begun in Scotland in the 1960s to provide medical consultations to explorer units in Antarctica, telemedicine is now established as a medical subspecialty with its own journals.[19] Telecare refers to the delivery of health care by means of telecommunications technology, and is relevant to community care.

An early example of telemedicine in the United States was a cable linkage that provided one-way transmission of black-and-white video and two-way audio between a medical facility at Logan Field, the major airport of Boston, MA, and physicians several miles away at the Massachusetts General Hospital. Patients could be interviewed and even their urines and simple blood stains examined microscopically through this electronic hook up. More complex equipment and communication alternatives helped to increase use of telemedicine in the U.S., and $85 million in federal funds was made available in 1995 for this field. Several states established telemedicine networks to link hospitals and clinics with correctional facilities for more readily available medical consultations, and to join universities for mutual support of medical student education.

Quick and accurate transfer of information between primary care physicians, the "gatekeepers" in health care nets, with tertiary care centers hundreds of miles away has resulted in telemedicine expansion especially in states with large rural populations or great distances between medical centers.[20] Telemedicine also enables the Cancer Treatment Center at the University of Kansas to follow-up patients undergoing medicinal treatment but who have returned home. When U.S. troops were deployed to Bosnia in 1996, telemedicine interchange between medical commands in field hospitals and their home bases in Landstuhl, Germany, and various military medical facilities in the U.S. was arranged through facilities at Fort Detrick, Maryland.

J.H. Sanders has described telemedicine services established in 1991 at the Medical College of Georgia to provide continuing medical education and consultations for rural practitioners, using a system described as "totally distance-insensitive." A digital communication link integrates modalities from twisted pair 56 k copper wire to cable, fiber, microwave, or to satellite in series or in parallel, Sanders notes.[21,22]

The efficacy of telemedicine in supervising continued care and patient education has been proven. Efforts are now underway to extend its value to medicine and personal health, including improved sensors that substitute for palpation of a tumor mass or body parts, electronic stethoscopes, cardiac ultrasound, and radiographs. Distant physicians can visualize the ocular retina and the tympanic membrane of the ear by scopes that have been adapted for telemedicine use.

Ethical and legal concerns also have arisen, given that telemedicine transmits patient data over publicly accessible lines to physicians in another state or country. Privacy and confidentiality are not yet fully protected, even though several varieties of patient consent forms have been used. Current requirements that physicians be licensed by the state or nation in which they practice must be modified to permit effective telemedicine examinations of patients in another state, as well as consultations between physicians similarly located in different jurisdiction, must be resolved.

DISASTER CARE

All health units are important when disasters such as floods, tornadoes, hurricanes, earthquakes, train derailments, major air crashes, or major acts of terrorism strike a community. Disasters of any size are accompanied by varying degrees of incoordination in rescue operations, and complicated by stress and fatigue of both trained and volunteer personnel who have responded to the emergency. Roads that are usually available for patient transport may be obstructed by fallen buildings, large trees, and other debris. Water supply, sewerage, communication, and power lines may have been damaged so that hospitals that respond to unusual casualty loads find that equipment may be inoperable, lighting inadequate, telephones out of order, and water supplies for cleansing or sterilization impaired.

Although many calamities are relatively small and localized, each community must prepare plans for rapid responses and quick resolutions. Relatively simple planning brings together the several agencies that will respond to disasters to discuss staffing and organization for recovery, define available resources, list locations of supplies and equipment, determine the availability of outside help in rescue work, clarify usable radio frequencies, and work out the multitude of other elements of a disaster plan. Ideally, representatives of community physicians, hospital information officers, and many other health care workers, even morgue attendants, should be brought into disaster drills to share information with fire and police units. As a plan develops, each department or agency should become familiar with the others' capabilities because in the throes of an emergency it may not be possible to refer to a written plan, page by page. "Planning is everything, the plan is nothing," General Dwight Eisenhower is reported to have said. In regional disaster planning, Eisenhower's dictum applies.

First responder agencies strive to develop a radio communication network with compatible radio frequencies that will link rescue units whose usual work does not require regular exchange of on-the-scene communications. Rescuers must learn to know where community supply depots and equipment pools are located, who owns or manages the materiel, plus knowledge of what is available in civilian depots. Responsibilities for establishing refugee centers with food and shelter, and locations of emergency medical sites should be defined. Repeated drills prepare a community's disaster workers for their duties in response to a disaster, and are the key to rescue and rapid resolution of disorder.

Training of health and allied workers begins with standard courses prepared for firefighters and police officers who, along with public works personnel, are often known as "first responders." Training is available from state agencies and the U.S. Federal Emergency Management Administration (FEMA), based at Emmitsburg, Maryland, who provide control and coordination of disaster relief efforts. Analysis of responses to previous disasters is also a major effort of governments in all of the Americas where an international networking exists to share information and prepare or a wide range of mutual support efforts by communities in emergencies. Easily accessible first aid stations provide initial care for injured and ailing persons and transfer critically ill persons to hospitals via designated medical transport. When catastrophes occur in limited areas, well-defined routes into and out of the disaster area must be defined for ambulances that are to be used to evacuate patients. Dispatchers should maintain accurate lists of patients and their destinations to provide timely information to communities of relatives and friends. Supplies of special protective clothing and personal equipment that support workers in the rescue operations must also be in the planning calculus. Schultz, Koenig, and Noji have described a modified simple triage and rapid treatment system aimed at reducing immediate mortality after earthquake.[23]

Hospitals may not be able to use telephone lines in some areas, for example, and should become part of the radio net for disasters. If hospitals are liable to be damaged by earthquakes, floods, or major neighborhood fires, all staff persons should be trained in rapid and safe patient evacuation techniques. When even modest numbers of casualties are expected, hospitals must prepare to receive worried relatives and curious friends as well as patients. Hospitals should be assured of effective highway traffic control for ambulances in addition to their own internal management of patient flow and necessary restriction of visitors. Administrators plan to meet needs of hospital personnel who may have difficulty in reaching the hospital, and may have to be housed and fed until the emergency has subsided.

Continued supplies of medications and surgical material, food, and laundry are additional concerns of municipal departments that support the hospital community. During early recovery and mop up periods, community health planners should anticipate that a disturbed environment may be a seed-bed for enteric diseases due to several bacteria, such as salmonella or staphylococci, and other pathogens that contaminate damaged water supplies and foodstuffs that lack refrigeration or safe storage. Increased numbers of respiratory diseases can be expected among the very young and old survivors, as well as exacerbated severity of many existing chronic diseases that require a continuous supply of remedies, like diabetes, heart or kidney disease, and seizure disorders.

A natural disaster or major industrial calamity often calls out the best in a community and may reveal unsuspected leadership strengths. Nonetheless, each jurisdiction should be prepared by careful community planning to meet the abrupt challenges of a flood, tornado, earthquake, building collapse, explosion, or fire.

CONCLUSIONS

Community health care is more than what is provided by an official public health department. It includes manifold basic structures and services that give a community an identity and allow it to serve the needs of its citizens. A catalog of core services would include both governmental and non-governmental agencies and range from police, fire, health and public works to housing, food supply and safety, sickness care, education, and entertainment—and more. A healthy community is a vital organization, continually growing by systematic or irregular alterations and adjustments to meet continuing challenges. Not all of its activities are planned by a central agency but each contributes to some degree to the general good. Whether hamlet or major city, a community must form friendly liaisons with its neighbors and eschew bitter relations if they are to prosper as a region. Councils of governments can bring together the elected officials of adjoining subdivisions conjointly to solve selected administrative problems such as fire and police protection along borders, trash disposal, recreation and entertainment, and public health. To keep a community in good health, the chief executive officer directs that each of its elements be alert to changes and any resultant needs to adjust, as well as continuing to be secure financially. A healthy community, like a healthy body, needs constant care even when it appears to be functioning almost automatically and with little direction.

REFERENCES

1. Health n. physical and mental well-being; freedom from disease, pain, or defect. HALE, WHOLE, HEAL. Webster's New World Dictionary, 1980. health [A.S. *health*]. The state of the organism when it functions optimally without evidence of disease or abnormality. *Stedman's Medical Dictionary,* 25th Ed. 1990.
2. Public health., the art and science of community health, concerned with statistics, epidemiology, hygiene, and the prevention and eradication of epidemic diseases. *Stedman's Medical Dictionary,* 25th Ed. 1990.
3. Foegge, W., Preventive Medicine and Public Health, *J. Amer. Med. Assoc.* 1995; **273:** 1712–1713.
4. Mirom, B.A., Hill, S.B., *Acupuncture, How It Works, How It Cures*. Keats Publishing Co. New Canaan, CT.
5. *Complementary Medicine: New Approaches to Good Practice*. British Medical Association. Oxford University Press, Oxford, UK. 1993. *Reader's Guide to Alternative Health Methods,* American Medical Association, Chicago 1993. Burroughs, H, Kastner, M. *Alternative Healing.* Halcyon Publishing, La Mesa, CA. 1993.
6. Alzraki, N.P., Nuclear Medicine, *J. Amer. Med. Assoc.*, 1995; **273:** 1697–1978.
7. Personal communication (1981), Alexander Langmuir, M.D., former Director of the Centers for Disease Control, Atlanta, Georgia.
8. *Prevalence* is the number of cases of a disease existing in a given population at a specific period of time or particular moment in time. *Incidence* is the number of new cases of a disease in a defined population over a specific period of time.
9. Killed polio virus, given by injection, was the first effective vaccine (IPV), followed by a live polio virus vaccine given orally (OPV). The smallpox vaccine was made from cowpox, with which cross-immunity exists.
10. Morbidity and Mortality Weekly Report. CDC. 19 Jan 96, pp. 38–41.
11. Ibid. p. 40
12. Brooks, S.M., *et al.* eds. *Environmental Medicine,* 780 pp. ISBN 0–8016–6469–1. St Louis, MO, Mosby, 1995.
13. Personal communication from Peale A. Burbugh, M.D., MPH, 1995.
14. This stops a rapid, totally irregular heart muscle twitching by use of a momentary electrical shock to the chest, enabling the heart to resume its normal rhythmic beating.
15. Dick, R.S., Steen, E.B., eds. The Computer-based Patient record: An Essential Technology for Health Care. Washington, D.C.: National Academy Press, 1991.

16. Woodward, B., The Computer-based Patient Record and Confidentiality, *N. Engl. J. Med.*, 1995; **333:** 1414–1422.
17. Lindberg, D.A.B., Humphreys, B.L., Computers in Medicine. *J. Amer. Med. Assoc.*, 1995; **273:** 1667–1668.
18. Johnson, I.D., Report to the National Practitioner Data Bank, Chicago IL. *J. Amer. Med. Assoc.*, 1991; **265:** 407–411.
19. Journal of Telemedicine and Telecare, Royal Society of Medicine, 1 Wimpole Street, London, WIM 8AE, UK.
20. Borzo, G., States Leading the Way with Telemedicine Projects. Chicago, IL. *Amer. Med. News.,* 21 Nov 1994.
21. Sanders, Jay H. Future Trends: Telemedicine. Bulletin of the Federation of State Medical Boards, Vol. 82; Number 4, 1995.
22. Sanders, J.H., Tedesco, F.J., Telemedicine: Bringing Medical Care to Isolated Communities. *J. Med. Assoc. Ga.,* 1993: **82:** 2370241.
23. Schultz, C.A., Koenig, K.L., Noji, E.K., A Medical Disaster Response to Reduce Immediate Mortality After an Earthquake, *N. Engl. J. Med.*, 1996: 438–444.

JOHN B. DE HOFF
Cockeysville, Maryland

COMPOSTING

INTRODUCTION

Composting is the biological decomposition of semi-dry organic waste by microorganisms under controlled conditions. During the process, a succession of microbial attacks by bacteria, fungi, and actinomycetes physically and chemically alters the character and structure of the organic material. The volume of the original material is reduced by 30 to 40% as the microorganisms convert the organic waste to nutrient-rich, nuisance-free soil-like humus resistant to further decomposition.

In nature, composting is a continually occurring, ubiquitous process. It is a soil-based, ecological system that recycles mineral plant nutrients in organic form[1] wherever residues of vegetation accumulate and whenever favorable environmental conditions (appropriate temperature and moisture) exist. In agriculture, composted green waste and manures have been applied to tilled soil to increase crop yield since land was first cultivated. While compost is still most widely used to improve soil fertility, its application also helps provide erosion control, restore wetlands and purify toxic sites through bioremediation.[2]

In the last half century, growth of the United States population and improvements in the country's standard of living, coupled with changing product consumption habits, has led to a massive solid waste stream. At present, nearly 2.25 tons of municipal solid waste (MSW) are generated per capita each year in the United States[3]—630 million tons per year for the country as a whole (figures cited do not include tonnage for industrial or agricultural wastes). Traditionally, this mountain of waste has been landfilled or incinerated; indeed, for the last forty years, efforts to produce a soil amendment or low grade fertilizer from composted MSW have been largely unsuccessful. Poor quality and higher than acceptable costs of the finished product have contributed to this failure.[4] Lack of public acceptance of composting as a valuable recycling method, nuisance problems and perceived but unfounded health concerns have also played a part in the closing of some composting facilities.

Over the course of the last 10 years, the study of the science and engineering of composting has intensified. The reconsideration of composting as an alternative disposal method is due, in part, to landfill shortages, rising transportation costs and tipping fees, and tougher federal and state regulations for siting and operating landfills and incinerators. This reconsideration has also been generated by significant improvements made in the design of facilities, equipment and process controls, and of equal importance, the wider dissemination and application of the mechanisms and best practices involved in successful composting. In communities throughout the country, large scale composting has begun to play a significant part in county and municipal integrated solid waste planning and management.

Renewed interest in composting began specifically with sewage sludge, expanded to yard wastes, and has now made inroads into the entire organic fraction of the municipal solid waste stream.[5] Yard waste composting, with a feedstock of leaves and yard/park debris mixed with animal manures, has become the most prevalent form of composting in the United States. Diverting yard waste from MSW reduces the waste stream—with seasonal fluctuations—on average, up to 30%. Several states now permit supermarket and/or restaurant organics to be included in yard waste composting.[6,7] As of 1992, 2,981 yard waste composting facilities and 12 source-separated (food, paper and other organics) composting facilities are in operation; as of 1993, 182 waste water sludge composting facilities are in operation; and as of 1994, 17 MSW composting facilities are in operation with 2 additional plants in the pilot stage and 3 facilities under construction.[8] While there are many biochemical, functional and operational similarities between various compost operations, this discussion will be limited to the fundamentals, the process mechanics and end products of MSW, source-separated and yard waste composting.

FUNDAMENTALS

Almost any solid waste with an organic component (e.g., vegetable matter, fats, leaves, grass-clipping, sawdust, wood chips, waste water sludges) can be composted by one of a variety of process modifications as long as an environment suitable for microbial growth on a particular feedstock is maintained. All composting regardless of the facility design or type of feedstock (substrate) share similar process and microbiological fundamentals. Figure 1 shows an overview of the composting process.

Substrate/Mixing and Blending

Composting begins with the substrate. Over the past decade, recycling and recovery requirements for municipal refuse have altered the composition of the waste being diverted to compost facilities. With much of the metal, plastic and glass

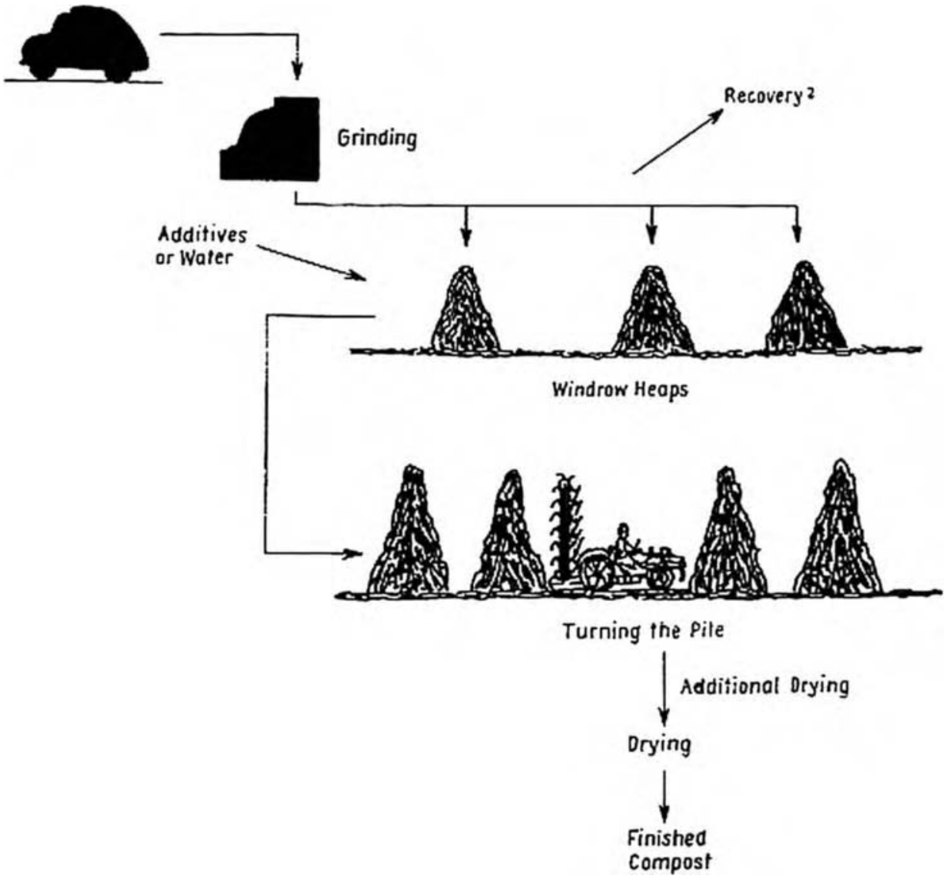


FIGURE 1 Batch composting using windrows.

now removed, the resulting source-separated organic waste (SSOW) is nutritionally richer, yet, because of increased consumption of paper products and packaging, it has a higher cellulosic content.

Most often, a number of different wastes are combined to achieve the nutrient balance necessary to sustain the microbial population. In general terms, the balance depends on nutritive waste fractions like food wastes, manures and waste water sludges (biosolids) combined with bulking agents like wood chips, leaves, bark, light paper and waxed cardboard. If necessary, adjustments are made to the substrate to achieve the proper balance of nutrients (e.g., adding ammonium or nitrate salts boosts nitrogen levels). Biosolids—because of their high moisture content (up to 98%), available nitrogen and an indigenous microbial population—are ideal for mixing (cocomposing) with MSW and SSOW. It must be noted, however, that potential problems characteristic of biosolids (e.g., the presence of pathogens, toxic metals and organics, and odor) must be recognized before beginning the co-composting process and producing an end product.

Preparation

Shredding, chopping or grinding reduces particle size and increases surface area, exposing more of the substrate to microbial attack. It also produces a uniform mass that facilitates mixing and aeration. Typically, the resulting particle size is smaller than one inch. Such preparation is not essential to all compost operations. For example, composting biosolids and leaves does not call for much preparation—equipment is needed only for mixing blending and moving the materials—while MSW and SSOW and yard wastes are routinely shredded. Optimal preparation of the feedstock includes front-end separation and recovery of ferrous metals, glass, plastic, textile and other recyclable materials (resource recovery and screening to remove recyclables and large objects may occur at any point in the process).

Environment and Supplements

In large scale composting operations, efficient decomposition and odor control are determined by the dependent relationship between moisture, oxygen, temperature and bulk density. The moisture content of the mixture is a key parameter in composting because the biochemical reactions that drive the compost process (providing nutrients that support microbial population growth) take place in the film of water enveloping each particle (hence the importance of particle size and surface area). Oxygen also plays a key role. It must be available in sufficient amounts to microorganisms throughout the composted feedstock in order to sustain the microbial attack under aerobic conditions. At least one investigator recommends 10–30 cubic feet of air per day per pound of volatile solids initial charge[9] while oxygen utilization was found to be 0.0244 parts of oxygen per minute per 100 parts of dry weight.[10]

Moisture content of a combined feedstock is determinant of the oxygen supply: If it's too dry, insufficient moisture exists for metabolism; if it's too moist, the pore spaces are filled with water, blocking oxygen from metabolic sites and causing the process to become anaerobic.[11] The optimal range for moisture content is 40–65% depending on the feedstock. Composting MSW and yard waste often requires the addition of water; it can be added at the beginning of the process (as noted, biosolids are an ideal source of moisture) and/or during biological stabilization. During this later stage, moisture content must be monitored on an on-going basis as particle size reduction from microbial and physical decomposition increases the density of the substrate.[12]

For metabolism and growth of microorganisms to occur in substrate, the macro-nutrients, carbon and nitrogen, should be present at an optimum ratio (C:N ratio, 30:1). Some municipal solid waste streams require the addition of nitrogen (in the form of ammonium sulfate, urea or protein) to bring the nutrient levels of the feedstock into balance (see Compost Microbiology).

The bulk density of a mixed feedstock is important in large scale windrow operations especially where food waste is a large fraction of the compost mix. A recent pilot study of grocery waste composting found that air-filled, macro pore space volume is the single most important component of the initial mix.[12] The more dense the mixture, the less oxygen, the more likely odor-producing anaerobic conditions will develop. The study concluded that ideal density is less than 900 pounds of SSOW per cubic yard.

The pH of composting material is another important parameter for microbial growth. While composting will occur with broad pH limits of 5 to 9, no one pH assures success. Rather, the pH changes during the process, reflecting the sequential decomposition of the substrate. Stated simply, at first, as carbohydrates present in the raw feedstock are degraded to organic acids and carbon dioxide, the pH drops; as the carbohydrate fraction is depleted, and the protein substrate falls to microbial attack, the pH, in turn, rises. The pH is seldom controlled during the process unless levels remain depressed due to low moisture content or other process upsets. In such situations, lime or caustic, carefully added, since buffers are characteristic of composting materials, is used to readjust the pH.

Biological Stabilization

The heart of the compost process is biological decomposition and stabilization of the substrate. Whether the feedstock is heaped into windrows (long haystack-shaped piles) or "digested" in-vessel (within an enclosed container or system), decomposition proceeds because of the metabolic activities of a diverse microbial flora. Compost stability (i.e., the point at which no further microbial attack will occur unless new feedstock is introduced) is measured by respirometric methods focusing on oxygen uptake, heat output and carbon dioxide evolution. These measurements give a good indication of microbial activity and odor potential.[13] Depending on facility operations and processes (see sections on Batch Composting and Continuous/Intermittent Composting) and the biological availability of the substrate, the time needed for stabilization (retention time) can be a matter of days, several weeks or several months.

Temperature is another critical parameter of the compost process. During decomposition, heat is generated by microbial metabolism. Due to the insulating capacity of the composted materials, the heat accumulates and temperatures rise rapidly to well above ambient. The temperature rise enhances the succession of microbial attack on the substrate; and the temperatures attained, typically 55–60°C, destroy weed seeds and fecal pathogens that may be present. Many reports have shown that 37°C to be within the optimal temperature range for activity of compost microorganisms,[13] while excessively high temperatures limit growth. Aerating the decomposing material by mechanized turning and mixing, forced aeration or some combination of the two methods moderates temperature and provides oxygen necessary for aerobic metabolism. Temperatures can rise more than once during the process: It is possible to induce a temperature cycle *de novo* if, for example, additional moisture is added to a substrate drying out too quickly, or if the materials are remixed and aerated thereby exposing new surfaces to microbial attack.

Eventually, the biological processes run their course; the temperature of the pile returns to ambient levels. In theory, as long as the moisture content is adequate, decomposition in a compost pile can go on indefinitely, or until no organic material remains. From an operational standpoint, however, when the temperature of the compost falls to about 2°C over ambient, the stabilization phase is complete. This final drop in temperature is probably the most reliable test for the compost's stability.

Biological availability refers to how readily a substrate can be used as food by microorganisms (i.e., how fast it degrades). When composting a municipal solid waste stream, the biological availability of the substrate will be as diverse as the incoming materials. Because of the variety of MSW and SSOW, it is difficult to estimate availability of even the most prevalent macro-nutrient, organic carbon; however, a first order approximation can be made on the basis of a BOD test (as used in testing waste waters). In general terms, decomposition proceeds sequentially with the most readily available substances: soluble carbohydrates, proteins and fats, attacked first. Polysaccharides, proteins and materials high in cellulose, hemi-cellulose and lignin (e.g. paper, cardboard, wood) only become available after a succession of microbial attacks breaks down compounds into their constituent units. Enzymes, synthesized by the microorganisms themselves, are the catalysts that enable these biochemical reactions to occur[14] (see section on Compost Microbiology).

It should be noted that decomposition can proceed in either aerobic (excess oxygen) or anaerobic (lack of oxygen) conditions. All large-scale facilities go to great lengths to maintain aerobic conditions throughout the composting process, for anaerobic conditions in decomposing organic matter can produce foul-smelling odors. A number of compost operations have failed solely because of nuisance problems with odors. Waste water sludge composting facilities often include a controlled anaerobic step in the processes, but further discussion of the anaerobic composting process is beyond the scope of this entry.

Compost Microbiology

The microorganisms that decompose MSW, SSOW and yard waste substrates are generally indigenous to the soil and refuse environment. Successful composting depends on a variety of organisms: bacteria, fungi, actinomycetes and to some extent, protozoa, attacking the substrate in successive waves of metabolic activity. In this succession, one group of organisms paves the way for a succeeding group with the metabolic end products produced by the first organism serving as starting nutrients for the second. It is thought that some microorganisms excrete small amounts of vitamins or amino acids that are essential for other's growth.[15] This cycle of microbial attack, characterized by a synergistic, nutritional and metabolic interaction between a variety of organisms, continues until the available organic material is virtually exhausted.

In the first wave, initiated by bacteria, the most available substrate (soluble carbohydrates, etc.) is consumed. The temperature of the composted material rises quickly, and bacteria and yeasts, multiplying at exponential rates, become the dominant population. At peak activity, they number as many as 10^8–10^9 per gram of compost. As decomposition continues and the pH rises, spore-forming fungi and actinomycetes, undergoing a succession of their own, become the predominant population. During each successive wave of metabolic activity, microorganisms synthesize and release specific enzymes, breaking down complex molecules through enzymatic reactions.[14] A variety of microbial species needs to be present in the composting mixture for complete "digestion" of the substrate as each group of microorganism produces only specific enzymes; i.e., it takes the combined metabolic activities of different types of organisms in the feedstock to produce finished compost.

The majority of microorganisms present in the compost environment are saprophytic (i.e., capable of living on dead organic matter), aerobic and facultatively anaerobic. Among the bacteria are aerobic, spore-forming rods, *Bacillus* spp., as well as *Achromobacter, Enterobacter, Pseudomonas* and *Xanthomona* spp., and numerous small gram-negative rods (coliforms, etc). Among the fungi are *Aspergillus, Fusarium, Penicillium, Rhizopus* and *Trichoderma.* Among the actinomycetes are *Steptomyces, Nocardia, Micromonospora* and *Streptosporangium.* These microorganisms exhibit optimal growth at different temperature ranges and, indeed, flourish at different times during the stabilization temperature cycle. They are mesophilic (thriving in 20°C to 45°C range) or thermophilic microorganisms (thriving in the 45°C to 70°C range). While mesophilic bacteria are more efficient than thermophilic groups for organic decomposition, thermophilic temperatures are necessary to kill pathogenic microorganisms and weed seeds. Operators strive to maintain temperatures in the decomposing material at an ideal range of 35°C to 55°C, enabling both types of microorganisms to flourish.

As noted, the most important microbial nutrients (macronutrients) are organic carbon and nitrogen. Available carbon is the major energy source for microorganisms, nitrogen is essential to microbial growth and reproduction. Phosphorous, and other minor nutrients (sulfur, iron, calcium) and trace elements (manganese, zinc, copper, cobalt, and molybdenum) are normally found in appropriate quantities in municipal refuse and are not usually determining factors. The optimum ratio of carbon to nitrogen is 30:1. However, this is an ideal ratio; the biological processes of composting will occur if the ratio is somewhat higher or lower. If the ratio of carbon to nitrogen is too great, reproduction and growth decline; and when nitrogen is unavailable, the compost process slows to a halt:[15] it is, in a sense, one of the limiting factors in composting. With a C:N ratio lower than 30:1 (e.g., organic grocery debris has an extremely high nitrogen content), the microorganisms eliminate unused nitrogen as ammonia, producing increased odors, reducing the available nitrogen content of the finished compost product and creating a potential health hazard.

While composting operations have occasionally added microbial "seed" to a composting unit, special inocula are not needed for composting. As mentioned above, the microorganisms responsible for decomposition are already present in the raw material: MSW, SSOW, yard wastes, waste water sludges, etc., and, in general, the microbial population most suited for substrate attack will develop on its own. On the other hand, seed recycle (i.e., adding microorganisms grown previously on the same substrate under the same environmental conditions) can help sustain continuous processes at peak activity, it can serve to shorten the retention time for batch processes (see sections on Continuous/Intermittent Composting and Batch Composting).

THE PROCESS

Two methods of large scale composting are in use in the United States: batch operations and continuous (or semi-continuous) operations. But methods require some degree of mechanization depending on facility siting, volume of incoming material, recovery operations and desired finished product. Community yard waste composting is a typical batch composting process; municipal solid waste composting is a typical continuous flow composting process. Three general steps are common to both operations: 1. Feedstock preparation, including particle size reduction and nutrient balancing; 2. Aeration and moisture content adjustment to maximize decomposition by microorganisms; 3. Curing and screening.

Batch Composting

In batch composting, solid waste materials are heaped into windrows. As noted, the feedstock is usually a combination of wastes chosen to provide the appropriate environment and ensure active decomposition within the windrow. Waste water sludges are often added to "seed" the pile with microorganisms (abundant in waste water sludges) and/or raise the moisture content of the mixture; water is often added to bring the moisture content up to optimum levels; and nutrients may be added, specifically nitrogen amendments, if it's necessary to balance the mixture's C:N ratio.

Periodically, the windrows are turned. In the first weeks after the feedstock has been windrowed, the piles are turned frequently (e.g., every fifth day), thereby aerating the interior sections of the pile. Adequate oxygen supply maintains aerobic conditions and ensures the microbial population reaches peak metabolic activity. After 1–2 months, the piles are turned less frequently. At some facilities, forced aeration helps keep the pile aerobic and speeds decomposition: Windrows are heaped on a composting pad constructed with a build-in ventilation system or, at more rudimentary facilities, simply piled over a ventilator pipe. Negative aeration pulls air down through the pile; positive aeration forces air up through the pile from below.

In batch operations, windrows are monitored for moisture content; long retention times are required to stabilize the feedstock. Kept in windrows for several months, the waste material undergoes a substantial change in character, becoming a dark, earthy, soil-like humus.

Continuous/Intermittent Composting

Continuous-flow or intermittently-fed composting processes (Figure 2) have been used as a method of processing MSW since the 1920s. Many European plants built since the 1950s are continuous flow operations. In the United States, a number of continuous-flow, in-vessel processes are currently

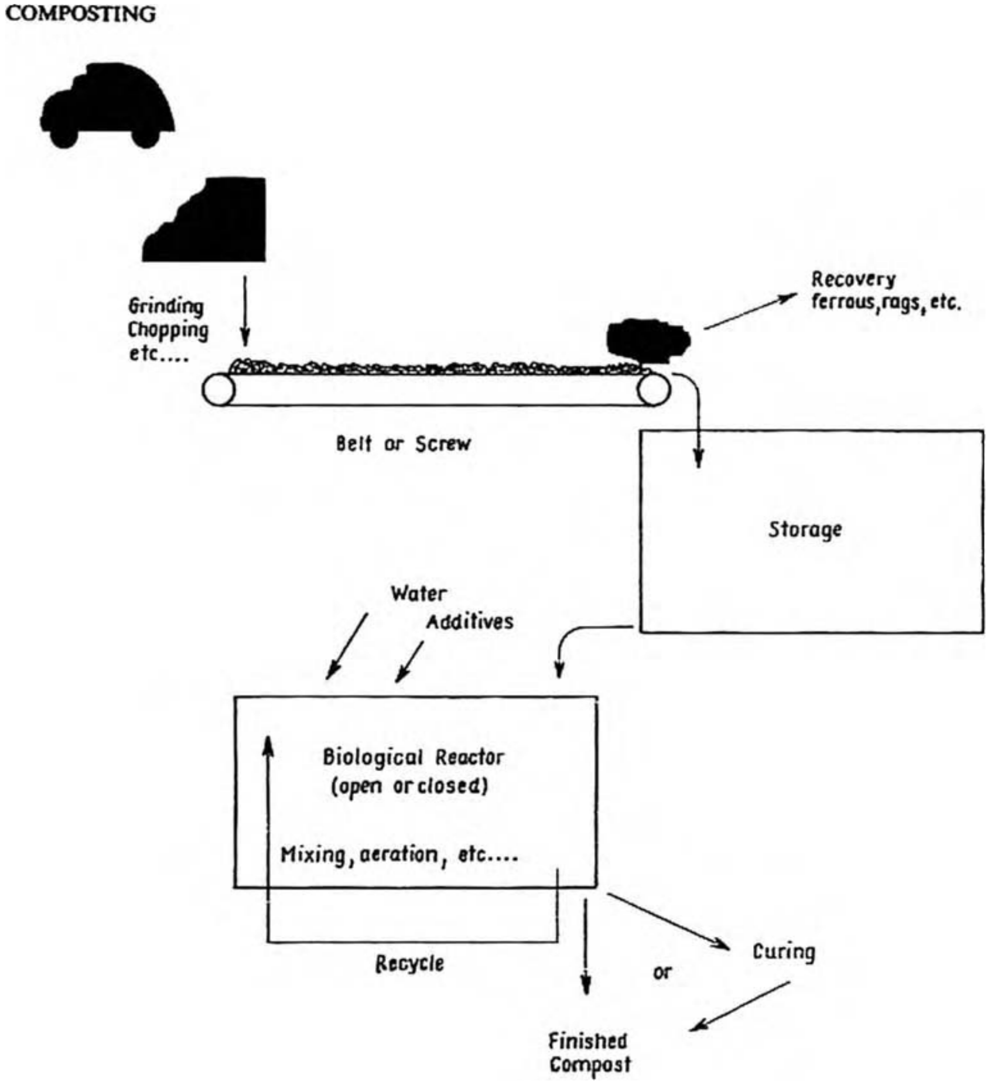


FIGURE 2 Continuous (or intermittently fed) compoting using biological reacter.

in use or under construction. The largest, a tunnel reactor in Baltimore, Maryland, has a throughput of 2,500 tons per week. Another facility, in Sevier County, Tennessee, using a continuous flow rotating tunnel digester, processes 1,100 tons of MSW per week; in St. Cloud, Minnesota, a digester with an agitated bed, processes 250 tons of MSW per week; in Easthampton, New York, a recently opened facility uses an in-vessel, agitated-bed digester to process 180 tons of source-separated and mixed MSW per week.[16,17]

Continuous flow operations have a number of common unit processes: feedstock preparation, storage, biochemical activation/stabilization and curing. After preparation (see Fundamentals: *Preparation,* above), the feedstock is transported to a storage vessel via mechanized conveyors; from there it is fed into the biological digester on a more or less continuous basis. The biochemical reactions that take place in-vessel are the same as those in batch operations (see Fundamentals): a succession of microorganisms attack the substrate, temperatures rise to 55–65°C, and the initial material undergoes chemical and physical transformation. At some facilities, the in-coming waste material is seeded with microorganisms from compost produced on site; in others, waste water sludges or animal manures provide the necessary microbial culture to maintain peak-activity within the digester. Nutritional supplements may also be added to accelerate the process or to supply a nutrient that can be used in the finished product.

The amount of time the feedstock remains in the digester depends on the design of the system and the incoming flow of the waste stream. For example, the Baltimore tunnel system (65' × 18' d.) "digests" material for 12–20 days; at the Sevier Country facility, the rotating digester (185' × 12' d.) stabilizes the MSW feedstock in 3 days; in Easthampton, the agitated-bed system (6' × 6' × 250' troughs) stabilizes waste materials in 21 days.[14] In continuous operations, the accelerated rates of microbial attack characteristic of in-vessel systems, require that moisture, temperature and oxygen be monitored carefully and maintained at as close to optimal levels as possible, and most process controls are now run by computer. It should also be noted that in the higher temperatures sustained in the digester, thermophilic microorganisms are the predominant population.

Curing and Final Processing

In continuous flow operations, biological stabilization is normally followed by some type of aerated curing. The curing lets the compost mature to a point where significant microbial activity has subsided. During this stage, phytotoxicity (characteristic of immature compost) disappears and available nitrogen (mostly nitrate-N) rises to higher levels. At the Baltimore facility, stabilized material is cured in static, negatively-aerated piles, while at the St. Cloud facility, stabilized material matures in mechanically turned, positively-aerated, 200' trenches; at the Easthampton plant, negatively aerated windrows are used.[17]

After a period of 1–3 months of curing, ambient temperatures are regained, and the finished compost is ready for application. Using a mechanized rotary drum screens, finished compost is screened to various particle sizes for use in agricultural, horticultural or landscape applications. While most field agricultural application does not require final processing, screening compost to a 3/4 inch size is common for landscaping; while a 1/4 inch screened compost product is more likely to be bagged and marketed at gardening centers. Generally, high in available nitrogen, low in ammonia and heavy metal concentrations, top-grade compost intended for use as a soil amendment, conditioner or fertilizer can be customized during the latter stages of the curing process to suit the users' needs. The parameters most commonly balanced for specific use are the NPK ratio (nitrogen, phosphorous, potassium) and soluble salt content (i.e., the conductivity of the finished product).

UTILIZATION

The recent growth of MSW composting facilities in the U.S. can be attributed to the necessity, in communities all over the country, to minimize the volume of waste being landfilled and incinerated. As composting takes a larger and larger bite out of the solid waste stream, and as a wider spectrum materials are composted, greater quantities of compost are being produced. To ensure product marketability and continued industry growth, solid waste managers and facility operators will now have to place greater emphasis on quality control, customer satisfaction and product utilization.[18,19] They will have to approach composting as a manufacturing process: The biodegradability of the incoming material must be well understood; systems and processes designed to maximize consistency and maturity; markets developed for a range of compost applications based on user's needs and product benefits (e.g., water conservation, fertilizer savings, increased yield/acre, stronger plants, pesticide savings, disease suppression[20] bioremediation of hazardous sites and land reclamation).

Quality

Currently, compost producers, Departments of Agriculture in a number of states and the Composting Council (a nonprofit industry association) are working together to develop compost quality standards and application (i.e., product use) guidelines. These standards will draw on the U.S. Environmental Protection Agency's Part 503 biosolids rule which established health, safety and environmental standards for composts from waste water sludges. When implemented these standards will enable operators to improve process design; they will provide end users with quantitative and qualitative data on specific parameters of a compost product and in general further awareness of the interrelationship of the composting process and the quality of the end product.[21] Quantitative standards will include: pH, soluble salt content, water holding capacity, bulk density, moisture content, organic matter content and particle size. Qualitative and unspecified standards will include concentrations of trace elements/heavy metals and compost stability (see Fundamentals: *Biological Stabilization*) and maturity.[19]

While compost maturity is a key parameter for end use applications, it is important to note that unlike other parameters, standard testing methods for stability and maturity do not yet exist.[19] Efforts by compost producers and scientists to devise a standard measurement for compost maturity are underway in a number of countries including the United States, Japan and Italy. Areas of consideration include a measure of the degree of humification via fractionation of organic carbon (i.e., the ratio between humified fractions—humic and fulvic acid—and total extractable carbon),[22] magnetic resonance scanning and kinetic modeling of the composting process.

Recent work on a kinetic model for commercial yard waste composting has enabled researchers, with some limitations, to better predict the biochemical behavior of different waste mixtures. With a further researched kinetic model in hand, producers will be able to adapt processes for a seasonally changing waste stream and design systems which reduce odor release, leachate production and other localized environmental problems. By optimizing the process, operators can produce a higher quality, more marketable finished product (e.g., a product that is more likely to function as a substitute for peat moss—a nonrenewable resource)[23,24] and, avoiding nuisance problems, gain greater public acceptance for the facility.

Application

Top quality composts from a variety of feedstocks are marketed to the agricultural, horticultural and landscaping industries as a soil amendments and low grade fertilizers for use in row crops, nursery stock, container plants, commercial sod production, tree farms, etc. Lesser quality composts (i.e., with some contamination by inerts or borderline heavy metal concentrations) are used for excavation and forest reclamation, hazardous waste bioremediation, DOT landscaping and landfill cover. Such composts are often products of co-composting facilities where an MSW waste stream is combined with biosolids (waste water sludges). It is widely known that in some MSW, as well as in biosolids from urban areas (i.e., where automotive contaminants enter waste water facilities through run-off into storm drains) heavy metal concentrations in the finished product will exceed unrestricted use criteria, limiting application to land reclamation and landfill cover.

Using compost to suppress soilborne plant diseases may be one of the most valuable, beneficial effects of compost application.[24] Agricultural scientists, plant pathologists and compost operators are currently field testing composts as biocontrol agents with specific applications in the nursery container plant and tree-farming industries.[25] Research over the past two decades has identified large numbers of microorganisms, isolated from mature composts, that induce suppression to soil-borne disease. There is now sufficient information available on compost's disease-suppressive properties that predictable, effective biological control can now be practiced. The mechanics of suppression, in the simplest terms, can be described as a competition between beneficial microbes and their disease-causing counterparts for the nutrients delivered to the soil by the plant's root system. In mature compost with a robust microbial population, the beneficial microbes dominate the nutrient source; biological control prevails; the disease is suppressed.[26,27]

Compost with its resident microorganisms is also useful for bioremediation of hazardous wastes (*in situ* as well as off-site). Recent tests and pilot programs have found that composts successfully degrade pesticide-contaminated soils[28] and crop residues; remediate oil field wastes (benzene, toluene, xylene) and petroleum sludges;[29] and detoxify other hydrocarbon pollutants such as PCBs and chlorinated aliphatic solvents. Composting hazardous wastes has distinct advantages to incineration: lower capital and operating costs, and it's safer in terms of impact on water, land and air resources. However, there are disadvantages as well: for instance, the preparatory work and precautionary measures (e.g., enclosing the site or facility) that might be required to compost a particular hazardous material can be economically prohibitive.[30]

Public Health Concerns/Environmental Impact

As composting gains wider acceptance, and becomes a part of more communities' integrated solid waste management plans, certain health and environmental issues have been raised and will continue to be raised by concerned facility neighbors and end users. The most common concerns are the fate of pathogenic organisms, bioaersol exposure, odors and cumulative loading of contaminants to soil groundwater. To ensure a continued role for composting in MSW disposal, producers are currently addressing, monitoring and managing these concerns more aggressively than ever before.

Questions often arise about the survival or reproduction of pathogenic organisms shown to be present in some MSW and waste water sludges: what happens to microorganisms such as *Salmonella,* enteric bacteria, flagellated and amebic protozoans and enteric viruses at a composting facility; can compost produced from the other pathogenic microorganisms do not compete well in enriched environments such as soil or waste water sludges, nor can they withstand for even short periods of time the elevated temperatures (140°F to 160°F) produced during the composting process. Under existing Federal regulations for handling sludges—and considering the temperature reached in composters for extended periods of time wastes can be safe for agricultural uses. In general, these from finished compost, is negligible. To meet EPA top quality, Class A criteria, compost must be maintained at 55°C (130+°F) for three days. This is referred to as PFRP (process to further reduce pathogens). The second EPA category, Class B, requires that compost be kept at 40°C for five days or 55°C for three hours. Class B is referred to as PSRP (process to significantly reduce pathogens).

Another public health concern often linked to compost facilities is aspergillosis, an infection of the lung caused by an invading fungus of the same genus name, mostly *Aspergillus fumigatus.* Although this organism is common in composting materials—its spores are small, buoyant bioaerosols—investigators have been unable to establish a causal relationship between compost operations and the disease.[31,32] With background levels of this ubiquitous fungus found in most households and backyards in the country, compost proponents state unswervingly that there has never been a case of *a. fumigatus* infection associated with composting facilities, either in workers or in the surrounding public.[33] By way of caution, however, it should be noted that any microorganism can become an invading opportunist given the right circumstances; and immunocompromised individuals, the elderly and workers with unprotected wounds should not risk exposure to microbial infection where microbiological processes are taking place.

Though the odors that can be generated during composting are more likely to be nuisance problems than health risks, offensive odors from poorly controlled compost operations have resulted in the closing of more facilities than any other health or environmental factor. The groups of odorous compounds that have been identified at composting facilities are reduced sulfur compounds, ammonia and amine compounds, fatty acids, terpenes, acetone, phenol and toluene.[34] The most common odors are for the most part the result of the formation, release and reactivity of hydrogen sulfide (i.e., a reduced sulfur compound), a natural end product of anaerobic decomposition. Hydrogen sulfide in itself is fairly malodorous, but in its reactions with available organic matter, a class of compounds are formed, called mercaptans or thioorganics, that are even more disagreeable. Typical mercaptans are dimethylsulfide (DMS) and dimethyldisulfide (DMDS). These compounds can then further combine with the natural odors of composting such as ammonia, or the alcohols, acids and aldehydes associated with fermentation, or with other vegetable-based hydrocarbons such as limonene to produce serious odor problems.

Odor control now plays a significant role in large and small composting operations alike. Aeration systems, discussed previously, are in use in many facilities, and the newer, enclosed facilities have odor control/treatment systems in place. The most common and effective treatment systems are multistage chemical scrubbers and biofilters. Both packed-tower adsorption and misting are operated in stages to allow different chemical reactions to remove different groups of compounds from the air stream. Gaining in popularity, biofilters are now used in over two dozen biosolids and MSW composting facilities in the U.S.[34] It should be noted that, in the near future, facilities' air treatment technology may have to have the capacity control VOC emissions. With the enactment of 1990 Clear Air Act Amendments, regulations are being developed to limit VOC emissions in nonattainment areas for the pollutant ozone, and such regulations would most likely have an impact on the operations of large scale MSW compost facilities.[34]

While virtually all land application of mature compost supports plant growth and improves soil properties, environmental concerns specific to compost use in field and horticultural crops have been raised by industry researchers and farmers. These concerns involve the cumulative loading of trace amounts of heavy metals to soils, the cumulative effect of nitrate leachate on ground water,[35] and to a lesser degree, the potential presence of lawn-care pesticide residues in finished yard waste compost.[36]

Despite the use of top quality compost and standard agricultural practices, trace amounts of heavy metals can accumulate in field soil if compost is applied at rates high enough to supply fertilizer requirements. After a 3–5 year period, the concentration of a specific metal in the soil could be toxic to certain agricultural or horticultural crops. It is interesting to note that some fruit tree farmers can repeatedly use compost with higher metal concentrations with benefits from the cumulative effect, for the longevity of some species of fruit trees depletes the surrounding soil of the necessary inorganic trace nutrients. Research into the concerns of nitrate-N contamination of ground water from compost amended field test plots indicates that after three consecutive years there is a cumulative effect in the soil. Because of the effect, it is suggested that lower rates should be applied in subsequent years to lessen nitrate leaching.[35] However, to date, there has been no indication of ground water contamination. Samples analyzed beneath several different test sites were in one case, well below 10 ppm drinking water standard, and in another case, showed no significant difference from ground water flowing under control plots.[37]

It is generally agreed that with rigorous monitoring, quality assurance and pH management to limit bio-availability, the use of compost containing low levels of trace minerals can be safely applied for an indefinite period of time without risk to agriculture or the environment. These concerns underscore the push for quality and utilization guidelines in the composting industry.

CONCLUSION

With the potential to divert millions of tons of MSW from landfills and incinerators and to provide beneficial end-products for horticultural and agricultural uses, composting is an obvious and natural method for recycling, reusing and conserving the organic fraction of the solid waste stream.

REFERENCES

1. Finstein, M.S. and M.L. Morris, "Microbiology of Municipal Solid Waste Composting," *Advances in Applied Microbiology,* D. Perlman, ed.; Vol. 19, Academic Press, Harcourt Brace Jovanovich, New York, 1975.
2. Anonymous, "Composting Basics," *Composting Council Fact Sheet,* The Composting Council, a compost industry association, Alexandria, VA, 1994.
3. Tchobanoglous, G., H. Theisen and S. Vigil, *Integrated Solid Waste Management,* McGraw-Hill, Inc., 1993.
4. Anonymous, "Cleaning Our Environment—The Chemical Basis," *Amer. Chem. Soc.,* Washington, D.C., 1969.
5. Diaz, L.F. and C.G. Golueke, "Composting of MSW in the USA," *Proceedings from the Recycling and Composting Sessions of the First United States Conference on Municipal Solid Waste Management,* Washington, D.C., June 1990.
6. Adams, N.E., "Composting Food Wastes on the Farm," *BioCycle,* June, 1993.
7. Rockwell, F., "Supermarket Stream Added to Composting Mix," *BioCycle,* Oct. 1994.
8. Goldstein, N. and R. Steuteville, "1994 Biocycle MSW Survey—Solid Waste Composting Seeks Its Niche," *BioCycle,* Nov. 1994.

9. Anonymous, "Municipal Refuse Disposal," publication of American Public Works Association, Public Administration Service, Chicago, IL.
10. Anonymous, "A Study of the Composting of Garbage and Other Solid Organic Wastes," Interim Report #1, Michigan State University, East Lansing, MI.
11. Cardenas, R.C. and L.K. Wang, "Composting Process," *Handbook of Environmental Engineering* Vol. 2, The HUMANA PRESS, 1980.
12. Croteau, G. and J. Alpert, "Low Tech Approaches to Composting Supermarket Organics," *BioCycle,* May 1994.
13. Frost, D.I., B.L. Toth and H.A.J. Hoitink, "Compost Stability," *BioCycle,* November 1992.
14. Diaz, L.F., G.M. Savage and C.G. Golueke, *Resource Recovery from Municipal Solid Wastes,* Vol. II, CRC Press, Inc. Boca Raton, FL, 1982.
15. Golueke, C.G., "Bacteriology of Composting," *BioCycle,* Jan. 1992.
16. Anonymous, "MSW Composting Facilities, Oct. 1994," *Composting Council Fact Sheet,* The Composting Council, a compost industry association, Alexandria, VA, 1994.
17. Personal communications.
18. Kashmanian, R.M. and R.L. Spencer, "Cost Considerations of Municipal Solid Waste Compost Production Versus Market Price," *Compost Science & Utilization,* Premiere Issue, 1993.
19. Alexander, R.A., "Standards and Guidelines for Compost Use," *BioCycle,* Dec. 1994.
20. Shiralipour, A., "Compost Effects on California Row Crops and Ornamentals," *Proceedings of the Composting Council's Fifth National Conference,* Nov. 1994.
21. Marugg, C., M.E., Grebus, R.C. Hansen, H.M. Keener and H.A.J. Hoitink, "A Kinetic Model of the Yard Waste Composting Process," *Compost Science & Utilization,* Premier Issue, 1993.
22. Ciavattaa, C., M. Govi and P. Sequi, "Characterization of Compost Produced with Municipal Solid Waste: An Italian Approach," *Compost Science & Utilization,* Premier Issue, 1993.
23. Whang, D.S. and G.F. Meenaghan, "Kinetic Model of Composting Process," *Compost Science/Land Utilization,* Vol 21: 3.
24. Hoitink, H.A.J. and M.E. Grebus, "Nurseries Find New Value in Composted Products," *BioCycle,* May, 1994.
25. Holusha, D.J., "Marking Compost Double as Pesticide," *NY Times,* February 27, 1994.
26. Hoitink, H.A.J., M.J. Boehm and Y. Hadar, "Mechanisms of Suppression of Soilborne Plant Pathogens in Compost Amended Substrates," *Science and Engineering of Composting,* Renaissance Publications, Worthington, OH, 1994.
27. Logsdon, G., "Plant Protection Through Compost," *BioCycle,* January, 1990.
28. Cole, M.A., X. Liu and L. Zhang, "Compost Accelerated Degradation of Pesticides in Contaminated Soil," *Poster Sessions of the Composting Council's Fifth National Conference,* Nov. 1994.
29. Dale, B.E., *Composting for Remediation of Oil Field Wastes,* submitted to Amoco Production Corporation, Sept. 1991.
30. Golueke, C.G. and L.F. Diaz, "Biological Treatment for Hazardous Wastes," *BioCycle,* Dec. 1989.
31. Millner, P.D., D.A. Bassett and P.B. Marsh, "Dispersal of *Aspergillus fumigatus* from Sewage Sludge Compost Piles Subjected to Mechanical Agitation in Open Air," *Applied and Environmental Microbiology,* May 1980.
32. Toomey, W.A., Jr., "Confronting Health Concerns," *BioCycle,* May 1994.
33. Anonymous, "*Aspergillus Fumigatus* & Composting," *Composting Council Fact Sheet* #5, The Composting Council, Alexandria, V.A.
34. Williams, T.O., "Odors and VOC Emissions Control Methods," *BioCycle,* May 1995.
35. Maynard, A., "Growth of Plants and Leaching of Nitrate from Compost-Amended Plots," *Research Symposium of the Composting Council's Fourth National Conference,* Nov. 1994.
36. Michel, F., "Fate of Certain Lawn Care Pesticides During Yard Waste Composting, *Research Symposium of the Composting Council's Fourth National Conference,* Nov. 1994.
37. Breslin, V.T., "Use of MSW Compost in Commercial Sod Production," *BioCycle,* May 1995.

JAMES K. McCARTHY
Rutgers University

RAUL R. CARDENAS, Jr.
City College of New York

CONSERVATION: see ENERGY SOURCES—ALTERNATIVES

D

DESALINATION

INTRODUCTION

Our planet, the earth or gaia, is a water place. Water occupies 70% of the total earth's area, about 360×10^{12} m^2, and the total water volume that covers the earth's surface is about 1.40×10^{18} m^3. This makes the planet earth a very large water reservoir, nevertheless practically, these huge amounts of water are not directly usable, as 97% of this water is the seawater of the open seas and oceans and only 3% consist of fresh and brackish water, i.e., 0.042×10^{18} m^3. From these 3.0% fresh water reserves only a portion (0.014%) is in liquid form in rivers, lakes and wells directly available to us for immediate use, the rest can be found as glacier, icebergs and very deep water of geological reservoirs.

Fresh water, for all of historical times has been an uncontrollable happening of nature wherever and however found. In the Bible "good land" is described as one "of brooks of water, of fountains, and depths that spring out of valleys and hills," for water is a precious source of development and civilization, because water and civilization are two inseparable conceptions.

From antiquity up to our times, rivers, seas, oases, and oceans have attracted man to their shores. As a rule, towns and countries have grown along rivers. Egypt for example was considered as "the gift of the river." Egypt is a typical historical example of the influence of water to the birth and development of a civilization.

As a largely developed agricultural country, Egypt was able to master the river and had most of the time such an abundant harvest, that it became the main wheat-exporting country in the whole Mediterranean. The Egyptians learned to determine the seasons of the year by the behaviour of the river. Inundation, Emergence of the fields and Drought were their seasons. The first calendar was created in this way and out of this was derived the modern calendar.

Unfortunately, fresh water, and even seawater has not had our proper attention, respect and treatment. Due to the increase of population, especially in certain regions, and the increase of living standards, water demand increased exponentially, and wells or other fresh water sources run dry. The great population increase multiplies the total withdrawal. In some areas twice or more as much water is being drawn out of the ground as sank into it, thus the water table drops every year a few meters and water shortage increases dramatically, especially in dry years.

In modern large civic centers, opening of the tap provides us with as much fresh water as we are able to waste. The notion of lack of water is usually a matter we very seldom think of. However, at the same time there are places where water is so scarce that there are serious problems of existence, as is the case these days in many places in Africa. Generally it is not realized that fresh water represents only 3% of the total reserves in the world and that 75% of it is immobilized as ice. Modern agriculture also requires considerable amounts of good water to meet increasing food requirements. There is a tremendous requirement of water to run industrial plants, to produce all kinds of goods but also all kinds of polluting effluent.

The quantity of water used varies from location to location throughout the day and throughout the year, as many factors influence this variation. The more important factors include the economics of a community, its geographic location, and the nearby availability of the water source or the transportation distance from the source.

Climate is the most common cause of water lack or of water insufficiency. Sparse rainfall to feed streams, wells and the soil for agricultural production of crops exhausts water reserves. The most arid areas are the deserts, where no rain exists and some underground waters most of the time are salty or brackish. About 19% of the total land surface of the earth on all continents but Europe, is covered by deserts, which are surrounded by semi-arid lands where existing water is insufficient.

Coastal deserts, where the lack of water is as high as in the interior deserts, cover about 33,000 km^2 around the world, the greatest part of which is found in the Middle East, along the Persian Gulf, adjoining parts of the Arabian Sea and the Indian Ocean. The coastal deserts are divided into four main categories, according to their climatic conditions.

The tropical regions, where the temperature is about 30°C in summer and 22°C during winter. The subtropical regions, where in warm periods the temperature is about 30°C, and it ranges between 10 and 22°C in the cold months. The regions, as in the Mediterranean coasts, where during warm periods the temperatures are 22 to 33°C, and in cold months 10 to 22°C. The cool coastal deserts have in summer temperatures under 22°C and in winter time 10 to 22°C. The last desert regions are the cold places where summer is under 22°C and winter under 10°C. The largest single coastal desert of the third type, with the moderate climatic conditions, is that of the Mediterranean Sea and covers about 2650 km^2.[1]

Coastal deserts have an advantage over the interior deserts. They are climatically more pleasant, because they are cooler in summer and warmer in winter. Further, they have advantages over the interior deserts from the desalination point of view. Coastal deserts are surrounded by abundant sea water supply which is in the same level as the desalting installation, and thus the intake of water can be pumped with less power consumption than the deep well salty or brackish waters of the inland deserts. The brine disposal is also easy, without problems, as it is discharged directly into the sea, whereas the disposal of brine in inland deserts may create serious problems. Also, coastal deserts are in favour over the inland ones concurring the transportation of the equipment and all other necessary supplies for a desalination plant. Figure 1 shows a modern desalination plant in Saudi Arabia.

Some of the most attractive areas and beaches of the world are almost devoid of water. Not only is this living space, and space for resort hotels lost, but, in some cases, profitable resources cannot be exploited. Thus, known minerals on Egypt's Red Sea coast cannot be mined, and fishing industries on South America's Pacific coast, and other places around the world, cannot be expanded for lack of water. These present major losses in the world supplies of minerals and foods.

WATER DEMAND AND USAGE

The water cycle leaves about 9000 km^3 of water worldwide per year. This amount is enough to provide, with good quality water, about 20 billion people, but this water is far from evenly divided, with major shortages in some regions and abundant quantities in other places.

In a modern urban agglomeration, supply of water may satisfy domestic, municipal and industrial demand, as well as agricultural needs. There are no standards of general acceptance for the quality of water required by each group of users. Domestic demand includes all water consumed in housekeeping and gardening. A limit of 500 mg/L (ppm) for total dissolved solids with a maximum of 250 mg/L for chloride and sulphate ions, respectively, is recommended by the World Health Organization (WHO).[2] Nevertheless, there is a large number of communities, which are still consuming water containing up to 1000 mg/L total dissolved solids and sometimes more. Physiological changes may result from the intake of large amounts of the main ions, as well as of some trace elements.

Municipal requirements, beside the supply of water for domestic use, include all water needed by offices, public and commercial establishments, fire-fighting and irrigation of municipal parks. Although the standards for the latter uses are not strictly the same, as for drinking water, in practice all municipal water requirements are identical to drinking water since it is nearly always supplied by the same piping system.

A large variety of quality standards is involved in the use of industrial water, depending on its specific use. They may vary from high-quality drinking water for food processing to completely demineralized water for specific uses. Limitations of salt content may be imposed in some cases for process water. Boiler feed water needs special treatment to minimize salt content and eliminate dissolved gases. Cooling water also needs some treatment to meet the process requirements. River water and sea water can be used for cooling purposes and this is the usual practice in plants located on a river or near the seashore.

About 70% of water withdrawn from the earth goes for agriculture purposes and the balance, 30%, for various uses, as household and industrial process water. Overirrigation the last years, brought salinization of the nearby water resources, affecting the soil and crop quality, as salts are accumulating in the soil.

Irrigation water quality, which includes also drinking water for animals, depends to a large extent on the nature of the soil, the crops and the climate. The yield and quantity of some crops can be affected, not only by the total amount of dissolved solids, but also by the presence of certain specific salts. Thus if desalinated water is to be used in certain places the make-up of the product water will be necessary.

The water withdrawn per year and per capita, concerning industry and agriculture is increasing by 8.5%, the main increase in the developed countries. The USA consumes 2500 m^3 per year per capita, Switzerland 500 and Ghana, a very poor African country, only 40 m^3.[3]

Meanwhile, the majority of fresh water streams are severely polluted, decreasing the quality water reserves. Self decontamination is not feasible in many cases and, thus, treatment methods have to be applied to degrade at least some of the pollutants in the water. On the other hand, sea water exist in huge amounts, given free. Although also polluted to some extent, it is a future source of fresh water as desalination is the future process to produce this valuable good quality water.

SEA WATER

The seas and oceans are great sources of material available to mankind, though their destiny is very low to be exploited, but high enough to make the water salty, unsuitable for drinking or processing purposes. Not all the seas around the world have the same amount of total dissolved solids, the amount of which range from 20,000 to 50,000 ppm.

FIGURE 1 Panoramic view of the Al-Jubail Saudi Arabia, phase II, MSF desalination plant. It is up to now the World largest desalination installation totaling a capacity of 947,000 m^3/d (250 Mgd) fresh water production. Each unit has a capacity of 23,500 m^3/d (6.2 Mgd). The plant was built for the Saline Water Conversion Corporation, of Saudi Arabia by the Japanese Companies of Sasakura and Mitsubishi. (Courtesy Sasakura Engineering Co., Japan)

Table 1 gives the total dissolved salt of various oceans and closed seas. Over seventy elements have been detected in seawater, some in very small to trace amounts. Their proportion in all oceans, independent of their concentration, is almost stable.[4]

The four main metals—sodium, magnesium, calcium and potassium—and their combining ions, chlorides, including the other halogens and bicarbonates, are presented in amounts beyond comprehension. In Table 2 are given the major and the minor elements in seawater.

TABLE 1
Total Dissolved Solids in Various Seas

Ocean/Sea	g/kg	ppm	Ocean/Sea	g/kg	ppm
Baltic Sea	7.0	7,000	Pacific Ocean	33.6	33,600
Caspian Sea	13.5	13,500	Atlantic Ocean	36.0	33,600
Black Sea	20.0	20,000	Mediterranean Sea	39.0	39,000
White Sea	28.0	28,000	Red Sea	43.0	43,000
Northern Adriatic	29.0	29,000	Kara Bogar (Caspian)	164.0	164,000
			Dead Sea	270.0	270,000

TABLE 2
Ionic Composition of Main Elements in Seawater[6]

Ions		g/kg	Ions		g/kg
Chlorides	Cl^-	18.980	Copper	Cu	0.003×10^{-3}
Sodium	Na^+	10.560	Uranium	U	0.003×10^{-3}
Sulfates	SO_4^{2-}	2.560			
Magnesium	Mg^{2+}	1.270	Total TDS	—	34.482
Calcium	Ca^{2+}	0.400	Water	H_2O	965.518
Potassium	K^+	0.380			
Hydrogen				Seawater characteristics	
Carbonates	HCO_3	0.143	Salinity	g/kg	34,330
Bromides	Br^-	0.065	Chlorinity	g/kg	19,000
Boric acid	H_3BO_3	0.026	Chlorocity	g/kg	19,950
Strondium	Sr^{2+}	0.014	Specific weight	N/m^3	10,243
Fluorides	F	0.001			

Desalination eliminates the main elements from seawater, producing fresh water and concentrated brine, almost saturated in the main salts, which are withdrawn to the sea. There are two main reasons that these salts are not exploited. The brine volumes are huge and cannot be handled easily. The present extraction technology is expensive for the relatively cheap materials. Nevertheless, there is some industry exploiting, in part, the concentrated brine.

Today throughout the world more than 1,900,000 m^3/d (500 MUSGPO) of fresh water is produced by the various desalination processes.[5] Usually twice to 2½ as much seawater is processed, so that the solids concentration of the brine is doubled. It is estimated that the recovery from the withdrawn brine can be:

Magnesium	2,306.000 t/y	Bromine	116.100 t/y
Calcium	728.000 t/y	Copper	5.385 t/y
Potassium	659.000 t/y	Uranium	5.385 t/y
Sulfate	4,855.000 t/y	Gold	7.2 kg/y

Calcium and magnesium are the main elements that cause scale formation. Scales are formed and precipitate inside desalination equipment simultaneously with other suspended solids content in the feed water sea or brackish. These materials precipitate in areas favored for deposition. In distillation plants these are the heat exchangers and, in reverse osmosis, the semipermeable membranes, cause the problems. These deposits are categorised in two main types, the sludge which is soft and can be easily washed out, and the scale which is hard, adheres to heat transfer surfaces and can be removed only by plant shutdown.

Brackish waters are classified as waters with total dissolved solids content ranging from 3,000 ppm to 20,000 ppm. The elements vary widely, depending on the rocks and soil coming in contact with the water. In some brackish waters large amounts of calcium sulfate are present up to saturation conditions, making the water bitter and unsuitable for any use.

DESALINATION PROCESSES

When all other possibilities to use existing natural water resources are exhausted or to augment fresh water supply by conventional methods fail, then desalting of seawater, or brackish water and/or of polluted water reserves might give the answer to local water problems. The cost of desalting has been drastically reduced over the past several years. This is the result of the combined effort of scientists and engineers. However, it should not be forgotten that desalted water is an industrial product and its cost can never compete with the cost of natural fresh water supplies.

The largest desalination plant is Nature. The hydrological cycle on earth begins by desalination of surface waters. As the sun's energy evaporates the water from the oceans and the land surface waters, the vapors condense again on the earth's surface, as desalted water, stored as snow, ice or through the soil returns to the rivers and seas. This water is the vital liquid for all creatures on the earth.

The importance of water, as a matter of life, is quoted as far back as there are records in history. We read in the Old Testament: "Moses brought the sons of Israel from the Red Sea and they went into the desert of Shur. They marched three days in the wilderness and could not find water to drink. And when they arrived to Merra they could not drink the waters of Merra, for they were bitter. Therefore, he named this place 'bitterness.' And the people murmured against Moses, saying: What shall we drink? And Moses cried onto the Lord. And the Lord sweded a wood, which when he had cast into the waters, the waters were made sweet."[7] Nobody has guessed what kind of wood this could be, but is the first known in history, technical desalination.

The effort of a desalination process is to separate one of the most common, and most useful and yet most unusual material—the water from the one of the next common material, the salt. Hundreds of processes have been proposed, based on the various properties of water and its saline solutions. Nevertheless only a few of these methods have reached such an advanced state of technology to be considered as safe processes for the commercial conversion of saline waters into fresh. Distillation processes, reverse osmosis and electrodialysis or in some cases combination of two processes. The expectations, connected with freezing processes, could not be met with current freezing technology in large scale industrial application.

The required separation may be of water from salt, or of salt from water. Thus the desalination processes can be classified, according to the operation reference parameter as follow:

1: Methods that separate water from salts
  1.1: All distillation methods
  1.2: Reverse osmosis
  1.3: Crystallization (freezing and hydrates)

2: Methods that separate the salts from the water
  2.1: Electrodialysis
  2.2: Ion-exchange
  2.3: Piezodialysis
  2.4: Osmionic methods

3: Methods with phase change
  3.1: All distillation methods
  3.2: Crystallization

4: Methods without phase change
  4.1: Reverse osmosis
  4.2: Electrodialysis

From the energy point of view, the methods are classified as follow:

5: Methods using heat (thermal methods)
  5.1: All distillation methods except mechanical vapor compression

6: Methods using mechanical energy
  6.1: Mechanical vapor compression
  6.2: Reverse-osmosis

7: Methods using electrical energy
  7.1: Electrodialysis

8: Methods depending on chemical energy
  8.1: Ion-exchange

The methods which found practical applications in large scale industrial plants are:

Distillation methods: which comprise the following modifications:

1:	Multiple-Effect Evaporation	ME
2:	Multi-Stage-Flash Evaporation	MSF
3:	Vapor-Compression methods	VC
4:	Solar distillation method	SD

Distillation is the most developed process of removing water from a saline solution. It is applied up to very large capacities with various types of evaporators and accounts for about 59.4% of the total world plant capacity.[5]

The latent heat of changing phase is an important factor in the overall process economics, but the degree of salinity of the raw water is of no importance. Multistage flash distillation and multi-effect evaporation are reducing considerably the economic effect of the latent heat of vaporization.

Reverse osmosis uses mechanical energy, as pressure, to drive the water out of the solution through semipermeable membranes. The applied pressure must be higher than osmotic pressure, its value depending from the salt content of the brackish or seawater solution. The necessary counterpressure in reverse osmosis depends greatly upon the salt content of the raw water and imposes constraints on membrane life and performance, but also varying energy consumption according to the salinity of the raw water. Membrane life is an important cost factor.

Today reverse osmosis plants account, worldwide, for 32.6% for plants having a capacity 100 to 4000 m^3/d and 19.5% for capacities over 4000 m^3/d.

Electrodialysis is the most developed process for eliminating salts from aqueous solutions. The economics depend closely on the salt content of the raw water, as the consumption of electric energy is related to the total dissolved solids removed from the solution. Electrodialysis may, therefore, preferably be applied for the purification of brackish waters. Reversal electrodialysis is a modification, by which poles are reversing every 20′ and which assures the production of high-quality water and minimizes the rejection of brine.

Today electrodialysis accounts for 5.7% of plants with capacities 100 to 4000 m^3/d and 2.5% for capacities over 4000 m^3/d.

Freezing processes found no commercial application though the simplicity of the method. They failed because the size of produced ice was very small and half of the fresh water was used to wash out the salt from ice surface, rendering the method uneconomical.

Independently from the method or procedure for sea or brackish water desalination the operation of a desalination plant includes some general steps to be followed. Figure 2 gives the procedures before and after the main desalination step.

ENERGY SOURCES

Running a desalination plant many expenses arise, the highest of which is energy cost. In a normal chemical plant energy cost is low, only 1 to 5% and, in some extreme occasion, 10% of the total operation cost. On the contrary, desalination is a high energy consuming procedure and the cost of necessary minimum energy to run the plant is 40% of the total cost.

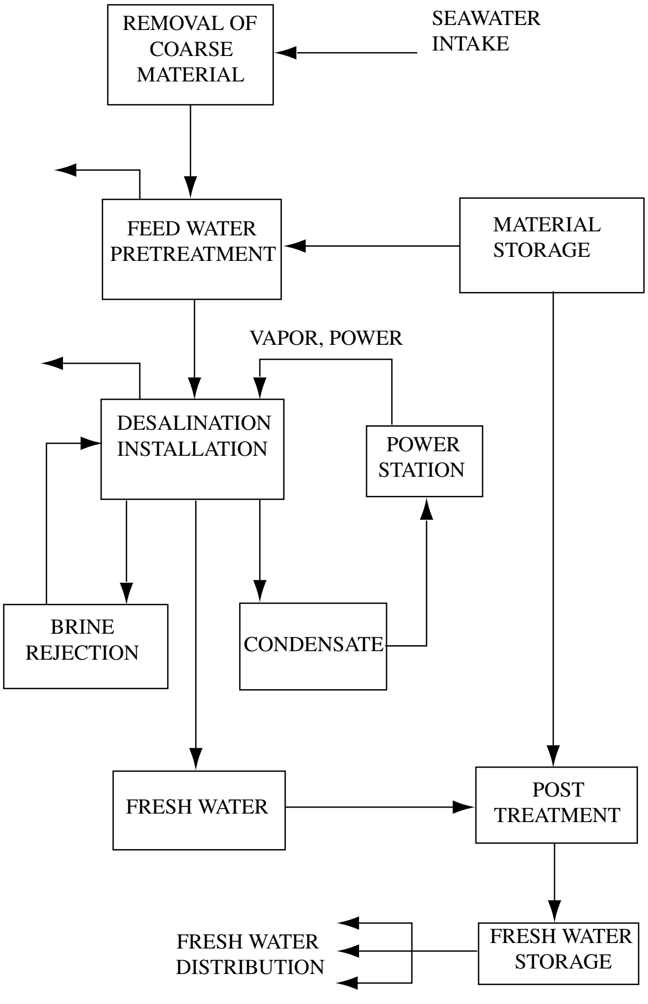


FIGURE 2 Flow diagram of the main procedures to be followed in the operation of a desalination plant.

The main energy sources, depending on the method, are low pressure stream and electricity, two energy sources easily available in any industrialized region. Few other energy sources are given at lower cost or free of charge. These alternative energies are suitable for small capacity plants and/or for remote and arid regions, where fuel and electricity are not available or the cost of fuel transportation renders its use uneconomic.

Alternative energy sources include geothermal energy when and where is available, all kinds of waste heat and waste heat from nuclear plants.

Renewable energy sources include wind energy, tidal energy, Ocean Thermal Energy Conversion (OTEC) and, above all, the abundant solar energy.

Waste heat is available from chemical industry, power plants and nuclear power plants in large amounts, but in low heat content. Wind and tidal energy are available in certain specified regions, transforming the corresponding energy into electricity. OTEC takes advantage of the temperature difference between the ocean surface and about the 500 m depth of the tropical regions. Solar energy is for the time being the most promising renewable energy. In the earth's sunny regions solar radiation is very intensive though also very spread out, thus the capture of solar energy depends on large areas.

Although solar, wind and tidal energy are natural forces given free, the corresponding equipment for transformation of these energies into a usuable form are yet very expensive and the yield very low.

DISTILLATION PROCESSES

Aristotle, the ancient Greek philosopher, wrote: "Salt water, when it turns into vapor, becomes sweet, and the vapor does not form salt water again when it condenses." Sailors have used simple evaporation apparatus to make drinking water for almost 400 years, and ocean going ships have traditionally used evaporators, often multiple-effect, as an accessory to steam boilers.

The simplest way to evaporate water is the natural one, using solar heat. Sun is a free inexhaustible source of energy. However, this energy has not been captured and stored at its most concentrated form as yet. The way to use solar energy for desalination purposes depends on the desalination process. The simplest and most common method is the direct use of the solar energy in specific equipment called "solar stills" which act simultaneously as converters of solar energy to heat and as distillers.[8]

Indirect use of solar energy, called "solar assisted" or "solar driven" desalination, captures the solar radiation using one of the modern procedures which transform the energy into either heat or electrical power. Horizontal tube, multiple-effect (HTME), multi-stage-flash (MSF) and thermal vapor compression (TVC) distillation methods are coupled to the heat source, though reverse osmosis (RO), electrodialysis (ED) and mechanical vapor compression (MVC) to the electrical power produced from the sun's radiation.[9]

As the incidence of solar radiation varies over the day, the time of the year, the degree of cloudy weather and the geographic location, conventional solar evaporation can never be a steady state operation. Moreover, convectional solar distillation is a single effect process and is characterized by the thermal disadvantages of single stage operation.

The intensity of solar radiation reaching the earth varies from zero to about 1047 W/m^2. Part of this radiation may come directly from the sun, but sometimes as much as 10% of it comes as scattered light, even when the atmosphere is unobstructed by clouds. In cloudy weather the total radiation is greatly reduced and most of the light that passes through may be scattered light.

The solar radiation striking a horizontal surface is greatest at noon, as the sun's rays pass through the atmosphere with a minimum length of passage through the air. In the morning and the afternoon the rays are subject to increased absorption and scattering. Considering the latitude, maximum radiation is at the equator. Hence the radiation intensity depends on the hour of the day, the day of the year, and the clarity of the atmosphere for a given location, as well as of the latitude of the earth at the point of observation. These limitations of the solar radiation render solar distillation method and solar driven desalination a nonsteady state operation except if solar energy storage is provided, which in general increases installation costs.

The daily production of conventional solar distillation is low, due to low performance of the stills. Depending on the intensity of solar radiation, the day of the month and the month of the year the fresh water production ranges from 1.5 to 5.51/m^{2}d (0.036 to 0.130 gal/ft^{2}d).[10]

Increasing feedwater temperature the daily productivity increases as well. This can be done by connecting a solar still with a solar collector or by using the condensate from low pressure steam. Many other methods have been proposed, to augment the efficiency of solar stills, nevertheless without any success due to increase of the corresponding costs.

To calculate the efficiency or the daily productivity of the solar stills have been proposed many mathematical models. Here two general equations are given: One concerns the operation of a conventional solar still and the second the productivity of a solar still connected to a solar collector.

The daily output of a stagnant solar still is given by the equation:[11]

$$M_{out} = F_1 H_d + F_2(T_{ad} - T_{wd}) + F_3 \tag{1}$$

and the daily output of a solar still connected to a solar collector is:[12]

$$M_{out} = F_p H_d + F(\overline{T}_{swd} - \overline{T}_{ad}) + F_2. \tag{2}$$

Both equations depend on construction and operational parameters.

Much material is required to construct a solar still: glass or plastic for the cover, black basin surface to absorb the solar radiation, material for the basin, usually concrete or plastic, pumps and piping—metal or preferably plastic, for the feed water and the fresh water distribution.[13,14,15]

Total cost of installation and operation of solar distillation plants is not very high if land is given free. They need large condensing areas and are vulnerable to storms. However, energy is free, except pumping, operation is simple, and maintenance cost very low.

Although the advantage of cost-free energy is partly offset by increased amortization cost and the large installation area, distillation with solar energy remains a favorable process for small-capacity water desalting at remote locations where there is considerable solar radiation. Most solar distillation plants are being (or will be) erected in less developed countries or in areas where there are limited maintenance facilities.

Solar energy for evaporation was first used on a major scale about 1872 in Chile, where a glass-roofed unit had 4,400 m^2 to make 22.4 m^3/d ($\sim$6000 gpd) in a mining camp.[16] Today many units, glass covered or plastic ones, are installed in small capacities world wide, mainly in arid and remote areas. Figure 3 is the photograph of a solar distillation plant, glass covered, yet in operation, in Porto Santo (Madeira) Portugal, with an installation area of 1200 m^2.[17]

It seems to be very simple as a method, and really it is, because theoretically solar energy can replace any other energy source. From a technical point of view this is not yet totally feasible because either the corresponding technology is not fully developed or the market is still very expensive.

Both procedures, solar distillation and solar driven desalination, depend on local insolation rates which vary from site to site for the same region, from the time of the day, the time of the year and the cloudy weather making desalination an unsteady state operation. Heat storage, if possible, improves productivity by extending operation during the nighttime or during cloudy days but also affects directly the economics of the method. However, for certain locations as remote, arid or semi-arid regions, where the small communities are poor and where the techniques and tools of water production and distribution developed in industrialized areas are not always appropriate to be used, solar desalination is admitted as the most suitable process.

The other way of using solar energy for desalination purposes is the collection of solar energy by solar collectors or concentrators, with subsequent conversion of the solar energy to heat or electricity. This solar assisted desalination is expanding rapidly and many installations have been erected in commercial but as yet small capacity sizes.

The simplest thermal conversion type of collector is a solar pond. A solar pond is a shallow body of water in which a stabilizing salinity gradient prevents thermal convection, thereby allowing the pond to act as a solar trap. The merit of solar ponds lies in their ability to collect solar energy in large scale and provide long-term heat storage. This long-term storage provides also increased flexibility of heat use. They can operate at all latitudes and are estimated to be less

FIGURE 3 Photograph of the Solar distillation plant in Porto Santo, Madeira, Portugal. It is the only solar plant in operation in Europe. Has a total evaporating area of 1,200 m^2, and consists from two different kinds of solar stills, of the assymetrical type. The Greek design developed at the T.U. of Athens and the design developed by the university of Berlin.

expensive than flat plate collectors per unit area installed and per unit of thermal energy delivered. Solar ponds, being low-grade heat source, can be competitive with convectional heat sources in many applications.

Flat-plate collectors, evacuated tube collectors and focussing collectors are used to produce hot water or steam as the heat medium for the distillation units. For reverse osmosis or electrodialysis units, photovoltaic devices are used or thermal conversion systems, e.g., central receivers, to drive the turbine generator.

A very important aspect of the solar assisted desalination process is the cost of energy and water produced. However, experience has shown that cost estimates are different everywhere. Labour, material cost, etc. depend on local circumstances, so the cost of water is not the same at all places.

Solar assisted desalination capacity is only a very small percentage, about 0.80%, of the total world capacity of convectional-fossil fuel fired desalination plants. A part, 0.60% is coupled to collectors or photovoltaic devices and 0.13% are wind-driven plants. The total capacity of worldwide solar-driven desalination plants is only about 15,250 m^3/d, and wind driven as low as 2,530 m^3/d.[5]

Figure 4 presents the first known sketch for solar distillation equipment.[18]

Distillation process, operated with conventional energy sources, i.e., low pressure steam, are applied up to very large capacities by using various types of evaporators and are classified accordingly as follow:

Multiple-effect evaporator (ME)

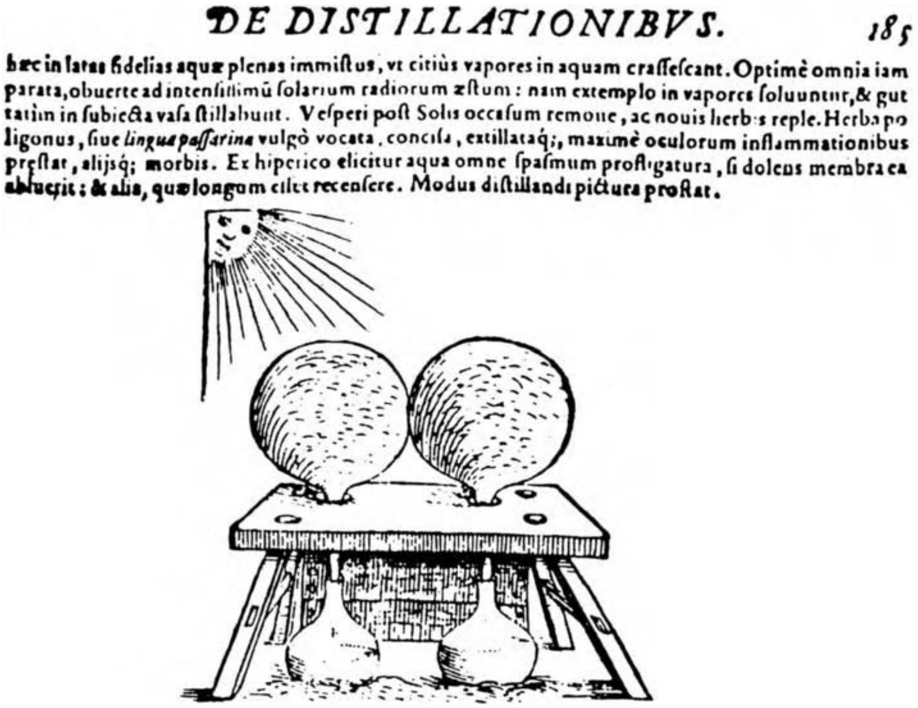
Vertical tube evaporators VTE, falling or climbing type
Horizontal tube evaporators HTE

Multi-stage-flash evaporator (MSF)

Vapor compression evaporator (VC)

Thermal vapor compression TVC
Vacuum vapor compression VVC
Mechanical vapor compression MVC

The term "evaporation" in the desalination refers especially to the vaporization of water from an aqueous saline solution, as brackish or seawater, where the solid constituents are practically nonvolatile, in the range of working



DE DISTILLATIONIBVS. 185

hæc in latas fidelias aquæ plenas immiſtus, vt citiùs vapores in aquam craſſeſcant. Optimè omnia iam parata, obuerte ad intenſiſſimũ ſolarium radiorum æſtum: nam extemplo in vapores ſoluuntur, & guttatim in ſubiecta vaſa ſtillabunt. Veſperi poſt Solis occaſum remoue, ac nouis herbis reple. Herba poligonus, ſiue *lingua paſſarina* vulgò vocata, conciſa, extillataq́;, maximè oculorum inflammationibus præſtat, alijsq́; morbis. Ex hiperico elicitur aqua omne ſpaſmum profligatura, ſi dolens membra ea abluerit; & alia, quæ longum eſſet recenſere. Modus diſtillandi pictura proſtat.

FIGURE 4 The first historically known solar distillation equipment, according to Giovanni Batista De La Porta. The sun evaporates the water inside the glass vessels and distilled water is collected beneath the vessels. "De distillations," Libri IX, Rome 1608.

temperatures and pressures. Thus, water alone is vaporized, which is the main product, and the dissolved solids remain in the residual liquid, the brine.

In the chemical industry, when an evaporation process is applied, the water vapors are usually discarded and the emphasis is given to the recuperation of the dissolved solids. In desalination the term "distillation" predominates over the correct term "evaporation." The process is performed in evaporators, where heat is supplied to the solution, to change phase.

The productivity is expressed either as the net evaporation or "gain output ratio" (GOR), i.e., the kilos of produced distilled water per kilo of boiler steam used, (kg/kg) or as performance ratio R. It is usually prefered, instead of the GOR, to use the term performance ratio which defines the mass of distillate produced per 2326 kJ (gal/1000 BTU) of heat input to the brine heater in case of MSF distillation or to the first effect in case of multiple-effect evaporators. The latter definition is thermodynamically more accurate, as it refers to the enthalpy of the steam instead to the mass.

Thermodynamic considerations lead to a common characteristic of all distillation process, that the percentage of evaporated water with respect to the circulating seawater is as much larger as is the difference between the maximum and minimum temperature of the saline solution. As the minimum temperature is defined by the temperature of the incoming seawater, enlarging of the temperature difference can only be obtained by increasing the initial maximum temperature of the salt water feed. Limitations due to the appearance of phenomena like scale formation and corrosion, which are becoming more important at higher temperatures, define an allowable maximum temperature for each distillation process. An appropriate pretreatment of the salt water is necessary to make an increase of the feed water temperature possible.

The economics of the distillation process might be affected by the following parameters:

- The chemical additives for feed water pretreatment
- Scale formation which decreases performance
- Increase of maintenance costs due to corrosion

Corrosion may increase fixed changes, when more expensive materials of construction must be used.

Multiple-Effect Distillation (ME)

Theoretically, in single-effect distillation 1 kg of distillate will be produced for every kg of steam consumed and the gain output ratio of the plant will be 1. In fact, despite preheating of the feed, a large part of the enthalpy of the vapors, evolved in the single-effect evaporator, is lost in the condenser. A better heat recuperation would be obtained if the heat, released by the condensing vapor, is not rejected in a condenser, but is used to heat the brine of a second evaporator and so on.

This leads to the concept of multiple-effect distillation, where the vapors from one effect are used as the heat source of the next effect, as long as the difference in temperature between the condensing vapor and the solution is high enough to act as the driving force in the evaporation process, each effect being at progressively lower temperature and pressure. Vapor condensing because of lower boiling temperature, in each effect, produces fresh water as distillate, whereas the vapor from the final effect is condensed by a circulating seawater cooling stream.

Theoretically, an additional kg of distillate would be obtained in each consecutive effect for the same kg of steam initially introduced into the first effect and the plant gain output ratio would be equal to the number of effects in operation. However, this is not true in practice. Part of the condensation heat to be recovered is lost to the atmosphere, in design features and in the differences of temperature used as the plant's driving force.

Multiple-effect distillation process uses evaporators which are modified successors of evaporators that have been used in sugar and other process industries for more than 100 years and have been in use for seawater distillation about 90 years. The latter were originally built for shipboard use, the main requirements being for compactness, simplicity in operation and reliability. In land based industrial evaporator plants the requirements are mainly directed to the cost of product water with emphasis on cheaper materials of construction, high boiling temperatures, efficient descaling methods and the use of the cheapest type of evaporator.

Previously multiple-effect distillation was second in importance of the distillation process, as medium capacity plants but day hardly is applied. Worldwide capacity of ME plants for units producing more than 100 m^3/d of fresh water, is only 765,143 m^3/day or 4.1% of total world capacity.[5]

Long Tube Vertical Evaporator, LTVE Long tube evaporators consist of a series of long tubes arranged vertically inside the evaporator shell. Seawater feed may be from the top or from the bottom, called respectively falling or rising film LTV evaporators.

In the falling film evaporator seawater is introduced at the top and the incoming seawater flows across an upper tube plate and is equally distributed to the tubes, and flows downward by gravity as a thin film. The principal advantage of the VTE process is that high heat-transfer can be achieved, which considerably reduces the required heat-transfer surface area. This forward feed is the usual method of feeding a multiple-effect-evaporator.

The VTE rising film is similar to falling film evaporator except that seawater is introduced at the bottom of the first effect, thus reducing the overall pumping requirements. Heat transfer in the VTE evaporators is increased by using fluted tubes, which enhance heat transfer.

Steam condenses outside the tubes, forming also a thin film of distillate. Surface tension forces are created, which are inversely proportional to the flute radius. This causes the condensate film to drain from the crests to the grooves, so that a very thin condensate layer is remaining on the crests, which promotes heat transfer.

The flow sheet of a typical multiple-effect vertical tube distillation plant is shown in Figure 5. In each effect is adapted a feed heater C which uses the product vapors as heating medium in the form of distilled water or vapor condensate. Vapors produced in the first effect condense outside the tubes of effect 2 and the brine is pumped from each effect to the top of the next. The average efficiency K of each effect is usually between 0.85 and 0.95. Concerning N effects in a LTE system, as in Figure 5, the GOR is given by the equation:

$$\mathrm{GOR} = K(1 - K)^N/(1 - K) \tag{3}$$

Thus when $K = 0.95$ and the number of effects N is 15, GOR = 10 kg/kg. Doubling the effects to 30, the GOR is only 14.9, and it will attain a maximum of 19.0 for an infinite number of effects, when $K = 0.95$.

There are some economic limitations increasing the number of effects. The investment costs and consequently the fixed charges are increasing almost linearly with the number of effects. The costs of steam and water fall off rapidly at first, but the savings diminish progressively. The total cost of operating an evaporator leads to an optimum number of effects, at the point where the sum of fixed costs and the cost of utilities shows a minimum. The most probable number of effects will be between 10 and 20.

The Multiple-Effect Horizontal-Tube Evaporator The (MEHT) type of evaporator operates on the same principle as the VTE evaporator, but the steam condenses on the inside of the horizontal tubes imparting its latent heat of condensation to the brine, which cascades and evaporates over the outside of the tubes. The brine falls to the next effect by gravity and the vapors formed in one effect are used in the next effect. The horizontal-tube evaporator eliminates the pumps required for each effect of the VTE brine circulation, by an arrangement in which the effects are stacked vertically on the top of each other. This compact arrangement of the ME evaporators, called also multiple-effect stack (MES), is constructed in low capacity units and though there are many advantages, it accounts for only 1% of the world capacity. In Figure 6 a typical HTE evaporator is presented.[19] Multiple-effect-horizontal-tube evaporators are suitable to operate with solar energy plants.

VAPOR COMPRESSION (VC)

Vapor compression (VC) distillation takes advantage of the latent heat of the vapors produced in the process. Vapor produced by evaporation from a salt solution is superheated because of the boiling point elevation of the solution and

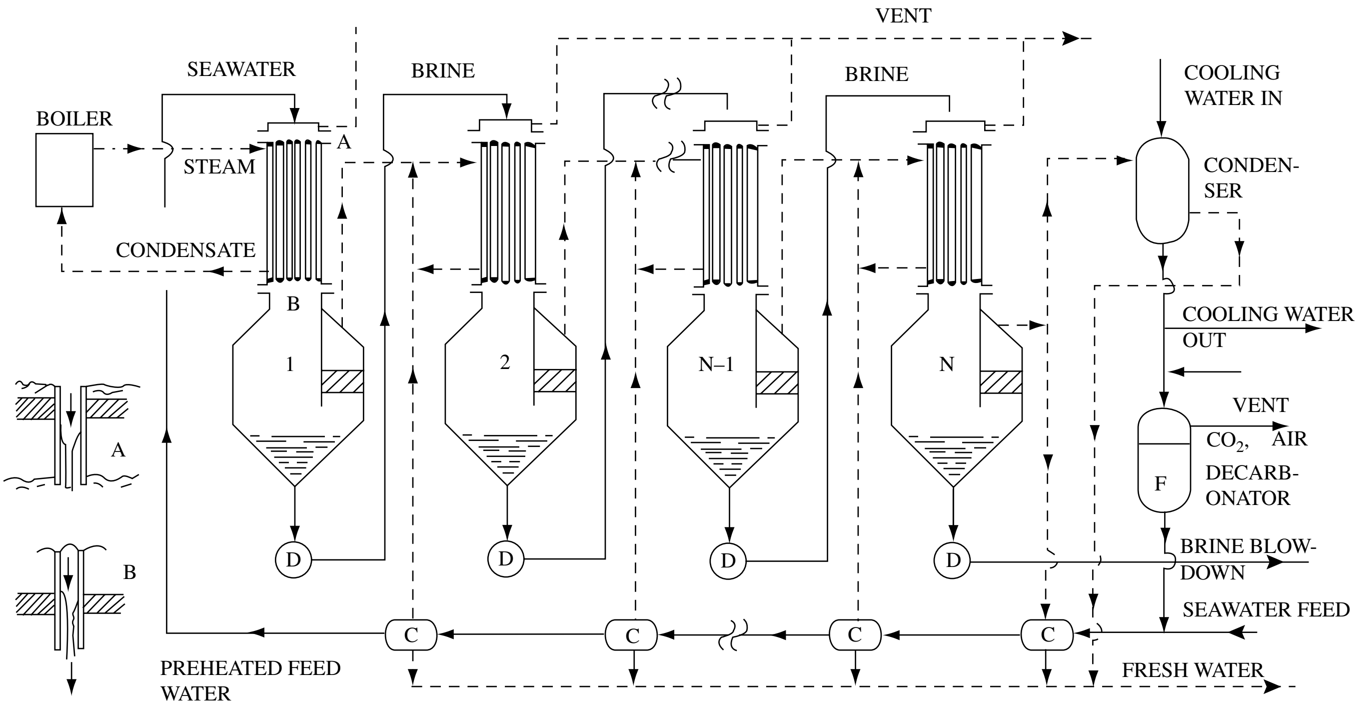


FIGURE 5 Flow diagram of a multiple-effect-evaporator for seawater desalination, of the falling type. Seawater is preheated in the heaters C and pumped on the top of the first evaporator No. 1, from where falls down inside the vertically oriented tubes B. A thin film of brine is formed inside (detail A and B). In the first effect steam from the boiler forms a thin film of condensate outside the tubes. In the following effects the vapors of each effect condenses outside the tubes of the next effect. The rest of the space of the evaporator is then filled with water vapors. The brine accumulates in the bottom of the evaporator from where is fed to the next effect by the pumps D. A distributor cap is fitted on each tube to ensure even distribution of the brine. The produced fresh water is used to preheat the seawater fed in the heaters C. Part of the seawater is acid treated and a decarbonator F, is used for the removal of air and CO_2.

has a lower pressure than the saturation pressure of pure water. It will, therefore, in losing the superheat, condense at a lower temperature than the boiling point of the solution. If this vapor is compressed to a higher pressure, the energy input results in a rise in temperature. With sufficient rise in pressure and temperature, the recompressed vapor might be used as a source of heat for evaporating the same salt solution.

Heat needs to be supplied to the system only at the start-up for elevating the temperature of the solution to the boiling point. Once boiling has started, it is maintained by the external supply of power and no more by the addition of heat as the cycle is repeated.

Vapor compression distillation accounts only 3.7% of total world wide desalination capacities and about 6.2% of distillation processes, for units producing 100 m^3/d or more fresh water. Daily productivity is 686,500 m^3/d.[5]

The energy source may be mechanical or electrical power to drive the compressors for mechanical vapor compression. Thermal vapor compression, or “thermocompression,” uses high pressure steam to compress the vapors to higher temperatures. Vacuum vapor compression uses electric or waste heat to reheat the vapors, circulating by the use of a blower.

Mechanical Vapor Compression (MVC) The MVC process uses compressors to reheat the vapors to higher temperatures. High or low pressure compressors are used, depending on the capacity of the system. As high capacity plants have many stages, the necessary temperature is higher and they are high pressure compressors. Low capacity plants use low pressure compressors. The higher the compression pressure, the smaller is the volume of the compressor and that of the vapor. Due to high temperatures, high capacity MVC plants are prone to scale formation.

Figure 7 presents the flow-sheet of a four-stage mechanical vapor compression evaporator and the T-S diagram of the thermodynamic operation.[20]

Thermocompression Thermocompression uses high pressure steam ejector to re-heat the vapors released from the last stage. Figure 8 gives a typical two stage thermal vapor compression diagram.

Another type of thermal vapor compression operates under vacuum inside the evaporation chamber. The low pressure vapors are circulated by a vapor blower and heated by electric heater, hot water or hot gas, according to the available heat source. A suitable adaptor for the various heat sources is necessary.

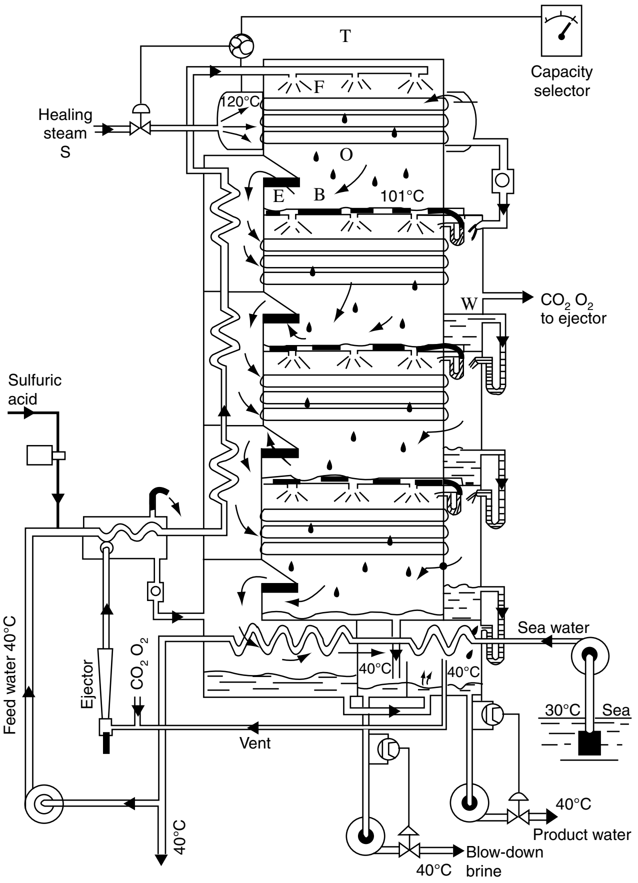


FIGURE 6 Flow-diagram of a multiple-effect-horizontal tube evaporator. The effects are vertically oriented, one on top of the other. This arrangement is compact and is called multiple-effect-stack type (MES) distillation equipment.[19] Preheated seawater feed is sprayed, F, onto the outer surface O of the evaporator tubes in the first effect at the top of the column, T, where a portion of seawater is evaporated by the heating steam S. The remaining seawater is collected at the bottom B, of the first effect and then sprayed onto the outer surfaces of the second effect where another portion of seawater is evaporated, being heated by the vapor generated in the first effect. The generated vapor is delivered through a mist eliminator section, E. The vapor itself condenses into fresh water in the side section W. The cycle is repeated in each successive effect up to the last one. Vapors generated in the last effect is condensed in a heat rejection condenser C.

Multi-Stage-Flash (MSF) Distillation

When saline water is heated to a temperature slightly below its boiling point at a given pressure and then introduced into a chamber where a sufficiently lower pressure exists, explosive boiling will occur. Bubbles are evolving from the whole mass of the liquid and part of the water will evaporate until equilibrium with its vapor at the prevailing pressure is reached.

This evaporation lowers the temperature of the remaining brine. The liquid may then be passed into another chamber at an even lower pressure, where it flashes again to vapor. If a higher rate of saline water circulation is supplied, an increased proportion of flash will occur. The increased flow rate may be considered as a means of obtaining increased evaporative yield in a system without increasing the evaporating surface. It is, therefore, equivalent to diminishing the evaporating surface.

Flashing of vapor requires a finite residence time of the liquid in the evaporation chamber in order to achieve near equilibrium conditions. For a given flow rate the residence time is determined by the chamber length. Mass-transfer rates in two-phase flow depend on the interfacial geometry of the two phases and on the degree of turbulence. They accordingly determine the residence time required and thus the size of the flashing chamber.

On the other hand, the length of the flashing chamber must be sufficient to achieve the required temperature rise of the incoming seawater under the acceptable maximum velocity inside the condensing tubes. As there are limitations for both brine and feed-water flow rates, the width of the flashing chamber becomes the important determinant for increasing the plant size. Figure 9 presents a flow diagram of a vacuum vapor compression unit.

MSF is the most widely applied distillation process, especially for large units, and despite the thermodynamic advantages of ME evaporation, all major plants installed are of the MSF principle because of the simplicity and reliability of the process. It accounts for 51.5% of the world desalination capacity and 86.9% of total distillation processes. The capacity of MSF plants capable of producing 100 m^3/d per unit or more fresh water was, by the end of 1993, about 9,640,000 m^3/d, 51.5% and 8,960,000 m^3/d, or 71.7% of world desalination capacity for desalting plants producing more than 4,000 m^3/d unit fresh water.

Multi-stage-distillation process, as applied in large scale desalting of seawater, may be considered as consisting of three sections in handling heat: the heat input section, usually named brine heater, by condensing external steam; the heat recovery section, in which the heat of the evaporation is recovered in the condensers at the various stages; and the heat rejection section, which maintains the thermodynamic process by reducing temperature and pressure and accounts for the last stages of the plant.

MSF distillation plants operate with recirculation part of the brine. Recirculation can be applied so far as the concentration of the scale-forming compounds does not reach, after the evaporation, the critical point. It is a disadvantage of this design that the brine concentration at the hottest stages of the plant is much higher than the concentration of dissolved solids in the seawater. The fact limits the maximum brine temperature of the process.

Operating with this cycle arrangement, the maximum operating temperature with acid injection is limited to 121°C (250°F), with brine concentration 1.5 times. The number of stages is usually limited by 2°C flashdown per stage, because of the low pressure differential available at the deep vacuum conditions prevalent in the last stages.

The total number of stages is affected by the initial and final brine temperatures, as well as by the necessary

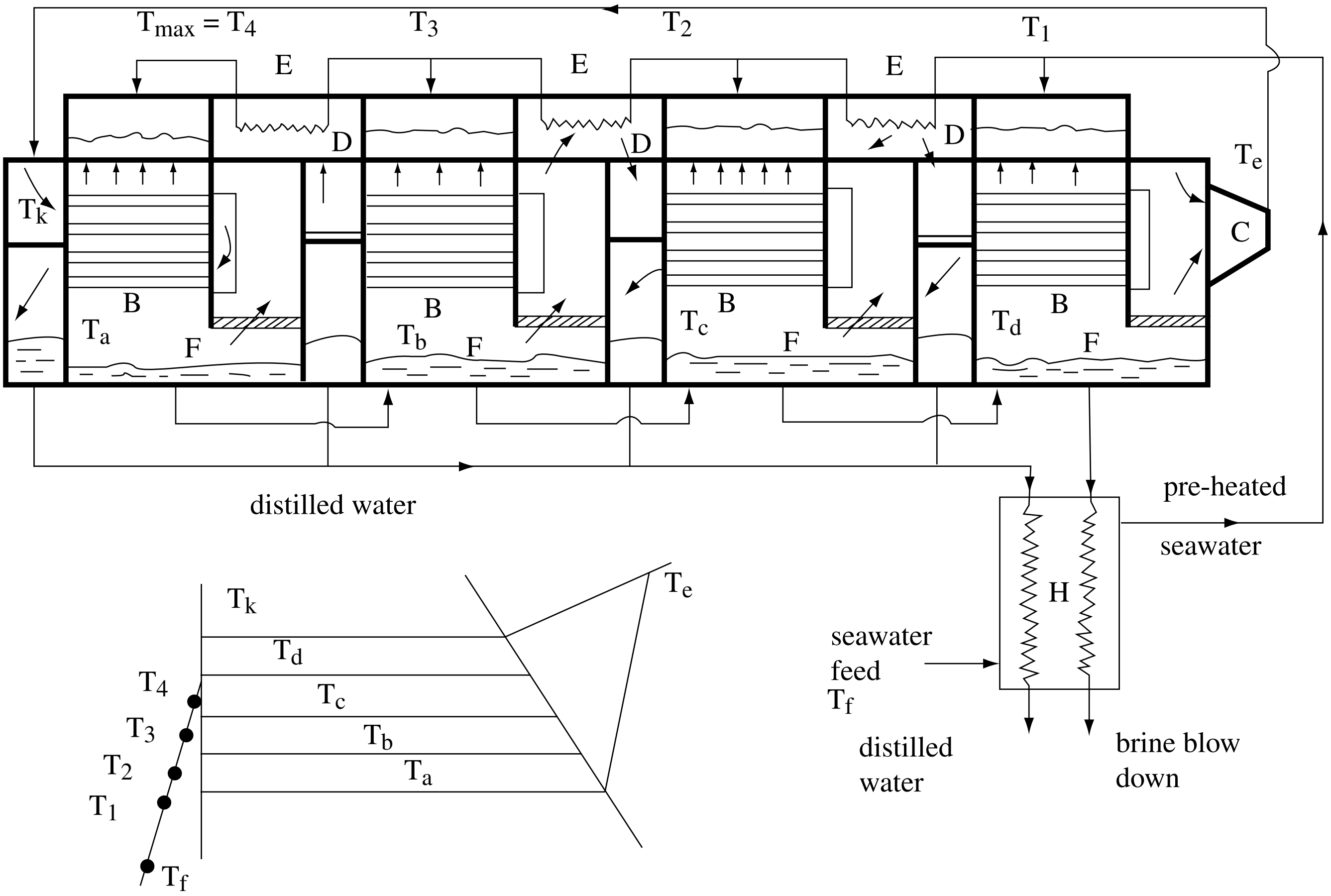


FIGURE 7 Flow diagram of a typical 4-stage, mechanical vapor compression plant. Incoming seawater feed is preheated in the heat-exchanger H, by the produced freshwater and the blowdown brine. The vapors released in the first stage are flashing to the second stage and so go on up to the 4th (or to the nth stage). The vapor from the last stage is compressed in the compressor C. Electric power is generated in a turbine to drive the compressor C. Compressed steam is circulated through the tubes of the condenser B, where it condenses, giving the heat to the evaporating seawater. Released vapors are used as heating medium in the second stage, etc. The condensate, i.e., the produced freshwater, leaving the flashing chambers F is collected after further cooling in the heat-exchanger H and brine is rejected.

temperature gradient between stages, to maintain the thermodynamic cycle. However, there are practical limits in the excessive increase of the number of stages. The additional investment to provide further stages to the plant should be reasonable with respect to the savings of heat obtained by the same stages. In large multi-stage flash evaporators for desalting of seawater, the cost of the heat transfer tubes is a very important part of the total construction cost. The temperature drop in each stage and the difference between brine-inlet temperature at the first stage and the discharge temperature at the last stage are the main controlling parameters of the MSF distillation process. In a cascading flashing stream of an MSF evaporator, the combined heat capacity of the flashing brine and distillate streams equals that of the recirculating brine and the temperature rise of the recirculating brine equals the temperature drop of the flashing stream.

The number of stages in an MSF distillation plant is related to the performance of ratio. An increase in the number of heat-recovery stages will generally result in a higher performance ratio for a given product-water output and in a decrease of steam consumption at the brine heater.

The performance ratio R can be correlated to the number of stages by the following relation:

$$R = \frac{\text{no recovery stages}}{\text{no reject stages}} = \frac{T_{max} - T_f}{T_o - T_F}\,\frac{\text{kg}}{2326\ \text{kJ}}. \tag{4}$$

Vapor bubbles almost explode to carry entrained brine to the product. Demisters D (Figure 5) reduce this but give additional temperature losses up to 0.11°C (0.2°F) per stage. Low flash temperatures per stage reduces flash violence and entrainment but this increases the number of stages to 40 or more and also increases equipment costs and inefficiency, which increase capacity and material costs.

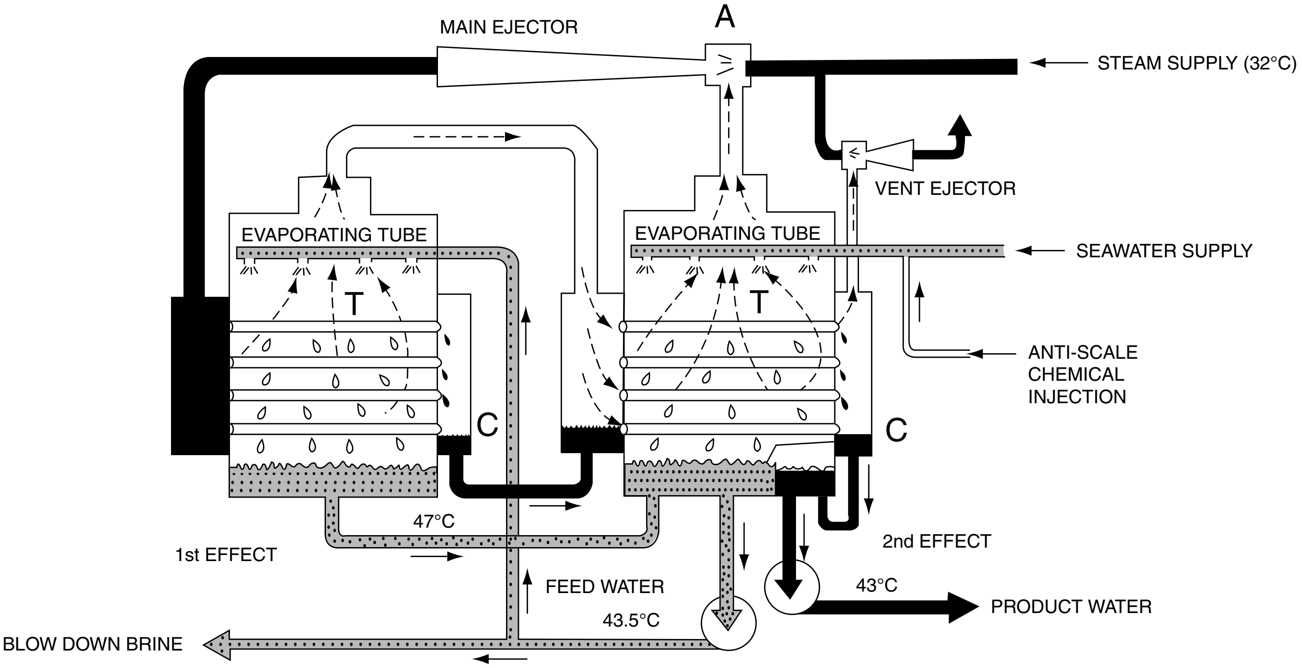


FIGURE 8 A two effect reheat or thermal compression unit. The low pressure, low temperature vapors from the second effect are sucked by the steam jet ejector A, driven by a small quantity of high pressure boiler steam, and delivers a hotter compressed mixture of stream and vapors to the condenser tubes T of the first effect. The seawater feed is sprayed onto the outside of the horizontal condenser tubes T. Part of the rejected brine from the last stage is used as feed to the first stage. Condensate, or product fresh water is collected in the last stage chamber C and distributed through the pump P. (Courtesy Sasakura Engineering Co., Ltd., Japan.)

Temperature differences between brine and vapor streams leaving stages may be from 0.45 to 1.7°C (~3/4 to 3°F) besides boiling point elevation.

In Figure 10 the flow diagram is given of a multi-stage-flash evaporator and the temperature profile across the plant. Figure 11 gives the two arrangements of the multi-stage-flash chambers. Configuration No. 1, with long tubes is preferred by American construction companies. The cross type No. 2, is usually preferred by European contractors.

Combined Distillation Plants

Significant economic advantages may be expected from combining different distillation processes, especially where desalting plants are designed with large capacities. Studies and design experience indicate that the combined system possesses substantial advantages in cost compared with the unit form of multi-stage plant of the same capacity. The savings in cost are due primarily to lower capital investment, and lower operational and maintenance costs. A further advantage of the combined plant is its high operational stability at varying loads.

Many designs have been proposed and many combinations have been tried, mainly for small capacity or pilot size plants. Commercial application used the vertical tube-multi-stage-flash (VTE/MSF) and the vertical tube-vapor compressor processes. Tubes with fluted surfaces are used in the vertical tube evaporator plants to obtain enhanced heat transfer performance. The heat recovery section of the multi-stage-flash plant is used as feed preheater of the vertical tube plant, which is the main evaporator for the distillate production. The combined vapor compression-vertical tube evaporator process uses the heat recovery section of the multi-stage-flash as preheater for the vapor compression plant as it is more efficient than the heat-exchangers in the single vapor compression process.

Scale Formation and Its Prevention

Formation of scale deposits on and fouling of heat transfer surfaces is one of the most serious problems of distillation equipment operating with sea or brackish water. As the scale deposits lower the efficiency of heat transfer surfaces and increase the pressure drop, pretreatment of feed water is necessary to prevent the deposition of scale.

As the salt concentration increases during progressive evaporation, the critical point may be reached at which the solubility limit of scale-forming compounds contained in the feed water, is exceeded and formation of scale occurs. The term scale is applied particularly to describe hard, adherent, normally crystalline deposits on the heat transfer

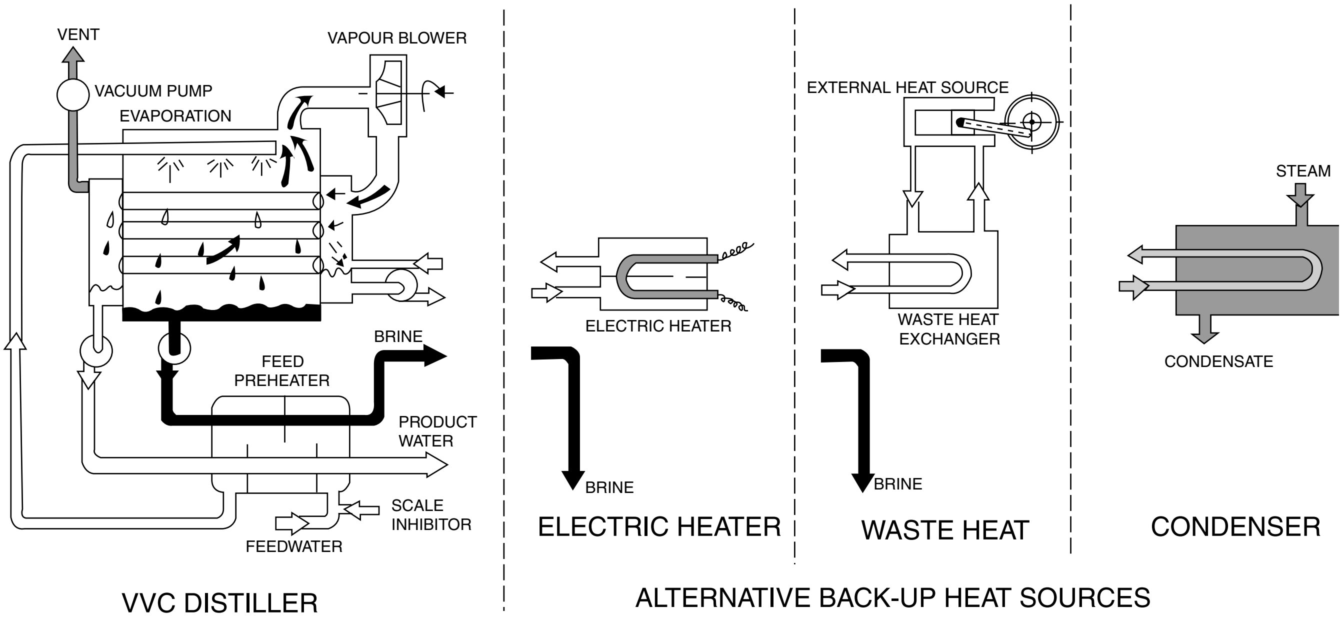


FIGURE 9 Flow diagram of a vacuum vapor compression (VVC) unit. The vacuum vapor compression distillers are small capacity units, and can produce fresh water from any kind of water, as dirty waters, with low energy consumption. It operates without acid treatment at 70°C (158°F) and uses, as heat source, electricity or can be combined with hot gas or hot water 80°C. The salt water feed is preheated and sprayed to the top of the evaporator A, where it is distributed as a thin film over the outside of the tubes T. The thin film boils at the tube surface due to condensation of the hotter vapor inside the tubes. The vapor thus created is compressed in the blower B and its temperature is raised, before it passes to the inside of the tubes T, where it condenses to form fresh water. Most of the heat is thus effectively recycled. C, D, E, represent the auxilliary parts of heating which can be adapted to the main distillation unit U, according to their availability. (Courtesy Sasakura Engineering Co., Ltd., Japan.)

surfaces. Three simultaneous factors are required for the formation of the scale:

1. Local supersaturation of the solution.
2. Nucleation, which when formed includes the rate of further scale deposition.
3. Sufficient contact time of the solution and the nucleus.

Under certain conditions a soft, amorphous material, called sludge, may be deposited or remain suspended in the brine and is generally more easily removed than hard scale.

If the ions contained in seawater are combined in the form in which they usually deposit, the resulting compounds will be approximately:

$CaCO_3$	109 mg/L
$CaSO_4 \cdot 2H_2O$	1,548 mg/L
$MgCl_2$	3,214 mg/L
$MgSO_4$	2,233 mg/L
NaCl	26,780 mg/L

assuming that hydrogen carbonate decomposes to carbonate before precipitation occurs. Alkaline scale, $CaCO_3$ and $Mg(OH)_2$, results from the decomposition of the hydrogen carbonate ion. On heating seawater up to 82°C (180°F), the hydrogen carbonate ion decomposes and calcium carbonate is formed.

A second type of scale, called acid scale, is due to three forms of calcium sulfate: the anhydrite $CaSO_4$, the hemihydrate and the dihydrate $CaSO_4 \cdot 2H_2O$, or gypsum. While the precipitation of $CaCO_3$ and $Mg(OH)_2$ is mainly affected by CO_3^{2-} concentration, pH and temperature affect the solubility of calcium sulfates in addition to the concentrations of other ions present. $CaSO_4$ has as well decreasing solubility in the temperature ranges of interest. The solubility increases in chloride solutions, as the concentration approaches 4 to 5% chloride and then decreases to values comparable to those in chloride free water as the chloride concentration becomes 10 to 15%.

Maximum brine temperature provided in the design, maximum allowable brine concentration and brine recirculation rate are also affected by the formation of scale. These operating variables and the plant availability are closely tied to the economics of the process as the production rate is generally lowered. Periodic plant shutdowns for descaling would be required either by an acid clean or, in extreme cases, by mechanical cleaning of the tubes. Incrustation allowances to reduce the frequency of shut-downs are made in designing evaporators, which are provided with a sufficient larger heat-exchange surface in order to maintain the design capacity. The term fouling is often extended to this type of admissible scaling.

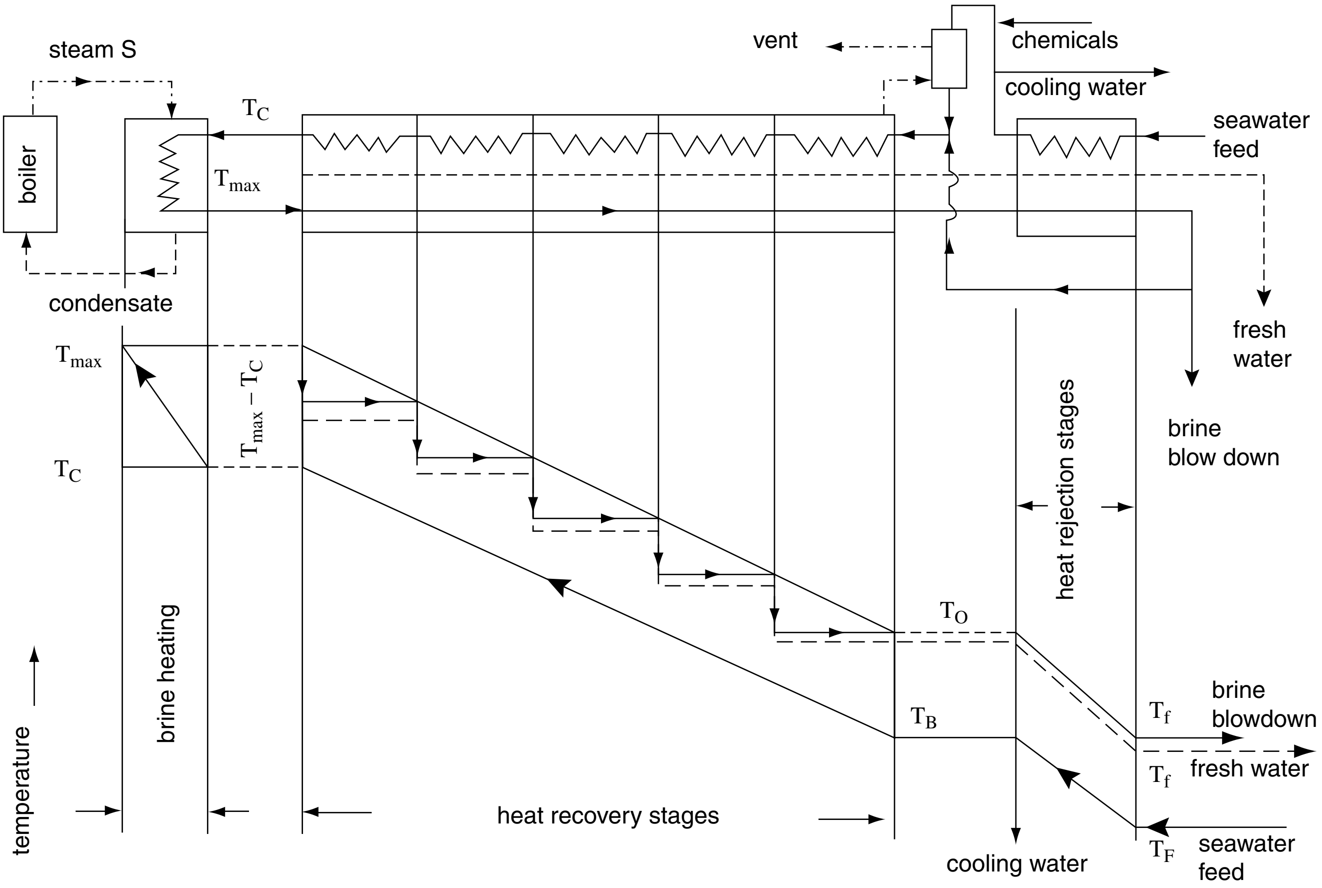


FIGURE 10 Flow diagram of a multi-stage-flash evaporator with brine recirculation, and temperature profile across the plant. Cold water feed is pre-heated inside the condenser tubes of the heat rejection stage A, then circulating through the condensers of the recovery stages, is heated to temperature T_c. Finally is heated by steam and reaches the highest temperature in the process T_{max}. By this temperature is fed to the 1st stage of the recovery section. The recovery and the rejection sections are enclosed in a single long vessel. The rejection stage removes excess heat from the flashing brine. Each stage is operating at a lower pressure than the preceding stage, with temperature fall from stage to stage. The vapor generated in the flash chambers condenses on the tubes of the condenser giving its latent heat of condensation to the heated seawater into the tubes. Leaving the heat rejection stage, brine is blow-down. Usually 50% to 75% of the brine recirculates after mixing with cold seawater. In the temperature profile, the stepped line shows the temperature fall at each stage of the heat recovery and heat rejection sections. T_{max} is the maximum temperature to which the seawater feed is heated. T_f is the discharge temperature of the brine in the last stage and T_o is the outlet temperature from the recovery stage to the rejection stage. T_F is the temperature of cold seawater feed. These temperatures are correlated to the performance ratio R.

There are several techniques used to avoid deposition of scale. Calcium carbonate and magnesium hydroxide formation can be controlled by acid injection and pH adjustment or by the addition of polyphosphates. Sulphuric acid transforms the carbonates to sulphates so that only one type of scale-forming salt remains present. Phosphates precipitate calcium and magnesium as sludge, minimizing the effects on the heat transfer surfaces. Either of the chemicals is introduced into seawater before its entry into the deaerator. In the deaerator the dissolved gases, together with carbon dioxide evolved during acidification, are eliminated. The pH of the treated seawater is controlled by the addition of dilute caustic soda, which binds the remaining carbon dioxide.

Calcium sulphate scale is more difficult to control. When formed on heat transfer surfaces, removal is difficult, if not impossible. To prevent deposits two methods may be applied. Seed crystals are injected into the hot seawater to promote precipitation of scale-forming compounds on these seeds, which then form a sludge rather than a deposit on the heat transfer surfaces. Ion exchange treatment is applied to eliminate completely both calcium and magnesium ions from the solution.

A technique of seawater pretreatment is the LMC or lime-magnesium carbonate process. 70 to 80% of the calcium originally present in seawater is removed and this permits operation at higher temperatures and concentration

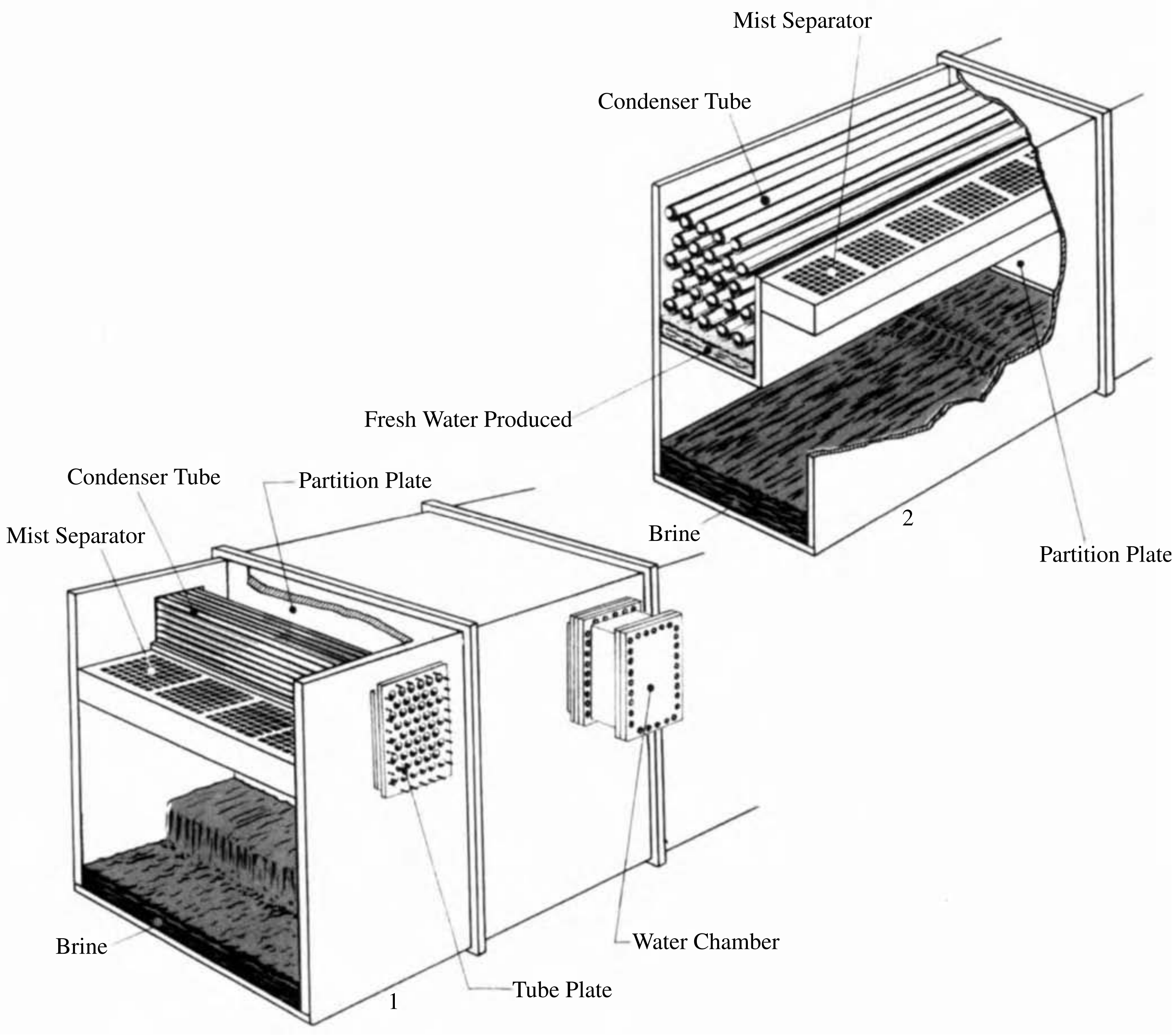


FIGURE 11 Drawing of a long tube multi-stage-flash (No. 1) and a cross tube multi-stage-flash chamber. The long tube configuration has a condenser running through each stage and the tubes are parallel to the flow direction of the flashing brine. High technical skill is necessary to seal the partition plates where the condenser runs through the partition plate. Long tube MSF plants are adopted for larger desalination plants. The cross tube type desalination units have the tubes across the top of the flash chamber and at right angles to the flow direction of the flashing brine. The cross tube configuration has condensers independent of each other and is mainly used for smaller desalination plants. (Courtesy of Sasakura Engineering Co., Ltd., Japan).

factors, thereby increasing the water recovery ratio for saline solutions and favourably affecting the economics of the desalting process. Treatment of the seawater with H_2SO_4, on the other hand, reduces the corrosiveness of the water by eliminating dissolved oxygen and carbon dioxide in the deaerator.

Acid treatment converts bicarbonate to carbon dioxide gas, whereas caustic treatment yields the carbonate ion, which combines with the calcium ion present in seawater to precipitate calcium carbonate. Nucleation of scale on the heat transfer surfaces is inhibited and precipitation of solids is dispersed in suspension by the addition of small amounts of certain commercial preparations. The sludge formed is then removed in the blowdown.

The first commercially available scale control compound was Hagevap, a mixture of sodium tripolyphosphate lignin sulfonic acid derivatives and various esters of polyalkylene glycols. Polyphosphates act as sequestrants for calcium and magnesium ions. Lignin sulfonic acid, starch, tannin, etc. act as dispersants in coating surfaces such that scale adherence and crystal growth are inhibited. Polyalkylene-glycols are surface active agents which tend to retard foaming of the seawater.

The use of polyphosphate based additives is limited to temperatures below 90°C (190°F). Above this temperature, polyphosphates undergo chemical changes which restrict their effectiveness as antiscaling agents. Other additives are low molecular weight polyacrylic acid or 10% ethyl acrylate-acrylic acid copolymer. An optimum polymer concentration of about 2 mg/L was observed to be most effective.

Materials of Construction: Corrosion

The selection of suitable metals to construct desalination plants is of prime importance as the use of inadequate materials may lead to shutdowns, increased replacement and maintenance costs and affect the overall economics of the plant.

The optimization of a distillation process has as a first object the minimization of the amounts of thermal or mechanical energy and of the amount of equipment and, hence, the amounts of materials used. Quite often, in such optimization studies, as energy goes down equipment goes up, and vice versa. Thus the most economical balance must be struck. Often this is at a point where the energy cost and the capital cost of the equipment are about equal.

Of the billion dollars per year to be spent for plants, materials for the equipment might be regarded as at least 50%, with costs of engineering, fabrication, transportation, installation, etc. accounting for the rest. Of materials used in equipment by far the largest amount will be for metals, particularly those metals which are least corroded by seawater. These are not the most abundant or least expensive. The specification of these metals for equipment—those which are suitable for withstanding the corrosion and other deleterious effects of this service—are of greatest importance. Their fabrication into sheets, tubes, shapes, and then finished vessels and parts and accessories will be a great test of the skill of the metallurgist, the chemical engineer, and the mechanical engineer.

Membrane Processes

Membrane processes are classified in two main categories:

- Methods which separate salts from the water, in their ionic form, called ionic processes. Electrodialysis and ion-exchange are the separation processes for desalination.
- Methods which separate water from a salt solution. Reverse osmosis, nanofiltration, ultrafiltration and microfiltration are the main processes of this type of separation which are applied to desalination or to water purification.

Ionic Processes Common salt, and other salts as well as acids and bases, are ionized in solution into the positive (for example, sodium ions) and the negative (for example, chlorine ions). Whereas in distillation processes the amount and kind of salts dissolved in the raw feed water are of no importance to the process and do not affect the economics, in all ionic processes the amount of dissolved salts is of primary importance. In electrodialysis, the amount of salts to be removed affects the consumption of electrical energy and in ion-exchange affect the amount and the cost of regenerant. Thus, ionic processes are much too expensive to use with the higher salt concentration of seawater, as compared with brackish, river or estuary water. In general ionic processes depend on specially developed polymeric resins in the form of membrane in the case of electrodialysis and as granules for ion exchange.

Ion Exchange The chemical system for removal of the ions of salt is called ion exchange and has long been used in the reverse process in which, for example, the sodium ion of salt is used to replace calcium ions to “soften” hard water. In this case the “bed,” of ion-exchanging material consists of granules of resin which have the property of removing positive or metal ions from a circulating aqueous solution by displacing them by a chemical bond with the resin of the bed. This replaces them with ions which are loosely bound in the molecular structure of the resin. When the positive ions of the resin are completely replaced, an aqueous solution, relatively concentrated in the other positive ion, is cycled, and the process is reversed. Other beds of resin have the properties of exchanging negative ions.

Similarly the sodium ion of a salt solution such as seawater passing through a suitable first bed may be interchanged with a hydrogen ion from the resin to leave the sodium ion in a loose combination with the resin and a resulting hydrochloric acid in the effluent solution. This solution is then passed through a second resin bed, in which a hydroxyl ion interchanges with the chlorine ion to give a resin chloride combination and an equivalent amount of chemically formed water which is added to the aqueous stream.

When the two resin beds have interchanged all of their respective hydrogen and hydroxyl ions, their activity ceases and they must be generated. This is done by stopping the flow and adjusting valves to take the beds off-stream and to pass through the first bed a dilute solution of sulfuric acid and then to pass through the second bed a dilute solution of caustic soda. The hydrogen ions of the sulphuric acid displace the sodium ions on the resin of the first bed to give sodium sulphate in its effluent which is run to waste and the hydroxyl ions of the caustic soda displace the chlorine ions on the second resin to give sodium chloride (salt) in the effluent of the second bed, which also runs to waste. Again, the ions on the beds are completely interchanged, the beds are thus generated and the controlling valves are adjusted, with those allowing the sequential flow of the original seawater being opened to allow the process to be repeated.

This process thus required both sulphuric acid and caustic soda in amounts which are chemically equivalent to the amount of salt present. Usually, at least 50 to 100% excess of each is used. This will represent a very high and impossible cost for large-scale desalination of seawater.

In special cases, particularly for desalinating brackish water, ion-exchange beds have been used, since only a relatively smaller amount of chemicals is required to interchange ionically with the salt in the feed solution. Thus water containing 1750 ppm of salt would be regarded as non-potable, but it would require only about 5% of the chemicals that seawater would require. On the other hand, the energy cost for its desalination by evaporation or RO would not be so greatly different from that required for seawater because in these processes it is the water which is being separated.

For emergency kits, packages of ion-exchange resins have been made to have only a single use, e.g. by aviators downed at sea. Seawater is passed through these small "beds" to make a small amount of drinking water. Provision for regeneration of the resins would be complicated and the resins are discarded when charged with sodium and chlorine, respectively.

The simplicity of ion exchange has attracted much effort to finding less expensive methods of generation of the ion-exchange resins. Carbon dioxide is a weak but cheap acid, as lime is a cheap alkali, and special resins and processes have been developed to use them and also to use the major differences of ion exchange equilibria at 80°C as compared to those at room temperature.

Ion exchange processes depend upon good resins. Excellent resins are available and they are not the limiting factor. The major cost is that of the chemicals required for the much less than stoichiometric replacement of salt by two chemicals. In desalination the ion-exchange process is used mainly for pre-treatment of brackish water for electrodialysis or reverse osmosis.

Electrodialysis, ED

Electrodialysis is the transport of ions through ion-selective membranes as a result of an electrical driving force. The process takes advantage of the ability of these membranes to discriminate between differently charged ions, allowing for free passage to either cations or anions and being impermeable to ions of the opposite charge. Electrodialysis is a desalination process of brackish water and, under certain circumstances for seawater as well.

The electrical mechanism of ion removal is much more complicated, and much cheaper, than ion exchange since it uses electrical energy rather than chemical equivalence to replace the two ionic changes of a molecule salt, since an electric current assists greatly the dialysis or movement of the ions through membranes permeable to the positive ions and to the negative ions, respectively. A membrane which is permeable to sodium ions forms the wall on the side of a channel of flowing saline water and a membrane permeable to chlorine ions forms the wall on the other side. The deionized water flows between the two membranes, and the electric current may be regarded as flowing at right angles. The other aqueous streams on the other side of the respective membranes may flow out counter-currently in the other direction (Figure 12).

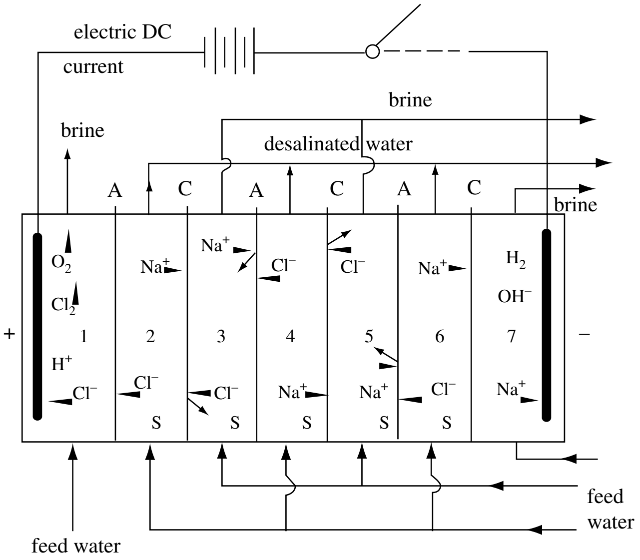


FIGURE 12 Flow diagram of a multicomponent electrodialysis cell, with 7 compartments showing the principle of the process operation. The saline water feed is pumped through the compartments, s, of the membrane stack and, when a direct current potential is applied, cations pass easily through the cation-permeable membrane and are stopped when they reach an anion-permeable membrane. Similarly anions have free passage through the anion-permeable membrane and are stopped at the cation-permeable membrane. The ion concentration increases in the alternate compartments.[3,5] Simultaneously the compartments between them become depleted of ions.[2,4] Hence two streams of water are extracted from the electrodialysis stack: one stream with low ion concentration, which is the product water, and one stream with high salt concentration, which is the reject brine. Separate feed and brine blowdown is provided for the first and last compartments containing the anode and cathode respectively, since corrosive oxygen, chloride and hydrogen gases are released.

The electrodialysis process is performed in cells consisting of many compartments formed alternatively by an anion and a cation exchange membrane placed between an anode and a cathode (Figure 12). Multicompartment electrodialysis cells are usually termed as electrodialysis stacks. A membrane pair is called a "cell pair" and consists of a:

- Cation transfer membrane
- Anion transfer membrane
- Demineralized water flow spacer
- Concentrate water flow spacer

A typical membrane stack contains 300 to 500 cell pairs, depending on feed salinity. If the feed water has a low salinity, it is possible to obtain an acceptable potable water in a single pass through the electrodialysis stack. If the feed water is high in salinity this is not normally practical. Feed velocities have to be maintained above a certain minimum value and this requirement is handicapped by the necessity to provide also for a minimum residence time of the fluid in the compartment. An increase of the current density may lead to concentration polarization, a phenomenon that can stop process operation.

Figure 13 is a view of an electrodialysis stack before assembly, that shows the main components. Typically the design is based on the configuration used in the plate and frame filter press. The end frames have provisions for holding the anode or cathode and are usually made relatively thick and rigid that pressure can be applied to hold the stack components together. The inside surfaces of the end frames are recessed to form an electrode-rinse compartment and provisions are made for introducing and withdrawing the solutions. Spacer frames with gaskets at the edges and ends are placed between membranes to form the solution compartments when ion exchange membranes and spacer frames are clamped together.

These are various electrodialysis systems:

- The conventional batch-type which was the first commercially developed system.
- The continuous-type unidirectional electrodialysis system.

Both types of the electrodialysis systems have some operation disadvantages and limitations. Ionic movement is unidirectional. In such systems cations are moving toward a fixed cathode and anion toward a fixed anode. Sealing forming salts, colloidal particles or slimes, slightly electronegative are accumulated on the surface of the anion-exchange membrane causing membrane fouling.

- Electrodialysis reversal (EDR) is a system designed for continuous operation. The polarity of the electrodes is reversed 3 to 4 times per hour.

This operation system reverses the direction of ion movement within the membrane stack, controlling thus scale formation and fouling.

To day almost all electrodialysis cells are constructed to operate with the reversal arrangement. Figure 14, gives a flow diagram of a reversal electrodialysis system, in a vertical arrangement.

The electrodialysis process is related to Faraday's law which states that the passage of 96,500 Ampere-s transfers theoritically one gram-equivalent of salt. The needed current for transferring a specific quantity of salt is given by the equation:

$$I = \frac{Fm_d \Delta c}{eN}. \tag{5}$$

If current efficiency e is 100%, 96,500 A-s will transfer one gram-equivalent of sodium ions, or 23 g Na^+, to the cathode and one gram-equivalent of chloride ions, or 35.5 g Cl^- to the anode.

Electrodialysis is applied usually to brackish water desalination up to 7 g/kg salinity. Sea water can be used also to produce water around 0.50 g/kg or less but energy consumption is very high. In reversal electrodialysis scaling is very low, almost zero, but in conventional ED pretreatment of feed water is necessary.

Energy input for electrodialysis depends on feed water salinity and varies from 19 to 20 kWh/m^3 (72 to 76 kWh per 1000 gallons) for seawater feed. For brackish waters the required energy is 2.4 kWh/m^3 (9 kWh per 1000 gallons) for

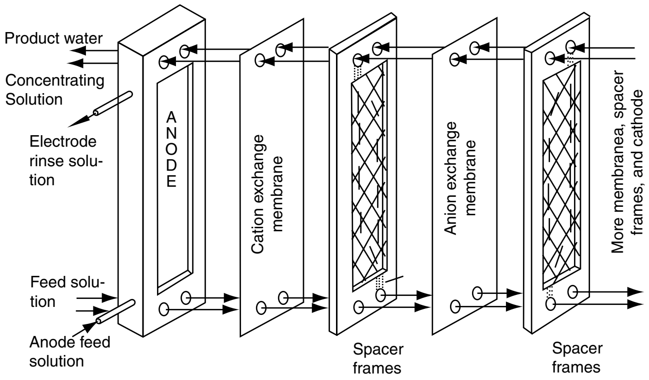


FIGURE 13 View of part of an electrodialysis stack before assembly. Component 1 is the one end frame, each of which holds an electrode and has opening for feeding and withdrawal of the depleting, the concentrating and the electrode rinse solutions. Compartments 3 and 5 are spacer frames which have gaskets at the edges and ends and form solution compartments with the membranes when they are clamped together. Many membranes and spacers clamped together form an electrodialysis unit.

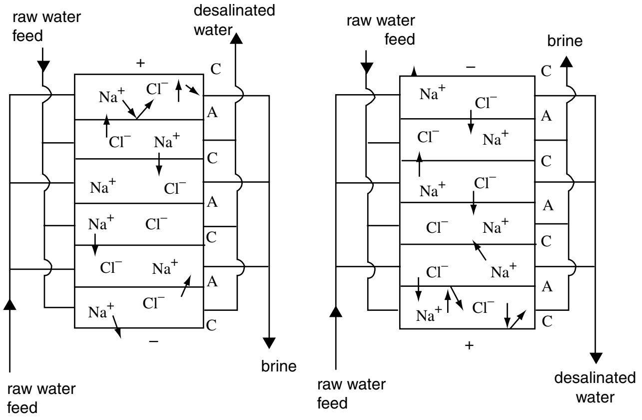


FIGURE 14 Flow diagram of an electrodialysis reversal cell. B represents the one-way operation and D the operation after the reversing of the polarity. Cathode or anode are reversing and become anode and cathode respectively. Ion movement reverses as well and concentration compartment acts as dilution compartment. Blowdown brine circulates in fresh water tubes and vice-versa.

feed water having 2,500 ppm and about 3.9 kWh/m^3 (14.8 kWh per 1000 gallons) for feed water containing 4,500 ppm total dissolved solids. The pumps operate at low pressure of 350 to 700 kPa (51 to 101 psi).

Electrodialysis has a variety of applications, in addition to that of fresh-water production, i.e., for the treatment of wastes, for the recovering of trace elements from effluents and for concentration of various solutions, such as fruit juices.

To the end of 1993 world capacity of electrodialysis plants producing more than 100 m^3/d/unit was 1,070,005 m^3/d covering only 5.7% of the total desalination capacity, and 14.8% of the membrane processes.

Ion Exchange Membranes The ion-exchange membranes are membranes selective either to cations or to the anions. When a cation-selective membrane is immersed in an electrolyte solution the cations in solution will enter in the resin matrix and replace the cations present. The anions are prevented from entering the matrix by the repulsion of the anions affixed to the resin. The opposite phenomenon takes place when an anion-selective membrane is immersed in an electrolyte solution. Ion exchange membranes are essentially ion-exchange resins cast in sheet form.

Membranes of synthetic resins have been developed which are highly selective to the passage of positive ions, and others have been developed which are highly selective to the passage of negative ions. Hundreds of membranes form passages in parallel between somewhat like a plate and frame filter press., which is the electrodialysis stack. A stack is presented in Figure 15.

Counter-ions within an ion exchange resin or membrane are ions with a charge opposite to the charges affixed to the membrane matrix. Co-ions are ions with the same charge as the fixed charge of the matrix. Hence, ion selective membranes are selectively permeable to counter-ions and selectively impermeable to co-ions. The selectivity, which might be expressed in terms of the transference number of counter-ions in the membranes, is not generally restricted to all ions of the charge. Membranes may well be more selective to some ionic species than to others. There are commercially available membranes that will selectively transport univalent ions in electrodialysis. With these membranes not only the concentrations but also the composition of electrolyte solutions can be altered. A membrane possessing specific selectivity between divalent and univalent ions would be useful for removing sulfate from a solution of chloride and sulfate, or to fractionate a mixture of the ions into two solutions, each one containing only one of the ions.

Transport numbers of ions in the ion exchange membrane are different from those in the solutions on both sides of the membrane. Because of the lower transport number of ions in the solution, the number of ions transported to the membrane surface by the electrical current is in deficiency to the ions removed from that surface and transferred through the membrane. The opposite phenomenon occurs on the other side of the membrane. A greater number of ions are transferred from the entering to the outgoing membrane surface than can be carried away by the electrical current. Two boundary layers with the opposite concentration gradients are formed at both sides of the membrane. This tendency for concentration and depletion is opposed by diffusion and physical mixing. Hence, the thickness of the boundary layers depend on hydrodynamic conditions and on the degree of turbulence. However, there remain layers adjacent to the membrane in which the solutions are completely static.

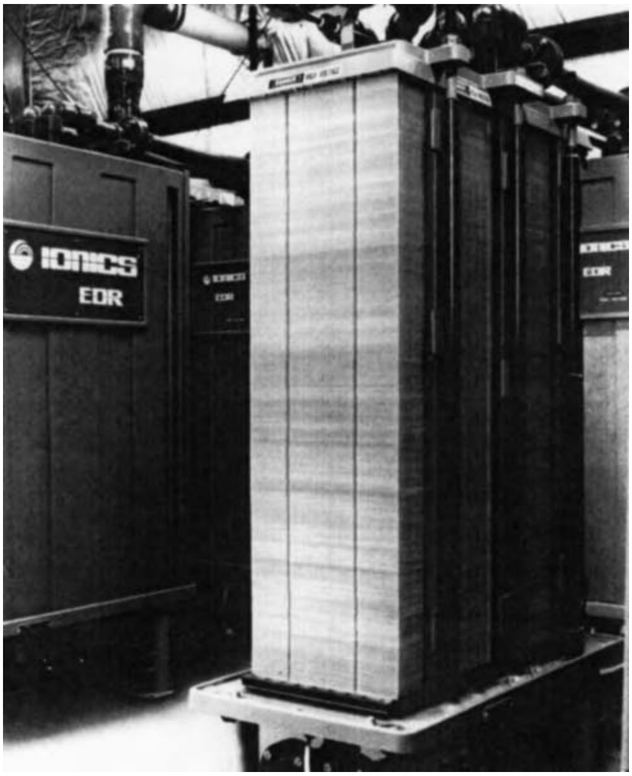


FIGURE 15 Photograph of an electrodialysis reversal stack with 500 cell pairs in vertical arrangement. (Courtesy Ionics Inc., Watertown, Mass. USA.)

Increasing the current density has the effect of increasing the concentration gradients at both surfaces of the membrane and the point may be reached at which the concentration of ions at the entering side of the membrane approaches zero. This is the limiting current density. When the limiting current density is exceeded, hydrogen and hydroxyl ions are transported through the solution and the membrane causing changes of pH inside the membrane and at the boundary layers of the solutions, as well as an increase in the overall electrical resistance. The desired ions participate with only a small amount in the transport. The phenomenon is termed concentration polarization and is the major limitation of the production rates achievable by electrodialysis.

The increase in the pH of the solution associated with polarization promotes the formation of alkaline precipitates such as calcium carbonate and magnesium hydroxide on the membrane surface. Membrane scaling causes additional electrical and flow resistance, a decrease in electrodialysis efficiency and an increase in pumping power requirement and mechanical damages in some occasions.

Ion exchange membranes are very thin, about 0.5 mm (0.020 inches), and they are supported by spacers as shown in Figure 13. Spacers are made usually from two layers of low density polyethylene having manifold cutouts which match the membrane cuts as it is shown in Figure 16.

Membrane life is short, from 3 to 6 years, depending on scale formation, fouling and the poisoning of the membranes. Poisoning of the membranes occurs from chemical agents such as chemicals for pre-treatment. The membranes lose their characteristic properties, their volume is increased, their water content decreased and they deteriorate easily. Scaling, fouling and poisoning of the membranes have influence not only on membrane life but also on membrane efficiency, which drops considerably after long period use.

Reverse Osmosis, RO Reverse osmosis separates pure water molecules from a salt containing solution through a semipermeable membrane having extreme fine pores.

Osmotic flow, direct or reversed, depends on the selective properties of some membranes to allow certain components of a solution, usually the solvent, to pass through the membrane. This intrinsic property of the membrane is termed as semipermeability. If two solutions of different concentration, or a pure solvent and a solution, are separated by a semipermeable membrane, the solvent will flow under normal conditions from the less concentrated department through the membrane into the concentrated solution, with the tendency that both solutions reach the same concentration. This flow is known as osmosis. Osmotic flow through the membrane will stop when the concentrated solution reaches a sufficiently higher pressure than prevailing in the less concentrated solution or the solvent compartment. The equilibrium pressure difference between solvent and solution, or the two solutions, known as osmotic pressure, is a property of the solution. Equilibrium can also be reached by applying an external pressure to the concentrated salt solution equal to the osmotic pressure. Further increase of the pressure on the concentrated solution, beyond the osmotic pressure, causes reversal of the osmotic flow. Pure solvent passes from the solution through the membrane into the solvent compartment. By applying a pressure higher than the osmotic the phenomenon is reversed and water flows from the concentrated solution to the dilute. This is the basis of reverse osmosis, the major attractions of which, from an economic point of view are its simplicity and the relatively low energy consumption.

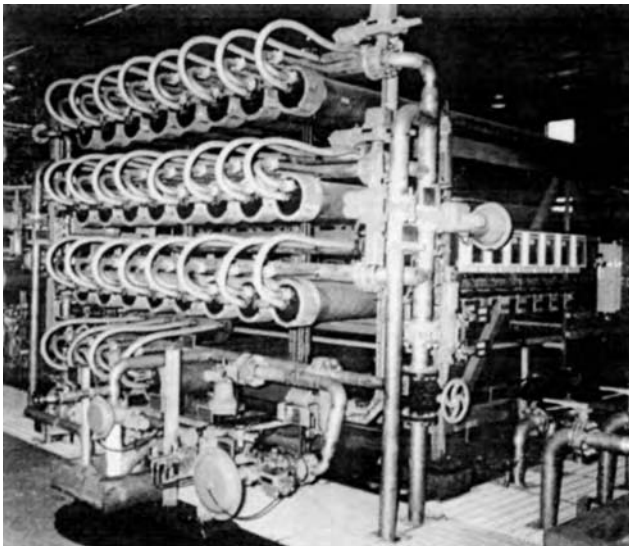


FIGURE 16 A cell pair spacers and anion and cation membranes with the corresponding manifolds. Water is circulating through the paths of the spacer. (Courtesy Ionics Inc., Watertown, Mass, USA).

Reverse osmosis has some analogies with filtration in that both remove substances from a liquid. As matter, removed by reverse osmosis is in solution, the process previously was termed "hyperfiltration," when the membranes used were highly semipermeable with respect to low molecular weight solutes. On the contrary, when the membranes used have little or no selectivity for such solutes but they may be applied for the separation of colloids or macromolecules from low molecules solutes, the method is called "ultrafiltration." In fact ultra, a Latin word, is identical to the Greek word hyper. Today high pressure, low molecular matter separation is called "reverse osmosis" and lower pressure low molecular separation is called "nanofiltration." By chemical engineering point of view reverse osmosis, nano-, ultra- and micro-filtration can replace, in some extent other unit-operation separation methods as these are shown in Figure 17. These four methods can be applied for the purification of water of solutions related or not to desalination methods and reverse osmosis and nanofiltration for desalination of seawater brackish or natural waters, as well.

Applied pressures are higher for the reverse osmosis method and low for microfiltration and, on the other hand, porosity of membranes is decreasing from micro-filtration to reverse osmosis as follow:

	Applied pressure, bars	Porosity
Reverse Osmosis, RO	30 to 70	5 to 20 Å
Nanofiltration, NF	20 to 40	10 to 20 Å
Ultrafiltration, UF	5 to 15	20 Å to 0.1 μm
Microfiltration, MF	1 to 4	0.1 to 2 μm

Reverse osmosis, for desalination of sea or brackish water, by the end of 1993 had a world capacity of 6,109,250 m^3/d per unit or 32.7% of total world capacity. This capacity gives RO second place in fresh water production worldwide.

The vapor gap osmotic distillation process is a forerunner of the reverse osmosis process. It is based upon the difference between the water vapor pressures over a saline solution and that over pure water.

Reverse osmosis as a process was developed first with the plate and frame concept, using the filter-press principle. This

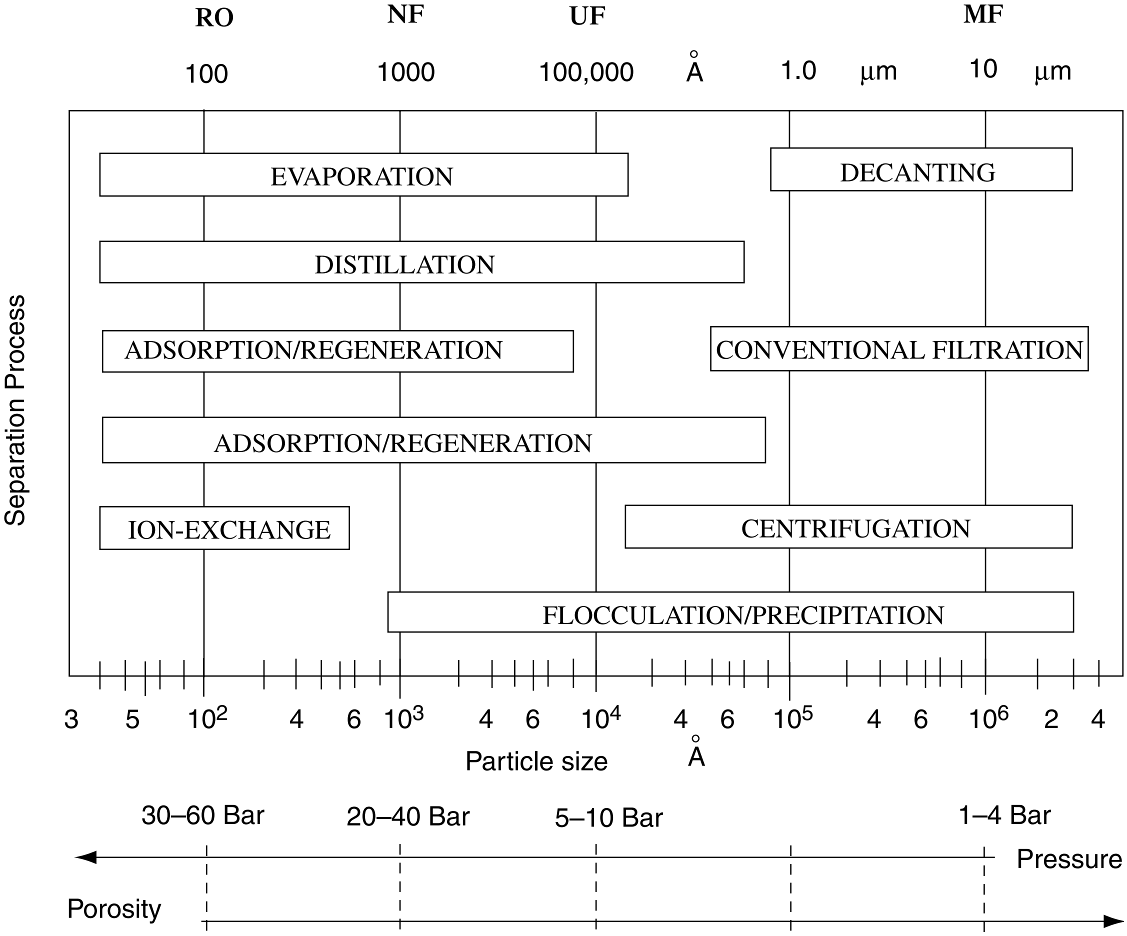


FIGURE 17 Particle size separation by various membrane and the range where they can replace the conventional chemical engineering separation methods. The particle size range for each membrane separation method depends on the porosity of the membrane, increasing with the particle size, and on the pressure which is decreasing proportionally with porosity increase.

configuration is now coming back, having many improvements and makes use of a rigid plate with the membranes mounted on opposite sides and sealed to the plate. Salt water under pressure is following on the outer side of the membrane and the product water is forced through the membranes into the interior of the porous plate, which serves as membrane support.

Two other configurations, are in use: The tubular and the spiral wound. Both, as the plate and frame concept, use flat sheet membranes, and another concept uses membranes in the form of hollow fibers.

All membranes are assembled in units called modules. Each module is self-contained and can be used as an independent reverse osmosis unit. Most often, many modules connected in series are used to increase capacity. Semipermeable membranes are very thin, as are ion-exchange membranes and fragile, and they need special support to withstand the pressure applied for the separation. Assembling the various types of modules special care must be taken to support the membranes properly. The fact that water flow through a membrane is proportional to the membrane area and inversely proportional to its thickness renders a module a device requiring the maximum possible area per unit volume of water passing through the membrane. The membranes are vulnerable to mechanical damage, fouling and scaling as well as to concentration polarization phenomenon. A perfect module has the following advantages.

- Minimizes the effects of concentration polarization.
- Has high resistance in seawater corrosion and in the cleaning chemicals.
- Continuous operation at high pressure.
- Almost constant quality of freshwater produced.

Reverse osmosis is a method used not only for desalination of sea or brackish water, but also for the purification of a large variety of various solutions in the chemical industry by using suitable semipermeable membranes.

Ultrafiltration and microfiltration do not desalinate waters directly but are used for pretreatment of various kinds of solution and water, depending on the molecular weight of the dissolved or suspended matter previous to reverse osmosis treatment.

Reverse Osmosis Membranes The reverse osmosis membranes are the main and more delicate component of the method. They are usually permeable to some species, as water and impermeable to other, as salts. They are characterized by two important parameters: The water flux, or water permeability and the salt rejection. Concentration polarization plays a significant role and influences both the above parameters.

The water flux at a given temperature is determined by the membrane properties and is defined by the equation:

$$J_v = L_p(\Delta P - \sigma\Delta\Pi)\ \mathrm{m^3/m^2s}. \quad (6)$$

Salt rejection is the ability of the membrane to reject the solution salts and leave only the solute, i.e., the water to pass the membrane mass. A perfect membrane would reject all salts contained in the feedwater and be highly permeable to the flux of water, or solute. Commercial membranes are not ideal and have certain amount of salts to move through the membrane. The salt rejection is defined as:

$$\%\ \text{Salt rejection} = \frac{\text{product concentration}}{\text{feedwater concentration}} \times 100. \quad (7)$$

A module characteristic is the recovery or yield which defines the fresh water product, m^3/d, to the feed water flow, m^3/d as well.

$$R = m_d/m_s \quad (8)$$

R is influenced by the required salinity in the product, fresh water.

The required mechanical energy to drive the RO pumps is determined by the required separation pressure, which in turn is proportional to the salt content of the feedwaters. It ranges between 8.0 to 12.0 kWh per m^3 freshwater produced, from seawater feed and 2 to 3.5 kWh/m^3 of brackish water feed. The corresponding pressure is 56 to 70 bars for seawater (812 to 1015 psi) and 20 to 28 bars (246 to 406 psi) for brackish water.

There is a large variety of membranes prepared from organic materials. For commercial use up to now the types of these composite membranes in general use are the ultrathin composite-spiral-wound membranes and the hollow fine fiber membranes. The best developed composite membranes are: Cellulose triacetate films deposited on a cellulose diacetate-cellulose nitrate support, furfuryl alcohol film on a polysulfone support and polyamide film on a polysulfone support. The polyamide-film membranes provide the best desalination performance. This development in membrane preparation brought the reverse-osmosis process next in importance to distillation during the last years so that distillation and reverse osmosis are today leading processes in seawater desalination.

The way by which the membranes are supported and the properties of the membranes define the reverse osmosis system.

A way to assemble membranes is the spiral wound module (SWM). Spiral membranes are cellulose acetate polymers or can be made of the thin film composite type. Spiral membranes have a diameter of 5.1 to 30 cm (2.0 to 11.8 inch) and a packing density of about 600 m^2/m^3.

Another module system is the tubular (TM) where the membrane surface is packed in a shell and tube arrangement. The membranes, usually cellulose acetate, are mounted inside the tubes made of metal or plastic. The active layer, i.e., the permeable layer, may be either on the inside or the outside of the tubes.

The hollow fine fiber module (FHFM) are very thin fibers, similar to human hair thickness. Their advantage is their high surface area. Inside the modules they present a packing area exceeding 30,000 m^2/m^3 of water produced. Millions of fibers are assembled inside a cylindrical bundle embedded in an epoxy resin tube sheet. Figures 18 to 20 present the reverse osmosis modules commercially available.

Freezing Processes

All variants of the freezing processes are based on the well-known phenomenon that, when a saline solution is cooled to its freezing temperature, ice crystals of pure water will form and the brine will be enclosed in the slurry.

The temperature of freezing is fixed by the concentration of the brine, while evaporation can operate over a wide temperature range. Freezing has basic advantages:

1. much lower latent heat of phase transition in the solid state than for evaporation (0.33 × 10^3 J/kg against 2.49 × 10^3 J/kg or 143.2 BTU/lb against 968 BTU/lb).
2. less heat losses (gains) because of working at temperature closer to the ambient;
3. no scale-formation from the usual impurities;
4. less corrosion of steel at the freezing point than at the boiling point of water; and
5. cheaper materials of construction may be used.

Against these there are basic disadvantages:

1. the time required for phase transition from liquid to solid is very much greater than that required from liquid to vapor;
2. handling the crystals of ice is very much more difficult than handling the fluids in evaporation processes;
3. separation of the pure water phase (ice versus steam) is extremely difficult;
4. the cost of removing heat energy is much more expensive than that of adding heat energy, and

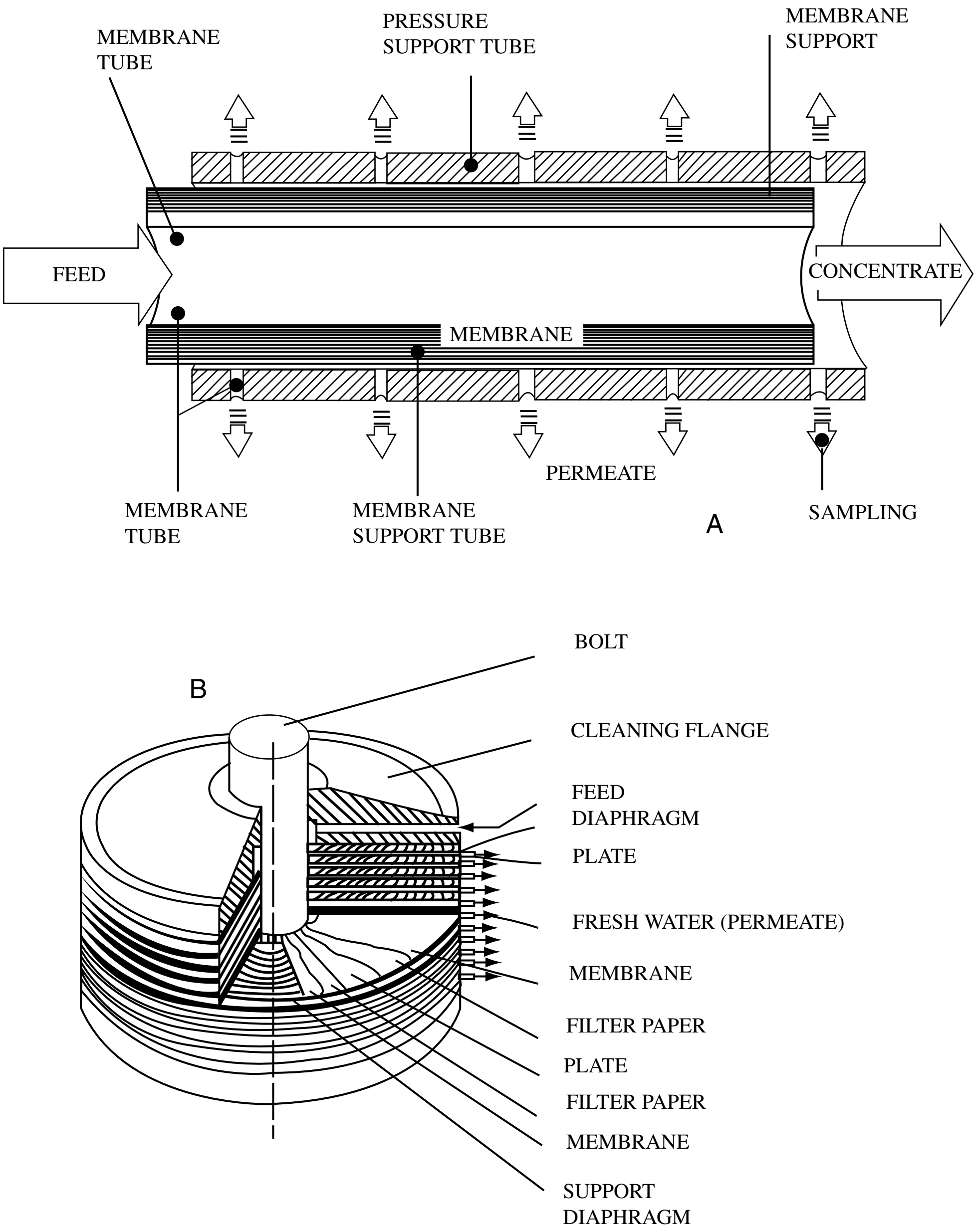


FIGURE 18 Schematic diagram of the reverse osmosis and ultrafiltration modules. A, the tube module arrangement. B, the plate and frame module arrangement. In the tube module membranes are formed as tubes with the active layer either inside the tube or the membrane support tube. In the plate and frame arrangement the membranes are flat sheets put onto the flat support diaphragms.

5. freezing operates, practically only at the freezing point of the final solution, and there can be no multistaging to reuse the latent heat of phase transition.

However, the major disadvantage of the freezing processes is the necessity of washing the ice crystals from adhering brine, an operation which inevitably consumes part of the product water.

As in all refrigeration, there are two types of processes for desalination of seawater:

1. mechanical compression of vapors, and
2. absorption of vapors by a hydrophylic liquid which then uses heat energy for desorption by evaporation. Both may use other refrigerants in either direct or indirect heat transfer relation.

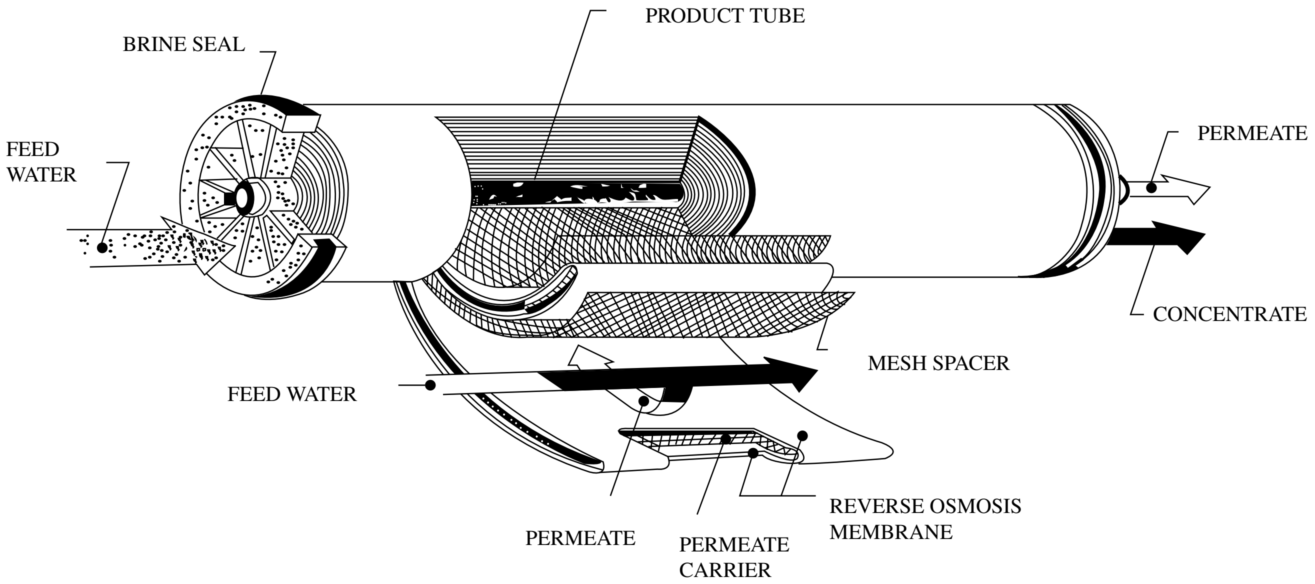


FIGURE 19 Reverse osmosis spiral wound arrangement, made from flat sheet, rectangular shaped membranes. A TORAY Industries, Inc. arrangement. Between the membranes is inserted a porous material, usually a dacron tricot, impregnated with a resin, and both are sealed together, forming the so called "sandwich" on three sides only, the fourth side joined to a slotted collection tube. More sealed membranes connected to the same tube compose the module. A spacer, mesh screen is inserted between the sandwiches and they are rolled to form the spiral wound module.

The vacuum flash vapor absorption system (VFVA) utilizes heat transfer by direct contact by means of an absorption refrigeration machine. For the above reasons, freezing processes found no commercial application. Worldwide there is only one unit producing 210 m^3/d.

Dual Purpose Plants

The economics of desalination plants can be considerably affected, and as a consequence the cost of product water, when electric-power generation is combined with a desalting plant in dual operation. This is especially true in locations where power demand and water requirements are in favorable relation, independent of the method applied in the desalting process. This system has been used for many years in various chemical plants, in which both electricity and low pressure steam are required for processing. This combination results in lower costs for both power and water than would be achieved by separate plants for the production of each.

The main part of the steam enthalpy used in distillation is the heat of condensation. On the other hand, turbines in a power plant use efficiently high temperature and pressure steam. The principle of a dual-purpose plant is to use high temperature steam for power production and low pressure steam leaving the turbine for the brine heater of the desalting plant. Thermodynamic considerations define dual-purpose plant profit when the ratio of water-to-power demand is between certain limits.

Electricity cannot be stored and its production has to meet the actual power demand. Water can be stored, up to the existing facilities, and this gives to the dual purpose plant a certain degree of flexibility in operation. The ratio of the water output to the power output is an important parameter in the economic analysis of dual-purpose plants, as the two products are not necessarily consumed by a single market.

To obtain this requirement the size and characteristics of each component of the dual-purpose plant must be selected in such a way, taking the power and water demand curves in consideration, as to arrive at the optimum cost. It should be noted that power and water demand may present daily and seasonal variations.

The ratio of the water output to the power output is an important parameter in the economic analysis of dual purpose plants. As the two products are not necessarily consumed by a single market, it is important to consider what product ratio can be marketed and how the costs of the two products are affected by varying the product ratio.

Many of the MSF desalination plants in the Arabian peninsula are dual-purpose plants producing electricity and desalinated water.

Water Reuse

The heavy pollution of river, lake or other natural land waters and the restricted sources of water in some places over the world led to the application of desalination methods to treat industrial or municipal effluents for reuse, either for drinking and household purposes or as process water.

The great variety of the raw water to be treated by a desalination process for reuse affects not only the selection of the

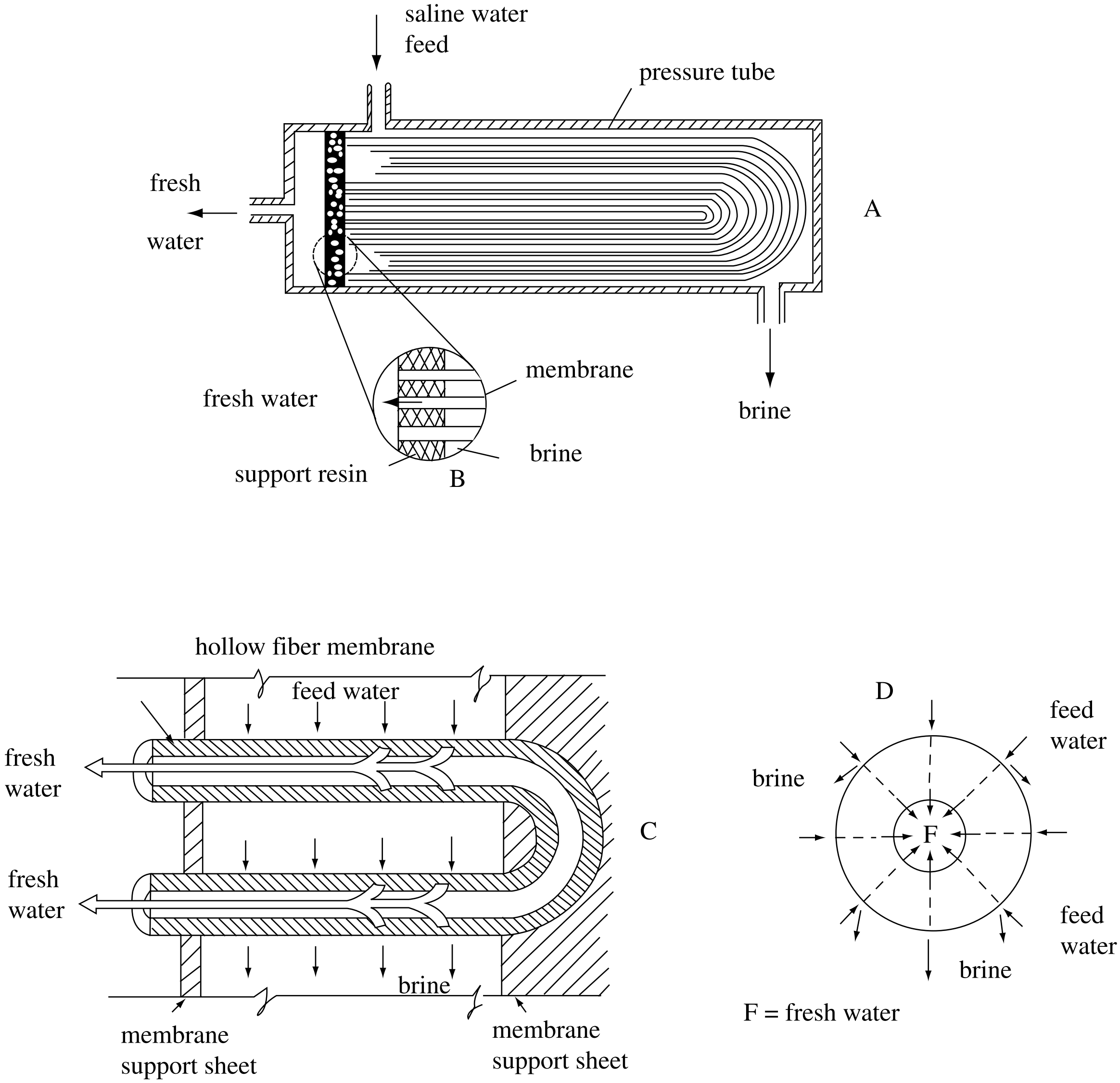


FIGURE 20 Reverse osmosis hollow fiber module, A. Millions of fiber loops, like hairpins, have both ends encapsuled in an epoxy tube sheet or short cylinder enclosed in the flanfed water end on right. Magnified section of the end of these threaded-fine fibers is indicated in B. C is a magnified hollow fiber and D its cross section. The saline water passes on the shell side from the feed at the left end to reject off the concentrated brine on the right. The permeate of fresh water discharges from the center of the hollow fiber, D, and from the center of the module A.

process but also the proper operation of the plant and the pretreatment needs of the desalination process. Reclamation of waste waters can be more efficient by combining desalination methods, i.e., electrodialysis and reverse osmosis, electrodialysis-distillation etc., depending on the raw water and the specifications.

Desalting processes are more expensive than conventional methods for purification of wastes but they provide the specific requirements for recycling. The recycling of the waste waters eliminates also the disposal of these waters, even partly purified, to already polluted streams, especially in land regions where the disposal can be limited by concentration and element content.

CONCLUSIONS

Fresh water, Man's greatest necessity, becomes more scarce every day, especially in certain places around the world, and this scarcity limits industry, agriculture and exploitation of various resources directly necessary for mankind and affects our lives in many ways.

Desalting processes, on the other hand, can provide freshwater in large quantities when and where it is needed. Nevertheless, desalinated water is an industrial product having higher cost than natural water resources, as all commercial products. Approximately 50% of fresh water costs are made up of variable charges, principally plant cost.

Large capacity plants usually produce, to some extent, cheaper water than small capacity plants.

Energy is usually an expensive item. The proposal to use cheap waste heat from nuclear plants failed, due to many reasons connected to the coupling of the two systems. Today the problem is reexamined and the trends are in favor of nuclear desalination. Nevertheless, no large plants have been erected.

The two leading processes of desalination are the MSF distillation and the reverse osmosis. In Table 3 the costs of the two processes are compared.

TABLE 3
MSF versus RO seawater desalting costs

	Plant capacity m^3/d (mgd)			
	3.8(1)	11.4(3)	18.9(5.0)	37.9(10.0)
Total capital cost $ × 10^3				
MSF	8.621	19.737	30.928	54.337
RO	5.887	15.581	24.082	43.01
Total operating $ × 10^3				
MSF	2.064	4.94	7.785	14.401
RO	1.758	4.675	7.404	13.806
Cost of product water				
MSF $/$m^3$ ($/kgallons)	1.65 (6.29)	1.32 (5.01)	1.25 (4.74)	1.15 (4.38)
RO $/$m^3$ ($/kgallons)	1.41 (5.35)	1.25 (4.74)	1.19 (4.51)	1.11 (4.20)

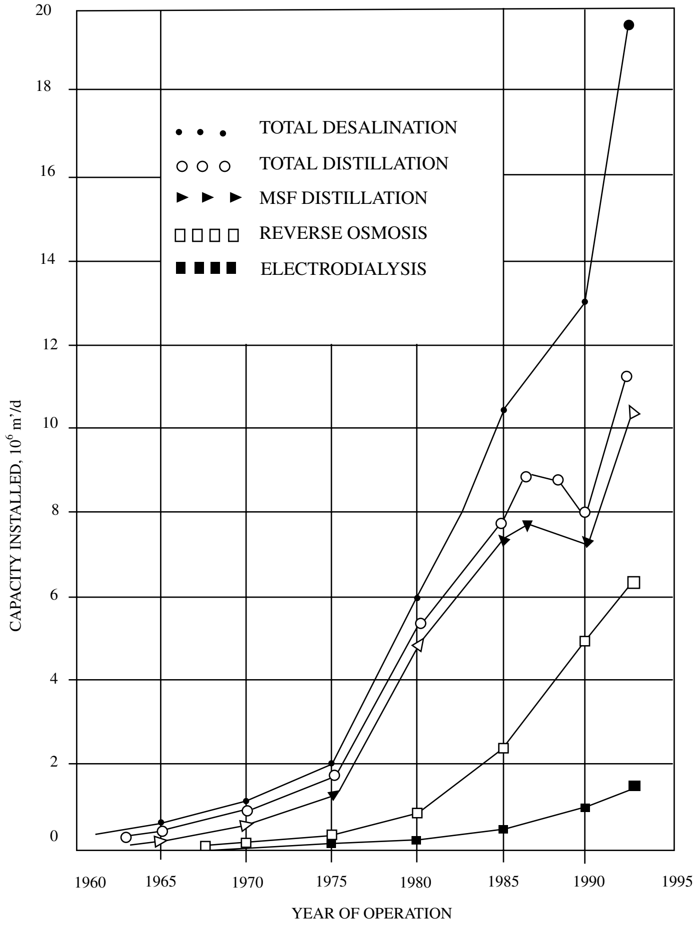


FIGURE 21

Table of SI Units for Desalination

	Symbol	Unit
Area	C	m^2
Average efficiency	k	—
Concentration	C	kg/m^3
Feed concentration	C_1	ppm
Permeate concentration	C_2	ppm
Conductance	S	s^3 A/kg m^2 or V/A
Cross flow velocity	u	m/s
Current density	—	A/m^2
Density	ρ	kg/m^2
Diffusion coefficient	D	m^2/s
Energy, Heat quantity, Work	E-W	J or N · m or kg · m^2/s^2
Energy density	—	J/m^3
Entropy	S	J/K
Electric potential	V	V
Heat capacity	C	J/kg K
Heat flux	Q	W/m^2
Heat flux density	q	W/m^2
Heat transfer coefficient	h	W/m^2 K
Heat transfer rate	Q″	W/m^2
Flow rate	q	m^3/s
Flux, mass	J	kg/m^2 s
Flux, volume	J	m/s
Force	F	N or kg/m^2 s
Gained output ratio	GOR	kg/kg
Insulation, per day	H	J/m^2 d
Insulation, instantaneous	I	J/m^2 s
Latent heat of condensation	ΔH_c, h_{fg}	J/kg
Latent heat of vaporization	ΔH_v, h_{fg}	J/kg
Membrane constant (solvent)	A	m s/Pa
Membrane constant (dissolved sub)	B	m/s
Membrane thickness	δ	μm
Molar energy	—	J/kg
Number of cell pairs	N	—
Number of effects	N	—
Osmotic pressure	π	Pa
Performance ratio	R	kg/2326kJ
Permeability, solvent	L_v	m/Pa s
Permeability, solute	L_s	l/m^{2}kmol s
Power	P	W or J/s or kg m^2/s^3
Potential difference	V	W/A
Pressure	p	Pa or N/m^2 or kg/m s^2
Specific, energy	u	J/kg
Specific, entropy	s	J/kg K
Specific, heat capacity	c	J/kg K
Surface tension	σ	N/m
Temperature, thermodynamic	T	K
Time	t	s
Thermal conductivity	λ	W/m s
Viscosity, dynamic	η	N · s/m^2 or Pa s
Viscosity, kinimatic	v	m^2/s
Volume	V	m^3

Oil fired distillation plants require about 0.5 barrels of oil per m^3 produced fresh water with a performance ratio around 10. Solar distillation or solar-driven desalination have free energy but are still very expensive. To produce one m^3 of fresh water by conventional solar distillation the solar energy equivalent to oil is about 1.3 barrels of oil with a performance ratio of about 5. Solar energy is now used indirectly to drive conventional small size desalination plants and one day the sun's energy may be trapped 100%, not 2% as is the reality today.

In Figure 21 the increases in capacity of the desalination processes are presented by process, starting from the years of their commercial application.[5]

BIBLIOGRAPHY

1. P. Meigs, World distribution of arid and semi-arid homoclimates, *Reviews of Research on Arid-Zone Hydrology,* UNESCO: Paris, Fold Maps, 1960.
2. World Health Organization, International Standards for Drinking Water, Geneva, a970,—Recommendations of the *Guidelines for Drinking Water Quality,* 139 pp., 1984.
3. V. Belessiotis and E. Delyannis, Solar Energy: Some Proposals for Future Development and Application to Desalination. Presented at the *Symposium on "Desalination for Europe,"* Amsterdam 15–16 May 1995 (in print in Desalination).
4. A. Delyannis and E. Delyannis, *Gmelin Handbuch der Anorganishen Chemie*, *Water Desalting,* 340 pp., Springer Verlag, 1974.
5. K. Wagnick, *IDA Worldwide Desalting Plants Inventory,* IDA, Topsfield, Mass., Rept. No. 13, 385 pp., 1994.
6. M.W. Kellog Co., *Saline Water Conversion Data Book*, Washington D.C., 270 pp., 1965.
7. *Exodus* 15, 22–25 (About 15th Century A.D.).
8. E. Delyannis and V. Belessiotis, Solar desalination: Part I and II, *Desalination and Water Reuse*, **4,** No. 4, 9/14, and **5,** 28/34, 1995.
9. A.A. Delyannis and E. Delyannis, Solar Desalination, *Desalination* **50,** 71/81, 1984.
10. N. Robinson, *Solar Radiation,* Elsevier Publ. Co., 345 pp., 1966.
11. V. Belessiotis, K. Voropoulos and E. Delyannis, Experimental method for the determination of the daily output of a solar still, *Desalination* **100**, Special Issue, 1995.
12. K. Voropoulos, E. Delyannis and V. Belessiotis, Thermohydraulic simulation of a solar distillation system under pseudo-steady-state conditions, 1997.
13. E. Delyannis and V. Belessiotis, Solar application in desalination, *Desalination* **100**, 1995.
14. A. Delyannis and E. Howe, Report of W.P. on recommended procedure for costing of solar stills. C.S.I.R.O., Int. Rept. No. 77, Melbourne, 1971.
15. T. Lawand, Systems for Solar Distillation, *Proc. Appropriate Technologies for Semi. Arid Areas,* 201/250, 1975.
16. E. Delyannis, V. Belessiotis and A. Delyannis, Collector Assisted Solar Distillation, *Proc. 7th Inter. Symposium on "Fresh Water from the Sea,"* **2,** 383/387, 1980.
17. H.J.T. Sousa and V. Janish, Dessalinizacao Solar, G.T.Z.-Gov. Reg. Madeira, Rept. Nov., 1986.
18. G.B. De La Porta, "De Distillatione," Libri IX, Roma, 1608.
19. M. Takada, *Proc. 5th International Symposium on "Fresh Water from the Sea"* **2**, 325/333, 1976.
20. M.A. Darwish, Thermal Analysis of VC Desalination System, *Desalination* **69**, No. 1, 275/295, 1988.

GENERAL BOOKS ON DESALINATION

1. G. Belfort, Editor, *Monograph on Synthetic Membrane Processes* (Academic Press, New York, 576 pp., 1984).
2. A. Delyannis, *Nuclear Energy Centers and Agro-Industrial Complexes* (I.A.E.A., Vienna, Technical Report Series No. 140, 138 pp., 1972).
3. A. Delyannis and E. Delyannis, *Water Desalting, Gmelin Handbuch der Anorganische Chemie, "Sauerstoff"* (in English), (Springer Verlag, Heidelberg, 338 pp., 1974, and Supplement to volume one 360 pp., 1979).
4. M. D'Orival, *Water Desalting and Nuclear Energy* (Verlag Karl Thiemig, Munich, Vol. 16, 194 pp., 1967).
5. H.G. Heitmann, *Saline Water Processing,* VCH 332 pp., 1990.
6. E.D. Howe, *Fundamentals of Water Desalination* (Marcel Dekker Inc., New York, 344 pp., 1974).
7. A.H. Khan, *Desalination Processes and MSF Distillation Practice* (Elsevier, Amsterdam, 504 pp., 1986).
8. B.Z. Khosaim and J.S. Williamson, *Solar Water Desalting* (Second SOLERAS Workshop, Denver, Colorado, USA, 1981, NTIS, MRI/SOL-0901, 364 pp., 1984).
9. R.E. Lacey and S. Loeb, *Industrial Processing with Membranes* (John Wiley and Sons, New York, 348 pp., 1972).
10. N. Lakshminarayanaiah, *Transport Phenomena in Membranes* (Academic Press, New York, 519 pp., 1969).
11. N. Lior, Editor, *Measurements and Control in Water Desalination* (Elsevier, Amsterdam, 480 pp., 1986).
12. M.A.S. Malik, C.N. Tiwari, A. Kumar and M.S. Sodha, *Solar Distillation* (Pergamon Press, Oxford, 175 pp., 1982).
13. R.F. Madsen, Hyperfiltration and Ultrafiltration in Plate and Frame Systems (Elsevier, Amsterdam, 367 pp., 1977).
14. U. Merten, *Desalination by Reverse Osmosis* (The M.I.T. Press, Cambridge, Mass., 288 pp., 1966).
15. A. Porteous, *Saline Water Distillation Process* (Longman Group Ltd., London, 150, pp., 1975).
16. R.G. Post and R.L. Searl, *Water Purification Using Nuclear Energy* (The Univ. of Arizona Press, Tuscon, 292 pp., 1966).
17. K.S. Spiegler, *Principles of Desalination* (Academic Press, New York, 566 pp., 1966).
18. K.S. Spiegler and Y.M. El-Sayed, *A Desalination Primer* (Balaban Desalination Publications, 215 pp., 1994).
19. E. Staude, *Membranen und Membrane Prozesse* (VCH 324 pp., 1992).
20. G. Tschobanoglous and E.D. Schroeder, *Water Quality* (Addison-Wosley Publ. Co., Menlo Park, Cal., 768 pp., 1985).
21. S. Sourirajan and K. Matsuura, *Reverse Osmosis/Ultrafiltration Processes Principles* (Nat. Res. Council Canada, 1113 pp., 1985).

ANNUAL REVIEWS ON DESALINATION

1. A. Delyannis and E. Delyaanis, Seawater and Desalting, Vol. 1 to 5, covering 1980 to 1985, about 15,000 abstracts.
2. E. Delyannis, Seawater and Desalting, Vol. 6, 1985 to 1987, (E. Delyannis, Editor, 323 pp., 1988).
3. A. Delyannis and V. Belessiotis, Editors. Desalination Abstracts, Vol. 1, Balaban Desalination Publications, in print, 1989/1990 and up to 1995.
4. Solar Desalting Abstracts, www Ariadne-t-gr, Phaethon.

E. DELYANNIS
V. BELESSIOTIS
Laboratory for Testing & Development of Solar
and other Energy Systems of the National
Center for Scientific Research "Demokritos"
Aghia Paraskevi, Greece.

DESULFURIZATION: see FOSSIL FUEL CLEANING PROCESSES; VAPOR AND GASEOUS POLLUTANT CONTROL FUNDAMENTALS

DISINFECTION

INTRODUCTION

Disinfection is a term which has for many years been used with different shades of meaning. It has frequently been confused with antisepsis which, although analogous to disinfection (see later), does not strictly have the same interpretation. This, particularly when considered in conjunction with other terminology, means that any article dealing with disinfection must clearly define the sense in which the term is being used. Davis (1968) rather vaguely defines a disinfectant as "a material having powerful germicidal activity and suitable for use as such."

Fortunately, at least some of the confusion relating to disinfection and to similar, but not identical, terms has now been resolved.

The terms defined in this report have been classified as shown in Table 1, but only those applicable to this chapter will be defined here, together with the term *chemosterilizer* (Borick, 1968).

Sterilization is the process of destroying or removing all microbial life.

A *sterilant* (sterilizer) is an agent used in sterilization which destroys microbial life, including bacterial spores, and is thus distinct from a disinfectant. The term "sterilant" may itself be somewhat confusing, however, for a "chemosterilant" is sometimes used in the Untied States to denote a "chemical substance used to sterilize insects and render them incapable of reproduction on mating with non-sterile partners" (Borick, 1968). Davis (1968) defines a sterilant as a disinfectant suitable for use in the food industry.

TABLE 1

Terminology[a] used in sterilization and disinfection

I. Definitive terms	II. Terms in common use
Sterile	Disinfectant
Sterilization	Disinfectant
Sterilizing agent	Antiseptic
-cide[b]	Antisepsis
-stat	Sanitizer
-statis	Sanitization

[a] British Standard Glossary of terms.

[b] Not germicide.

A *sporicide* is a chemical agent that kills bacterial spores.

A *chemosterilizer* (Borick, 1968) refers to a chemical compound which is used to destroy all forms of microbial life, and is thus the same as a sterilant defined above. The term has not been widely used.

Disinfection is the destruction of microorganisms but not usually bacterial spores. Commercially, the term applies solely to the treatment of inanimate objects, and does not necessarily imply that all microorganisms are killed, but rather that they are reduced to a level not normally harmful to health.

Antisepsis is the destruction of microorganisms, but not bacterial spores, on living tissues—not necessarily killing all microorganisms, but reducing them to a level not normally harmful to health. The term is thus analogous to disinfection.

A *sanitizer* is a disinfectant with the connotation also of cleansing; it is used mainly in the food and catering industries.

The suffices "-cide" and "-stat" may be added to various words to give a precise meaning, e.g., bactericide means a substance which kills bacteria but not spores, bacteriostat (bacteristat) a substance which inhibits the growth of bacteria, thereby producing the state of bacteriostasis. Other terms which are frequently used in this context include the following: sporicide (see earlier), fungicide, fungistat, virucide, microbiocide and biocide.

Not all authorities would agree with all of the definitions listed, and one term in particular which might be hotly disputed is "antiseptic." This is often used to denote a chemical agent usually applied to human skin and acting either by destroying microorganisms or inhibiting their growth (Olivant and Shapton, 1970).

Another term which is frequently employed is "detergent-sterilizers" or "detergent-sterilants"; these consist of two components, one of which has a cleansing action, and the other an antimicrobial activity. Unfortunately, "sterilizer" or "sterilant" has an absolute meaning (see above) and this would imply that a detergent sterilizer (sterilant) is sporicidal as well as being lethal to other microorganisms, whereas those compounds which comprise the "sterilizer" or "sterilant" component are usually not sporicidal. There is probably, however, need of a term which includes the word "detergent." Foster *et al.* (1953) use "sanitization" to denote the application of a bacterial process sufficient to render dairy equipment approximately sterile, this also implying that pathogens likely to be associated with such equipment and with eating and drinking vessels will be killed. Davis (1968) has employed the name "detergent-sterilant" throughout his review, although, as pointed out earlier, his definition of a sterilant differs from that above. Throughout the present chapter, "sanitizer," without quotes, will be employed rather than these other terms.

KINETICS

It is tempting to imagine that the application of an appropriate disinfection procedure will result in immediate elimination of all microorganisms from the site of interest. This temptation is often fostered by various advertising interests in pursuance of their sales campaigns. However, a cursory inspection of the literature soon dispels this cosy, over-simplified view. Disinfection has been shown repeatedly to be not only a gradual or even prolonged, process, but also a complex one.

Almost invariably, investigation into the course of disinfection processes have involved the study of purified cultures of microorganisms (usually bacteria) under specified conditions. This has led to certain criticisms that such systems are too far removed from reality to be of practical significance. While it is true that considerable caution must be exercised in applying the results of these studies to practical situations, the experimental systems are still far from simple, and have yielded much useful information.

Survivor Curves

The most common method of monitoring the progress of a disinfection process is by means of viable counting techniques. These suffer from certain inherent limitations. In particular, the absolute values obtained are dependent on the specific technique and the experimental conditions associated with it; and in addition, cells which have been exposed to the process, may respond quite differently from those examined prior to exposure. In order to obviate this difficulty, alternative methods of assessing "vital activity" have been suggested, usually biochemical in nature. Unfortunately, while the greater simplicity of these methods allows more precise measurements to be made, the killing of microorganisms usually involves a whole series of complex reactions, which makes correlation of the results rather difficult. Despite their faults, viable counting methods do reflect the complexity of the killing process.

The usual scheme of events is to expose the chosen culture of microorganisms to the disinfection process of interest, under controlled experimental conditions. Estimates of the viable population density of the system are made by performing viable counts on representative samples removed from the system: see, for example, Prince *et al.* (1975). For convenience, these estimates are usually plotted graphically against time of exposure or occasionally dosage of the disinfection agent employed. While the estimated numbers of organisms may be plotted directly, they are usually converted to a proportional basis such as "surviving fraction" or "percentage survivors," since this facilitates visual comparison of the results.

The simplest graph so obtained is the arithmetic plot which invariably exhibits a curve of similar general form to Figure 1. The main point of interest about this curve is that it indicates that the rate of disinfection varies inversely with the number of surviving organisms. This is interpreted as an indication that the individual cells of the culture exhibit differing sensitives to the process, i.e., there is a distribution of resistances. Unfortunately, curves of this type are difficult to analyze or to compare visually, and so the survivors are often plotted in a logarithmic fashion. This results in a whole "family" of possible results, as shown in Figure 2.

Figure 2(a) shows the simplest result, the familiar straight line which is often prized for its ease of characterization. It also possesses the sometimes dubious advantage of ease of extrapolation; a property which should be utilized only with extreme caution. This graph indicates that the rate of disinfection is inversely proportional to the logarithm of the number of surviving organisms. The similarity between this situation and the kinetics of a first-order chemical reaction has caused this type of response to be described as unimolecular or monomolecular. It is important to stress, however, that the description applies to the graphical response of the system; for it would be extremely naive to assume from this that the mode of death of the cells is attributable to a first-order chemical reaction. The straight line may be described mathematically by the equations:—

$$k = \frac{1}{t}\log\left(\frac{N_0}{N}\right)$$

where

k = rate constant or slope of line
t = time elapsed
N_0 = number of viable cells initially
N = number of viable cells at time t

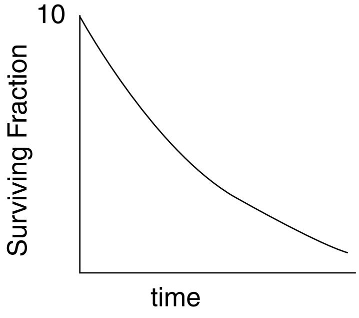


FIGURE 1

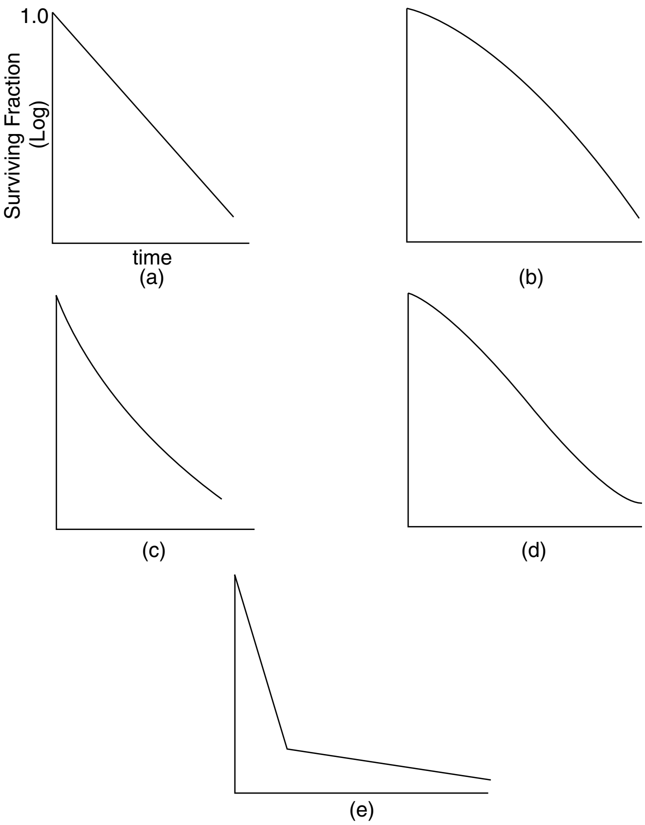


FIGURE 2

or, if natural logarithms are used:—

$$\frac{N_0}{N} = \exp(kt).$$

These equations have led to the alternative and preferred descriptions of the response as being exponential or logarithmic. Since the number of viable cells decreases throughout the process, then the rate constant, *k,* is always negative in character.

The two curves shown in Figures 2(b) and 2(c) illustrate similar, though opposite, deviations in response from the straight line case. These are described, according to their shape, as concave or convex, and qualified by the additional designation upward or downward, to indicate orientation. The graphs indicate that the rate of disinfection changes gradually in the early stages, but then assumes a more steady state of change similar to that in the straight case. In Figure 2(b), the rate of disinfection is relatively slow at first, but gradually increases to a steady, limiting value. Various interpretations have been suggested to account for this change; among them, that the distribution of resistances between the individual cells exhibits a relative deficiency of cells of low resistance; or alternatively, that the cells must pass through one or more intermediate stages before becoming sensitive to the disinfection process in question. The weakness of such interpretations is underlined by the fact that changes in the experimental conditions often result in a change in the shape of the graphical response. Figure 2(c) illustrates a relatively fast rate of disinfection initially, which gradually decreases to a steady, limiting value.

In Figure 2(d) is illustrated the most common deviation from the straight line response, the sigmoid curve. Responses of this type are more easily demonstrated in systems employing a moderate rate of disinfection. As the rate of disinfection is increased, the limitations of viable counting techniques make it more and more difficult to monitor the progress of the process with any precision. This results in the apparent response becoming indistinguishable from the straight line case. It is sometimes suggested that the sigmoid response is the most common situation encountered, but that it is often unrecognized due to the practical difficulties experienced. The sigmoid response is usually interpreted as an indication that the distribution of resistances between the individual cells is of the log normal type. Complete agreement with this model distribution is indicated when the sigmoid curve is symmetrical.

Figure 2(e) illustrates a response of particular interest. The graph consists of two parts, both of which are linear but of different slope, with a fairly sharp transition between the two. This would appear to indicate a fairly rapid rate of disinfection initially, followed by a fairly sharp transition to a lower, but steady, rate. Such a sudden transition naturally engenders interest, if not suspicion. It has been suggested that this type of response indicates the presence of two distinct groups of cells, each of which exhibits its own characteristic distribution of resistances. Experiments with a mixture of two bacterial cultures of different identity, whether obtained different species, or consisting of spores and vegetative cells of the same species, can be shown to yield this type of response. The first part of the graph corresponds to the usual response of the more sensitive component, and the second part to that of the more resistant component. However, at relatively low temperatures and humidities, exposure of nominally homogeneous cultures to ethylene oxide gas often yields this type of response. While it is sometimes suggested that this indicates the presence of two distinct groups of cells, as discussed above, it must also be considered that not only can this phenomenon be demonstrated with cultures apparently homogeneous to other sterilization methods, but also that this two part response reverts to the exponential type on increasing the temperature or humidity of the system.

As indicated in the foregoing discussion, consideration of the shapes of survivor curves may provide useful circumstantial evidence on which to base hypotheses relating to the response of cell populations to disinfection processes. However, as also indicated in the discussion, the operative word is "circumstantial."

Empirical Parameters

While a survivor curve illustrates the response of a cell population in terms of variation in number of survivors with time (or dose) of disinfection treatment, this is essentially a static situation since all other factors are held constant. Often, it is the variation in response of the system to changes in experimental conditions which is of particular interest. This variation in response may be monitored by constructing families of survivor curves, one curve for each level of the factor being varied. Comparison of curves within these families indicates the nature of the change in response.

The two factors most prominent in the regulation of disinfection processes are temperature and concentration of disinfectant. For both of these factors, it is found that their influence on the disinfection process varies in a regular manner over a fairly wide range of conditions. As a result, empirical parameters have been derived which enable the influence of these two factors to be characterized in a convenient form.

Temperature Coefficient

In general, it is found that the activity of a disinfectant varies directly with the temperature, i.e., the higher the temperature the greater the activity. This change in activity may be quantified by expressing the disinfection rates observed at two different temperatures as a ratio. The value of this ratio is found to remain reasonably constant over a wide range of temperature. This may be expressed mathematically:

$$\theta^{T_2 - T_1} = k_2/k_1$$

where k_1 and k_2 are the rate constants at temperatures T_1°C and T_2°C, respectively, with T_2 greater than T_1. The constant, θ, is termed the temperature coefficient, and assumes a numerical value which is characteristic of the agent and, to a certain extent, of the organism, employed. The superscript $(T_2 - T_1)$°C is necessary to indicate the temperature difference. Since the disinfection rate is inversely related to killing or extinction time, the latter may be used instead. The expression then becomes

$$\theta^{T_2 - T_1} = t_1/t_2$$

where t_1 and t_2 are the extinction times at temperatures T_1°C and T_2°C, respectively. At high value for θ indicates that the process is relatively sensitive to temperature changes.

It is usual to find two versions of the temperature coefficient in most common use: the coefficient for a 1°C temperature change, θ, when the superscript is usually omitted; and the coefficient for a 10°C temperature change, θ_{10}, which may be sometimes expressed as Q_{10}. The popularity of the 10°C coefficient follows from its use to characterize changes in reaction rates in chemical systems. This enables interesting comparisons to be made between disinfection processes and non-living chemical reactions. Attempts have been made to deduce modes of death of the organisms by such comparisons. However, subsequent biochemical investigations have tended to disagree with these deductions.

Dilution Coefficient

The activity of disinfectants under otherwise constant conditions is found to vary directly with the concentration of disinfectant employed over a considerable concentration range. As before, killing or extinction times are usually employed as a measure of disinfection rate. The effect of concentration may be expressed mathematically as:

$$c^{\eta}t = \text{a constant}$$

or:

$$\eta \log c + \log t = \text{a constant}$$

where

c = concentration of disinfectant
t = extinction time
η = dilution coefficient.

A high value for the dilution coefficient indicates that the process is relatively sensitive to changes in concentration of disinfectant. The dilution coefficient is sometimes referred to as the concentration exponent.

Other Factors Influencing Activity

The antimicrobial activity of several disinfectants is influenced to a considerable extent by changes in pH. For example, a rise in pH results in a decrease in the activity of phenols (Bennett, 1959), organic, acids, compounds liberating chlorine, benzoic acid and iodine, although iodine is less affected by acidity than is chlorine. An increase in pH increases the dissociation of phenols and benzoic acid (Wedderbern, 1964); chlorine in water forms HClO, which is dissociated with a rise in pH with a concomitant loss of activity. In contrast to the above, however, there is an increased microbicidal activity of the quaternary ammonium compounds (QACs) and of acridines (Foster and Russell, 1971); in the case of QACs, the effect of pH is considered by Salton (1957) to be on the cell rather than the disinfectant molecule, since the number of negatively-charged groups on the bacterial surface will be increased as the pH rises, thus influencing the number of positively charged molecules which can be attached.

Another factor which influences the activity of certain antimicrobial agents is organic matter, e.g., the presence of blood, serum, pus, etc. In general terms, the more chemically reactive a compound, the greater the effect of organic matter on its activity. This is particularly true with the hypochlorites. other examples are provided under individual compounds later.

Mathematical Models

A considerable number of mathematical models have been derived at various times, in attempts to reconstruct the disinfection process. These have utilized deterministic, probabilistic, and thermodynamic approaches to the problem, and have been reviewed in detail by Prokop and Humphrey (1970).

In general, these models are based on attempts to reconstruct the survivor curves obtained, even though, as already discussed, these are liable to be changed by changes in experimental conditions. In addition, they usually involve preliminary assumptions concerning the nature of the disinfection process. Consequently, attempts to deduce mechanisms of killing from these models tend to be rather disappointing. They are, however, useful in terms of concise descriptions of the results of the process; and as further biochemical information on modes of killing becomes available, their usefulness and reliability should increase further.

DISINFECTANT TESTING

The testing of disinfectants is a topic with a long history of controversy. Differing opinions on the merits and relevance of various methods are legion, and have been the cause of many heated exchanges between those holding them. The wide variety of methods may most conveniently be considered under the headings of screening, standardization and "in-use" tests, although there is a degree of overlap between the three categories. The interested reader should consult the papers by Forsyth (1975), Miner *et al.* (1975) and Reybrouck (1982) for further information.

Screening Tests

These are usually of the simplest type, since they are speculative by nature, and often involve the testing of large numbers of disinfection agents or formulations; both time and cost usually dictate this simplicity. The actual methods employed vary according to the physical characteristics of the disinfectant and the type of use envisaged. Ideally, any test should be as realistic as possible, although for screening purposes this will be subject to the requirements of simplicity discussed above.

Disinfectants which are soluble or miscible in water are often incorporated in microbiological culture media which are then inoculated with suitable microorganisms. After incubation under appropriate conditions, the inoculated media are examined for growth of the organisms, absence of growth indicating that the inoculum has been inhibited. Where the continuous presence of the disinfectant is inappropriate, then the method is modified to include a suitable means of removing or inactivating the disinfectant after a predetermined exposure time.

Other disinfectants may be tested in a similar manner, by arranging for inocula of suitable organisms to be subjected to standardized exposure to them. After exposure, the organisms are inoculated into suitable culture media, and incubated as before. Since the end-point in these tests is the death of all the organisms involved, this type of test is often referred to as an Extinction Test.

The phrase "suitable organisms" used above, represents one of the key factors in the testing of disinfectants. A disinfectant is expected to kill all undesirable organisms, which usually refers to organisms injurious to health (see previous section). Test organisms may be chosen, therefore, either for their resistance to disinfection or, like *Salmonella typhi,* for their medical significance. Although the designation "undesirable organisms" covers a wider range of life-forms, most disinfectant testing has employed various species of bacteria for reasons of convenience. In retrospect, this does not appear to have been a significant disadvantage, since activity against bacteria usually coincides with that against the other life-forms. However, although it is recognized that bacterial spores are much more resistant to disinfection than the vegetative forms, the latter have, with few exceptions, invariably been used in testing. This is partly due to convenience, but it should be noted that there are many disinfectants marketed which under normal conditions of use are incapable of killing bacterial spores. This has been countered, in certain quarters, by re-defining a disinfectant as a chemical agent capable of killing bacteria but not necessarily bacterial spores (see previous section).

Standardization Tests

This is the area which has given rise to the greatest amount of contention. In most cases, this has been due to misinterpretation and misapplication of the results of testing in situations where they have little, if any, relevance. The standardization of disinfectants usually requires the use of microbiological techniques rather than chemical assay. Even when chemical characterization of the active agent(s) involved is possible, the activity of a disinfectant will usually be significantly influenced by factors concerned with the formulation and method of use of the product. However, although the disinfectant may be standardized in terms of its intended biological effect, this is carried out under specific, controlled conditions. Depending on the similarity of their circumstances, the results from standardization tests may, or may not, be capable of extrapolation to practical situations. This is where the controversy tends to arise.

The most popular methods of testing for standardization purposes have been extinction tests. These are basically similar to the general method discussed in the previous section, except that the procedure and materials used are rigidly standardized in order to achieve best reproducibility of results. In addition, the most widely used tests employ the pure chemical, phenol, as a control of the resistance of the test organisms used. Since results are expressed by comparison of the activities of phenol and the disinfectant being tested, these tests are referred to as Phenol Coefficient methods. The most popular and official versions in use are the Rideal–Walker Method (as modified by B.S. 541), the Chick–Martin Method (as modified by B.S. 808), and the Association of Official Agricultural Chemists (AOAC) Phenol Coefficient Method (1970). Details of these tests may be found in the appropriate publications.

While the use of phenol as a control on the resistance of the test organisms is extremely valuable, it has led to widespread assumptions that the ratio of activities of disinfectant and phenol indicated by the phenol coefficient will hold true in all circumstances. This, of course, is far from true. The AOAC (1970) attempted to improve this situation by introducing a Use-Dilution Method. This test is based on the hypothesis that disinfectants in use shall be at least as efficient as 5% phenol, and that a dilution of twenty times the phenol coefficient should achieve this. Accordingly, this dilution is tested against standardized cultures in order to confirm this, or alternatively, in order to determine a correction factor.

Whereas the extinction methods discussed above evaluate the extinction concentration corresponding to predetermined exposure times, Berry and Bean (1954) devised a test for evaluating extinction times for chosen disinfectant concentrations. While the test is only applicable to phenolic and other easily inactivated disinfectants, it has been claimed to be at least as reproducible as other methods in common use (Cook and Wills, 1954). The method of assessing the end-point of the reaction has been further improved by Mathei (1949).

Methods other than extinction tests for standardizing disinfectants include various methods based on assessment of cessation of vital enzyme activities. The enzymes involved have generally been certain oxidases or dehydrogenases (Sykes, 1939; Knox *et al.*, 1949). In a less specific manner, inhibition of respiration has been used as a method of assessment (Roberts and Rahn, 1946). The controversy surrounding these methods centers around the problem of correlating enzyme activity with viability.

Other methods which have been proposed include measurement of post-incubation opacity corresponding to standard survivor levels (Needham, 1947), and also measurement of cell volume increase following post-exposure incubation (Mandels and Darby, 1953).

The range of tests discussed above is by no means exhaustive, and only covers general purpose tests applicable to water-misible disinfectants. Any number of alternative tests could be devised by appropriate selection and standardization of the various parameters. In addition, there are numerous tests which can be, and have been, devised in order to standardize disinfectants intended for specific uses such as sporicidal or tuberculocidal duties (AOAC 1970).

Methods of standardizing disinfectants other than those to which the above discussion applies are usually designed more closely around the particular use envisaged. Thus, a considerable element of actual, or simulated, in-use testing is usually involved. For convenience, they are best discussed in the following section.

A recent test is the Kelsey–Sykes test (Kelsey and Maurer, 1974), which is a form of capacity test. In this, incremental additions of test organisms are made to appropriate dilutions of test disinfectant and aliquots are removed for detecting survival immediately prior to the next addition of organisms. On the basis of this method, use-dilutions of the test disinfectant (which need not necessarily be a phenolic) under clean and dirty conditions can be recommended to hospitals, which should then check them during in-use tests (see also Coates, 1977; Cowen, 1978).

In-Use Tests

The previous two sections have dealt with testing methods applied in the laboratory, which yield information primarily of value to the disinfectant manufacturer. The "consumer," however, is almost solely concerned with the performance of disinfectant materials under the conditions of use which are associated with the application envisaged. To this end, he is more interested in testing methods which resemble, as closely as possible, these practical conditions. However, the foregoing remarks should not be taken to imply that laboratory standardization methods are completely arbitrary. In view of the greater difficulty experienced in killing organisms which are present as dried surface films, as opposed to those in fluid suspension, many standardization tests have employed films on various surfaces as their inocula. Surfaces used have included silk thread (Koch, 1881), garnets (Kronig and Paul, 1887), glass cover-slips (Jensen and Jensen, 1933), glass cylinders (Mallmann and Hanes, 1945), glass slides (Johns, 1946), stainless steel cylinders (AOAC Use-Dilution Confirmatory Test, 1960), and glass tablet tubes (Hare, Raik and Gash, 1963). It should be noted that while these surfaces represent a step toward the practical situation, they nevertheless comprise a collection of laboratory "artifacts" when compared with real situations. A nearer approach was achieved by use of surfaces such as rubber strips (Goetchins and Botwright, 1950) and glazed, waxed, and rubber tiles (Rogers, Mather and Kaplan, 1961).

Where the physical size of articles required to be disinfected is fairly small, it is quite feasible to carry out in-use testing in the laboratory. Hence metal trays were used by Neave and Hoy (1947), 10 gallon milk churns by Hoy and Clegg (1953), and small drinking glasses by Gilcreas and O'Brien (1941). Similarly, scalpels, syringes and similar small items may conveniently be subjected to in-use testing in the laboratory. Where the physical size of the system is somewhat greater, then the laboratory must be forsaken in order to carry out on-site testing. Typical examples of such situations include hospital walls and floors, and industrial or dairy processing machinery. The outstanding value of in-use testing arises from the fact that results obtained are directly applicable to the system without the need for interpretation and extrapolation, and that practical difficulties such as short contact times or inaccessibility of certain areas can be accounted for.

The choice of test organisms usually reflects either of two main approaches. The simplest approach is to inoculate the system artificially with organisms considered to be of practical significance in the particular application. This significance may be due to the resistance to disinfection of the organism, or alternatively, to its practical effect on the system should it survive the disinfection process. In order to simulate the practical system, the organisms may be suspended in appropriate materials before inoculation. An example would be the use of milk as a suspending fluid in tests on dairy disinfection. A slight variation on this approach is to inoculate with indeterminate mixtures of likely organisms obtained from natural sources, such as low quality, raw milk. The second approach involves the use of the normal, preexisting flora of the system which has arisen during normal use. Investigations of this flora before and after the disinfection process would provide direct evidence as to whether the process has achieved its object. One difficulty with this approach is that of providing suitable growth conditions for all of the possible species of organisms which may be present. Also, the normal flora is likely to vary with time, and so, ideally, this method is best applied in the form of a routine monitoring procedure. Both of these difficulties produce an increase in the financial cost of this type of approach, but the reliability of the results obtained is correspondingly high. The ideal method would involve a combination of these two approaches: an initial test with known resistant organisms in order to indicate the upper levels of activity which may be required; this to be followed by routine monitoring tests to guard against both unsuspected resistance amongst the normal flora, and unforeseen breakdowns in the method of application.

Any testing method which involves the use of surface films is subject to the problem of physical recovery of the organisms. In the case of small surfaces, the articles can often be placed in or on a suitable nutrient medium, and provided this allows contact between the medium and the organisms in the film, then growth takes place. Where larger or less accessible surfaces are concerned, then the simple direct method will have to be replaced by some kind of sampling technique. Sampling techniques vary widely and not all are applicable in all situations. In the case of accessible surfaces, simple methods such as wiping with sterile swabs may be used; or alternatively, flooding the surface with a sterile liquid, some or most of which is then removed by swab or pipette. An interesting alternative consisting of a sterile, agar medium, "sausage" was devised by Ten Cate (1965). The exposed transverse surface of this sausage is pressed against the surface to be sampled. After sampling a slice is cut from the sausage and incubated with the exposed side uppermost. A simple method which is used for sampling skin surfaces may also be used in other applications. This method employs adhesive cellophane tape. The adhesive surface is pressed onto the surface to be sampled, and then removed for transfer to a suitable medium. Where surfaces are inaccessible, as in pipes and processing machinery, then a rinsing technique is usually most convenient. The resulting liquid may be added directly to suitable nutrient media, or, if present in large bulk, may be filtered through membrane filters to remove the organisms. The resulting filter membrane is then transferred to a suitable medium.

One final problem which is of importance in all methods of testing which involve assessment of viability, is that of recovery and growth of the test organisms following exposure to the disinfectant. Organisms which have survived a disinfection process often show altered requirements for optimal growth. Response to physical conditions such as incubation temperature, as well as to biochemical conditions such as dependence on certain nutrients, may be completely altered from that of unexposed cells. While considerable effort has been made to derive efficient recovery methods (Flett *et al.*, 1945; Jacobs and Harris, 1960; Harris, 1963; Russell, 1964) the problem is so variable and so many combinations of factors must be considered, that it is far from being solved. There is general agreement that recovery methods should be selected which will allow maximum recovery of exposed organisms; but putting this into effect can be extremely difficult.

LIQUID DISINFECTANTS

Several chemical agents have long been employed for destroying microorganisms, although it is frequently asserted that such substances are without effect on bacterial spores. With many agents, however, this is untrue (Sykes, 1970; Russell, 1971, 1982). The most important substances are phenols and cresols, biguanides, chlorine-releasing compounds and other halogens, aldehydes, alcohols, quaternary ammonium compounds, mercury compounds, strong acids and alkalis and hydrogen peroxide. The majority of these are considered in detail below. Further information is provided by Hugo and Russell (1982).

Phenols and Cresols

Although Kronig and Paul (1897) and Chick (1908) showed that phenol was active against spores, the concentrations, 5%, employed, were considerably higher than those needed to kill vegetative bacteria. More recent studies have indicated that bacterial spores are not killed even after long exposure to phenols (Sykes, 1958, 1965; Loosemore and Russell, 1963, 1964; Russell and Loosemore, 1964; Russell, 1965, 1971; Rubbo and Gardner, 1965; Briggs, 1966). Of the bacterial spores. *Bacillus stearothermophilus* is the most resistant to phenol and *B. megaterium* the most sensitive (Briggs, 1966). However, in contrast to its lack of sporicidal activity, phenol is sporostatic at low concentrations.

Several factors influence the antimicrobial activity of phenols and cresols (Bennett, 1959; Cook, 1960; Bean, 1967):

1) Concentration. These compounds have high concentration exponents, η, which, as described above, indicates that they rapidly lose their antibacterial activity on dilution. This also means that dilution procedures can be used to prevent the carryover of inhibitory concentrations into recovery media when viable counts or sterility tests are being carried out (Russell, 1982; Russell *et al.*, 1979). Studies from Bean's laboratory (Bean and Walters, 1955; Bean and Das, 1966; Bean, 1967) are of interest, for they show that with dilute solutions of disinfectants with high intrinsic activity, e.g. benzylchlorphenol, the high proportion taken up by the cells means that the concentration remaining is only weakly bactericidal, so that the surviving cells do not meet lethal conditions.
2) Temperature. The bactericidal activity of the phenols and cresols increases rapidly with an increase in temperature. Examples of temperature coefficient (θ^{10}) against *E. coli* at 30–40° (Tilley, 1942) are: phenol 8.4, *o*-cresol 6.9, *p*-cresol 5.6.

Of importance, also, is the finding that these substances are sporicidal at elevated temperatures (Berry *et al.*, 1937; Russell and Loosemore, 1964). As a result of the findings of Berry *et al.* (1937), one method of sterilizing certain injections by heating them with 0.2% w/v chlorocresol is still an official method in Britain (British Pharmacopoeia, 1980).

3) pH. The phenols are more active at an acid pH than in alkaline solution, as phenates (phenoxides) are formed at high pH. Acid pH also results in a more effective, although still slow, sporicidal action (Sykes and Hooper, 1954).
4) Organic matter. The presence of organic matter may decrease the antimicrobial activity of these compounds; this was early recognized in the design of the Chick–Martin test for evaluating phenolic disinfectants (Chick and Martin, 1910; Garrod, 1934, 1935). The results do, however, depend on the actual phenol used and on the kind of organic matter, and the interference is less than with other disinfectants such as the quaternary ammonium compounds (Cook, 1960).
5) Oxygen tension. Anaerobic bacteria are generally more resistant than aerobes to phenols. Moreover, facultative organisms, e.g., *E. coli,* are more resistant when grown under anaerobic conditions.
6) Type of organism. As described above, these compounds are bactericidal and sporostatic at low concentrations, and sporostatic and not sporicidal even at high concentrations. As a group, however, they are also fungicidal to several moulds, and use is made of this in the inclusion of cresol and chlorocresol as preservatives in creams which are liable to fungal contamination (Wedderbern, 1964).

   Morris and Darlow (1971) have pointed out that phenolic compounds with high R-W coefficients are effective against some viruses, but that they are generally too variable in their activity to be suitable as general virucidal agents.
7) Chemical nature. Dihydric and trihydric phenols (Figure 3) are generally less active than phenols, and alkylation of monohydric phenols to give cresols potentiates the antimicrobial activity. Also halogenation of the phenols increases their activity (although to a lesser extent if the halogenation is in the ortho- than in the paraposition) and this is even more pronounced when accompanied by the introduction of aliphatic or aromatic groups into the nucleus, e.g., *p*-chloro-*m*-cresol (chlorocresol). This increase in antimicrobial activity is, however, paralleled by a decrease in water solubility.

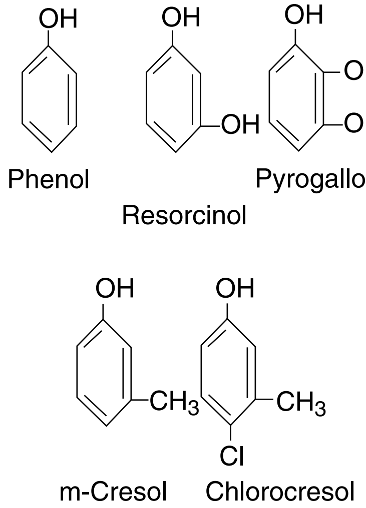


FIGURE 3

To overcome this decrease in aqueous solubility, various soaps have been used to render water-soluble (solubilize) these substances. However, the effect of soap on the biological efficacy of the phenols depends on two factors, firstly the nature of the soap, and second the proportion of soap to phenol. Solution of cresol with soap (Lysol, BP), for example, contains 50% of cresol in a saponaceous solvent and has a bactericidal activity which depends on the nature and amount of soap used. The ratio of cresols to soap may be critical, the optimal cresol–soap ratio being of the order of 2:1. In lysol, the soap content is *c.* 22%, and the ratio is thus 2.2:1. The soap solutions are able to solubilize insoluble phenols in the micelles; the critical micelle concentrations (cmcs) of different soaps vary, and this explains differences in bactericidal action noted above.

Lysol and the so-called "black-fluids," which consist of the lower coal tar phenols, are formulated in sufficient soap so that they are retained in solution when diluted with water. In contrast, the white fluids consist of concentrated emulsions of high boiling phenols stabilized with protective colloids; they can be diluted with hard or soft waters, whereas black fluids should be diluted with soft waters only.

Micelles have been considered as being reservoirs of phenols, so that when the concentrated solution is diluted before use, the phenols are released by dilution below the cmc to give a highly active solution. Two types of system have been investigated experimentally in attempts to assess the exact role of the micelles; these are (a) constant phenol concentration, and (b) a constant phenol/soap ratio (Berry, 1951; Berry and Briggs, 1956; Berry, Cook and Wills, 1956; Cook, 1960). With a constant phenol system, there is a rapid increase in bactericidal activity below the cmc of the soap which could be the result of an increased uptake of phenol together with an increased permeability of the bacterial surface (Mulley, 1964); however, above the cmc in this system, there is a decrease in bactericidal activity, as the increasing numbers of micelles being formed compete for the phenol with the cell surface. In systems where there is a constant phenol/soap ratio, there is likewise an increase in activity below the cmc, and a decrease in activity at the cmc. However, in this system the activity again increases at higher soap concentrations; thus, this increased bactericidal activity parallels a saturation of the soap micelles with the phenol. The reasons for this remain unclear, for although the soap itself could contribute to or be responsible for the increased activity, systems in which the soap has a low activity give similar results. The actual role of the micelles is thus difficult to assess. Of the two systems described, it is apparent that the system with the constant phenol/soap ratio is more important from the viewpoint of practical disinfection. It is necessary to use the lowest possible proportion of soap to phenol, thus giving a comparable situation to bactericides in two-phase systems, in which the bactericidal efficiency is related to the concentration in the aqueous phase (Bean and Heman-Ackah, 1965).

Phenol itself is an effective bactericide and anti-fungal agent, which is used as a preservative in some injections and creams; it is also the standard reference substance in some methods of testing phenolic bactericides (see earlier). Cresol, a mixture of *o*-, *m*- and *p*-cresol, in which the *meta*-isomer predominates, and of other phenols obtained from coal tar, is highly bactericidal and fungicidal. Apart from being the active constituent of Lysol, it is also used as a preservative in certain injections and creams. Chlorocresol is a powerful antimicrobial agent which, in conjunction with heat, is employed for the sterilization of certain injections. It is also used as a preservative in certain cosmetic creams and lotions, and is included in Sodium Benzoate and Chlorocresol Solution, which may be used for the storage of sterilized surgical instruments, the sodium benzoate delaying rusting. Chloroxylenol has low water solubility and is solubilized by means of soap, the solution being known as Roxenol. Its antimicrobial activity is markedly reduced in the presence of organic matter. Thymol is a potent bactericide and fungicide which, in the form of Glycerin of Thymol, is employed as an oral antiseptic. Hexachlorophane [2,2′-methylenebis(3,4,6-trichloro-phenol)] is most frequently used as a soap containing 2% of the substance; the effectiveness of this soap as a skin disinfectant may depend upon the accumulation of hexachlorophane on the skin. Pentachlorophenyl dodecanoate is extensively used as a fungicide and insecticide in the textile and packing industries. Unlike the parent molecule, pentachlorophenol, it is nontoxic to humans.

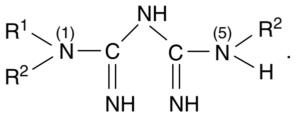
None of the above compounds can be relied upon to kill bacterial spores at ordinary temperatures.

Another group, related to the phenols, consists of the esters (parabens) of *p*-hydroxybenzoic acid (3-hydrobenzoic acid). Unlike benzoic acid, the dissociation of which increases with increasing pH with a corresponding decrease in activity, the parabens are active against bacteria and fungi over a fairly wide pH range. Their bactericidal and fungicidal properties increase within the homologous series, but this is paralleled by a decrease in aqueous solubility. The parabens are not sporicidal, and although on their discovery they were heralded as being the ideal preservatives, they are now used (often in combination of two or more) mainly as preservatives in various pharmaceutical and cosmetic products (Russell, Jenkins and Harrison, 1967; Parker, 1982).

The phenols and parabens have an effect on the cytoplasmic membranes of bacteria and fungi (Hugo, 1976a,b).

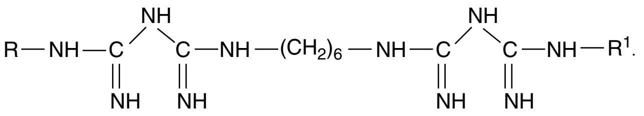
Biguanides and Bisbiguanides

Biguanides have the general formula



Three distinct antimicrobial actions of N^1, N^5-substituted biguanides have to date been recognized (Weinberg, 1968): germicidal, antiviral and antimalarial. Unfortunately, no one compound is generally active in more than one of these three categories. The requirements for each type of activity are fairly unique, e.g., for maximum broad-spectrum germicidal activity, both N^1 and N^5 should have a halogen-substituted aralkyl substituent.

Bisbiguanides have the general formula



For maximum broad spectrum germicidal activity R and R^1 should consist of a halogen-substituted aryl group. A number of bisbiguanides are known to be germicidal, and the most important member of this group is chlorhexidine (R and R^1 are both C_6H_5Cl).

Chlorhexidine was first described in 1954 (Davies *et al.*, 1954), and it is bacteriostatic in low concentrations; higher concentrations are bactericidal to several bacterial species, including strains of the genus *Proteus,* of which *Pr. mirabilis* is the most resistant, and *Pr. rettgeri* and *Pr. morganii* are the most sensitive to chlorhexidine, with *Pr. vulgaris* occupying an intermediate position. Some strains of *Ps. aeruginosa* may be highly resistant. Chlorhexidine also possesses antifungal activity. It is a membrane-active agent (Hugo, 1976a,b, 1982).

Chlorhexidine, the active constituent of "Hibitane," is used in surface disinfection, as an antimicrobial agent in eye-drops, and, in the presence of sodium nitrite to prevent corrosion, for the storage of surgical instruments.

Chlorine-Releasing Compounds

It was observed by Dakin (1915, 1916) that the commercial hypochlorites then in use were not of constant composition and contained free alkali and sometimes free chlorine. He thus developed a solution (Dakin's Solution or Chlorinated Soda Solution, Srugical) which is still in use today. The stability of free available chlorine in solution is dependent on a number of factors, in particular on the chlorine concentration, pH of the solution, the presence of catalysts, temperature, the presence of organic matter, and light (Dychdala, 1977).

The types of chlorine compounds which are frequently used are (1) Hypochlorites. These are cheap and convenient to use, and have a wide antibacterial spectrum (Davis, 1963). They possess potent sporicidal activity (Truman, 1971; Kelsey *et al.*, 1974; Waites, 1982) which may be potentiated by alcohols (Coates and Death, 1978; Death and Coates, 1979).

The hypochlorites are moderately effective against animal viruses.

The antibacterial activity of the hypochlorites decreases with increasing pH (Charlton and Levine, 1937; Weber, 1950; Ito *et al.*, 1967; Hays *et al.*, 1967), e.g., whereas 99% of spores of *B. cereus* are killed after 2.5 min at pH 6 by a solution containing 25 ppm available chlorine, nearly 8 hours are required for a comparable kill at a pH of *c.* 13 (Rudolph and Levine, 1941). At constant pH, the time to kill bacteria depends on the concentrations of available chlorine. The sporicidal activity of sodium hypochlorite may be potentiated by various compounds, e.g., by the addition of ammonia (Weber and Levine, 1944) or 1.5–4% sodium hydroxide (Cousins and Allan, 1967), notwithstanding the earlier comment about pH. In the presence of bromide, hypochlorite has an enhanced effect in bleaching cellulosic fibres as compared with hypochlorite alone, possibly because of a continuous generation of hypobromite when hypochlorite is in excess. A potentiation of the bactericidal effect of hypochlorite has been achieved by the addition of small amounts of bromide (Farkas–Himsley, 1964).

The antimicrobial activity of hypochlorites is considerably reduced by organic matter. However, the hypochlorites are used in the disinfection of water, dairy equipment and eating utensils. (2) Chloramine-T (sodium *p*-toluene sulphonchloramide). Dakin *et al.* (1916) considered that chloramine-T had a powerful germicidal action. It is bactericidal and sporicidal, although the rate of kill is slower than with the hypochlorites. Its activity is considerably higher at acid than at alkaline pH (Weber, 1950), and a drop of 10°C in the reaction temperature results in a 3–4 fold increase in the time necessary to kill microorganisms (Weber and Levine, 1944). Chloramine-T is employed as a wound "disinfectant," and as a general surgical disinfectant. It is nonirritant and nontoxic, in contrast to Dichloramine-T (toluene-*p*-sulphon-dichloramide) which although a powerful disinfectant is not used because of its toxicity and instability.

The mode of action of chlorine compounds is unknown, although several proposals have been made, e.g., the information of chloramines as a result of combination of chlorine with bacterial protoplasm, halogenation or oxidation reactions of chlorine with bacterial cells, changes in cellular permeability and an effect on enzyme systems. It has also been found, however (Bernarde *et al.*, 1967) that chlorine dioxide causes a marked and immediate cessation of protein synthesis in growing cells.

Iodine and Iodophors

Iodine in aqueous or alcoholic solution is considered by most authors (Gershenfeld and Witlin, 1950; Gershenfeld, 1956; Report, 1965; Sykes, 1970) to be a reliable and effective germicide which is lethal to vegetative bacteria, bacterial spores and acid-fast bacilli. Spaulding *et al.* (1977), however, consider that alcoholic iodine (0.5% iodine in 70% alcohol) possesses good activity against non-sporing bacterial and *M. tuberculoses* but none against bacterial spores, whereas Rubbo and Gardner (1965) state that bacterial spores are moderately resistant to iodine. Viruses are considered by Rubbo and Gardner (1965) to be moderately sensitive to iodine.

Iodine has a high fungicidal or fungistatic activity against yeasts and various moulds, but its antimicrobial properties are to a great extent inhibited in the presence of organic matter, since iodine is a highly reactive element.

Iodine is sparingly soluble in cold water, but more soluble in hot water. Stronger solutions can be made in potassium iodide solution or in aqueous alcohol. Iodine is more effective as a germicide at acid than at alkaline pH, but is less affected by pH than are chlorine compounds. The concentration of iodine to disinfect does not vary greatly with different types of microorganisms. Various types of iodine solution are used for the first-aid treatment of small wounds and abrasions, and in pre-operative skin "disinfection." Iodine has also been employed for the sterilization of surgical catgut (although this method is now little used) and is nowadays used for the disinfection of drinking and swimming pool water, the disinfection of instruments and of clinical thermometers, and the sanitization of eating and drinking utensils.

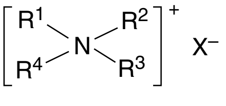
Unfortunately, iodine solutions stain fabrics and tissues and tend to be toxic. However, certain non-ionic surface-active agents can solubilize iodine to form compounds, the iodophors (Blatt and Maloney, 1961; Davis, 1962, 1963, 1968) which retain the germicidal activity of iodine, but not its undesirable properties; these iodophors are literally "iodine-carriers." They are active against bacterial spores, including pathogenic anaerobic spores (Lawrence *et al.*, 1957; Gershenfeld, 1962). It is the concentration of free iodine in an iodophor which is responsible for its microbial action; this has been well demonstrated by Allawala and Riegelman (1953) who made a log-log plot of killing time against amount of free iodine, and found that the 99% killing time of *B. cereus* spores was a function of the concentration of free-iodine in the presence or absence of added surface-active agent. The bactericidal properties of the iodophors are increased at low pH values, but their stability is unaffected (cf. hypochlorites). They may thus be employed with acids, e.g., phosphoric acid, to enhance their microbial action and also to assist in preventing the formation of film or milkstone (see later) in the dairy industry. The iodophors are considered (Davis, 1968) to be powerful detergents, although they do not dissociate protein as readily as do alkalis. The formulation of acidic solutions of iodophors is particularly useful when calcium or magnesium scale is encountered, but they can be corrosive, especially with galvanized iron.

Surface-Active Agents

Surface-active agents have 2 regions in their molecular structure, one a hydrocarbon, water-repellent (hydrophobic) group, and the other a water-attracting (hydrophilic or polar) group. Depending on the basis of the charge or absence of ionization of the hydrophilic group, surface-active agents are classified into (James, 1965):

1) Anionic agents, which usually have strong detergent, but weak antimicrobial properties, e.g., sodium lauryl sulphate.
2) Cationic agents, which have strong bactericidal, but weak detergent, properties. The term "cationic detergent" usually signifies a quarternary ammonium compound (QAC, onium compound), but this is not strictly accurate, as the concentration at which a QAC is germicidal is so low that its detergent activity is negligible.
3) Non-ionic agents, which consist of a hydrocarbon chain attached to a non-polar water-attracting group, which is usually a chain of ethylene oxide units, e.g., cetomacrogols. Non-ionic agents have no antimicrobial properties.
4) Ampholytic (amphoteric) agents, which are compounds of mixed anionic–cationic character, in which the charge can be similarly positive and negative. They are effective antimicrobial compounds, which are extensively used in the dairy industry.

The QACs are organically substituted ammonium compounds, in which the nitrogen atom has a valency of 5, four of the substituent radicals (R^1–R^4) are alkyl or heterocyclic radicals, and the fifth is a small anion. The sum of the carbon atoms in the 4R groups is more than 10. For a QAC to have a high activity, at least one of the four organic radicals must have a chain length in the range C_8 to C_{18}. The germicidal activities of the QACs were originally recognized in 1916, but they did not attain prominence until Domagk's work in 1935. They are primarily active against Gram-positive bacteria, with concentrations as low as about 1 in 200,000 being lethal; higher concentrations (*c.* 1 in 30,000) are lethal to Gram-negative bacteria, although *Pseudomonas aeruginosa* is highly resistant. The QACs possess antifungal properties, are sporostatic and not sporicidal (Russell, 1971) and are inactive against the tubercle bacillus. Their antimicrobial activity is markedly affected by organic matter, and they are incompatible with anionic surface-active agents, some of the non-ionic agents (such as Lubrols and Tweens) and the phospholipids, e.g., lecithin and other fat-containing substances (Russell, 1982; Russell *et al.*, 1979). The QACs exert their maximal bactericidal activity under alkaline conditions. Although the QACs are effective bactericidal and fungicidal agents, it has been found (Grossgebauer, 1970) that viruses are rather more resistant than bacteria or fungi.



Although several hundred QACs have been prepared and tested for antimicrobial activity, only a few are regularly used. These are:

1) Cetrimide (cetyltrimethylammonium bromide, CTAB), a mixture of dodecyl, tetradecyl- and hexadecyl-trimethylammonium bromide. In addition to it uses for pre-operative skin disinfection and treatment of seborrhoea of the scalp, cetrimide is also employed, in conjunction with sodium nitrite, which delays or prevents rusting, for the storage of sterilized surgical instruments, although this practice should be discontinued (Rubbo and Gardner, 1965).
2) Benzalkonium chloride (Zephirol, Zephiran) is the active constituent of a general purpose and skin sterilizing agent, "Roccal." It is also used to alleviate or prevent napkin rash caused by urea-splitting organisms, as a preservative in eye-drop preparations (Pharmaceutical Codex, 1979) and for the disinfection of blankets.

Other important QACs are domiphen bromide and cetylpyridinium chloride.

The value of QACs in the disinfection of woollen blankets has been observed by various authors (Frisby, 1957; International Wood Secretariat, 1961), although they cannot be incorporated in the wash with an anionic compound, because of incompatibility. In addition, it must be emphasized that residues of anionic agents on blankets from previous washings could interfere with the action of QACs. To overcome this, blankets may be washed in an appropriate non-ionic detergent, with a final prolonged rinse in a QAC, or they may be treated with a non-ionic detergent incorporating a QAC.

The QACs are also of considerable value as disinfectants in food and dairy plant. If an alkali detergent containing anionic surface-active wetting agents is used prior to sanitizing, then utensils and equipment must be thoroughly rinsed (Barrett, 1969). The QACs at their normally used concentrations are odourless and noncorrosive, but many are not free-rinsing, and undesirable traces may remain on equipment or even be present in dairy food (Clegg, 1967, 1970). However, a combination of a free-rinsing type of QAC and a suitable non-ionic agent may be usefully employed for washing instruments and cutlery, etc. (Barrett, 1969).

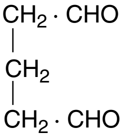
The cytoplasmic membrane of bacteria and fungi is the site of action of the QACs. The membrane is composed of lipoprotein, and it is considered (Russell, 1971) that the lipid moiety is involved in the lethal action of these compounds (for more recent information, see Hugo, 1976a,b).

There appears to be a clear relationship between the thermodynamic and antibacterial activities of QACs, with solutions having equal antimicrobial activity against an organism having surface concentration values of the same order of magnitude. Thermodynamic activities of QACs at two bacterial survivor levels (1% and 0.01%) have been shown to be sufficiently constant (Laycock and Mulley, 1970) to support the Ferguson (1939) principle that compounds with the same thermodynamic activity will have an equal bactericidal activity.

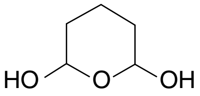
Amphoteric compounds, as already stated, are of mixed anionic-cationic character, and they combine the detergent properties of anionic compounds with the bacterial properties of the cationic substances; their bactericidal properties remain virtually constant over a wide pH range (Barrett, 1969) and they are less readily inactivated by proteins than are the QACs (Clegg, 1970). Examples of amphoteric surface-active agents are dodecyl-β-alanine, dodecyl-β-aminobutyric acid and dodecyldi(aminoethyl)-glycine (Davis, 1960a,b, 1968), the last named being a "Tego" compound. The Tego series of compounds have a high molecular weight, and in addition to being recommended for use in pre-operative hand cleansing and pre-operative skin preparation it has also been found that they are suitable for the cleansing of surgical operating theatre floors, walls and equipment and for ward cleansing (Frisby, 1959, 1961). It has, however, recently been shown that Tego 103S in 1% solution is less active than a 0.5% solution of chlorhexidine in 70% alcohol (Kuipers and Dankert, 1970). Amphoteric surface-active agents are inactivated by soaps and other anionic compounds (Frisby, 1959), but they are non-irritating and non-corrosive. Unfortunately, they tend to be expensive.

Aldehydes

The two most important aldehydes are glutaraldehyde (Pentanedial) and formaldehyde (methanal).



$$\begin{array}{l} CH_2 \cdot CHO \\ | \\ CH_2 \\ | \\ CH_2 \cdot CHO \end{array}$$



Hydrated Ring Structure

Glutaraldehyde is a dialdehyde which has been used for several years as a fixative in electron microscopy investigations and its antimicrobial activity has been comparatively recent (see Rubbo and Gardner, 1965; Rubbo *et al.*, 1967; Borick, 1968), but they do indicate that this substance has a valuable role to play. A 2% solution of glutaraldehyde buffered with sodium bicarbonate (0.3% w/v is considered to be the optimum bicarbonate concentration) is effective in killing nonsporing bacteria within 2 min, *M. tuberculosis,* fungi and viruses in 10 min, and spores of *Bacillus* and *Clostridium* spp. in 3 hours. Aqueous solutions of glutaraldehyde are acid, and are considerably less active against microorganisms than are alkaline ones (Pepper and Chandler, 1963; Stonehill *et al.*, 1963; Snyder and Cheatle, 1965; Lane *et al.*, 1966; Rubbo *et al.*, 1967; Munton and Russell, 1970a,b), but solutions become progressively less stable at pHs above 7. Concentrated solutions of glutaraldehyde (25%) can be purchased, diluted to the required concentration (2%) and "activated" by the addition of sodium bicarbonate. (Alternatively, 2% solutions ready for use when "activated" can also be purchased.) When made alkaline, glutaraldehyde solutions gradually undergo polymerization with a consequent loss of activity, this polymerization proceeding rapidly at pH values above 9. At pH 7.5–8.5, however, activity is maintained for at least 2 weeks.

Serum does not affect the antimicrobial activity of glutaraldehyde, but the dialdehyde is considerably less active in nutrient broth at pH 7.5 than it is at the same pH in buffer (Rubbo *et al.*, 1967; Munton and Russell, 1970a), the reason being that glutaraldehyde combines with the peptone present in broth (which is thereby discolored).

Glutaraldehyde is used as a fixative in the preparation of microbial cells for electron microscopy. It is a useful hospital disinfectant, particularly for articles which cannot be sterilized by physical means (Report, 1965). It has been employed in the sterilization of cytoscopes in urology (Lane, McKeever and Fallon, 1966) and of endoscopic instruments, such as bronchoscopes (Snyder and Cheatle, 1965), as it has no deleterious effect on the cement or lens coating. Glutaraldehyde is also employed as a tanning agent in preference in glyoxal and formaldehyde (Fein *et al.*, 1959), and has been shown to inactivate rapidly influenza virus and a coliphage in mouse tissue blocks (Sabel *et al.*, 1969).

Glutaraldehyde is non-corrosive, and does not affect rubber and plastic articles or the sharpness of cutting instruments; because it does not coagulate protein matter, such as blood and mucus, it does not render the cleaning of blood-covered instruments more difficult. It is obvious, therefore, that glutaraldehyde is most useful.

Rubbo *et al.* (1967) have proposed that the microbicidal activity of glutaraldehyde is due to the presence of two free aldehyde groups in the molecule. In solution, glutaraldehyde exists in an equilibrium between the open chain structure and the hydrated ring structure (see above), and there is a complete loss of activity if one or both of the aldehyde groups is altered, whereas a substitution elsewhere in the molecule reduces, but does not abolish, its activity. It is thus essential to have free aldehyde groups, which may react with cell sulphydryl or amino groups. Glutaraldehyde is about 10 times as active as glyoxal, with succinaldehyde occupying an intermediate position (Pepper and Chandler, 1963). Certain bacteria treated with glutaraldehyde become pink in color (Munton and Russell, 1970b) as a result of cell-aldehyde interaction.

Formaldehyde has long been employed as a disinfectant. Formaldehyde solution is rapidly sporicidal to *B. subtilis* (Ortenzio *et al.*, 1953; cf. Klarmann, 1956, 1959) but not to various *Clostridia* (Ortenzio *et al.*, 1953; Klarmann, 1956, 1959). Ethanol (Rubbo *et al.*, 1967) and methanol (Willard and Alexander, 1964) cannot be recommended as vehicles for formaldehyde, as there is a reduction in antibacterial activity.

Formaldehyde is an important virucidal agent which finds its greatest use in the preparation of certain sterile vaccines, e.g., Poliomyelitis Vaccine (Inactivated). As a result of the experimental evidence accumulated over several years, a considerable amount of information is now available on the kinetics of the virus inactivation by formaldehyde. This particular vaccine consists of poliovirus Types I, II and III, and each is inactivated separately and then blended to give the trivalent vaccine. It is thus essential that formaldehyde treatment be sufficient to destroy the viruses without affecting their antigenicity; prolonged exposure to the aldehyde will, in fact, destroy the antigenic potency also (Morris and Darlow, 1971). The inactivation of poliovirus by formaldehyde has been considered to be a first-order reaction so that extrapolation of the death curve would give a point at which the probability of any infective particles remaining would approach zero. However, first-order kinetics cannot be used with any degree of safety to extrapolate the inactivation curve. Similarly, although the inactivation of SV_{40} virus by formaldehyde has been shown by Sweet and Hilleman (1960) to be linear, it is now known that this linear inactivation is followed by a flattening of the curve indicating the persistence of a residual fraction which resists inactivation.

Formaldehyde is rapidly lethal to vegetative bacteria, and is sometimes used for this purpose in the preparation of inactivated bacterial vaccines.

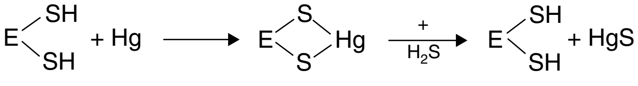
Metals

Because of their antibacterial and antifungal activity, compounds of mercury, silver, copper and tin are of importance from both medical and industrial points of view (Hugo and Russell, 1982).

Mercury Compounds These are of two types, the inorganic mercuric and mercurous salts and the organic substances. Mercuric salts are primarily bacteriostatic and fungistatic and contrary to earlier findings are not sporicidal (see Russell, 1971). Because of their toxicity, the mercuric salts do not find widespread use in modern medicine, but are extensively employed as industrial preservatives, e.g., in the preservation of wood, leather and paper, and in the control of fungal infections in seeds and bulbs. Mercurous salts have no application as preservatives.

The most important organic mercury compounds are the phenylmercuric salts (nitrate, acetate and borate) and thiomersal. Phenylmercuric nitrate (PMN) and acetate (PMA) are now mainly employed as preservatives in various pharmaceutical and cosmetic products. PMN is also used as a spermicide in certain contraceptive formulations, as a plant fungicide and for the disinfection of leather and timber. However, because of their lack of sporicidal activity at ordinary temperature, the organic mercury compounds cannot be recommended as sterilizing agents. Some plasmid-containing gram-negative bacteria are resistant to mercury compounds, which are vaporized (Chopra, 1982).

Various sulphydryl compounds, such as cysteine and thioglycollic acid, can reverse mercury-induced bacteriostasis, which led Fildes (1940) to propose that these compounds combined with, and displaced, mercury from its combination with the —SH group of an enzyme (E).



Silver Compounds Silver compounds have long been used in medicine for their antimicrobial activity, which extends to Gram-positive and Gram-negative bacteria and fungi. Of the silver compounds available, silver protein and silver nitrate are the most important. The latter, in the form of compresses, is highly effective in preventing the colonization of burns with *Ps. aeruginosa* and *Proteus* species.

Copper Compounds These are bactericidal and fungicidal. They have been used for the latter purpose for more than 200 years, and their sole use nowadays is as industrial preservatives against fungal spoilage. The most frequently used substances are copper naphthenate, oxinate, 1-phenylsalicylate and sulphate; the last-named, in combination with a lime mixture, is known as Bordeaux mixture.

Dialkyldithiocarbamates are considered (Albert, 1963) to be converted into active bactericides and fungicides in the presence of copper. Such salts are highly successful, widely used, agricultural fungicides (Owens, 1969).

Tin Compounds Stannous and stannic salts have little antimicrobial ctivity. However, when tin is coupled with organic radicals, forming what are known as the organotins, potent antimicrobial activity results. If R represents the organic radical linked directly to a tin atom, by C—Sn bond, and X an inorganic or organic radical not so linked, various types of compounds can be obtained, of which $R—SnN_3$ is most active. Gram-negative bacteria are less sensitive than Gram-positive bacteria to organotin compounds. Triphenyltin acetate and hydroxide are important agricultural fungicides.

Dyes

The acridines have held a valuable place in medicine for several years, although with the advent of the antibiotics and other chemotherapeutic agents, they are now less widely used than hitherto. The acridines are active against several Gram-positive and Gram-negative bacteria, and have been used mainly in treating infected wounds. Their uses and mode of action have been reviewed (Foster and Russell, 1971).

The most important members of the triphenylmethane group of dyes are crystal violet, brilliant green and malachite green. These have mainly been employed for local application to burns and wounds.

Some members of a third group of dyes, the quinones, are important agricultural fungicides. The quinones are natural dyes which impart color to many forms of plant and animal life. Chemically, the quinones are diketocyclohexadienes, the simplest of which is 1,4-benzoquinone (Figure 4). Naphtaquinones are the most toxic to bacteria, moulds and yeasts, followed (in this order) by phenanthrenequinones, benzoquinones and anthraquinones (Figure 4). Antimicrobial activity is increased by halogenation, and two powerful antimicrobial agents employed as fungicides are chloranil (tetrachloro-1, 4-benzoquinone) and dichlone (2,3-dichloro-1,4-naphthaquinone).

Alcohols

Ethyl alcohol, although active against Gram-positive and Gram-negative bacteria, is devoid of lethal activity against bacterial spores, and thus cannot be relied upon as a sterilizing agent (Russell, 1971). Methyl alcohol is likewise not

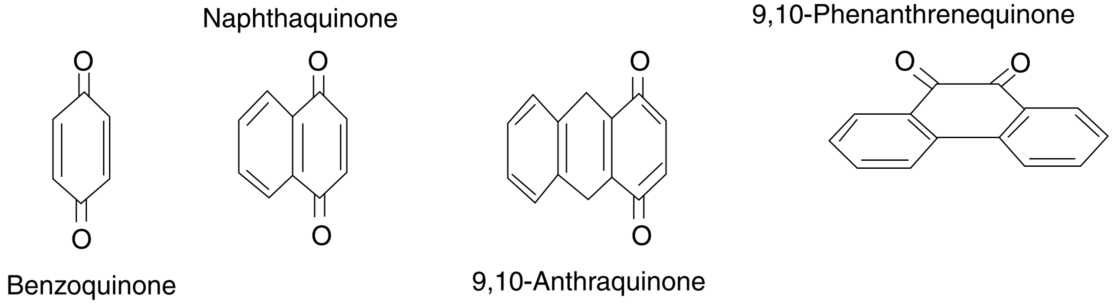


FIGURE 4

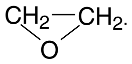
sporicidal. Moreover, alcohols may reduce the sporicidal activity of aldehydes.

GASEOUS ANTIMICROBIAL AGENTS

Gaseous agents may play an important role in sterilization of certain types of medical equipment and of components used in outer space research. However, only two gases (ethylene oxide and formaldehyde) are used extensively for the sterilization of medical products. Other gases (methyl bromide, β-propiolactone and propylene oxide) are not used as routine methods. Appropriate measures must be taken to counteract toxicity to humans (Christensen and Kristensen, 1982).

Ethylene Oxide

Ethylene oxide (EO) exists as a gas which is soluble in rubber, water, oils and several organic solvents. Chemically, it is



$$\underset{\diagdown\, O \,\diagup}{CH_2\!-\!CH_2}.$$

Its inflammability when in contact with air is overcome by using mixtures of EO with carbon dioxide or with fluorocarbon compounds (Phillips and Kaye, 1949; Barwell and Freeman, 1959; Freeman and Barwell, 1960). EO diffuses freely through paper, cellophane, cardboard and some plastics, but less rapidly through polythene; it cannot penetrate crystalline materials (Opfell and Miller, 1965). The antimicrobial activity of EO has been dealt with in many papers and reviews (Phillips, 1949, 1958, 1961, 1977; Phillips and Kaye, 1949; Kaye, 1949, 1950; Kaye and Phillips, 1949; Kaye *et al.,* 1952; Phillips and Warshowsky, 1958; Thomas, 1960; Bruch, 1961; Sutaria and Williams, 1961; Russell 1971; 1976; Kelsey, 1961, 1967; Sykes, 1970; Hoffman, 1971; Kereluk, 1971; Ernst, 1974, 1975; Hugo and Russell, 1982).

Several factors are known to influence the antimicrobial activity of EO (Christensen and Kristensen, 1982):

1) Concentration. As would be expected, the higher the concentration of EO (expressed as mg/l, which refers to the actual amount present in the sterilizer) the more rapid is the rate at which microorganisms are killed. However, even at high concentrations, EO is only slowly lethal, and long periods may, therefore, be needed for sterilization to be achieved.
2) Temperature. The lethal activity of EO increases with a rise in temperature. It has a temperature coefficient of 2.74 for each 10°C rise in temperature.
3) Type of organism. EO gas will kill bacteria and their spores, yeasts, moulds and their spores, and viruses (Griffith and Hall, 1938), and resistant strains have not been developed. In contrast to many other chemical substances which are considerably less effective against spores than against vegetative cells, bacterial spores are only about 2–10 times as resistant as the latter to EO (Toth, 1959; Phillips, 1958) and in some cases, e.g., *B. stearothermophilus* (Thomas *et al.,* 1969) even less resistant. These results imply that EO can freely penetrate the outer layer(s) of the bacterial spore, although experimental results in support of this contention are sadly lacking.

   In addition to its antibacterial and antifungal activity, ethylene oxide is also effective against rickettsiae and viruses (Hoffman, 1971).
4) Relative humidity (RH). Of all the factors which influence the activity of EO, RH is probably the most important. The optimum RH is considered to be *c.* 28–33% (Schley *et al.,* 1960), and EO is considerably less microbicidal at high RH and in relatively dry air. The correct RH may be achieved to prehumification of the sterilization chamber (Halowell *et al.* 1958; Ernst and Schull, 1962). However, the moisture content of microorganisms themselves, as well as the RH of the environment, is also important. Bacterial cells which have been desiccated and then equilibrated to successively high RH values contain less water and are more resistant to EO than cells which have not been desiccated but have instead been allowed to dry naturally until equilibrated to the same RH values. To overcome this resistance

to EO produced by dehydration, the cells have to be actually wetted (Perkins and Lloyd, 1961; Winge-Heden, 1963; Gilbert *et al.,* 1964).

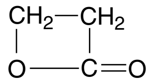
Alkylating agents act through alkylation of sulphydryl, amino, carboxyl, hydroxyl and phenolic groupings, with the loss of a hydrogen atom and the production of an alkyl hydroxyethyl group (Phillips, 1952, 1958; Kelsey, 1967), and it seems likely that EO kills microorganisms by an alkylation of protein molecules (Gilbert *et al.*, 1964). The influence of RH on the microbicidal activity of EO is related to this, since the presence of insufficient water prevents alkylation, whereas excess water causes hydrolysis of EO to ethylene glycol, CH_2OHCH_2OH (Wilson and Bruno, 1950).

5) Effect of drying medium on spore resistance. Organisms dried from salt solution always show some survivors after exposure to EO (Royce and Bowler, 1961; Beeby and Whitehouse, 1965). Moreover, organisms enclosed within crystals are protected from the action of EO (Abbott *et al.*, 1956; Royce and Bowler, 1961) as a result of the inability of the gas to penetrate crystalline materials.

EO has several uses in the pharmaceutical and medical fields. These include the sterilization of ophthalmic instruments, anaesthetic equipment, heart-lung machines, disposable syringes (Rubbo and Gardner, 1968) and hospital blankets (Kaye, 1950), and as a decontamination procedure for articles handled by tuberculous patients. There may, however, be a "crazing" of disposable syringes, and EO treatment is a slow, costly process, with a "quarantine" period necessary at the end of the process to ensure that complete dissipation of the gas has been achieved.

β-Propiolactone

β-Propiolactone (BPL) is a colorless liquid at room temperature. It has been employed as a chemosterilizer in both the liquid and gaseous states, and is microbicidal in both forms (Allen and Murphy, 1960). Its antimicrobial activity depends on its concentration and on the temperature and RH at which the vapor is used (Spiner and Hoffman, 1960), a temperature above 24°C being required for optimal activity. As with EO, RH is of considerable importance in determining the activity of BPL vapor, although with the latter a RH in excess of 70% is required for rapid microbicidal acivity (Hoffman and Warshowsky, 1958). Again, however, it is not necessarily the atmospheric RH which is of greatest importance but the location and content of water within the microbial cell. *B. globigii* spores equilibrated to 98% are rapidly killed by BPL at 45% RH, whereas they are more resistant if preconditioned at 75% RH before treatment with BPL at 45% RH, and a small proportion of spores equilibrated at 1% RH are subsequently highly resistant to BPL at 75% RH (Hoffman, 1968).



Bacterial spores are more resistant to BPL than vegetative cells, viruses or fungi (Lo Grippo *et al.*, 1955; Trafas *et al.*, 1954; Bruch and Koesterer, 1962; Toplin, 1962) although some strains of *Staph. aureas* may be almost as resistant as spores (Hoffman and Warshowsky, 1958).

BPL is also highly active against viruses and rickettsiae (Hoffman, 1971).

BPL has been used for the chemical sterilization of regenerated collagen sutures (Ball *et al.*, 1961), for the decontamination of enclosed spaces (Bruch, 1961b) and for the sterilization of a variety of instruments contaminated with various sporing and non-sporing bacteria (Allen and Murphy, 1960). However, its reported carcinogenic effects in rats and mice (Walpole *et al.*, 1954) mean that a considerable degree of caution is needed before BPL is employed as a chemosterilizer.

Formaldehyde

Formaldehyde vapor may be obtained by evaporating appropriate dilutions of standardized batches of commercial formalin (a 40% solution of formaldehyde in water) with 10% methanol added to prevent polymerization (Report, 1958). Temperature affects the activity of the gas, as does the RH, there being an increase in activity with increasing RH up to 50%, but little further increase in killing rate as the RH rises from 50 to 90% (Nordgren, 1939; Report, 1958). In contrast, gross wetting retards killing.

Bacteria protected by organic matter, such as blood and sputum are less rapidly killed by formaldehyde vapor (Nordgren, 1939; Bullock and Rawlins, 1954). Microorganisms may also be protected from it when they are included in a crystal mass, in contrast to surface-contaminated crystals (Abbot *et al.*, 1956).

Although bacterial spores are more resistant than vegetative cells to formaldehyde vapor, the degree of difference is not high (Phillips, 1952; Report, 1958; Sykes, 1965). The vapor only has weak penetration, and its application is thus normally limited to surface sterilization (Borick, 1968; Davis, 1968). However, the addition of formaldehyde vapor to steam under sub-atmospheric pressure at temperatures below 90°C results in deep penetration into fabrics with destruction of heat-resistant microorganisms (Alder *et al.*, 1966; Alder and Simpson, 1982).

Formaldehyde vapor has long been used for the disinfection of blankets, and is considered to be one of the best methods available for disinfecting woolen blankets that have not received a shrink-resist treatment (International Wool Secretariat, 1961). It is also used to decontaminate rooms, buildings and instruments (Hoffman, 1971).

PRACTICAL USES

This final section brings together some of the data presented in the preceding sections, and also provides information in certain specific instances. Disinfectants will be considered from two points of view, first their medical, and second their nonmedical uses. Some brief information on antiseptic and preservative use will also be supplied.

Medical Uses

The use and choice of disinfectants in hospitals have been extensively considered in the last decade in Britain by a special committee (Report, 1965) set up for this purpose. Our comments here are thus based on the recommendations of this report and on the findings presented by Kelsey and Maurer (1967). The Report (1965) recommended that two classes of disinfecting agents were needed, (a) for general disinfection and (b) for surface disinfection of clean objects. Agents for general disinfection should have a wide spectrum, and at appropriate dilutions should remain active in the presence of organic matter; the main purpose of such disinfectants, e.g., phenolic disinfectants based on coal-tar acids, is not necessarily to kill *all* bacteria but to ensure that an object is free of significant numbers of organisms. Chemicals for surface disinfection must be quick-acting, have a wide spectrum, be non-harmful to materials and leave no objectionable odours. Such disinfectants, e.g., hypochlorites, should be used for the rapid disinfection of clean surfaces such as trolley tops, kitchen tables and clinical thermometers.

Kelsey and Maurer (1967, 1972) have presented a list of the steps to be taken in drawing up a policy for the use of general purpose disinfectants in hospitals, and among the points they make is the non-usage of disinfectants in certain cases, especially where sterilization is the objective or where other more reliable means are available. For further information, see Lynn (1980), Ayliffe and Collins (1982) and Lowbury (1982). Preoperative disinfection of the skin (including surgeon's hands), disinfection of operation sites and topical prophylaxis, i.e., antisepsis in burns, are discussed by Lowbury (1982). Ayliffe and Collins (1982) provide a rational approach to hospital disinfection.

Nonmedical Uses

The main nonmedical uses of disinfectants occur in the food, dairy, brewing and fermentation industries (Foster *et al.*, 1958; Frazier, 1967). The maintenance of equipment for use in these industries in a proper sanitary condition cannot be overemphasized. This therefore means that the cleaning of such equipment is of considerable importance, since the presence of organic matter can reduce or virtually eliminate the effect of many disinfectants (page 164). In the dairy industry, milk stone—resulting from milk drying on equipment, and thus consisting of fat, protein and minerals—and milk film are a well-known problem in disinfection (Clegg, 1967). Chemicals which are of use against microorganisms in liquid suspension in laboratory tests may be of little use against such organisms on a soiled surface if they are poor detergents (Cousins, Hoy and Clegg, 1960).

To counteract the unwanted effects of organic matter, one of the following two methods may be employed (see also Davis, 1972a,b and BSI 1977):

1) detergent first, followed by a disinfectant;
2) combination of detergent and disinfectant (this corresponds to a sanitizer, or to the detergent-sterilant of Davis (1968), as described in the Introduction).

Harris (1969) stresses the need for using two operations, i.e., the use of cleaning before disinfection. Cleaning is the first essential in the sanitary care of food equipment, and approximate sterility the last (Foster *et al.*, 1958). Steam under pressure is an obvious method of sanitization, but this is limited only to closed systems which can withstand pressure (Frazier, 1967). Theoretically, separate procedures would be expected to give a better result (Clegg, 1970), because of the inactivating effect of organic matter on disinfectants; however, a finding made several years ago (Neave and Hoy, 1947) suggests that a detergent-disinfectant combination would be of greater use, because "the effect of the detergent on milk solids more than outweighs the effect of milk residues inactivating the disinfectant" (Clegg, 1970). An effective detergent should dissociate organic and inorganic solids, emulsify, saponify or suspend grease, fats and oils, have good wetting ability, be easily rinsed and be non-corrosive (Olivant and Shapton, 1970). Detergents are thus of considerable importance in this field, because they can also be responsible for the mechanical removal of bacteria (Gilbert, 1969). The most commonly used detergents are strong and mild alkalis, alkali salts, strong acids, anionic alkyl sulphates and aryl sulphonates and non-ionic condensates (Davis, 1968). An excellent descaling agent is nitric acid which can be used hot at concentrations of 0.25–0.5% or cold at 0.5–1%, and which is, in addition, a powerful disinfectant. It is, however, less effective than alkali in the removal of hardened protein films, and is normally employed with a corrosion inhibitor. In actual fact, many detergents are good disinfectants and vice versa, e.g., a detergent such as sodium hydroxide possesses considerable germicidal power (Whitehouse and Clegg, 1963), whereas hypochlorites have a useful detergent effect by disintegrating protein matter (Davis, 1968). Cleaning is an essential first part in high-speed food and beverage processing plant, and considerable economic benefit is achieved as a result of the production of stable liquid detergents and disinfectants which can be delivered by tanker and then distributed to the cleaning areas by pipeline (Hill, 1969). However, even with good precleaning, traces of inactivating protein material may remain on equipment, and it is thus important to choose a disinfectant which has a high protein tolerance (Harris, 1969). The design of equipment to facilitate cleaning must also be stressed, and stainless steel is an obvious example, with glass pipelines to give a high degree of visual cleanliness (Harris, 1969).

The disinfection of fermentation laboratories has been described by Darlow (1969). Gaseous disinfectants may be employed, and aerosols appear to have an important role to play. Local disinfection of bench tops, floors, etc. is also a standard practice. The importance of disinfectants in water conservation is emphasized by Fielden (1969), and in the pharmaceutical industry by Underwood (1980).

The control of airborne microorganisms is of particular importance in the fermentation industry, in laboratories where strict asepsis is essential (e.g., in the production of various sterile products in the pharmaceutical industry and in hospital pharmacy departments, as well as in the rearing of germ-free animals), in hospital wards to reduce the incidence of cross-infection, and in special wards set aside for patients with a rare disease (hypogammaglobulinaemia) who are particularly sensitive to infection.

This control is normally achieved by the use of special air filters, often in conjunction with ultraviolet lamps to irradiate the upper atmosphere. Disinfectants in the form of aerosols are also of importance in aerial disinfection. To be effective for air disinfection, a chemical should ideally possess the following properties:

1) be odorless, cheap and stable
2) be without toxic or irritant properties
3) be capable of being dispersed in the air, with consequent complete and rapid mixing of infected air and chemical
4) be capable of maintaining an effective concentration in the air
5) be highly and rapidly effective against airborne organisms
6) be unaffected by relative humidity (RH).

Aerosols consist of a very fine dispersed liquid phase in a gaseous (air) continuous phase. The germicide must be nebulized in sufficiently fine spray (aerosols droplets of <1 μm are the accepted standard) so that it will remain airborne and thus have ample opportunity to collide with any microorganisms present in the air. At low RH, particles are too dry for adequate condensation of the disinfectant that such organisms enclosed in particles, and thus bacteria occurring on dust or on surfaces are much less susceptible to the aerial disinfectant than such organisms enclosed in droplets (Sykes, 1965). The optimum RH is usually *c.* 40–60%.

Chemical aerosols are often generated in the following manner: if the chemical is liquid, it may be sprayed directly into the air from an atomizer; if the chemical exists as a solid, it may be dissolved in an appropriate solvent, e.g., propylene glycol, and atomized, or alternatively the solid may be vaporized by heat from a thermostatically-controlled hot-plate.

Chemicals which have been used as aerial disinfectants include hexylresorcinol, lactic acid, propylene glycol (this possesses antimicrobial activity in its own right), hypochlorous acid, formaldehyde gas and sulphur dioxide.

Other Uses of Antimicrobial Agents

Antimicrobial agents are widely employed as preservatives in pharmaceutical and cosmetic products. Factors influencing their activity, as well as those governing their choice in different classes of sterile or non-sterile products have been well considered by Bean (1972), Croshaw (1977), Parker (1978, 1982), Kazmi and Mitchell (1978a,b), Allwood (1980) and Beveridge (1980).

Preservation is also required in other specialized areas, e.g., in cutting oils (Hill, 1982a), fuels (Hill, 1982b), paper and pulp (Weir, 1982), wood (Richardson, 1982), paint (Springle and Briggs, 1982) and textiles (Hugo, 1982b) and in the construction industry (Bravery, 1982). A most important aspect concerns the use of chemical food preservatives and this has been comprehensively reviewed by Sofos and Busta (1982).

REFERENCES

1. Abbot, C.F., J. Cockton, and W. Jones, 1956, *J. Pharm. Pharmac.,* **8,** 709.
2. Albert, A., 1963, *Adv. Appl. Microbiol.,* **5,** 1.
3. Alder, V.G., A.M. Brown, and W.A. Gillespie, 1966, *J. Clin. Path.,* **19,** 83.
4. Alder, V.G. and R. Simpson, 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
5. Allawala, N.A. and S. Riegelman, 1953, *J. Am. Pharm. Ass., Sci. Ed.,* **42,** 396.
6. Allen, H.F. and J.T. Murphy, 1960, *J. Am. Med. Ass.,* **172,** 1759.
7. Allwood, M.C., 1980, in *Pharmaceutical Microbiology,* 2nd Edition, Ed., W.B. Hugo and A.D. Russell, Blackwell Scientific Publications, Oxford.
8. Association of Official Agricultural Chemists, 1970, Official Methods of Analysis of the AOAC, Washington, DC.
9. Ayliffe, G.A.J. and B. Collins, 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
10. Ball, E.L., A.C. Dornbrush, G.M. Sieger, F.E. Stirn, J.C. Vitucci, and J.F. Weidenheimer, 1961, *Appl. Microbiol.,* **9,** 269.
11. Barrett, M., 1969, *Process Biochem.,* **4,** 23.
12. Barwell, C.F. and M.A.R. Freeman, 1959, *Lancet,* **i,** 917.
13. Bean, H.S., 1967, *J. Appl. Bact.,* **30,** 6.
14. Bean, H.S., 1972, *J. Soc. Cosmet. Chem.,* **23,** 703.
15. Bean, H.S. and A. Das, 1966, *J. Pharm. Pharmac.,* **18,** 107S.
16. Bean, H.S. and S.M. Heman-Ackah, 1964, *J. Pharm. Pharmac.,* **16,** 58T.
17. Bean, H.S. and V. Walters, 1955, *J. Pharm. Pharmac.,* **7,** 661.
18. Beeby, M.M. and C.E. Whitehouse, 1965. *J. Appl. Bact.,* **28,** 349.
19. Benarde, M.A., W.B. Snow, V.P. Olivieri, and B. Davidson, 1967, *Appl. Microbiol.,* **15,** 257.
20. Bennett, E.O., 1959, *Adv. Appl. Microbiol.,* **1,** 49.
21. Berry, H., 1951, *J. Pharm. Pharmac.,* **3,** 689.
22. Berry, H., and H.S. Bean, 1954, *J. Pharm. Pharmac.,* **7,** 224.
23. Berry, H., and A. Briggs, 1956, *J. Pharm. Pharmac.,* **8,** 1143.
24. Berry, H., A.M. Cook, and B.A. Wills, 1956, *J. Pharm. Pharmac.,* **8,** 425.
25. Berry, H., E. Jensen, and F.K. Silliker, 1937, *Quart. J. Pharm. Pharmac.,* **11,** 729.
26. Beveridge, E.G., 1980, in *Pharmaceutical Microbiology,* 2nd Edition. Edition. Ed., W.B. Hugo and A.D. Russell, Blackwell Scientific Publications, Oxford.
27. Blatt, R. and J.V. Moloney, 1961, *Surg. Gynaec. Obstet.,* **113,** 699.
28. Borick, P.M., 1968, *Adv. Appl. Microbiol.,* **10,** 291.
29. Borick, P.M., 1970, in *Disinfection,* Ed., M.A. Benarde, Marcel Dekker, New York.

30. Boucher, R.M.G., 1975, *Can. J. Pharm. Sci.,* **10,** 1.
31. Bravery, A.F., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
32. Briggs, A., 1966, *J. Appl. Bact.,* **29,** 490.
33. British Pharmacopoeia, 1980, Pharmaceutical Press, London.
34. British Standards Institution, 1934, British Standard techniques for determining the Rideal-Walker Coefficient of Disinfectants: BS 541.
35. British Standards Institution, 1938. British Standard specification for the modified technique of the Chick-Martin Test for Disinfectants: BS 808.
36. British Standards Institution, 1977, Recommendations for sterilization of plant and equipment used in the dairying industry: BS 5305.
37. Bruch, C.W., 1961a, *Ann. Rev. Microbiol.,* **15,** 245.
38. Bruch, C.W., 1961b, *Am. J. Hyg.,* **73,** 1.
39. Bruch, C.W., and M.G. Koesterer, 1962, *Appl. Microbiol.,* **10,** 123.
40. Bullock, K. and E.A. Rawlins, 1954, *J. Pharm. Pharmac.,* **6,** 859.
41. Cadet, J. *et al.,* Photosensitized Reactions of Nucleic Acids, *Biochimie,* **68,** 1986.
42. Charlton, D.B. and M. Levine, 1937, *Iowa Eng. Exp. Sta. Bull.,* No. 132, Iowa, USA.
43. Chick, H., 1908, *J. Hyg. Camb.,* **8,** 92.
44. Chick, H. and C.J. Martin, 1908, *J. Hyg. Camb.,* **8,** 698.
45. Chopra, I.C., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
46. Christensen, E.A. and K. Christensen, 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
47. Clegg, L.F.L., 1967, *J. Appl. Bact.,* **30,** 117.
48. Clegg, L.F.L., 1970, in *Disinfection,* Ed., M.A. Benarde, Marcel Dekker, New York.
49. Coates, D. and J.E. Death, 1978, *J. Clin. Path.,* **31,** 148.
50. Coates, D., 1977, *Pharm. J.,* **219,** 402.
51. Cook, A.M., 1960, *J. Pharm. Pharmac.,* **12,** 19T.
52. Cook, A.M. and B.A. Wills, 1954, *J. Pharm. Pharmac.,* **6,** 638.
53. Cousins, C.M. and C.D. Allan, 1967, *J. Appl. Bact.,* **30,** 168.
54. Cousins, C.M., W.A. Hoy, and L.F.L. Clegg, 1960, *J. Appl. Bact.,* **23,** 359.
55. Cowen, R.A., 1978, *Pharm. J.,* **220,** 202.
56. Croshaw, B., 1977, *J. Soc. Cosmet. Chem.,* **28,** 3.
57. Dakin, H.D., 1915, *Br. Med. J.,* **2,** 809.
58. Dakin, H.D., 1916, *Br. Med. J.,* **2,** 318.
59. Dakin, H.D., J.B. Cohen, and J. Kenyon, 1916, *Br. Med. J.,* **1,** 160.
60. Darlow, H.M., 1969, *Process Biochem.,* **4,** 15.
61. Davies, G.E., J. Francis, A.R. Martin, F.L. Rose, and G. Swain, 1954, *Br. J. Pharmac.,* **9,** 192.
62. Davis, J.G., 1960a, *J. Pharm. Pharmac.,* **12,** 29T.
63. Davis, J.G., 1960b, *J. Appl. Bact.,* **23,** 318.
64. Davis, J.G., 1962, *J. Appl. Bact.,* **25,** 195.
65. Davis, J.G., 1963, *The Sanitarian,* October, p. 3.
66. Davis, J.G., 1968, *Progr. Industr. Microbiol.,* **8,** 141.
67. Davis, J.G., 1972a, *J. Soc. Cosmet. Chem.,* **23,** 45.
68. Davis, J.G., 1972b, *Dairy Ind.,* **37**(4), 212.
69. Death, J.E. and D. Coates, 1979, *J. Clin. Path.,* **32,** 148.
70. Dychdala, G.R., 1977, in *Disinfection, Sterilisation and Preservation,* 2nd Edition, Ed., S.S. Block, Lea and Febiger, Philadelphia.
71. Ernst, R.R., 1974, *Biotech. Bioeng. Symp. No. 4,* p. 858.
72. Ernst, R.R., 1975, *Acta Pharmaceutica Suecica,* **12** (Supplement), 44.
73. Ernst, R.R. and J.J. Shull, 1962. *Appl. Microbiol.,* **10,** 337.
74. Farkas-Himsley, H., 1964, *Appl. Microbiol.,* **12,** 1.
75. Fein, M.L., E.H. Harris, J. Naghski, and E.M. Filachione, 1959, *J. Am. Leather Chem. Ass.,* **54,** 488.
76. Ferguson, J., 1939, *Proc. Roy. Soc. B,* **127,** 387.
77. Fielden, T.B., 1969, *Process Biochem.,* **4,** 35.
78. Forsyth, M.P., 1975, *Dev. Ind. Microbiol.,* **16,** 37.
79. Flett, L.H., C. Haring, A.F. Guiteras, and R.L. Shapiro, 1945, *J. Bact.,* **50,** 591.
80. Foster, E.M., F.E. Nelson, M.L. Speck, R.N. Doetsch, and J.C. Olson, 1958, *Dairy Microbiology,* Macmillan and Co. Ltd.
81. Foster, J.H.S. and A.D. Russell, 1971, in *Inhibition and Destruction of the Microbial Cell,* Ed., W.B. Hugo, Academic Press, London and New York.
82. Frazier, W.C., 1967, *Food Microbiology,* 2nd Edition, McGraw-Hill, New York.
83. Freeman, M.A.R. and C.F. Barwell, 1960, *J. Hyg., Camb.,* **58,** 337.
84. Frisby, B.R., 1959, *Lancet,* **2,** 57.
85. Frisby, B.R., 1961, *Lancet,* **2,** 829.
86. Garrod, L.P., 1934, *J. Hyg., Camb.,* **34,** 322.
87. Garrod, L.P., 1935, *J. Hyg., Camb.,* **35,** 219.
88. Gershenfeld, L., 1956, *Am. J. Pharm.,* **128,** 335.
89. Gershenfeld, L., 1962, *Am. J. Pharm.,* **134,** 78.
90. Gershenfeld, L. and B. Witlin, 1950, *Ann. NY Acad. Sci.,* **53,** 172.
91. Gilbert, G.L., V.M. Gambill, D.R. Spiner, R.K. Hoffman, and C.R. Phillips, 1964, *Appl. Microbiol.,* **12,** 496.
92. Gilbert, J., 1969, *Process Biochem.,* **4,** 37.
93. Gilireas, F.W. and J.E. O'Brien, 1941, *Am. J. Publ. Health,* **31,** 143.
94. Goetchins, G.R. and W.E. Botwright, 1950, *J. Milk Technol.,* **13,** 63.
95. Griffith, C.L. and L.A. Hall, 1938, *US Patent 2,* **189,** 947.
96. Grossgebauer, K., 1970, in *Disinfection,* Ed., M.A. Benarde, Marcel Dekker, New York.
97. Halowell, P., J.T. Murphy, and A.B. Mangiaracine, 1958, *Anaesthesiology,* **19,** 665.
98. Hare, R., E. Raik, and S. Gash, 1963, *Br. Med. J.,* **1,** 496.
99. Harris, C., 1969, *Process Biochem.,* **4,** 19.
100. Harris, N.D., 1963, *J. Appl. Bact.,* **26,** 387.
101. Hays, H., P.R. Elliker, and W.E. Sandine, 1967, *Appl. Microbiol.,* **15,** 575.
102. Hill, E.C., 1982a, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
103. Hill, E.C., 1982b, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
104. Hill, R.P., 1969, *Process Biochem.,* **4,** 31.
105. Hoffman, R.K., 1968, *Appl. Microbiol.,* **16,** 641.
106. Hoffman, R.K., 1971, in *Inhibition and Destruction of the Microbial Cell,* Ed., W.B. Hugo, Academic Press, London and New York.
107. Hoffman, R.K. and B. Warshowsky, 1958, *Appl. Microbiol.,* **6,** 358.
108. Hoy, W.A. and L.F.L. Clegg, 1953, *Proc. Soc. Appl. Bact.,* **16,** 1.
109. Hugo, W.B., 1976a, in *The Survival of Vegetative Microbes,* Ed., T.R.G. Gray and J.R. Postgate, 26th Symposium of the Society for General Microbiology, Cambridge University Press, Cambridge.
110. Hugo, W.B., 1976b, in *The Inactivation of Vegetative Bacteria,* Ed., F.A. Skinner and W.B. Hugo, Society for Applied Bacteriology Symposium Series No. 5, Academic Press, London and New York.
111. Hugo, W.B., 1982a, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
112. Hugo, W.B., 1982b, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
113. Hutchinson, F., A Review of Some Topics Concerning Mutagenesis by Ultraviolet Light, *Photochem. Photobiol.,* **45,** 1987.
114. International Wool Secretariat, 1961, The Sterilization of Hospital Woollen Blankets, *Wool Science Review,* No. 19 and No. 20.
115. Ito, K.A., D.J. Seslar, and W.A. Mercer, 1967, *in Botulism,* **66.** Ed., M. Ingram and T.A. Roberts, Chapman and Hall, London.
116. Jacobs, S.E. and N.D. Harris, 1960, *J. Appl. Bact.,* **23,** 294.
117. James, A.M., 1965, *Surface Activity and the Microbial Cell,* Soc. Chem. Ind., London.
118. Jensen, V. and E. Jensen, 1933, *Dansk. Tidss. Farm.,* **7,** 77.
119. Johns, C.K., 1946, *Am. J. Publ. Health,* **37,** 1322.
120. Johns, C.K., 1954, *Can. J. Technol.,* **32,** 71.
121. Kaye, S., 1949, *Am. J. Hyg.,* **50,** 289.
122. Kaye, S., 1950, *J. Lab. Clin. Med.,* **35,** 823.
123. Kaye, S. and C.R. Phillips, 1949, *Am. J. Hyg.,* **50,** 296.
124. Kaye, S., H.F. Irminger, and C.R. Phillips, 1952, *J. Lab. Clin. Med.,* **40,** 67.
125. Kazmi, S.J.A. and A.G. Mitchell, 1978a, *J. Pharm. Sci.,* **67,** 1260.

126. Kazmi, S.J.A. and A.G. Mitchell, 1978b, *J. Pharm. Sci.,* **67,** 1266.
127. Kelsey, J.C., 1961, *J. Clin. Path.,* **14,** 313.
128. Kelsey, J.C., 1967, *J. Appl. Bact.,* **30,** 92.
129. Kelsey, J.C., and I.M. Maurer, 1967, *Mon. Bull. Min. Health,* **26,** 110.
130. Kelsey, J.C. and I.M. Maurer, 1972, *The Use of Chemical Disinfectants in Hospitals,* Public Health Laboratory Service Monographs No. 2, HMSO, London.
131. Kelsey, J.C. and I.M. Maurer, 1974, *Pharm. J.,* **207,** 528.
132. Kelsey, J.C. and Sykes, 1969, *Pharm. J.,* **202,** 607.
133. Kelsey, J.C., M.M. Beeby, and C.M. Whitehouse, 1965, *Mon. Bull. Min. Health,* **24,** 162.
134. Kelsey, J.C., I.H. Mackinnon, and I.M. Maurer, 1974, *J. Clin. Path.,* **27,** 632.
135. Kereluk, K., 1971, in *Progress in Industrial Microbiology,* **10.** Ed., D.J.D. Hockenhull, Churchill Livingstone, Edinburgh.
136. Klarmann, E.G., 1956, *Am. J. Pharm.,* **128,** 4.
137. Klarmann, E.G., 1959, *Am. J. Pharm.,* **131,** 86.
138. Knox, W.E., V.H. Auerbach, K. Zarudnaya, and M. Spirtes, 1949, *J. Bact.,* **58,** 443.
139. Koch, R., 1881, *Arb. Kaiserl. Gesundh.,* **1,** 1.
140. Kronig, B., and T. Paul, 1897, *Z. Hug. Infekt. Kr.,* **25,** 1.
141. Kuipers, J.S. and J. Dankert, 1970, *J. Hug., Camb.,* **68,** 343.
142. Lamy, P.P. and H.L. Flack, 1962, *Am. J. Hosp. Pharm.,* **19,** 473.
143. Lane, V., J.D. McKeever, and M. Fallon, 1966, *J. Irish Med. Ass.,* **58,** 131.
144. Lawrence, C.A., C.M. Carpenter, and A.W.C. Naylor-Foote, 1957, *J. Am. Pharm. Ass., Sci. Ed.,* **46,** 500.
145. Laycock, H.H. and B.A. Mulley, 1970, *J. Pharm. Pharmac.,* **22,** 1515.
146. Lo Grippo, G.A., P.R. Overhulse, D.E. Szilagyi, and F.W. Hartman, 1955, *Lab. Invest.,* **4,** 217.
147. Loosemore, M. and A.D. Russell, 1963, *J. Pharm. Pharmac.,* **15,** 558.
148. Loosemore, M. and A.D. Russell, 1964, *J. Pharm. Pharmac.,* **16,** 817.
149. Lowbury, E.J.L., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
150. Lynn, B., 1980, in *Pharmaceutical Microbiology,* 2nd Edition, Ed., W.B. Hugo and A.D. Russell, Blackwell Scientific Publications, Oxford.
151. Mallman, W.L. and M. Hanes, 1947, *J. Bact.,* **49,** 526.
152. Mandels, G.R. and R.T. Darby, 1953, *J. Bact.,* **65,** 16.
153. Mather, R.L., 1949, *Biometrics,* **5,** 127.
154. Miner, N.A., E. Whitmore, and M.L. McBee, 1975, *Dev. Ind. Microbiol.,* **16,** 23.
155. Morris, E.J. and H.M. Darlow, 1971, in *Inhibition and Destruction of the Microbiol Cell,* Ed., W.B. Hugo, Academic Press, London and New York.
156. Mulley, B.A., 1964, *Adv. Pharm. Sci.,* **1,** 87.
157. Munton, T.J. and A.D. Rusell, 1970a, *J. Appl. Bact.,* **33,** 410.
158. Munton, T.J. and A.D. Russell, 1970b, *Experientia,* **27,** 109.
159. Neave, F.K. and W.A. Hoy, 1947, *J. Dairy Res.,* **15,** 24.
160. Needham, N.V., 1947, *J. Hyg., Camb.,* **45,** 1.
161. Nordgren, G., 1939, *Acta Path. Microbiol. Scand. suppl.,* No. 40, 1.
162. Olivant, D.J. and D.A. Shapton, 1970, in *Disinfection,* Ed. M.A. Benarde, Marcel Dekker, New York.
163. Opfell, J.B. and C.E. Miller, 1965, *Adv. Appl. Microbiol.,* **7,** 81.
164. Ortenzio, L.F., L.S. Stuart, and J.L. Friedl, 1953, *J. Ass. Off. Agric. Chem.,* **36,** 480.
165. Owens, R.G., 1969, in *Fungicides,* **2,** Ed., D.C. Torgeson, Academic Press, New York.
166. Parker, M.S., 1978, *J. Appl. Bact.,* **44,** Sxxix.
167. Parker, M.S., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russel, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
168. Pepper, R.E. and V.L. Chandler, 1963, *Appl. Microbiol.,* **11,** 384.
169. Perkins, J.J. and R.S. Lloyd, 1961, in *Recent Developments in the Sterilizatino of Surgical Materials,* Report of a Symposium, p. 76, Pharmaceutical Press, London.
170. Pharmaceutical Codex, 1979, Pharmaceutical Press, London.
171. Phillips, C.R., 1949, *Amer. J. Hyg.,* **50,** 280.
172. Phillips, C.R., 1952, *Bacteriol. Rev.,* **16,** 135.
173. Phillips, C.R., 1958, Becton, Dickinson Lectures on Sterilization: Seton Hall, New Jersey, USA.
174. Phillips, C.R., 1961, in *Recent Developments in the Sterilization of Surgical Materials,* Report of a Symposium, p. 59, Pharmaceutical Press, London.
175. Phillips, C.R., 1977, in *Disinfection, Sterilization and Preservation,* 2nd Edition, Ed., S.S. Block, Lea and Febiger, Pennsylvania, USA.
176. Phillips, C.R. and S. Kaye, 1949, *Amer. J. Hyg.,* **50,** 270.
177. Phillips, C.R. and B. Warshowsky, 1958, *Ann. Rev. Microbiol.,* **12,** 525.
178. Prince, J., C.E.A. Deverill, and G.A.J. Ayliffe, 1975, *J. Clin. Path.,* **28,** 71.
179. Prokop, A. and A.E. Humphrey, 1970, in *Disinfection,* Ed., M.A. Benarde, Marcel Dekker, Inc., New York.
180. Report, 1958, Report by the Committee on Formaldehyde Disinfection*, J. Hyg. Camb.,* **56,** 488.
181. Report, 1965, Report by Public Health Laboratory Service Committee on the Testing and Evaluation of Disinfectants*, Br. Med. J.,* **1,** 408.
182. Reybrouck, G., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
183. Richardson, B.A., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
184. Rittenburgy, M.S. and M.E. Hench, 1965, *Ann. Surg.,* **161,** 127.
185. Roberts, M.H. and O. Rahn, 1946, *J. Bact.,* **52,** 639.
186. Rogers, M.R., K. Mather, and A.M. Kaplan, 1961, *Appl. Microbiol.,* **9,** 497.
187. Royce, A. and C. Bowler, 1961, *J. Pharm. Pharmac.,* **13,** 87T.
188. Rubbo, S.D. and J.F. Gardner, 1965, *A Review of Sterilization and Disinfection,* Lloyd-Luke, London.
189. Rubbo, S.D. and J.F. Gardner, 1968, *J. Appl. Bact.,* **31,** 164.
190. Rubbo, S.D., J.F. Gardner, and R.L. Webb, 1967, *J. Appl. Bact.,* **30,** 78.
191. Rudolph, A.S. and M. Levine, 1941, *Iowa Eng. Exp. Sta. Bull.,* No. 150, Iowa, USA.
192. Russell, A.D., 1964, *Lab. Pract.,* **13,** 114.
193. Russell, A.D., 1971, in *Inhibition and Destruction of the Microbial Cell,* Ed., W.B. Hugo, Academic Press, London and New York.
194. Russell, A.D., 1976, in *The Inactivation of Vegetative Bacteria,* Ed., F.A. Skinner and W.B. Hugo, Society for Applied Bacteriology Symposium Series No. 5, Academic Press, London and New York.
195. Russell, A.D., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
196. Russell, A.D., I. Ahonkhai, and D.T. Rogers, 1979, *J. Appl. Bact.,* **46,** 207.
197. Russell, A.D., J. Jenkins and I.H. Harrison, 1967, *Adv. Appl. Microbiol.,* **9,** 1.
198. Russell, A.D. and M. Loosemore, 1964, *Appl. Microbiol.,* **12,** 403.
199. Sabel, F.L., A. Hellman, and J.J. McDade, 1969, *Appl. Microbiol.,* **17,** 645.
200. Salk, J.E., U. Krech, J.S. Younger, B.L. Bennett, and L.J. Lewis, 1954, *Amer. J. Publ. Health,* **44,** 563.
201. Salton, M.R.J., 1957, *2nd Int. Congr. Surface Activity,* Butterworths, London.
202. Schley, D.G., R.K. Hoffman, and C.R. Phillips, 1960, *Appl. Microbiol.,* **8,** 15.
203. Snyder, R.W. and E.L. Cheatle, 1965, *Amer. J. Hosp. Pharm.,* **22,** 321.
204. Sofos, J.N. and F. Busta, 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
205. Spaulding, E.H., 1958, *Becton, Dickinson Lectures on Sterilization,* p. 51.
206. Spaulding, E.H., K.R. Cundy, and F.J. Turner, 1977, in *Disinfection, Sterilization and Preservation,* 2nd Edition, Ed., S.S. Block, Lea and Febiger, Philadelphia.
207. Spiner, D.R. and R.K. Hoffman, 1960, *Appl. Microbiol.,* **8,** 152.
208. Springle, W.R. and M.A. Briggs, 1981, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
209. Stonehill, A.A., S. Krop, and P.M. Borick, 1963, *Am. J. Hosp. Pharm.,* **20,** 458.
210. Sutaria, R.H. and F.H. Williams, 1961, *Pharm. J.,* **186,** 311.
211. Sykes, G., 1939, *J. Hyg. Camb.,* **39,** 463.

212. Sykes, G., 1958, *J. Pharm. Pharmac.,* **10,** 40T.
213. Sykes, G., 1965, *Disinfection and Sterilization,* 2nd Edition, Spon, London.
214. Sykes, G., 1970, *J. Appl. Bact.,* **33,** 147.
215. Sykes, G. and M.C. Hooper, 1954, *J. Pharm. Pharmac.,* **11,** 235T.
216. Ten Cate, L., 1965, *J. Appl. Bact.,* **28,** 221.
217. Thomas, C.G.A., 1960, *Guy's Hospital Rep.,* **109,** 57.
218. Thomas, C.G.A., B. West, and H. Besser, 1959, *Guy's Hospital Rep.,* **108,** 446.
219. Tilley, F.W., 1942, *J. Bact.,* **43,** 521.
220. Toplin, I., 1962, *Biotechn. Bioeng.,* **4,** 331.
221. Toth, L.Z.J., 1959, *Arch. Mikrobiol.,* **32,** 409.
222. Trafas, P.C., R.E. Carlson, G.A. LoGrippo, and C.R. Lam, 1954, *Arch. Surg.,* **69,** 415.
223. Truman, J.R., 1971, in *Inhibition and Destruction of the Microbial Cell,* Ed., W.B. Hugo, Academic Press, London and New York.
224. Underwood, E., 1980, in *Pharmaceutical Microbiology,* 2nd Edition, Ed., W.B. Hugo, and A.D. Russell, Blackwell Scientific Publications, Oxford.
225. Waites, W.M., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
226. Walpole, A.L., D.C. Roberts, F.L. Rose, J.A. Hendry, and R.F. Homer, 1954, *Br. J. Pharmac.,* **9,** 306.
227. Weber, G.R., 1950, *Publ. Health Rep.,* **65,** 503.
228. Weber, G.R. and M. Levine, 1944, *Am. J. Publ. Health,* **34,** 719.
229. Wedderbern, D.L., 1964, *Adv. Pharm. Sci.,* **1,** 195.
230. Weinberg, E.D., 1968, *Ann. NY Acad. Sci.,* **148,** 587.
231. Weir, B., 1982, in *Principles and Practice of Disinfection, Preservation and Sterilisation,* Ed., A.D. Russell, W.B. Hugo, and G.A.J. Ayliffe, Blackwell Scientific Publications, Oxford.
232. Whitehouse, R.L. and L.F.L. Clegg, 1963, *J. Dairy Res.,* **30,** 315.
233. Willard, M. and A. Alexander, 1964, *Appl. Microbiol.,* **12,** 229.
234. Wilson, A.R. and P. Bruno, 1950, *J. Exp. Med.,* **91,** 449.
235. Winge-Heden, K., 1963, *Acta Path. Microbiol. Scand.,* **58,** 225.

A.D. RUSSELL
P.J. DITCHETT
Welsh School of Pharmacy
University of Wales
Institute of Science & Technology
Cardiff, UK

DRAINAGE-SURFACE: see URBAN RUNOFF

E

EARTH–SPACE ORGANIZATION: see THE TERRESTRIAL SYSTEM
ECOLOGY: see ECOSYSTEM THEORY
ECOLOGY OF FISH: see POLLUTION EFFECTS ON FISH; THERMAL EFFECTS ON FISH ECOLOGY
ECOLOGY OF HUMANS: see EFFECTS OF CHEMICALS; HAZARDOUS WASTES

ECOLOGY OF PLANTS

Ecology is the study of the relationship between organisms and the environment and the interrleationships between organisms. The response of an organism to the environment is determined by the effects of the environment on the processes in the organism. Stated in another way, all environmental factors affect the growth and distribution of organisms only by affecting rates of processes in the organisms. The study of the effects of environmental factors on processes in organisms is *physiological ecology,* or in other words, it is the study of the physiological basis of ecological behavior. This field obviously represents a marriage between ecology and physiology, and is sometimes called *environmental physiology.* The field of physiological ecology is extremely broad: thus, this paper is restricted to certain aspects of physiological *plant* ecology.

ENVIRONMENTAL FACTORS

The environment is very complex, consisting of all influences external to an organisms. Many systems have been devised to classify the environment but space does not allow a discussion here. The following list of *specific* environmental factors is a satisfactory one although the list could obviously be subdivided in many ways:

1) water
2) temperature
3) radiant energy
4) essential elements
5) aeration
6) toxins
7) wind
8) topography
9) nature of geologic strata
10) altitude
11) organic matter content of soil
12) texture and structure of soil
13) base exchange capacity of soil
14) color of soil
15) plant growth regulators
16) soil organisms
17) parasites
18) plants and animals other than above.

Even cursory examination of this list makes it clear that the effects of all these factors on processes in plants could not be discussed in a brief survey. For example, thorough coverage of the effects of the various essential elements on plant processes would require a sizeable manuscript. I have decided, therefore, to discuss a few factors in sufficient depth to give the reader a feeling for the area.

INTERRELATIONSHIPS BETWEEN ENVIRONMENTAL FACTORS

In addition to having a direct effect on many plant functions, the water factor affects other factors through the amount and type of precipitation involved. It affects: (1) amounts of radiant energy, (2) temperature of air and soil, (3) amount of nitrogen brought down by rain, (4) amounts of minerals leached from the soil, (5) availability of various elements to plants, (6) topography of the region, (7) numbers and types of organisms in the soil, (8) toxins in the soil, (9) aeration of soil, and (10) numbers and types of parasites in a region.

All other environmental factors have many indirect effects also.

EFFECTS OF WATER FACTOR

Warming (1909) stated that "no other influence impresses its mark to such a degree upon the internal and external structures of the plant as does the amount of water present in the air and soil." This is due primarily, of course, to the following roles of water: (1) it is important as a solvent system in the soil and in cells of all living organisms on earth, (2) it serves as the dispersion medium for the colloidal systems present in cells, (3) there is a close interaction between the nature of the hydration shell surrounding protein molecules and the physicochemical properties of the proteins (Klotz, 1958), (4) it is important as a raw material in photosynthesis, (5) it is important as a raw material in all hydrolytic processes, and (6) it has some influence on leaf temperatures through transpirational cooling.

Klotz (1958) suggested that protein molecules in cells have hydration shells of lattice-ordered water which act like ice-shells around the molecules and that the maintenance of an active configuration is due to the shells. The effect of heat is thought to reduce the extent of the ice-shell, whereas lower temperatures increase the thickness of the shell. Urea is thought to be an effective denaturizer of proteins because its strong hydrogen bonding characteristics may break down the frozen structure of the water envelope. The observed effects of water stress on the viscosity of protoplasm suggests that reduced hydration has a definite effect on the ice-shell.

Jacobson (1953) offered considerable evidence for the presence of an ordered-water lattice around DNA molecules also. He suggested that such an ice-shell would make separation of the two chains of DNA during replication possible with low energies.

The models of Klotz and Jacobson have apparently been well received (Stocker, 1960; Slatyer, 1967). If these models are indeed correct, it is easy to see how water deficiency can have pronounced effects on all sorts of processes, because of the role of DNA in protein synthesis and the role of proteins in enzymes. It is probable that the integrity of specific water-protein structures is necessary for the continuance of organic processes at optimum rates. It is likely that the variations in reductions to different species to water stress are due chiefly to different degrees of sensitivity of essential metabolic systems.

Buckman and Brady (1960) state that a slight lowering of the available water content in the soil below field capacity stimulates growth of plants generally. Blair *et al.* (1950) found that the rate of stem elongation in sunflower (a dwarf type from an inbred line, apparently of *Helianthus annuus*) is markedly reduced before half the available water is used, and that zero growth results during the use of the last one-fourth of the available water. A reduction of 30–40% between field capacity and the wilting point probably slows growth in most plants.

Probably the most direct effect of water on plant growth is its effect on the turgor pressure of the individual cells. Turgor pressure affects cell enlargement and stomatal closure. Closing of the stomates due to reduced turgor pressure in the guard cells reduces both transpiration and photosynthesis and can ultimately reduce growth. Probably, many of the detrimental effects of moderate water deficits are due to stomatal closure. One important effect is that the temperature of leaves can rise to detrimental levels under some conditions when the transpiration rate is lowered markedly (Slatyer, 1967).

It appears that cell division is influenced less by water deficits than cell enlargement (Slatyer, 1967). Gates and Bonner (1959) found that the amount of DNA in water stressed tomato leaves remained constant during the period of their experiments whereas the amount increased steadily in control leaves. This indicated that chromosomal multiplication was continuing to occur in the control leaves but had ceased in the test leaves. Apparently, cell number is linearly related to DNA content (Slatyer, 1967) so it appears that there was a reduction in rate of cell division in the water stressed leaves. Gardner and Nieman (1964) found that water stress in cotyledonary leaves of *Raphanus sativus* (radish) reduced the rate of increase of DNA.

Gates and Bonner (1959) found that water stress prevented a net increase in RNA in *Lycopersicum esculentum* (tomato) leaves also. The water stressed leaves were able to incorporate P^{32}-labeled phosphate into RNA apparently at the same rate as in control leaves. They concluded, therefore, that the rate of destruction of RNA was greater in water stressed leaves than in controls. Kessler (1961) reported that drought stress causes a pronounced decrease in the RNA content of several plants. Despite this, the rate of incorporation of uracil-C^{14} was the same as in control leaves but had ceased in the test leaves—drought stressed plants. This was confirmed because the activity of RNase was found to increase markedly with water stress.

A reduction in the water content of leaves generally results in a decrease in the rate of photosynthesis (Meyer and Anderson, 1939). This may seem strange because less than one per cent of the water absorbed by a plant is used as a raw material in photosynthesis. According to Slatyer (1967) however, there are two main modes of action of water stresses on photosynthesis: (1) stomatal closure and reduced rates of CO_2 exchange can influence the supply of CO_2 and (2) there is probably a direct effect of the water deficit on the biochemical processes involved in photosynthesis. Gates (1964) maintained that the rate of photosynthesis is more closely related to the water content of leaves than to the diffusive capacity of the stomates. Certainly, if Klotz's model of protein hydration and behavior is correct, one would expect Gate's contention to be logical.

Gates (1964) reported that during water stress, the respiration rate is accelerated but later declines. There is an increase in starch hydrolysis and an increase in sucrose which later declines. The respiration rate appears to be correlated with the sucrose content. Protein synthesis is impaired and active proteolysis may occur with prolonged wilting, especially in older leaves. Finally, if the degree of protoplasmic desiccation reaches critical levels, individual cells, tissues and even the entire organism may die.

Flooding lowers the oxygen supply to the roots or roots and tops if the tops are covered with water also. Under these conditions, Crawford and Tyler (1969) found that plants intolerant of flooding have an increased rate of glycolysis including alcohol dehydrogenase induction and the greatest production of acetaldehyde. On the other hand, malic acid accumulates in flood tolerant species and falls in intolerant species. Organic acids can accumulate in cells in much greater quantities than ethanol without damaging the plant. In fact, malic acid helps preserve the electrical neutrality of the cell.

EFFECTS OF TEMPERATURE

The cardinal temperatures (i.e., minimum, optimum, maximum) often differ for the same function in different plants and different functions in the same plant. They often differ also for the same function at different stages of development, and often in different organs of the same plant at a given, stage. Moreover, they may differ with the general physiological condition of the plant, the duration of temperature regime, and changes in other environmental factors.

Daubenmire (1959) states that the minimum temperature for growth of melons (*Citrullus*) and sorghums (*Sorghum*) is in the range of 15 to 18°C; and for peas (*Pisum*), rye (*Secale*), and wheat (*Triticum*), −2 to 5°C. He further states that roots usually have a lower minimum for growth than shoots of the same plants. The optimum temperature for photosynthesis is below that for respiration in all known cases. This is probably the reason peaches, apples, white potatoes, and many other plants do not accumulate normal food reserves in low latitudes.

Seeds of many native plants of temperature or colder regions have a low temperature requirement for germination. The value of such a mechanism in seedling survival is obvious. Work in my laboratory at the University of Oklahoma has shown that seeds of *Ambrosia artemisiifolia* (common ragweed), *A. psilostachya* (western ragweed) and *Sporobolus pyramidatus* (drop seed) require exposure to temperatures in the range of 4–7°C for about 6–10 weeks in moist conditions to break dormancy. Lane (1965) reported that seeds of *Helianthus annuus* (annual sunflower) require 90 days at 4°C in moist sand to completely break dormancy. Donoho and Walker (1957) reported that Elberta peach (*Prunus persica*) seeds usually required stratification for 60–100 days at 7°C or below to break dormancy.

Overwintering buds of many, if not most, perennial plants of temperature or colder regions also have low temperature requirements for the breaking of dormancy. Weinberger (1956) found that Elberta peach buds require 950 hours at 7°C or below to break dormancy. Daubenmire (1959) stated that peaches in general require 400 or more hours of temperatures at or below 7°C, blueberries (*Vaccinium* spp.) require 800 hours at this level, and apples (*Pyrus malus*) require even more.

The chemical changes which are necessary for the breaking of dormancy in seeds and buds, and which are stimulated by exposure to low temperatures, are not definitely known. In fact, this is a fertile field for research. There is growing evidence, however, that the breaking of dormancy may be related to a change in balance between growth inhibitors and stimulators. Sondheimer, Tzou, and Galson (1968) reported that abscisic acid inhibits germination of excised nondormant embryos and this can be reversed with a combination of gibberellic acid and kinetin. Natural amounts of abscisic acid in seeds and pericarps of several species of *Fraximus* (ash) were determined, and it was found that the amount in *F. americana* decreased 37% in the pericarp and 68% in the seed with chilling treatment. They felt that it is unlikely that the amount in the pericarp influences dormancy but that the inhibitor in the seed does appear to play a regulatory role. Junttila (1970) found that dormancy of seeds of *Betula nana* (birch) was effectively broken by low temperatures or treatment with gibberellic acid. Haber, Foard, and Perdue (1969), reported 84% germination in dormant lettuce (*Lactuca sativa*) seeds treated with a 5×10^{-4} M gibberellic acid solution, 1.5% germination in seeds treated with a 5×10^{-4} M gibberellic acid solution plus 100 mg/l of abscisic acid, and 0% in water controls.

Donoho and Walker (1957) reported a 98% breaking of dormancy of Elberta peach buds after only 164 hours at 7°C when sprayed with a 4000 ppm gibberellic acid solution, despite the usual requirement of 950 hours at 7°C or below. Aung, Hertogh, and Staby (1969) found free gibberellic acid in *Tulipa gesneriana* (tulip) increased by 67% over the initial level in bulbs grown for 4 weeks at 18°C. Bound gibberellic acid showed a slight initial increase followed by a rapid decrease. In bulbs treated at 13°C, a marked decline, occurred in free gibberellic acid and a 2-fold increase occurred in the bound form. The rates of floral, shoot, and root development were accelerated at 18°C as compared with 13°C.

Cold hardening of plants (i.e., improving the ability of plants to withstand low temperatures without severe damage or death) is probably just as important to plant survival in regions with periods of relatively low temperatures as dormancy. Actually, some of the same internal changes may be responsible for both. Irving (1969) found a buildup during cold hardening in *Acer negundo* (box elder) of an inhibitor similar or identical to abscisic acid which has, of course, been associated with dormancy. He found that it or abscisic acid increased hardiness when applied to branches after a hardening period of 3 weeks at 5°C. There are several cellular chemical changes associated with cold hardening (Klages, 1947); (1) increase in bound water, (2) decrease in total water, (3) increase in osmotic pressure, (4) change of starch to sugars, (5) increase in pentosans, and (6) change of certain proteins to amino acids. Nothing new has been added to this list since Klages' publication in 1947.

Temperature obviously affects rates of all organic processes in plants, but a comprehensive discussion of such effects is outside the scope of this discussion. The relationship of temperature conditions at the geographic point of origin of plants to the optimum temperatures for photosynthesis and respiration in those plants is very important in explaining their distribution. Mooney and Billings (1961) studied relationships of temperature to photosynthesis and respiration in *Oxyria digyna* (alpine sorrel), a species with a circumboreal distribution of wide latiudinal range. The sources of the populations studied varied from an elevation of 12,300 feet at 39° 40′N latitude to a zero elevation at 68° 56′ N. They found that plants of northern populations had higher respiration rates at all temperatures than did plants of southern alpine populations. Moreover, they found that plants of northern populations had higher photosynthetic rates at lower temperatures and attained maximum rates at lower temperatures than did plants of the southern alpine populations. Obviously, there are definite temperature ecotypes in this species.

Mooney, Wright, and Strain (1964) investigated the gas exchange capacity of plants in relation to elevational zonation in the White Mts. of California. Field measurements over an elevational span of 4000 to 14,000 feet indicated that the ratio of photosynthesis to dark respiration decreased with altitude primarily due to increased respiration rates with increased elevation. When similar measurements were made on plants grown in uniform greenhouse conditions, from seeds collected at different elevations, no differences in respiration rates correlated with elevation were obtained indicating considerable environmentally conditioned plasticity in gas exchange capacity. However, the high elevation plants generally displayed peak photosynthetic activity at temperatures lower than the low elevation plants even though all were grown in the same conditions.

Mooney and West (1964) studied the photosynthetic and respiratory acclimation of five species in the White Mts. of California (Table 1). Seeds were collected at the indicated elevations, planted in pots in the greenhouse, and after establishment of the seedlings, they were taken to stations at various elevations in the mountains for acclimation (Table 2).

They were kept in the same pots and soil and subirrigated at each place. Water was thus not limiting and the soil type was constant. Gas exchange measurements were made after a minimum of 3 weeks acclimation. The dark respiration rates of plants of all species from all stations were similar. Plants of all species acclimated to the colder subalpine environment were more efficient in photosynthesis at colder temperatures than were plants grown at the warmer environment of the lowest elevation. Conversely, plants grown in the desert were more efficient in photosynthesis at high temperatures than were plants grown in the subalpine region. Plants of the two species with the widest natural distributional range in elevation in the White Mts. showed the greatest plasticity in their photosynthetic response as indicated by shifts in the optimum photosynthetic temperatures dependent on treatment.

TABLE 1

Natural distribution of experimental plants

Species	Elev. range (m)	Elev. origin of seeds used (m)
Artemisia tridentata	1829–3200	3078
Chamaebatiaria millefolium	1980–3200	3200
Haplopappus apargioides	3048–3962	3658
Artemisia arbuscula	3200–3719	3475
Encelia virginensis ssp. *actoni*	1372–1829	1402

TABLE 2

Acclimation stations

Natural vegetation	Elev. (m)	Mean air temp. °C[a] Max.	Mean air temp. °C[a] Min.
Desert scrub	1402	38.9	11.7
Pinyon woodland	2408	33.4	10.3
Subalpine forest	3094	19.5	–2.4

[a] Based on weekly maximum and minimum temperatures during period of June 17 to Aug. 12, 1963.

Miller (1960) compared the effects of temperature on the photosynthetic rate in *Agrostis palustris* (creeping bent) which is adapted to cooler regions of the US, and *Cynodon dactylon* (bermuda grass) which is adapted to the warmer regions. The relative rate of apparent photosynthesis in creeping bent increased from 64.6% at 15°C to a maximum of 100% at 25°C, then decreased to 62.2% at 40°C. On the other hand, the relative rate of apparent photosynthesis in bermuda grass increased form 54.9% at 15°C to a maximum of 100% at 35°C, and only dropped to 97.7% at 40°C.

In summary, it is evident that plants vary genetically in their photosynthetic and respiratory responses to temperature changes, and they vary also in such responses depending on the temperature conditions during acclimation.

EFFECTS OF LIGHT

According to Johnson (1954) solar radiation at the outer surface of the earth's atmosphere has a virtually constant intensity of 1400 watts m^{-2} or 2cal cm^{-2} min^{-1}. About 98% of the energy is in the wavelength interval of 0.2–4.5 μ, including about 40–45% in the visible range of 0.4–0.7 μ. About one-third of the energy is reflected back to space by the earth's atmosphere and smaller amounts are absorbed and scattered by various components of the atmosphere. As a result, only about 50% of the total radiation reaches the earth on the average. This figure may be greater than 70% in dry regions with little cloud cover and lower than 40% in tropical rainy areas (Slatyer, 1967). Obviously, on very cloudy days the energy reaching the surface of the earth may be very low. Of the solar energy reaching the earth, about 40–45% is still present in the visible range.

Radiation in the visible range is most important in affecting organic processes in plants, so the rest of this discussion will be concerned primarily with visible light. The chief processes in plants which are affected by light are: (1) chlorophyll synthesis, (2) photosynthesis, (3) phototropism, (4) photoperiodism, and (5) transpiration.

Light Intensity Effects. Many plants are able under most conditions to develop well in less than full sunlight. Red kidney beans (*Phaseolus vulgaris*) grow and fruit well in growth chambers in a light intensity of 700–1000 ft-c. Intensity of light has a definite effect on the distribution of plants. Many ferns grow well under trees because such ferns attain their maximum rate of photosynthesis at light intensities considerably below full sunlight. Species which characteristically grow in shaded locations are able to remain alive at very low light intensities. Redwood (*Sequoia sempervirens*) seedlings normally grow in the deep shade of the forest floor and they are able to grow at intensities of as little as 100 ft-c (Bonner and Galston, 1952). *Pinus taeda* (loblolly pine) and *Pinus echinata* (shortleaf pine) seedlings are found in great profusion along road cuts and other open areas in the southeastern United States, but they fail to survive under dense forest canopies. Many hardwood seedlings such as those of *Quercus alba* (white oak) and *Quercus rubra* (red oak) survive nicely under such canopies and thus replace the pines in succession. Kramer and Decker (1944) found that photosynthesis in loblolly pine increased with light intensity up to the highest intensity used which was almost that of full sunlight. On the other hand, photosynthesis in white oak and red oak reached a maximum at one-third or less of full sunlight and showed slight decreases at higher light intensities. These results indicate that lack of sufficient light for photosynthesis may be a significant factor in the failure of pine seedlings to become established under forest stands. Seedlings of many other climax forest dominants such as *Fagus grandifolia* (beech) and *Acer saccharum* (sugar maple) have very low light requirements for photosynthesis also (Daubenmire, 1959). After a long series of studies, Blackman (1950) concluded that the distribution of *Scilla nonscripta* (bluebell) in English forests is determined primarily by light intensity.

Björkman and Holmgren (1963) investigated the photosynthetic response to light intensity of sun and shade plants of *Solidago virgaurea* (golden-rod). They found that the shade plants had a higher chlorophyll content per unit leaf area in weak light (3×10^4 ergs cm^{-2} sec^{-1}) than in strong light (15×10^4 ergs cm^{-2} sec^{-1}) and more than did sun plants in weak light. On the other hand, the sun plants had lower concentrations of chlorophyll in weak light than strong light. The gross rate of photosynthesis (apparent rate plus respiration rate) in sun plants at light saturation was 25.5 ± 2.0 mg CO_2 dm^{-2} h^{-1} when the plants were grown in strong light (same as above) prior to testing whereas it was only 17.9 ± 0.6 mg CO_2 dm^{-2} h^{-1} in shade plants grown in strong light. The gross rate of photosynthesis at light saturation was higher in the shade plants than in the sun plants when the plants were grown in weak light prior to testing, 2.93 ± 0.08 in shade plants versus 2.31 ± 0.04 in sun plants. The difference of the means in the plants grown in low intensity was highly significant statistically. The authors concluded that the photosynthetic behavior was consistent with light intensities prevailing in the natural environment indicating that the behavior was a result of a genetic adaptation to the habitat. Hybridization experiments supported this conclusion.

Björkman (1966) reported on additional comparative studies of sun and shade plants of several species. Plants of shaded habitats consistently reached a higher percentage of the maximum photosynthetic rate at light saturation, when tested in a low light intensity of 10^4 ergs cm^{-2} sec^{-1} (Table 3). *Plantago lanceolata* (plantain) was found only in open meadows and *Lamium galeobdolon* (henbit) only in dense beech forests. Björkman found that the activity of Photosystem I (far red, 700 mμ) in *Solidago virgaurea* was about the same when grown in weak or strong light, regardless of the original habitat. On the other hand, the action of Photosystem II (>625 mμ used) was impaired in shade plants grown in high light intensity and in sun plants grown in low light intensity. The Emerson enhancement effect (increased photosynthesis from combination red–far red light over sum of photosynthesis in each light quality separately) was drastically reduced in shade plants grown in high light intensities.

Milner and Hiesey (1964) studied the photosynthetic responses to light intensity of several races of *Mimulus cardinalis* (monkey-flower) collected from an elevation of 45 m to 2220 m in California. They found that the saturating light intensity for the races increased with elevation of their original habitats. At 0°C, the saturating light intensity varied from 290 to 570 ft-c for different races, and at 40°C from 3500 to 5900 ft-c. Light intensity, of course, generally increases with elevation. Mooney and Billings (1961) in their study of ecotypes of *Oxyria digyna* found that the high elevation, low latitude plants attained photosynthetic light saturation at

TABLE 3

Photosynthesis at a low light intensity of 10^4 ergs cm^{-2} sec^{-1} in percent of the photosynthetic rate at light saturation[a]

Species	Sun plants	Shade plants
Plantago lanceolata	8	—
Solidago virgaurea	9	17
Rumex acetosa	11	16
Geum rivale	14	16
Lamium galeobdolon	—	25

[a] From Björkman (1966).

a higher light intensity than did low elevation, high latitude plants even when preconditioned under the same conditions.

Light Quality All visible light is effective in the synthesis of chlorophyll except wavelengths above 680 mμ (Sayre, 1928). Chlorophyll has a maximum absorption peak in the blue-violet (430 mμ) and a secondary maximum in the red (660 mμ), and chlorophyll b has a maximum absorption peak in the blue (455 mμ), and a secondary maximum in the orange-red (640 mμ) (Salisbury and Ross, 1969). Hoover (1937) reported that wheat (*Triticum* sp.) has a maximum rate of photosynthesis at 655 mμ in the red and a secondary maximum at 440 mμ in the blue. Warburg and Negelein (1923) gave a general curve for photosynthesis showing a maximum rate in the red and a secondary maximum in the blue. It is now known that photosynthesis actually consists of two photosystems which work together to complete the light reactions which result in the production of ATP and NADPH. Photosystem I has an absorption peak at 700 mμ and Photosystem II has an absorption peak around 672 mμ (Salisbury and Ross, 1969).

Phototrophic responses (movements in response to light) are chiefly due to wavelengths in the blue-violet, 400–490 mμ, but there may be some response to UV light (Johnston, 1934).

The most efficient light quality to interrupt the dark period to prevent a short-day photoperiodic response is in the wavelength range of 620–640 mμ.

Effects of Daylength The response of organisms to the relative lengths of the light and dark periods is termed photoperiodism. Garner and Allard (1920) discovered this phenomenon while working with Maryland Mammoth Tobacco (*Nicotiana tabacum*). They extended the work to many other species and found that flowering in many plants is regulated by the relative lengths of the light and dark periods. They classified plants into three categories on the basis of their photoperiodic flowering responses: (1) short-day plants—those plants which flower only when the day length is below a *certain critical maximum,* (2) long-day plants—those which flower only when the day length is above a *certain critical minimum,* and (3) day-neutral plants—those in which flowering is not determined by day length. Another category has been added, the intermediate-day plants—those which flower only between two critical day lengths. Larsen (1947) found that the strains of *Andropogon scoparius* (little bluestem) from the area south of 36°N latitude were intermediate-day plants because they failed to flower on a 13-hour photoperiod and also on a 15-hour photoperiod, but flowered well between those daylengths.

It is now known that numerous phenomena in plants are regulated by day length in many species, bulb formation, type of vegetative growth, dormancy, leaf abscission, seed germination, etc. The timing mechanism is still obscure, but it seems to involve a pigment called phytochrome which was discovered and isolated by Borthwick and his associates at the USDA laboratories in Beltsville, Maryland (Borthwick, Hendricks, and Parker, 1952). They found that phytochrome exists in two forms which can readily be converted from one form to another. One form has an absorption peak in the red at about 660 mμ and is designated as P_r. The other form has an absorption peak at a longer wavelength of red (730 mμ) and is designated as P_{fr}, for far-red. When P_{fr} absorbs light at 730 mμ it is converted to P_r, and when P_r absorbs light at 660 mμ it is converted to P_{fr}. In addition to these rapid changes brought about by light, there is a slow conversion of P_{fr} to P_r in darkness. In sunlight and broad spectrum artificial light the conversion of P_r to P_{fr} predominates. The P_{fr} form appears to be inhibitory to the flowering process, so darkness is important to eliminate all or part of this inhibitory form. Salisbury (1958) feels that the plant possibly synthesizes a postulated flowering hormone only during darkness also. Borthwick and Hendricks (1960) reported that the basic difference in long- and short-day plants may be in the ratio of P_r to P_{fr} required to initiate the flowering or other response.

The presence of photoperiodic ecotypes has been well documented in many geographically wide-ranging species of plants. Olmsted (1944) was the first to demonstrate such ecotypes in his work with *Bouteloua curtipendula* (sideoats grama). He found that plants of this species from southern Texas and southern Arizona were generally short-day or intermediate-day plants, with an upper critical photoperiod for flowering between 14 and 16 hours. These strains flowered more vigorously on a 13-hour than on a 9-hour photoperiod. The North Dakota strains consisted largely of long-day plants with a critical photoperiod of about 14 hours. The strains from Oklahoma, Kansas, and New Mexico included mostly long-day individuals although the length of the critical photoperiod decreased with decrease in latitude of origin.

Larsen (1947) found that only strains of *Andropogon scoparius* (little bluestem) originating south of 36°N latitude produced visible inflorescences on a 14-hour photoperiod. The northern strains flowered on a 15-hour photoperiod but not the southern strains. The northern strains were apparently long-day plants because they were able to initiate flowers on a long photoperiod but were inhibited from flowering on 13- and 14-hour photoperiods. As was previously indicated at the beginning of this section on effects of day-length, the strains originating south of 36°N latitude were apparently intermediate-day plants.

Vaartaja (1959) collected seeds from 19 genera of trees, 38 species and 82 origins of widely different latitudes in Europe and North America. Plants derived from these seeds were grown in uniform conditions except for photoperiod. Photoperiods of 12, 14, 16, and 18 hours were used with the total light energy being identical in all cases. He found that the farther north the origin of the strain, the longer was the critical daylength below which height growth of seedlings was retarded. Moreover, the more northerly the source, the shorter was the critical dark period which induced dormancy in several species. In some plants from the far north, photoperiods that inhibited elongation permitted great increases in weight of stems and roots, and in numbers of buds. Cambial growth thus would continue later in the growing season than elongation.

Irgens-Moller (1957) collected approximately 100 plants (2–6 year old) of *Pseudotsuga menziesii* (douglas-fir) from each of seven locations in Oregon at elevations of 60 feet to

4000 feet. The plants were potted and kept in a cold frame at Corvallis, Oregon, during two winters after collection so that all were exposed to the same winter conditions. On February 1 after the second winter, half of the plants from each locality were placed in a 9-hour photoperiod and half on an 18-hour photoperiod with all other conditions the same. There were no effects of photoperiod on time of bud bursting in the low altitude plants, but the long days hastened bud bursting strikingly in high altitude plants as compared with effects of short days. The selection probably results from the killing of seedlings by low night temperatures at high altitudes if they break dormancy in the short days of winter or early spring.

EFFECTS OF PLANT PRODUCED TOXINS

De Candolle first suggested the existence of naturally-occurring plant growth-inhibitors in 1832 (Bonner 1950). He noted from field observations that *Euphorbia* (spurge) was apparently inhibitory of flax (*Linum usitatissimum*), thistles (probably *Cirsium*) to oats (*Avena sativa*) and rye to wheat. Very little research was done on the subject for almost 100 years at which time Cook (1921) described the characteristic wilting of potato (*Solanum tuberosum*) and tomato plants grown near *Juglans nigra* (black walnut). Davis (1928) extracted juglone from mature hulls and roots of black walnut and found it to be very toxic to tomato and alfalfa (*Medicago sativa*) plants when injected into the stems. The juglone was identified as 5-hydroxynaphthaquinone.

Proebsting and Gilmore (1940) related the problem of re-establishing peach (*Prunus persica*) trees in old peach orchards to the presence of toxic substances in the soil. Bonner and Galston (1944) reported that the edge rows in guayule (*Parthenium argentatum*) plantings in California had much larger plants than the center rows and the differences could not be eliminated by heavy watering and mineral application. Experiments indicated that leachates from pots of year-old guayule plants were very inhibitory to guayule seedlings. The inhibitor was subsequently identified as transcinnamic acid.

Muller, Muller, and Haines (1964) observed areas free of vegetation around certain shrubs in California grasslands. They found that at least three shrubs, *Salvia leucophylla* (clearleaf sage), *S. apiana* (bee sage), and *Artemisia californica* (California sagebush) produced volatile materials which inhibited various test plants. Muller and Muller (1964) identified several terpenes which were produced by leaves of three species of *Salvia,* and two of these, camphor and cineole, were found to be very inhibitory to test plants. Muller, Hanawalt, and McPherson (1968) related the disappearance of herbaceous species in the open California chaparral during the fire cycle to the production of inhibitory chemicals by chaparral species.

My students and I have obtained good evidence that the rapid disappearance of the pioneer weed stage (2 to 3 years) in our revegetating old-fields in central Oklahoma is due to the inhibition of seedlings of species of that stage by several plants of the pioneer stage (Abdul-Wahab and Rice, 1967; Parenti and Rice, 1969; Wilson and Rice, 1968). Triple-awn grass (*Aristida oligantha*), the dominant of the second stage, is generally not inhibited by species of the pioneer weed stage and can grow well in the infertile soil, so it is able to invade the area.

There is some evidence that antibiotics (substances produced by microorganisms which inhibit other microorganisms) may be important in the growth and distribution of the higher plants as well as microorganisms. Certain soil bacteria inhibit the root-nodule bacteria which are important in adding nitrogen to the soil (Konishi, 1931). Iuzhina (1958) reported that many soil bacterial, fungi and actinomycetes are antagonistic to *Azotobacter,* a free living bacterium that adds nitrogen to the soil. Some soil microorganisms have been found to be inhibitory to certain microorganisms that cause diseases of higher plants (Cooper and Chilton, 1950).

Another level of inhibition is that of microorganisms against higher plants. There is no question that many, if not most, pathogenic microorganisms produce abnormal symptoms in the host plants through the production of toxins. Very few of these compounds have been identified and this remains an important and fruitful field for research. Microorganisms are sometimes responsible for changing certain non-inhibitory metabolites of higher plants into inhibitory compounds (Börner, 1960).

The remaining level of inhibition is that of higher plants against microorganisms. There is considerable evidence that resistance of many plants to fungal, viral, and bacterial diseases may be associated with the production by resistant varieties of inhibitors of the pathogens (Farkas and Kiraly, 1962; Hughes and Swain, 1960; Schaal and Johnson, 1955).

Ferenczy (1956) found that the seeds or fruits of 52 species and varieties from 19 plant families contained antibacterial compounds and that seeds of some species were attacked intensely by molds whereas others were attacked slightly or not at all. Lane (1965) found that fruits of the native sunflower, *Helianthus annuus,* rotted more frequently during germination tests if certain inhibitors were leached from them before the tests. Those species in which decay of the seeds during germination is less likely to occur would certainly stand a better chance of becoming established.

I have found that many of the low-nitrogen requiring early plant invaders of abandoned cultivated fields produce inhibitors of the nitrogen-fixing bacteria (Rice, 1964; 1965b; 1965c). Triple-awn grass, the dominant of the second stage, is one of the inhibitory species. This would give such plant species a selective advantage in competition over plants with higher nitrogen requirements. This probably slows succession in old-fields, because Rice, Penfound, and Rohrbaugh (1960) found that the order in which species invade revegetating fields in central Oklahoma, beginning with the second stage, is the same as the order based on increasing nitrogen requirements.

Muller (1965) found that the volatile terpenes from *Salvia leucophylla* reduce cell elongation and cell division in the radicles and hypocotyls of germinating *Cucumis sativus* (cucumber) seeds. Muller *et al.* (1969) reported that two volatile terpenes, cineole and dipentene, which emanate from

Salvia leucophylla leaves markedly reduce oxygen uptake by suspensions of mitochondria from *Avena fatua* (wild oats) or cucumber. The inhibition appears to be localized in that part of the Krebs cycle where succinate is converted to fumarate, or fumarate to malate. They also reported that the permeability of cell membranes appears to be decreased on exposure to terpenes.

Most of the inhibitors that my students and I have identified from our old-field species are phenolics (Abdul-Wahab and Rice, 1967; Parenti and Rice, 1969; Rice, 1965a, 1965b; Rice and Parenti, 1967; Wilson and Rice, 1968). Some of the inhibitors identified are: chlorogenic acid, p-coumaric acid, ferulic acid, gallic acid, p-hydroxybenzaldehyde, isochlorogenic acid complex, scopoletin, scopolin, sulfosaliyclic acid, and several tannins.

Sondheimer (1962) reported that chlorogenic acid is a strong inhibitor of several enzyme systems and it is possible that this may be its chief role in growth inhibition. Schwimmer (1958) found that a 10^{-3} M concentration of chlorogenic acid caused a 50% inhibition of potato phosphorylase activity. Mazelis (1962) reported that a peroxidase-catalyzed oxidative decarboxylation of methionine is strongly inhibited by a 5×10^{-6} M concentration of chlorogenic acid. Sondheimer and Griffin (1960) reported that indoleacetic acid oxidase is inhibited by chlorogenic acid and its isomers, and by isochlorogenic acid and dihydrochlorogenic acid.

Zenk and Müller (1963) reported that p-coumaric and ferulic acids increase decarboxylation of indoleacetic acid (IAA), resulting in reduced plant growth. Henderson and Nitsch (1962) found a pronounced inhibiting effect of p-coumaric acid on IAA induced growth also.

Bendall and Gregory (1963) found that phenol oxidases which are important in the formation of lignin-like polymers from coniferyl alcohol and oxidation products of tannins such as quinones react with each other to produce a complex that may make phenol oxidase catalytically inactive, or may produce an active protein of modified properties. Hulme and Jones (1963) and Jones and Hulme (1961) found that tannins caused considerable reduction of Krebs cycle succin oxidase and malic enzyme activity. Gallic and tannic acids were shown also to inhibit IAA induced growth (Zimsmeister and Hollmuller, 1964). Hall (1966) and Williams (1963) reported that relatively low concentrations of tannic acid were very inhibitory to pectolytic enzymes. Gallic and tannic acids are very inhibitory to the nodulation of heavily inoculated legumes and to hemoglobin formation in the nodules (Blum and Rice, 1969). The hydroxyl groups of tannins are known to have a strong affinity for the peptide linkages of proteins, so it is likely that reactions readily occur between the two compounds when they come in contact at a time when the ice-shell around the protein is appropriately modified. This would undoubtedly change the nature of the protein.

Einhellig *et al.* (1970) found that sunflower (*Helianthus annuus*) and tobacco plants treated with 10^{-4} and 5×10^{-4} scopoletin concentrations had significant increases in scopoletin and scopolin in their tissues. Net photosynthesis in tobacco plants treated with a 10^{-3} M concentration of scopoletin was depressed to as low as 34% of that of the controls. Reduced growth in leaf area over a 12 day experiment correlated well with the significant reduction in the rate of net photosynthesis.

The surface has just been scratched in identifying plant produced inhibitors of other plants. Moreover, the field is wide open for research on the mechanisms of action of the presently known inhibitors.

GENERAL CONCLUSIONS

I have discussed briefly the presently known physiological basis for some of the ecological responses of plants to the environmental factors of water, temperature, radiant energy, and toxins. It would obviously require most of the present volume to discuss the presently known effects of all environmental factors. Even so, our knowledge of the effects of most factors is superficial. Many of the known physiological effects may be indirect ones also.

In closing, I emphasize that the physiological ecologist is searching for the physiological basis for ecological behavior. It is essential that research in this area be vigorously pursued.

REFERENCES

1. Abdul-Wahab, A.S. and E.L. Rice, *Bull. Torrey Bot. Club,* **94**, 486, 1967.
2. Aung, L.S., A.A. de Hertogh and G. Staby, *Plant Physiol.,* **44**, 403, 1969.
3. Bendall, D.S. and R.P.F. Gregory, Purification of phenol oxidase, in *Enzyme Chemistry of Phenolic Compounds,* MacMillan Co., New York, 1963.
4. Björkman, O. and P. Holmgren, *Physiol. Plant,* **16**, 889, 1963.
5. Björkman, O., *Brittonia,* **18**, 214, 1966.
6. Blackman, G.E., *Ann. Bot.,* **14**, 487, 1950.
7. Blair, G.Y., L.A. Richards and R.B. Campbell, *Soil Sci.,* **70**, 431, 1950.
8. Blum, U. and E.L. Rice, *Bull. Torrey Bot. Club,* **96**, 531, 1969.
9. Bonner, J. and A.W. Galston, *Bot. Gaz.,* **106**, 185, 1944.
10. Bonner, J. and A.W. Galston, *Principles of Plant Physiology,* W.H. Freemand and Co., San Francisco, 1952.
11. Bonner, J., *Bot. Rev.,* **16**, 51, 1950.
12. Börner, H., *Bot. Rev.,* **26**, 393, 1960.
13. Borthwick, H.A., S.B. Hendricks and M.W. Parker, *Proc. Nat. Acad. Sci.,* **38**, 929, 1952.
14. Borthwick, H.A. and S.B. Hendricks, *Science,* **132**, 1223, 1960.
15. Buckman, H.O. and N.C. Brady, *The Nature and Properties of Soils,* The MacMillan Co., New York, 1960.
16. Cook, M.T. *Phytopathology,* **11**, 346, 1921.
17. Cooper, W.E. and S.J.P. Chilton, *Phytopathology,* **40**, 544, 1950.
18. Crawford, R.M.M. and P.D. Tyler, *J. Ecol.,* **57**, 235, 1969.
19. Daubenmire, R.F., *Plants and Environment,* John Wiley and Sons, Inc., New York, 1959.
20. Davis, R.F., *Amer. J. Bot.,* **15**, 620, 1928.
21. Donoho, C.W., Jr. and D.R. Walker, *Science,* **126**, 1178, 1957.
22. Einhellig, F.A., E.L. Rice, P.G. Risser and S.H. Wender, *Bull. Torrey Bot. Club,* **97**, 22, 1970.
23. Farkas, G.L. and Z. Kiraly, *Phytopathologische Z.,* **44**, 105, 1962.
24. Ferenczy, L., *Acta Biol. Acad. Sci. Hung.,* **6**, 317, 1956.
25. Gardner, W.R. and R.H. Nieman, *Science,* **143**, 1460, 1964.
26. Garner, W.W. and H.A. Allard, *J. Agr. Res.,* **18**, 553, 1920.
27. Gates, C.T., *J. Aust. Inst. Agr. Sci.,* **30**, 3, 1964.
28. Gates, C.T. and J. Bonner, *Plant Physiol.,* **34**, 49, 1959.
29. Haber, A.H., D.E. Foard and S.W. Perdue, *Plant Physiol.,* **44**, 463, 1969.

30. Hall, C.B., *Nature,* **212**, 717, 1966.
31. Henderson, J.H.M. and J.P. Nitsch, *Nature,* **195**, 780, 1962.
32. Hoover, W.H., *Smithsonian Misc. Coll.,* **95**, 1, 1937.
33. Hughes, J.C. and T. Swain, *Phytopathology,* **50**, 398, 1960.
34. Hulme, J.D. and A.C. Jones, Tannin inhibition of plant mitochondria, in *Enzyme Chemistry of Phenolic Compounds,* MacMillan Co., New York, pp. 97–120, 1963.
35. Irgens-Moller, H., *For. Sci.,* **3**, 79, 1957.
36. Irving, R.M., *Plant Physiol.,* **44**, 801, 1969.
37. Iuzhina, Z.I., *Microbiol.,* **27**, 452, 1958.
38. Jacobson, B., *Nature,* **172**, 666, 1953.
39. Johnson, F.S., *J. Met.,* **11**, 431, 1954.
40. Johnston, E.S., *Smithsonian Inst. Misc. Coll.,* **92**, 1, 1934.
41. Jones, J.D. and A.C. Hulme, *Nature,* **191**, 370, 1961.
42. Junttila, O., *Physiol. Plant,* **23**, 425, 1970.
43. Kessler, B., Nucleic acids as factors in drought resistance in higher plants, in *Recent Advances in Botany,* University of Toronto Press, Toronto, Canada, 1961.
44. Klages, K.H.W., *Ecological Crop Geography,* MacMillan Co., New York, 1947.
45. Klotz, I.M., *Science,* **128**, 815, 1958.
46. Konishi, K., *Mem. Col. Agr., Kyoto Imp. Univ.,* **16**, 1, 1931.
47. Kramer, P.J. and J.P. Decker, *Plant Physiol.,* **19**, 350, 1944.
48. Lane, F.E., Dormancy and germination in fruits of Sunflower, Ph.D. Dissertation, University of Oklahoma, Norman, 1965.
49. Larsen, E.C., *Bot. Gaz.,* **109**, 132, 1947.
50. Mazelis, M., *J. Biol. Chem.,* **237**, 104, 1962.
51. Meyer, B.S. and D.B. Anderson, *Plant Physiology,* D. Van Nostrand Co., Inc., New York, 1939.
52. Miller, V.J., *Proc. Amer. Soc. Hort. Sci.,* **75**, 700, 1960.
53. Milner, H.W. and W.M. Hiesey, *Plant Physiol.,* **39**, 208, 1964.
54. Mooney, H.A. and W.D. Billings, *Ecol. Monogr.,* **31**, 1, 1961.
55. Mooney, H.A. and M. West, *Amer. J. Bot.,* **51**, 825, 1964.
56. Mooney, H.A., R.D. Wright and B.R. Strain, *Amer. Midland Naturalist,* **72**, 281, 1964.
57. Muller, C.H., W.H. Muller and B.L. Haines, *Science,* **143**, 471, 1964.
58. Muller, W.H. and C.H. Muller, *Bull. Torrey Bot. Club,* **91**, 327, 1964.
59. Muller, W.H., *Bot. Gaz.,* **126,** 195, 1965.
60. Muller, C.H., R.B. Hanawalt and J.K. McPherson *Bull. Torrey Bot. Club,* **95**, 225, 1968.
61. Muller, W.H., P. Lorber, B. Haley and K. Johnson, *Bull. Torrey Bot. Club,* **96**, 89, 1969.
62. Olmsted, C.E., *Bot. Gaz.,* **106,** 46, 1944.
63. Parenti, R.L. and E.L. Rice, *Bull. Torrey Bot. Club,* **96**, 70, 1969.
64. Proebsting, E.L. and A.E. Gilmore, *Proc. Amer. Soc. Hort. Sci.,* **38**, 21, 1940.
65. Rice, E.L., W.T. Penfound and L.M. Rohrbaugh, *Ecology,* **41**, 224, 1960.
66. Rice, E.L., *Ecology,* **45**, 824, 1964.
67. Rice, E.L., *Physiol. Plant,* **18**, 255, 1965a.
68. Rice, E.L., *Proc. Okla. Acad. Sci.,* **45**, 43, 1965b.
69. Rice, E.L., *Southwestern Naturalist,* **10**, 248, 1965c.
70. Rice, E.L. and R.L. Parenti, *Southwestern Naturalist,* **12**, 97, 1967.
71. Salisbury, F.B., *Scientific Amer.,* **198**, 108, 1958.
72. Salisbury, F.B. and C. Ross, *Plant Physiology,* Wadsworth Publishing Co., Inc., Belmont, Calif., 1969.
73. Sayre, J.D., *Plant Physiol.,* **3**, 71, 1928.
74. Schaal, L.A. and G. Johnson, *Phytopathology,* **45**, 626, 1955.
75. Schwimmer, S., *J. Biol. Chem.,* **232**, 715, 1958.
76. Slatyer, R.O., *Plant-Water Relationships,* Academic Press, New York, 1967.
77. Sondheimer, E. and D.H. Griffin, *Science,* **131**, 672, 1960.
78. Sondheimer, E., The chlorogenic acids and related compounds, in *Plant Phenolics and Their Industrial Significance,* V.C. Runeckles, Ed., Imperial Tobacco Co. of Canada, Montreal, 1962.
79. Sondheimer, E., D.S. Tzou and E.C. Galson, *Plant Physiol.,* **43**, 1443, 1968.
80. Stocker, O., *UNESCO Arid Zone Res.,* **15**, 63, 1960.
81. Vaartaja, O., *Ecol. Monogr.,* **29**, 91, 1959.
82. Warburg, P. and E. Negelein, *Zeirschr. Phys. Chem.,* **106**, 191, 1923.
83. Warming, E., *Oecology of Plants; An Introduction to the Study of Plant Communities,* The Clarendon Press, Oxford, England, 1909.
84. Weinberger, J.H., *Proc. Amer. Soc. Hort. Sci.,* **67**, 107, 1956.
85. Williams, A.H., Enzyme inhibition by phenolic compounds, in *Enzyme Chemistry of Phenolic Compounds,* MacMillan Co., New York, 1963.
86. Wilson, R.E. and E.L. Rice, *Bull. Torrey Bot. Club,* **95**, 432, 1968.
87. Zenk, M.H. and G. Müller, *Nature,* **200**, 761, 1963.
88. Zimsmeister, H.D. and W.H. Hollmuller, *Planta,* **63**, 133, 1964.

GENERAL REFERENCES

1. Alexander, M., *Introduction to Soil Microbiology,* 2nd Ed. John Wiley, New York, 1977
2. Bainbridge, R., G.C. Evans and O. Rackham, *Light as an Ecological Factor,* John Wiley, New York, 1966.
3. Baker, F.B. and W.C. Snyder, Ed., *Ecology of Soil-borne Plant Pathogens,* Univ. of Calif. Press Berkeley, 1965.
4. Burges, A. and F. Raw, *Soil Biology,* Academic Press, New York, 1967.
5. Evans, L.T., Ed., *Environmental Control of Plant Growth,* Academic Press, New York, 1963.
6. Fried, M. and H. Broeshart, *The Soil-Plant System,* Academic Press, New York, 1967.
7. Geiger, R., *The Climate Near the Ground,* Harvard Univ. Press, Cambridge, 1965.
8. Gray, T.R.G. and D. Parkinson, Ed., *The Ecology of Soil Bacteria,* Univ. of Toronto Press, 1968.
9. Hilman, B., Crop Losses from Air Pollutants, *Environ. Sci. and Tech.*, 16, 495, 1982.
10. Hoagland, D.R., *Inorganic Nutrition of Plants,* Chronica Botanica, Waltham, Massachusetts, 1948.
11. Jeffrey, D.W., *Soil-Plant Relationships,* Croom Helm, London, 1987.
12. Keller, T., *Air Pollutants Deposition and Effects on Plants* (B. Ulrich and J. Pankrath, Eds.), Reidel, Dordrecht, 1983.
13. Kononova, M.M., *Soil Organic Matter,* Pergamon Press, New York, 1966.
14. Kramer, P.J., *Plant and Soil Water Relationships,* McGraw-Hill, New York, 1949.
15. Leopold, A.C., *Plant Growth and Development,* McGraw-Hill Book Co., New York, 1964.
16. Nason, A. and W.D. McElroy, Modes of action of the essential mineral elements, in *Plant Physiology, A Treatise, Vol. III,* Academic Press, New York, p. 451–536, 1963.
17. Stewart, W.D.P., *Nitrogen Fixation in Plants,* Athlone Press, London, 1966.
18. Sutcliffe, J.F., *Mineral Salts Absorption in Plants,* Pergamon Press, New York, 1962.
19. Treshow, M., *Environment and Plant Response,* McGraw-Hill Book, Co., New York, 1970.
20. Treshow, M., Effects of toxicants on terrestrial plants, in *Environmental Toxicology,* WHO, Copenhagen, 1983.
21. Whyte, R.O., *Crop Production and Environment,* Faber and Faber, London, 1946.

ELROY L. RICE
University of Oklahoma

ECOLOGY OF PRIMARY TERRESTRIAL CONSUMERS

BASIC CONCEPTS

There are many approaches to the study and appreciation of the natural world. The ecologist looks at it with his interest focused on the relations between living things and their surroundings. For purposes of quantitative analysis, he finds it useful to think of nature as organized into ecological systems (*ecosystems*), in which the living units interact with their environments to bring about the *flow of energy* and the *cycling of matter* wherever life is found. In this conceptual framework, organisms can profitably be considered according to their major roles in the handling of matter and energy. Thus, living things have ecologically classified (Thienemann, 1926) as *producers* if they are *autotrophic,* i.e., able to manufacture their own food from simple inorganic substances with energy obtained from sunlight (photosynthetic green plants) or from the chemical oxidation of the inorganic compounds (chemosynthetic bacteria), or as *consumer* if they are *heterotrophic,* i.e., required to depend on already synthesized organic matter as the source of food energy. A special and very important group of consumers are the *decomposers,* which break up the complex organic substances of dead matter, incorporating some of the decomposition products in their own protoplasm and making available simple inorganic nutrients to the producers. Decomposers consist chiefly of fungi and bacteria, which absorb their food through cell membranes and thus differ significantly from the larger consumers, which ingest plant and animal tissue into an alimentary tract. There are, however, many different modes of nutrition, and it has recently been suggested (Wiegert and Owen, 1970) that energy flow and the cycling of matter may be better understood if heterotrophic consumer organisms are classified on the basis of their energy resources rather than in terms of their feeding habits. Thus we may recognize two major groups of consumers: *biophages* if they obtain their energy from living matter, and *saprophages* if they derive their energy from dead and decaying materials. These basic classifications do not accommodate such organisms as Euglena and the Venus fly-trap which are capable of both photosynthesis and heterotrophy, but they provide a reasonable framework for the majority of plant and animal species.

FOOD CHAINS AND FOOD WEBS

The concept of *food-chain* was developed by Elton (1927) to represent the series of interactions that occur between organisms in their efforts to obtain nourishment. Thus a hierarchy of relationships is formed as green plans or their products are eaten by animals and these are in turn eaten by other animals. Organisms, then, are functionally related to one another as links in a chain, i.e., as successive components in a system for the transfer of energy and matter. An organism can be characterized ecologically in terms of the position it occupies in its food-chain, and consumer organisms that are one, two and three links removed from the producer are referred to as *primary, secondary* and *tertiary consumers,* respectively. Thus primary consumers are those which subsist on green plants or their products, and are broadly termed *herbivores,* in contrast to secondary and higher-order consumers, which are termed *carnivores.* (Note again, as with the producer–consumer classification, that the herbivore–carnivore dichotomy is not perfect; it does not allow for *omnivores,* which eat both plant and animal tissue, or for those organisms which are herbivorous at one state in their life history and carnivorous at another.)

With each successive transfer, some of the energy that was incorporated by the producer organism (the initial link in the chain) is lost as heat, and for this reason food-chains generally do not involve more than four or five links from the beginning to the ultimate large consumer. A typical terrestrial food-chain consists of a green plant, e.g., grass, eaten by an insect, e.g., a grass-hopper, which is in turn eaten by a small bird, e.g., a sparrow, and this in turn by a larger bird, e.g., a hawk. However, food chains are generally linked to other chains at almost every point; the grass is likely to be eaten by numerous species of herbivore, each of which has its own set of predators, etc. The result is that the community ecosystem is made up, not of a set of isolated food-chains but of an interconnected *food-web* whose structure rapidly becomes very complex when more than a few species are considered. This complexity, however, is restricted in part by the limited length of food-chains and in part by the unidirectional flow of energy in these chains. Food-webs are made up of two principal types of food-chains: *grazing food-chains,* which involve the direct consumption and transformation of living tissue, as in the grazing of rangeland by cattle or sheep, and *detritus food-chains,* which involve the disintegration and conversion of dead matter, both plant and animal, by a sequence of decomposer organisms.

TROPHIC LEVELS AND ECOLOGIGAL PYRAMIDS

The concept of the food-chain permits us to equate, in terms of ecosystem function, all organisms which occupy the same feeding position, i.e., which are the same number of links removed from the producer organism. Organisms which are similar in this respect are said to occupy the same *trophic level.* Such levels form a natural hierarchy arranged from the producer level at the bottom, through the primary consumer (herbivore) level, to one or more successive levels of subsidiary consumers at the top. The trophic level concept was developed by Lindemann (1942) to compare the energy content of the different feeding groups in natural communities and to evaluate the efficiency with which energy is transferred from level to level.

If the number of individuals, the total *biomass* (weight of tissue), or the energy content of the organisms of successive trophic levels in a natural community is examined, the quantity is generally found to decrease as one goes upward from producer to ultimate consumer levels. Diagrammatic models of this trophic structure thus tend to be pyramidal and have given rise to the concept of *ecological pyramids.* Comparisons of trophic levels based on numbers of individuals can be misleading, however, especially when species of very different size and rate of growth are involved: herbivorous insects are likely to be much more numerous per unit of suitable habitat than are grazing or browsing mammals (Evans and Lanham, 1960). This difficulty is partly resolved by measuring total weight of organisms present, thus replacing a *pyramid of numbers* with a *pyramid of biomass.* If they are constructed for parasite food-chains (in which one host can support many parasites) or for communities whose producers (such as diatoms and other minute algae) have a more rapid rate of replacement, or *turnover,* than its consumers, such pyramids may appear inverted or may show a very narrow base. Furthermore, because the number of calories varies considerably from tissue to tissue because of differences in composition (fat averages about 9500, and carbohydrate and protein about 4000, calories per gram), biomass may prove a poor indicator of energy content. It is therefore desirable, whenever possible, to replace weights with calorific values and to convert the pyramid of biomass into a *pyramid of energy.* If the total amount of energy utilized at each trophic level over a set period of time is taken into account the quantity will always be less at each succeeding level and the upright pyramidal shape of the model will be maintained. Thus energy units provide the best basis for comparing the *productivity* (the rate at which energy and matter are stored in the form of organic substances) of different organisms, of different trophic levels, and of different ecosystems. They also offer the best means of evaluating the *efficiency* (the ratio of energy stored or put out in a process to that put in, usually expressed as a percentage) of organisms and ecosystems in carrying out their activities of transferring and transforming energy and matter.

STANDING CROPS, PRODUCTION, AND ENERGY FLOW

The quantity of living organisms present at a given time may be referred to as the *standing crop* or *stock.* For reasons already explained, this quantity is best expressed in terms of its energy content (in calories). The magnitude of the standing crop will vary from place to place, depending basically on the available quantities of energy and nutrients, and in almost all places from season to season, being influenced by all the factors affecting growth and reproduction. Relatively long-lived consumers like the large herbivorous mammals may accumulate and store considerable quantities of matter and energy in their bodies and thus achieve a large standing crop, but surprisingly high values can be reached by much smaller organisms, such as insects, which feed more efficiently and reproduce more rapidly. Unusually high standing crops are seen at times of population explosions, such as those of defoliating insects and of periodically fluctuating species like snowshoe hares and lemmings, but these levels cannot be sustained for more than a short time. Inverted pyramids of biomass illustrate the possibility of a relatively large standing crop of consumers supported by a small standing crop of producer plants, when the latter are smaller, grow more rapidly, and are replaced more frequently than the consumers.

Because organisms may be eaten or move away from an area, the standing crop often fails to be a good measure of the total quantity of tissue they have produced over a given period of time. This total quantity, or *production*, includes not only the new tissue added as *growth* to the bodies of individual organisms but also that resulting from the reproduction of new individuals. That part of the production that is removed by man (or some other species) is known as the *yield* or harvest. Because of the limited efficiency of the metabolic processes required in the formation of new tissue, more energy and matter are needed by the organism than are stored in production. Thus, to evaluate the complete functioning of an ecosystem, we need a measure of the total energy (or matter) involved in metabolism; for consumer organisms, this is referred to as *assimilation* or *energy flow.* (For producer organisms, the total energy or matter used in their metabolism is called *gross primary production*, while that incorporated as new tissue is called *net primary production.*)

ENERGY BUDGETS AND EFFICIENCIES

Full understanding of increases or decreases in standing crop or energy flow requires quantitative knowledge of the biological processes involved in the transfer of matter and energy. For consumer organisms, these consist of the processes of *ingestion* or intake, *respiration*—a measure of metabolic activity, and *egestion* (generally taken to include the elimination of both feces and urine). Use is made of the *energy budget* or balance described by the equations.

$$\text{Ingestion} = \text{Production} + \text{Respiration} + \text{Egestion}$$

$$\text{Assimilation} = \text{Production} + \text{Respiraton}$$

and other expressions to calculate the values for processes whose quantities are not known or to determine the net change of energy transfer. Such calculations enable the ecologist to estimate the efficiency with which these processes are carried out, as for example

$$\text{assimilation efficiency} = \frac{\text{calories assimilated}}{\text{calories ingested}}$$

$$\text{growth efficiency} = \frac{\text{calories in growth}}{\text{calories ingested}}$$

$$\text{yield efficiency} = \frac{\text{calories to man (or other exploiter)}}{\text{calories ingested}}$$

These and other efficiencies provide the means for comparing different trophic levels or other ecological units with respect to their functional roles in the ecosystem. For primary consumers, the ratio

$$\frac{\text{calories consumed by herbivore population}}{\text{calories of available plant food}}$$

measures their *food-chain efficiency,* while the ratio

$$\frac{\text{calories of herbivores consumed by carnivores}}{\text{calories of plant food consumed by herbivores}}$$

measures their *efficiency of energy transfer.*

DIVERSITY OF PRIMARY CONSUMERS

All parts of plants—leaves, buds, flowers and their products (seeds, pollen, nectar), stems, sap, roots, bark and wood—are eaten by primary consumers, who have evolved a great diversity of life form and habits in response to their food. Large grazers and browsers, such as elephants and deer, can ingest plant tissue in bulk, but others are specialized to particular products, e.g., the mylabrid weevils that feed on beans and peas, and the sap-sucking aphids and leaf-hoppers. The specialization may extend to particular kinds of plants; the Australian koala "bear," for example, is limited to leaves of *Eucalyptus*, and the larva of the swallowtail butterfly *Papilio marcellus* feeds exclusively on the foliage of the prickly ash (*Zanthoxylua*). Much less is known of the food habits of detritus-feeders, which never attain the large size of some grazers, but Petrusewicz and Macfadyen (1970) point out that "remarkably few species restrict their diets at all narrowly."

All of the principal phyla of land animals have developed forms which are partly or entirely herbivorous, and it is likely that the majority of terrestrial heterotrophic species are primary consumers. About half of the known kinds of insects are plant-eaters (Brues, 1946), and in some habitats the proportion may be considerably higher: Menhinick (1967) found that herbivores made up about 80% of the insect species of a bush clover (*Lespedeza*) stand in South Carolina, and Evans and Murdoch (1968) reported that 85% of approximately 1500 insect species encountered on a 30-year-old abandoned field in Michigan were herbivorous. With such *diversity,* it is unlikely that any species of plant has escaped exploitation by herbivores. Indeed, most plants are host to a wide range of consumers; for example, at least 227 species of herbivorous insects are known to be associated with the oaks (*Quercus robur* and *Q. petraea*) of British woodlands (Elton, 1966, 197).

The most important insect consumers of live vegetation belong to the orders Orthoptera (grasshoppers), Hemiptera (true bugs), Homoptera (leaf-hoppers, aphids, scales), Coleoptera (beetles), Lepidoptera (moths, butterflies), Diptera (flies), and Hymenoptera (bees, wasps, ants). Insects as a group share dominance as herbivores with the mammals, including such larger types as horses, pigs, deer, antelope, goats, sheep and cattle, and such smaller ones as hares, rabbits, squirrels, marmots, voles and lemmings. (In tropical regions, various monkeys and apes are also important herbivores, as are fruit-eating and nectarivorous bats.) There are significant herbivores among terrestrial birds, e.g., the Galliformes (grouse, quail, pheasants) and the "specialist" seed-eating finches and sparrows, the fruit-eating parrots, and the nectar-sucking hummingbirds, and among land mollusks, such as the "garden" snails and slugs.

Much less is known of the herbivores which feed on roots, underground stems, and other living plant parts in the soil. Studies by Bornebusch (1930), Van der Drift (1951), Cragg (1961) and Macfadyen (1961) indicate that the most important groups of herbivores in north temperate grassland and forest soils are the parasitic nematode worms, the molluscs, and the larvae of Diptera, Coleoptera, and Lepidoptera. They occur in close association with other primary consumers, the detritus feeders, which include the Oligochaeta (earthworms and enchytraeid worms), Isopoda (wood lice), Diplopoda (millipedes), free-living Nematoda, Collembola (spring tails), and Acari (mites). The mechanical comminution of dead matter by these organisms provides a substrate of small particles which can be more easily attacked by fungi, protozoa, bacteria and other microorganisms (Phillipson, 1966).

MEASUREMENT OF PRIMARY CONSUMPTION

The measurement of primary consumption involves the assessment, in terms of number of individuals, biomass, and energy equivalence, of the quantity of herbivorous animals present or produced over a period of time, the *energetic cost* of producing and maintaining that product, and the fate of the matter and energy that becomes incorporated in herbivore tissue. A detailed account of methods and techniques employed for these purposes is beyond the scope of this article, but a brief treatment is presented below.

The mobility, abundance, and small size of many primary consumers make it difficult to assess their numbers. Total counts or *censuses,* as of large mammals by aerial photography (Parker, 1971) or of small ones by intensive trapping (Gliwicz *et al.,* 1968), can only rarely be achieved, and it is general practice to rely on *sample collections* (e.g., Wiegert, 1964, 1965). The marking of individual organisms for subsequent recapture has been employed to estimate population size in such herbivores as grasshoppers (Nakamura *et al.,* 1971), and *indices of relative abundance* based on fecal pellet counts (Southwood, 1966, 229) or on damage to vegetation (Odum and Pigeon, 1970, 1–69) are sometimes used. Estimation of change in numbers with time involves a knowledge of birth, death and migration rates, the calculation of which requires considerable *lifetable information.*

Biomass values are commonly derived by multiplying the number of individuals by average weights obtained from sample specimens or calculated from formulas relating weight to body dimensions (Petrusewicz and Macfadyen, 1970, 51). More detailed studies require knowledge of the growth of individuals and the rate of weight gain. Ultimately, biomass should be converted to its energy equivalent by determining its calorific value; this requires the use of a *calorimeter,* such as that developed by Phillipson (1964), and because of the difficulty of obtaining complete combustion, it is generally carried out in the laboratory. Published tables of the caloric content of various plant and animal tissues (e.g., Golley, 1961) show a considerable range of values even for the same species, reflecting differences in life history stage, season, and environmental conditions.

It is also clear that the dimensions of primary consumption cannot be accurately gauged without knowledge of the processes (*ingestion, egestion, assimilation, respiration*) associated with it. It is important, therefore, to investigate the food habits and feeding rates of herbivores, as well as to determine the quality of the food consumed, the proportion rejected as feces, the digestive efficiency, and the rates of respiration (oxygen consumption can be measured by a *respirometer* such as that of Smith and Douglas, 1949; see also the book on manometric techniques by Umbreit *et al.,* 1957). Much of this work has had to be done on animals in confinement, and little is known as yet of activity levels and metabolic rates of herbivores in nature. Recent developments in the use of *radioactive isotope tracers* (Williams and Reichle, 1965) and the *telemetric measurement* of heart rate or other characteristics related to respiration (Adams, 1965) give promise that the study of metabolism under filed conditions will eventually be feasible.

EFFICIENCY OF PRIMARY CONSUMERS

With careful management and appropriate stocking large mammalian herbivores can consume a fairly high proportion of the available plant food. In Great Britain, sheep stocked at a density of 10 ewes per hectare on lowland pasture may ingest up to 70% of the annual production of grass (Eadie, 1970), and beef cattle on management rangeland in the United States, when stocked at the maximum recommended *exploitation rate,* will consume from 30% to 45% of the forage (Lewis *et al.,* 1956). The efficiency of feeding will of course vary with the situation: sheep on hill pastures, where a stocking rate cannot exceed 0.8 ewes per hectare, may utilize no more than 20% of the available food (Eadie, 1970), and Paulsen (1960) indicates that on alpine ranges in the Rocky Mountains the proportion of herbage production removed by sheep was only 7%. This latter figure is close to estimates of feeding efficiency obtained for large mammals in the wild: 10% for the Uganda kob (Buechner and Golley, 1967) and 9.6% for the African elephant (Wiegert and Evans, 1967). Where large, multispecies herds of ungulates occur, as on the savannas of Africa, their combined efficiency may be much higher: estimates of 60% for Uganda grassland and of 28% for Tanganyika grassland were obtained from observations by Petrides and Swank (1965) and Lamprey (1964), respectively.

At ordinary densities, small herbivores, both vertebrate and invertebrate, are much less efficient in their utilization of available food. Golley (1960) reports an efficiency of 1.6% for meadow voles (*Microtus*) in a Michigan grassland, and values of less than 0.5% are estimated for a variety of other small mammals and granivorous birds (Wiegert and Evans, 1967). Similar low efficiencies apparently characterize insect herbivores except when these are present in plague proportions. Wiegert (1965) found that grasshoppers (23 species of acridids and tettigoniids) consumed 1.3% of net plant production in a field of alfalfa, and Smalley (1960) obtained efficiency values of 1.6–2.0% for the meadow grasshopper *Orchelimum* in a *Spartina* salt marsh. Even if all invertebrate herbivores are considered together, their total consumption of net primary production seems rarely to exceed 10%; the following values appear to be representative:

Nature of ecosystem	% net plant production used by invertebrates	References
Bush-clover stand (S.C.)	0.4–1.4	Menhinick, 1967
Festuca grassland (Tenn.)	9.6	Van Hook, 1971
Spartina salt marsh (Ga.)	7.0–9.2	Teal, 1962
mesophytic woodland (Tenn.)	1.5	Reichle and Crossley, 1967
Mature deciduous forest (southern Canada)	1.5–2.5	Bray, 1964

These values do not necessarily indicate the total damage done to the plant crop. For example, Andrzejewska *et al.* (1967) report that grasshoppers may destroy 4.8 times the amount of plant material they ingest, by gnawing the grass blades so that part of the leaf falls off. Such material does not enter the grazing food-chain, however, but drops to the ground and is consumed by detritus feeders.

The nutritive value of most higher plants varies with age of the plant, season and environmental conditions, which also affect the *palatability* of the food. Few herbivores have the ability to digest cellulose without the assistance of *symbiotic bacteria* or *protozoa,* and much of the food that is eaten fails to be assimilated and is eliminated as feces. Assimilation therefore involves an important split in the flow of matter and energy through the ecosystem.

High assimilation/consumption ratios can be achieved by herbivores which are selective feeders on concentrated foods such as nuts, seeds and fungi; assimilation efficiencies of 85–95% have been recorded for such small mammals as *Clethrionomys, Apodemus* and *Microtus* (Drozdz, 1967; Davis and Golley, 1963, 81). Somewhat lower values are reported for large ruminants, ranging from 60–80% for sheep, cattle and deer down to 40% for moose and elephant (Graham, 1964; Dinesman, 1967; Davis and Golley, op. cit.). Interestingly, insect herbivores appear to be considerably less efficient feeders, reaching assimilation levels of 29% for the caterpillar *Hyphantria,* 27% for the grasshopper *Orchelimum,* 36% for the grasshopper *Melanoplus,* and 33% for the spittlebug *Philaenus* (Gere, 1956; Smalley, 1960; Wiegert, 1964, 1965). Detritus feeders seem to have even lower efficiencies, as witness values of 20% for oribatid mites and 10% for the millipede *Glomeris* (Engelmann, 1961; Bocock, 1963). Thus, grazing food-chains and detritus food-chains appear to be characterized by quite different assimilation rates.

Part of the energy assimilated by herbivores is stored as growth or production, while the rest is respired, dissipated as heat in the oxidation of organic matter. These respiratory losses may amount to as much as 98% of the energy assimilated (Petrusewicz and Macfadyen, 1970) but vary greatly, depending on such factors as environmental temperature, level of activity, and age of the individual. In mammals, there is a tendency for the *weight-specific respiration rate* to rise as body weight decreases, and this, in combination with the fact that growth and reproduction rates tend to be greater in small species than in large ones, apparently results in a rather constant low production/assimilation ratio of 1–2% for mammalian populations; values of this magnitude have been calculated for both elephants and mice (Wiegert and Evans, 1967). In contrast, herbivorous insect populations seem to show much higher assimilation efficiencies, on the order of 35–45% and, when only young life stages (larvae, nymphs) are involved, even of 50–60% (Macfadyen, 1967). Data are still too few to permit satisfactory generalizations, but the importance of further studies is evident when it is remembered that energy lost in respiration is irrevocably removed from the ecosystem and cannot be recycled.

Synthesis of these process ratios leads to an evaluation of the overall efficiency with which herbivores convert the energy of primary (plant) production into their own tissue. The value of the production/consumption ratio for a fairly broad spectrum of vertebrate and invertebrate herbivores seems to range from less than 1% to 15–20% as a maximum (Petrusewicz and Macfadyen, 1970), with some evidence that the insects are generally more efficient than the non-domesticated mammals; this difference may be due in large part to the necessity for the latter to maintain a steady, high body temperature. When these values are considered along with the proportions of available food ingested (see above), the efficiency of energy transfer of herbivores is rarely found to exceed 15%, and this has been suggested as a likely maximum level for natural ecosystems (Slobodkin, 1962).

REGULATORY MECHANISMS OF PRIMARY CONSUMERS

Although terrestrial herbivores are occasionally so abundant as to deplete the local supply of plant food, as sometimes happens when an introduced species like the Japanese beetle (*Popilia japonica*) or the Europena gypsy moth (*Porthetria dispar*) enters a new biotic community, such cases seem to be the exception rather than the rule, and, in contrast to many aquatic ecosystems, the bulk of net annual primary production on land is not eaten in the living state by primary consumers but dies and is acted on by decomposer organisms. Thus it appears that, on the whole, land herbivores usually occur at densities well below the level of their available plant food, and the reason or reasons for this are of great ecological interest.

Despite the fact that sound generalizations about abstract concepts such as "trophic levels" are not easily arrived at (Murdoch, 1966), several hypotheses about the *regulation of herbivore numbers* have been suggested. The apparent rarity of obvious depletion of vegetation by herbivores, or of its destruction by meteorological catastrophes, has led Hairston, Smith and Slobodkin (1960) to the belief that herbivores are seldom limited by their food supply; after rejecting weather as an effective control agent, they conclude that herbivores are most often controlled by their predators and/or parasites, interacting in the classical density-dependent manner. These views have been questioned on such grounds as (1) that much green material may often be inedible, unpalatable or even unreachable by the herbivores present, so that food limitation might occur without actual depletion (Murdoch, 1966) or (2) that native herbivores such as forest Lepidoptera and grasshoppers will often increase and cause serious defoliation even in the presence of their predators (Ehrlich and Birch, 1967). Despite these and other criticisms, Slobodkin, Smith and Hairston (1967) find it unnecessary to modify their views in any essential respect.

The possibility that herbivore (and other animal) populations have some capacity for *self-regulation* has not been overlooked. Pimentel (1961, 1968) has suggested the concept of "*genetic feedback*," according to which population density influences the intensity of selective pressure, selection influences the genetic composition of the surviving individuals, and genetic composition influences the subsequent population density. The models thus far proposed to explain how this system works appear to be untenable (Lomnicki, 1971), but the general validity of the concept seems to gain support from the numerous instances of the *co-evolution* of plants and herbivorous animals, e.g., the association between certain orchids and their insect pollinators (Van der Pijl and Dodson, 1966), or that between certain species of *Acacia* and ants (Janzen, 1966), which have been interpreted as involving reciprocal selective interaction.

Another suggested mechanism for self-regulation involves the elaboration of *social behavior patterns,* e.g., territoriality, social hierarchies, and warning displays, which tend to maintain animal populations at relatively low densities, thereby reducing the possibility of depletion of food supplies or other resources (Wynne Edwards, 1962, 1965). The evolution of such a mechanism seems to require that natural selection act on groups rather than on individual organisms, and for this reason Wynne Edwards' hypothesis has been heavily criticized (Williams, 1966; Maynard Smith, 1964).

The basic importance of self-regulatory mechanisms and other density-dependent interactions in limiting populations of animals has also been questioned. Ehrlich and Birch

(1967), for example, deny that "the numbers of all populations are primarily determined by density-regulating factors" and emphasize the significant limiting effects that weather conditions often have on natural populations of herbivores. This view maintains that limiting factors which act on populations without relation to their density commonly maintain such populations at levels where self-regulatory, *density-dependent mechanisms*, if they exist at all, are not called into play.

Wiegert and Owen (1971) suggest that the precise kinds of mechanisms limiting the population of a particular species of herbivore will depend firstly on (a) whether the population, in making use of its plant food resources, directly affects the rate of food supply (either by destroying or by stimulating its capacity to produce new plant tissue), and secondly on (b) the life history characteristics of both the herbivore and its plant food. Most stable terrestrial ecosystems are dominated by relatively large, structurally complex plants that are long-lived, slow-growing and have low rates of population increase, and by herbivores that are smaller, more numerous and faster-growing than their food plants. In contrast, open-water aquatic communities are often composed of small and structurally simple plants (phytoplankton) that are more abundant and multiply more rapidly than their consumers. In such aquatic communities, herbivores can attain high rates of consumption—approaching 100%—of net primary production without danger of completely exhausting their nutritional resources, and their populations are likely to be limited by direct *competition* to levels dictated by the rate at which food is supplied to them. As already noted above, in terrestrial ecosystems, such high rates of consumption by herbivores are rare, and when they occur their influence on the food plants is usually severe. In forest, and to a lesser extent in grassland, communities the herbivores are less likely to be limited by direct competition for food; they tend, rather, to be subject to the effects of predation and parasitism and to have evolved behavioral patterns such as migrations which result in reducing grazing pressure. At the same time, the plants of these systems have developed an array of toxic compounds, unpalatable tissues, thorns and other protective devices to resist grazing. Thus Wiegert and Owen's model of population control stresses the importance of trophic structure (grazing food-chains vs. detritus food-chains) and the biological properties of the interacting populations.

MANAGEMENT OF PRIMARY CONSUMERS

Although energy flow studies suggest that man should become predominantly herbivorous in order to make the most effective use of the solar radiation captured by plants, his need for protein, which he can obtain in more concentrated form from animal tissue, is likely to continue to motivate him toward increasing the production of primary consumers such as cattle, chickens, and fish. In fact, the majority of man's domesticated animals are herbivores.

It is clear that if the conversion of solar energy into animal protein for human use is to be maximized herbivores with high growth efficiencies should be cropped. The energy that man himself must spend in order to secure his food must also be assessed, and the time, effort and other costs of cropping have therefore to be taken into consideration. Over the several thousand years in which man has attempted to domesticate plants and animals, his selective efforts have been remarkably successful in developing efficient herbivores for his own use. Modern methods of *animal husbandry,* where animals are kept in specially constructed buildings and are fed specially processed foods, have greatly increased the amount of food energy reaching the animals (Phillipson, 1966). However, these methods require the expenditure of much energy that goes as hidden cost, in the activities involved in the construction and maintenance of facilities and in the production of the processed food. Are there ways of increasing the more direct conversion of plant tissue to animal protein useful to man?

Macfadyen (1964) has indicated that in Great Britain beef cattle raised on grassland commonly consume only one-seventh of the net annual primary production, the rest going to other herbivores and to decomposers. More care in stocking and better management of range are two of the principal ways in which possible improvement can be sought. Less obvious is the extension of husbandry to other herbivorous species, perhaps even to insects, that may be more efficient feeders than the large warmblooded mammals and that might, through selection, be developed as a productive source of high protein nourishment. The use of faster-growing herbivores, such as rabbits, would yield greater efficiency in terms of meat production per unit of time (Phillipson, 1966). In East Africa, increasing use for human food is being made of antelope, zebra and other native ungulates which are relatively immune to the parasites and diseases that afflict European cattle, and other unproductive areas are being managed for their propagation.

Increased production of plant food as a base for greater herbivore production is also possible. In this effort, man has tried particularly to exploit tropical regions with their higher average temperatures and longer growing seasons. Petrusewicz and Macfadyen (1970) point out, however, that primary productivity does not apparently increase proportionately to the increased light regime of tropical climates, largely due to higher respiration rates in tropical plants and to more rapid rates of decomposition; the tenfold increase in solar radiation experienced by some tropical forests seems not to result in any significant increase in energy available to primary consumers over that found in a temperate deciduous forest. It has long been apparent that agricultural practices developed in temperate regions are poorly adapted for use in the tropics. A great deal more needs to be learned about *tropical ecology* before marked improvement in environmental management can be expected (Owen, 1966).

CONCLUDING REMARKS

This article has attempted to outline the current view of nature as an interaction system in which organisms play a variety of roles in facilitating the circulation of matter and the flow of energy within that thin layer of the earth's surface known as the *biosphere.* Despite the relatively small quantities of material and energy that are channeled through primary consumers, as compared with autotrophic plants and decomposer organisms, the former play a vital role in the maintenance of such a system, in that they accumulate and concentrate inorganic nutrients such as nitrogen, phosphorus and potassium, as well as many trace elements, which are otherwise thinly dispersed in the environment. It has not been possible here to present more than a superficial account of the role of primary consumers, and indeed the study of these animals is still in its early stages. But further observation and experimentation in this aspect of ecology are clearly needed, if man is ever to understand and manage wisely the world he lives in.

REFERENCES

Adams, L. (ed.), *Bio. Science,* **15,** 83–157, 1965.

Andrzejewska, L., A. Breymeyer, A. Kajak, and Z. Wojcik, pp. 477–495 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1967.

Bocock, K.L., pp. 85–91 in *Soil Organisms,* edited by J. Doeksen and J. Van der Drift, Amsterdam, 1963.

Bornebusch, C.H., *Forstl. Forsogskomission* (Copenhagen), **11,** 1–224, 1930.

Bray, J.R., *Ecology,* **45,** 165–167, 1964.

Brues, C.T., *Insect Dietary: An Account of the Food Habits of Insects,* Harvard University Press, Cambridge, 1946.

Buechner, H.K., and F.B. Golley, pp. 243–254 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1967.

Cragg, J.B., *J. Anim. Ecol.,* **30,** 205–234, 1961.

Davis, D.E., and F.B. Golley, *Principles of Mammalogy,* Reinhold Publishing Corp., New York, 1963.

Dinesman, L.G., pp. 261–266 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1967.

Drozdz, A., pp. 323–330 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1976.

Eadie, J., pp. 7–24 in *Animal Populations in Relation to Their Food Resources,* edited by A. Watson, Blackwell Scientific Publications, Oxford, 1970.

Ehrlich, P.R., and L.C. Birch, *Amer. Nat.,* **101,** 97–107, 1967.

Elton, C., *Animal Ecology,* Sidgwick and Jackson Ltd., London, 1927.

Elton, C.S., *The Pattern of Animal Communities,* Methuen and Co. Ltd., London, 1966.

Engelmann, M.D., *Ecol. Monogr.,* **31,** 221–228, 1961.

Evans, F.C., and U.N. Lanham, *Science,* **131,** 1531–1532, 1960.

Evans, F.C., and W.W. Murdoch, *J. Anim. Ecol.,* **37,** 259–273, 1968.

Gere, G., *Acta Biol. Acad. Sci. Hungary,* **7,** 43–72, 1956.

Gliwicz, J., R. Andrzejewski, G. Bujalska, and K. Petrusewicz, *Acta Theriol.,* **13,** 401–413, 1968.

Golley, F.B., *Ecol. Monogr.,* **30,** 187–206, 1960.

Golley, F.B. *Ecology,* **42,** 581–584, 1961.

Graham, N.M., *Austral. J. Agric. Res.,* **15,** 969–973, 1964.

Hairston, N.G., F.E. Smith, and L.B. Slobodkin, *Amer. Nat.,* **94,** 421–425, 1960.

Hutton, M., The Role of Wildlife Species in the Assessment of Biological Impact from Chronic Exposure to Persistent Chemicals, *Ecotoxicol. Environ. Saf.,* **6,** 471, 1982.

Janzen, D.H., *Evolution,* **20,** 249–275, 1966.

Lamprey, H.F., *East African Wildlife J.,* **2,** 1–46, 1964.

Legge, A.H. and S.V. Krupa, *Air Pollutants and Their Effects on the Terrestrial Ecosystem,* John Wiley, New York, 1986.

Lewis, J.K., G.M. Van Dyne., L.R. Albee, and F.W. Whetzal, *Bull. S. Dak. Agric. Expt. Sta.,* **459,** 44, 1956.

Lindemann, R.L., *Ecology,* **23,** 399–418, 1942.

Lomnicki, A., *Amer. Nat.,* **105,** 413–421, 1971.

Macfadyen, A., *Ann. Appl. Biol.,* **49,** 216–219, 1961.

Macfadyen, A., pp. 3–20 in *Grazing in Terrestrial and Marine Environments,* edited by DJ. Crisp, Blackwell Scientific Publications, Oxford, 1964.

Macfadyen, A., pp. 383–412 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1967.

Maynard Smith, J., *Nature,* **201,** 1145–1147, 1964.

Menhinick, E.F., *Ecol. Monogr.,* **37,** 255–272, 1967.

Murdoch, W.W., *Amer. Nat.,* **100,** 219–226, 1966.

Nakamura, K., Y. Ito, M. Nakamura, T. Matsumoto, and K. Hayakawa, *Oekologia,* **7,** 1–15, 1971.

Odum, H.T., and R.F. Pigeon, (eds.), *A Tropical Rain Forest,* US Atomic Energy Commission, Washington, DC, 1970.

Owen, D.F., *Animal Ecology in Tropical Africa,* Oliver and Boyd, Edinburgh, 1966.

Pandy, S.D., V. Misra, and P.M. Viswanathan, Environmental Pollutants and Wildlife—A Survey, *Int. J. Env. Studies,* **28,** 2/3, 1986.

Parker, G.R., Occas. Pap. No. 10, Canad. Wildlife Serivces, Ottawa, 1971.

Paulsen, H.A., Jr., *Iowa State J. Sci.,* **34,** 731–748, 1960.

Petrides, G.A., and W.G. Swank, *Proc. Ninth Internat. Grasslands Congress,* San Paulo, Brazil, 1965.

Petrusewicz, K., and A. Macfadyen, *Productivity of Terrestrial Animals: Principles and Methods,* Internat. Biol. Program Handbook No. 13, Burgess and Son, Abingdon, Berkshire, England, 1970.

Phillipson, J., *Oikos,* **15,** 130–139, 1964.

Phillipson, J., *Ecological Energetics,* St. Martin's Press, New York, 1966.

Pimentel, D., *Amer. Nat.,* **95,** 65–79, 1961.

Pimentel, D., *Science,* **159,** 1432–1437, 1968.

Rasmussen, L., *Ecotoxicology,* Swedish Research Council, Stockholm, 1984.

Reichle, D.E., and D.A. Crossley, Jr., pp. 563–587 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1967.

Slobodkin, L.B., *Adv. Ecol. Res.,* **1,** 69–101, 1962.

Slobodkin, L.B., F.E. Smith, and N.G. Hairston, *Amer. Nat.,* **101,** 109–124, 1967.

Smalley, A.E., *Ecology,* **41,** 672–677, 1960.

Smith, A.H., and J.R. Douglas, *Ann. Ent. Soc. Amer.,* **42,** 14–18, 1949.

Southwood, T.R.E., *Ecological Methods,* Methuen and Co. Ltd., London, 1966.

Teal, J.M., *Ecology,* **43,** 614–624, 1962.

Thienemann, A., *Verh. deutsch. Zool. Ges.,* **31,** 29–79, 1926.

Umbreit, W.R., R.H. Burris, and J.F. Stauffer, *Manometric Techniques,* 3rd ed., Burgess, Publishing Co., Minneapolis, 1957.

Van der Drift, J., *Tijdschr. Ent.,* **94,** 1–68, 1951.

Van der Pijl, L., and C.H. Dodson, *Orchid Flowers: Their Pollination and Evolution,* University of Miami Press, Coral Gables, Florida, 1966.

Van Hook, R.I., *Jr., Ecol. Monogr.,* **41,** 1–26, 1971.

Walker, C.H., Pesticides and Birds; Mechanisms of Selective Toxicity, *Agricultrue, Ecosystems and Environment,* **9,** 211, 1983.

Wiegert, R.G., *Ecol. Monogr.,* **34,** 225–241, 1964.

Wiegert, R.G., *Oikos,* **16,** 171–176, 1965.

Wiegert, R.G., *Energy Transfer in Ecological Systems,* Patterns of Life Series, Rand McNally and Company, Chicago, 1970.

Wiegert, R.G., and F.C. Evans, pp. 499–518 in *Secondary Productivity of Terrestrial Ecosystems,* edited by K. Petrusewicz, Institute of Ecology, Warsaw, 1967.

Wiegert, R.G., and D.F. Owen, *J. Theoret. Biol.,* **30,** 69–91, 1971.

Williams, E.C., Jr., and D.E. Reichle, *Oikos,* **19,** 10–18, 1968.

Williams, G.C., *Adaptation and Natural Selection,* Chapter 4, Princeton University Press, Princeton, NJ, 1966.

Wynne Edwards, V.C., *Animals Dispersion in Relation to Social Behaviour,* Oliver and Boyd, Edinburgh, 1962.

Wynne Edwards, V.C., *Science,* **147,** 1543–1548, 1965.

FRANCIS C. EVANS
University of Michigan

ECOLOGY RADIATION: see RADIATION ECOLOGY

ECOSYSTEM THEORY

The concept of the ecosystem is not only the center of professional ecology today, but it is also the most relevant concept in terms of man's environmental problems. During the mid 70s, the public seized on the root meaning of ecology, namely "oikos" or "house," to broaden the subject beyond its previously rather narrow academic confines to include the "totality of man and environment," or the whole environmental house, as it were. We are witnessing what is called a historic "attitude revolution" (Odum, 1969, 1970c) in the way people look at their environment for the very simple reason that for the first time in his short history man is faced with ultimate rather than merely local limitations. It will be well for all of us to keep this overriding simplicity in mind as we face the controversies, false starts and backlashes that are bound to accompany man's attempts to put some negative feedback into the vicious spiral of uncontrolled growth and resource exploitation that has characterized the past several decades.

Man has been interested in ecology in a practical sort of way since early in his history. In primitive society every individual, to survive, needed to have definite knowledge of his environment, i.e., of the forces of nature and of the plants and animals around him. Civilization, in fact, began when man learned to use fire and other tools to modify his environment. It is even more necessary than ever for mankind, as a whole, to have an intelligent knowledge of the environment if our complex civilization is to survive, since the basic "laws of nature" have not been repealed; only their complexion and quantitative relations have changed, as the world's human population has increased and as man's power to alter the environment has expanded.

Like all phases of learning, the science of ecology has had a gradual, if spasmodic, development during recorded history. The writings of Hippocrates, Aristotle, and other philosophers of the Greek period contain material which is clearly ecological in nature. However, the Greeks literally did not have a word for it. The word "ecology" is of recent coinage, having been first proposed by the German biologist Ernst Haeckel in 1869. Before this, many of the great men of the biological renaissance of the eighteenth and nineteenth centuries had contributed to the subject even though the label "ecology" was not in use. For example, Anton von Leeuwenhoek, best known as a pioneer microscopist of the early 1700s, also pioneered the study of "food chains" and "population regulation" (see Egerton, 1968), two important areas of modern ecology. As a recognized distinct field of biology, the science of ecology dates from about 1900, and only in the past decade has the word become part of the general vocabulary. Today, everyone is acutely aware of the environmental sciences as indispensable tools for creating and maintaining the quality of human civilization. Consequently, ecology is rapidly becoming the branch of science that is most relevant to the everyday life of every man, woman, and child.

As recently as 1960 the theory of the ecosystem was rather well understood but not in any way applied. The applied ecology of the 1960s consisted of managing components as more or less independent units. Thus we had forest management, wildlife management, water management, soil conservation, pest control, etc., but no ecosystem management and no applied human ecology. Practice has now caught up with theory. Controlled management of the human population together with the resources and the life support system on which it depends as a single, integrated unit now becomes the greatest, and certainly the most difficult, challenge ever faced by human society.

As we have seen, an "anthropocentric" definition of the ecosystem might read something as follows: Man as a part of, not apart from, a life support system composed of the atmosphere, water, minerals, soil, plants, animals, and microorganisms that function together to keep the whole viable.

Any unit including all of the organisms (i.e., the "community") in a given area interacting with the physical environment so that a flow of energy leads to a clearly defined trophic structure, biotic diversity and material cycles (i.e. exchange of materials between living and nonliving parts) within the system is an *ecological system or ecosystem.*

The following formal definition is the one used in the third edition of *Fundamentals of Ecology* (E.P. Odum, 1971)

> The word ecology is derived from the Greek *oikos,* meaning "house" or "place to live." Literally, ecology is the study of organisms "at home." Usually ecology is defined as the study of the relation of organisms or groups of organisms to their environment, or the science of the interrelations between living organisms and their environment. Because ecology is concerned especially with the biology of *groups* of organisms and with *functional* processes on the lands, in the oceans, and in fresh waters, it is more in keeping with the modern emphasis to define ecology as the study of the structure and function of nature, it being understood that mankind is a part of nature.

One of the definitions in Webster's Unabridged Dictionary seems especially appropriate for the closing decades of the 20th century, namely, "*the totality or pattern of relations between organisms and their environment.*" In the long run the best definition for a broad subject field is probably the shortest and least technical one, as for example, "environmental biology."

To understand the scope and relevance of ecology, the subject must be considered in relation to other branches of biology and to "*ologies*" in general. In the present age of specialization in human endeavors, the inevitable connections between different fields are often obscured by the large masses of knowledge with the fields (and sometimes also, it must be admitted, by stereotyped college courses). At the other extreme, almost any field of learning may be so broadly defined as to take in an enormous range of subject material. Therefore, recognized "fields" need to have recognized bounds, even if these bounds are somewhat arbitrary and subject to shifting from time to time. A shift in scope has been especially noteworthy in the case of ecology as general public awareness of the subject has increased. To many, "ecology" now stands for "the totality of man and environment." But first let us examine the more traditional academic position of ecology in the family of sciences.

For the moment, let us look at the divisions of biology, "the science of life." Morphology, physiology, genetics, ecology, evolution, molecular biology, and developmental biology are examples of such divisions. We may also divide the subject into what may be called "taxonomic" divisions, which deal with the morphology, physiology, ecology, etc., of specific kinds of organisms. Zoology, botany, and bacteriology, are large divisions of this type, and phycology, protozo-ology, mycology, entomology, ornithology, etc., are divisions dealing with more limited groups of organisms. Thus ecology is a basic division of biology and, as such, as also an integral part of any and all of the taxonomic divisions. Both approaches are profitable. It is often very productive to restrict work to certain taxonomic groups, because different kinds of organisms require different methods of study (one cannot study eagles by the same methods used to study bacteria) and because some groups of organisms are economically or otherwise much more important or interesting to man than others. Ultimately, however, unifying principles must be delimited and tested if the subject field is to qualify as "basic."

Perhaps the best way to delimit modern ecology is to consider it in terms of the concept of *levels of organization* visualized as a sort of "biological spectrum." Community, population, organization, organ cell, and gene are widely used terms for several major biotic levels.

Interaction with the physical environment (energy and matter) at each level producing characteristic functional systems. By a *system* we mean just what Webster's College Dictionary defines as "regularly interacting and interdependent components forming a unified whole." Systems containing living components (biological systems or biosystems) may be conceived at any level in the hierarchy. For example, we might consider not only gene systems, organ systems, and so on, but also host-parasite systems as intermediate levels between population and community.

HISTORICAL REVIEW OF THE ECOSYSTEM CONCEPT

Although the term ecosystem was first proposed by the British ecologist, A.G. Tansley in 1935, the concept is by no means so recent. Allusions to the idea of the unity of organisms and environment (as well as the oneness of man and nature) can be found as far back in written history as one might care to look, and such an idea has been a basic part of many religions (less so in Christian religions as recently pointed out by historian Lynn White, 1967). Anthropologists and geographers have long been concerned with the impact of man on his environment and early debated the question: To what extent has man's continuing trouble with deteriorated environments stemmed from the fact that human culture tends to develop independently of the natural environment? The Vermont prophet George Perkins Marsh wrote a classic treatise on this theme in 1864. He analyzed the causes of the decline of ancient civilizations and forecast a similar doom for modern ones unless "man takes what we would call today an ecosystematic" view of man and nature. In the late 1800s biologists began to write essays on the unity of nature, interesting enough in a parallel manner in German, English, and Russian languages. Thus Karl Möbius in 1877 wrote about the community of organisms in an oyster reef as a "biocoenosis," while in 1887 the American S.A. Forbes wrote his classic essay on "The Lake as a Microcosm." The Russian pioneering ecologist V.V. Dokuchaev (1846–1903) and his disciple G.F. Morozov (who specialized in forest ecology) placed great emphasis on the concept of "biocoenosis," a term later expanded to geobiocoenosis (or biogeocoenosis) (see Sukachev, 1944), which can be considered a synonym of the word "ecosystem."

No one has expressed the relevance of the ecosystem concept to man better than Aldo Leopold in his essays on the land ethic. In 1933 he wrote: "Christianity tries to integrate the individual to society, Democracy to integrate social organization to the individuals. There is yet no ethic dealing with man's relation to the land" … which is "still strictly economic entailing privileges but not obligations." Thus, man is continually striving, with but partial success so far, to establish ethical relationships between man and man, man and government, and, now, man and environment. Without the latter what little progress has been made with the other two ethics will surely be lost. In the context of the 1970 scene Garrett Harden (1968) says it in another way when he points out that technology alone will not solve the population and pollution dilemmas; ethical and legal constraints are also necessary. Environmental science is now being called upon to help determine a realistic level of human population density and rate of use of resources and power that are optimum in terms of the quality of human life, in order that "societal feedback" can be applied before there are serious overshoots. This requires diligent study of ecosystems, and, ultimately, a judgment on the carrying capacity of the biosphere. If studies of natural populations have any bearing on the problem, we can be quite certain that the optimum density in terms of the individual's options for liberty and the pursuit of happiness is something less than the maximum number that can be sustained at a subsistence level, as so many domestic "animals" in a polluted feed lot!

My advanced ecology class recently attempted to determine what might be the optimum population for the State of Georgia on the assumption that someday the state would have to have a balanced resource input-output (i.e., live within its own resources). On the basis of a per capita approach to land use the tentative conclusion was that a density of one person per five acres (2 hectares) represented the upper limit for an optimum population size when the space requirements for quality (i.e., high protein) food production, domestic animals, outdoor recreation, waste treatment and pollution-free living space were all fully considered. Anything less than five acres of live support and resource space per capita, it was concluded, would result in a reduction in the individual person's options for freedom and the pursuit of happiness, and, accordingly, a rapid loss in environmental quality. Since the 1970 per capita density of Georgia is 1 in 8 acres and for the United States as a whole, 1 in 10 acres no more than double the present US population could be considered optimum according to this type of analysis. This would mean that we have about 30 years to level off population growth. The study also suggested that permanent zoning of at least one-third of land and freshwater areas (plus all estuarine and marine zones) as "open space" in urbanizing areas would go a long way toward preventing overpopulation, overdevelopment and social decay that is now so evident in many parts of the world today. The results of this preliminary study have been published (E.P. Odum, 1970b). See also the provocative symposium on *The Optimum Population for Britain* edited by Taylor (1970).

THE TWO APPROACHES TO ECOSYSTEM STUDY

G. Evelyn Hutchinson in his 1964 essay, "The Lacustrine Microcosm Reconsidered," contrasts the two long-standing ways ecologists attempt to study lakes or other large ecosystems of the real world. Hutchinson cites E.A. Birge's (1915) work on heat budgets of lakes as pioneering the *holological* (from holos = whole) or *holistic* approach in which the whole ecosystem is treated as a "black box" (i.e., a unit whose function may be evaluated without specifying the internal contents) with emphasis on inputs and outputs, and he contrasts this with the merological (from meros = part) approach of Forbes in which "we discourse on parts of the system and try to build up the whole from them." Each procedure has obvious advantages and disadvantages and each leads to different kinds of application in terms of solving problems. Unfortunately, there is something of a "credibility gap" between the two approaches. As would be expected, the merological approach has dominated the thinking of the biologist-ecologist who is species-oriented, while the physicist-ecologist and engineer prefer the "black box" approach. Most of all, man's environmental crisis has speeded up the application of systems analysis to ecology. The formalized, or mathematical model, approach to populations, communities, and ecosystems has come to be known as *systems ecology* which is rapidly becoming a major science in its own right for two reasons: (1) extremely powerful new formal tools are now available in terms of mathematical theory, cybernetics, electronic data processing, etc. (2) Formal simplication of complex ecosystems provides the best hope for solutions of man's environmental problems that can no longer be trusted to trial-and-error, or one-problem one-solution procedures that have been chiefly relied on in the past.

Again we see the contrast between merological and holological approaches in that there are systems ecologists who start at the population or other component level and "model up," and those who start with the whole and "model down." The same dichotomy is evident in the very rewarding studies of experimental laboratory ecosystems. One class of microecosystems can be called "derived" systems because they are established by multiple seeding from nature in contrast to "defined" microcosms which are built up from previously isolated pure cultures. Theoretically, at least, the approaches are applicable to efforts to devise life support systems for space travel. In fact, one of the best ways to visualize the ecosystems concept for students and laymen is to consider space travel, because when man leaves the biosphere he must take with him a sharply delimited enclosed environment that will supply all vital needs with sun energy as the only usable input from the surrounding very hostile space environment. For journeys of a few weeks (such as to the moon and back), man does not need a regenerative ecosystem, since sufficient oxygen and food can be stored while CO_2 and other waste products can be fixed or detoxified for short periods of time. For long journeys man must engineer himself into a complete ecosystem that includes the means of recycling materials and balancing production, consumption and decomposition by biotic components or their mechanical substitutes. In a very real sense the problems of man's survival in an artificial space craft are the same as the problems involved in his continued survival on the earth space ship. For example, detection and control of air, and water pollution, adequate quantity and nutritional quality of food, what to do with accumulated toxic wastes and garbage, and the social problems created by reduced living space are common concerns of cities and spacecrafts. For this reason the ecologist would urge that national and international space programs now turn their attention to the study and monitoring of our spaceship earth. As was the case with Apollo 13, survival becomes the mission when the limits of carrying capacity are approached.

THE COMPONENTS OF THE ECOSYSTEM

From the standpoint of trophic energy an ecosystem has two components which are usually partially separated in space and time, namely, an *autotrophic component* (autotrophic = self nourishing) in which fixation of light energy, use of simple inorganic substances, and the buildup of complex substances predominate; and secondly, a *heterotrophic component* (heterotrophic = other-nourshing) in which utilization, rearrangement and decomposition of complex materials predominate. As viewed from the side (cross section) ecosystems consist of an upper "green belt" which receives incoming solar energy and overlaps, or interdigitates, with a lower "brown belt" where organic matter accumulates and decomposes in soils and sediments.

It is convenient for the purposes of first order analysis and modelling to recognize six structural components and six processes as comprising the ecosystem as follows:

A. Components

1) *Inorganic substances* (C, N, CO_2, H_2O, etc.) involved in material cycles.
2) *Organic compounds* (proteins, carbohydrates, lipids, humic substances, etc.) that link biotic and abiotic.
3) *Climate regime* (temperature, rainfall, etc.).
4) *Autotrophs* or *producers,* largely green plants able to manufacture food from simple substances.
5) *Phagotrophs* (phago = to eat) or *macro-consumers,* heterotrophic organisms, largely animals which ingest other organisms or particulate organic matter.
6) *Saprotrophs* (sapro = to decompose) or *micro-consumers* (also called osmotrophs), heterotrophic organisms, chiefly bacterial, fungi, and some protozoa that break down complex compounds, absorb some of the decomposition products and release inorganic substances usable by the autotrophs together with organic residues which may provide energy sources or which may be inhibitory, stimulatory or regulatory to other biotic components of the ecosystem.

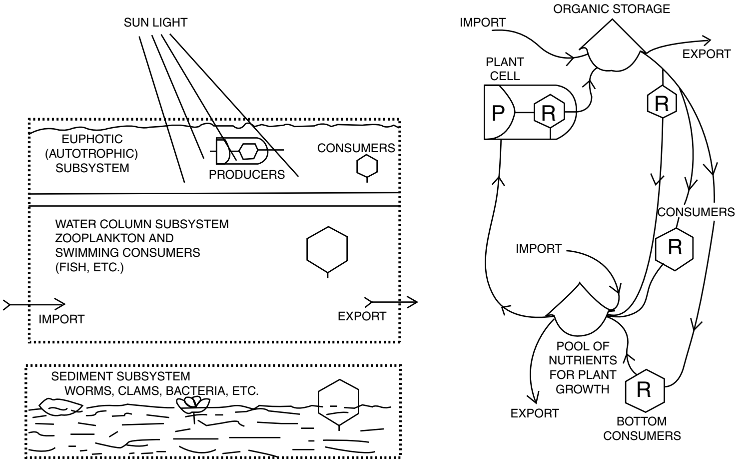
(Another useful division for heterotrophs: *biophages* = organisms that feed on other living organisms; *saprophages* = organisms that feed on dead organic matter.)

B. Processes

1) Energy flow circuits.
2) Foods chains (trophic relationships).
3) Diversity patterns in time and space.
4) Nutrient (biogeochemical) cycles.
5) Development and Evolution.
6) Control (cybernetics).

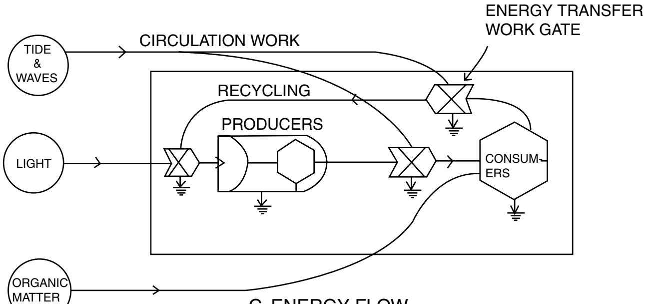
Subdivision of the ecosystem into these six "components" and six "processes," as with most classification, is arbitrary but convenient, since the former emphasize structure and the latter function. From the holistic viewpoint, of course, components are operationally inseparable. While different methods are often required to delineate structure on the one hand and to measure rates of function on the other, the ultimate goal of study at any level or organization is to understand the relationship between structure and function. It is not feasible to go into any detailed discussion of these component-processes in this brief introduction (one can refer to textbooks and review papers), but we can list a few key principles that are especially relevant to human ecology. Figure 1 is a schematic diagram that may be useful in picturing the basic arrangement and functional linkage of ecosystem components.

1) The living (items A4–6 above) and non-living (items A1–3) parts of ecosystems are so interwoven into the fabric of nature that it is difficult to separate them, hence operational classifications (B1–6) do not make a sharp distinction between biotic and abiotic. Elements and compounds are in a constant state of flux between living and non-living states. There are very few substances that are confined to one or the other state. Exceptions may be ATP which is found only inside living cells, and humic substances (resistant end-products of decomposition) which are not found inside cells yet are characteristic of all ecosystems.
2) The time-space separation of autotrophic and heterotrophic activity leads to a convenient classification of energy circuits into (1) a *grazing food chain* (where term grazing refers to direct consumption of living plants or plant parts) and (2) an *organic* detritus (from *deterere* = to wear away) food chain which involves the accumulation and decomposition of dead materials. To build up a stable biomass structure there must be negative feedback control of grazing—a need too often neglected in man's domesticated ecosystems.
3) As in well known available energy declines with each step in the food chain (so a system can support more herbivores than carnivores; if man wants to keep his meat-eating option open there will have to be fewer people supported by a given food base). On the other hand, materials often become concentrated with each step in the food chains. Failure to anticipate possible "biological magnification" of pollutants, such as DDT or long-lived radionuclides, is causing serious problems in man's environment.
4) It is becoming increasingly evident that high biological productivity (in terms of calories per unit area) in both natural and agricultural ecosystems is almost always achieved with the aid of *energy subsidies* from outside the system that reduces the cost of maintenance (thus diverting more energy to production). Energy subsidies take the form of wind and rain in a rain forest, tidal energy in an estuary (see Figure 1), or fuel, animal or human work energy used in the cultivation of a crop. In comparing productivity of different systems it is important to consider the *complete budget*—not just light input.
5) Likewise it is increasingly evident that *both harvest and pollution are stresses* which reduce the



A. VERTICAL ZONES

B. MINERAL CYCLE



C. ENERGY FLOW

FIGURE 1 Three aspects of the structure and function of ecosystems as illustrated by an estuarine system. A. Vertical zonation with photosynthetic production above (autotrophic stratum) and most of the respiration and decomposition below (heterotrophic stratum). B. Material cycle with circulation of plant nutrients upward and organic matter food downward. C. Energy flow circuit diagram showing three sources of energy input into the system. The bullet-shaped modules represent producers with their double metabolism, that is P (production) and R (respiration). The hexagons are populations of consumers which have storage, self maintenance and reproduction. The storage bins represent nutrient pools in and out of which move nitrogen, phosphorous and other vital substances. In diagrams B and C the lines represent the "invisible wires of nature" that link the components into a functional network. In diagram C the "ground" symbols (i.e., arrow into the heat sink) indicate where energy is dispersed and no longer available in the food chain. The circles represent energy inputs. The work gate symbols (large X) indicate where a flow of work energy along one pathway assists a second flow to pass over energy barriers. Note that some of the lines of flow loop back from "downstream" energy sources to "upstream" inflows serving various roles there including control functions (saprotrophs controlling photosynthesis by controlling the rate of mineral regeneration, for example). The diagram (C) also shows how auxilliary energy of the tide (energy subsidy) assists in (1) recycling of nutrients from consumer to producer and (2) speeding up the movement of plant food to the consumer. Reducing tidal flow by dyking the estuary will reduce the productivity just as surely as cutting out some of the light. Stress such as pollution or harvest can be shown in such circuit models by adding circles enclosing negative signs linked with appropriate heat sinks to show where energy is diverted away from the ecosystem. Both subsidies (+) and stress (−) can be quantitated in terms of Calories added or diverted per unit of time and space. (From E.P. Odum, 1971 after H.T. Odum, and B.J. Copeland, 1969.)

energy available for self-maintenance. Man must be aware that he will have to pay the costs of added anti-thermal maintenance, or "disorder pumpout," as H. T. Odum, 1971 (Chapter 5) calls it. It is dangerous strategy to try to force too much productivity, or yield, from the landscape (as is being attempted in the so-called "green revolution") because very serious "ecological back-lashes" can result from the following (1) pollution caused by heavy use of fertilizers and insecticides and the consumption of fossil fuels, (2) unstable or oscillating conditions created by one-crop systems, (3) vulnerability of plants to disease because their self-protection mechanisms have been "selected out," in favor of yield, and (4) social disorder created by rapid shift of rural people to cities that

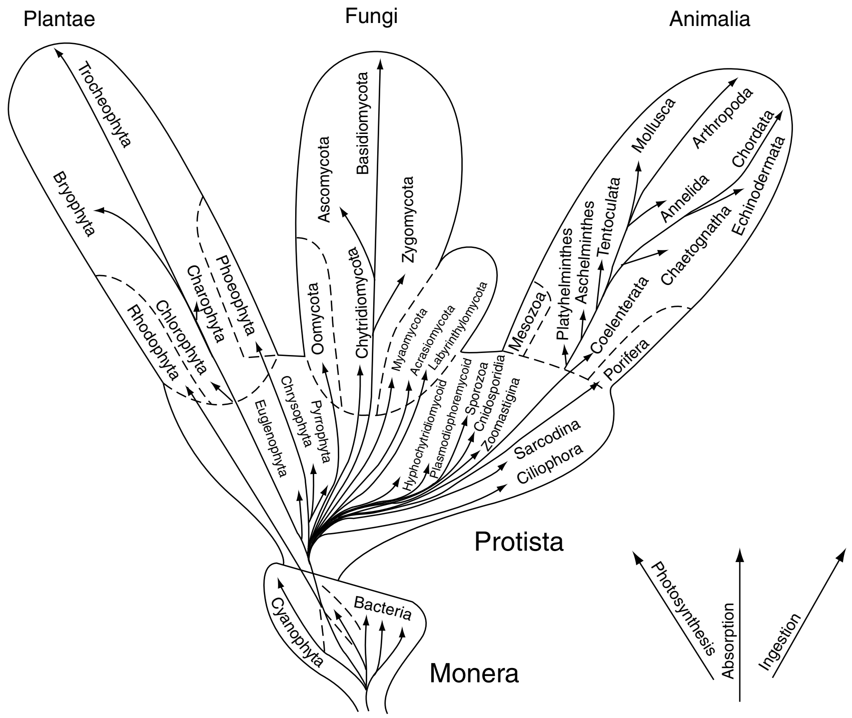


FIGURE 2 A five-kingdom system based on three levels of organization—the procaryotic (kingdom Monera), apearyotic unicelluar (kingdom Protista), and eucaryotic multicelluar and multinucleate. On each level there is divergence in relation to three principal modes of nutrition—the photosynthetic, absorptive, and ingestive. Many ecology and microbiology texts list four kingdoms by combining the "lower Protista" (i.e., Monera) with the higher "Protista" to form "Protista." Evolutionary relations are much simplified, particularly in the Protista. Only major animal phyla are entered, and phyla of the bacteria are omitted. The Coelenterata comprise the Cnidaria and Ctenophora; the Tentaculata comprise the Bryozoa, Brachiopoda, and Phoronida, and in some treatments the Batoprocta. (From Whittaker, 1969.)

are not prepared to house or employ them (the tragedy here is that industrialized agriculture *can* result in increased food per acre but it *can* also widen the gap between rich and poor so that there are increasing numbers of people unable to buy the food!).

6) While we generally think of production and decomposition as being balanced on the biosphere as a whole, the truth is that this balance has never been exact but has fluctuated from time to time in geological history. Through the long haul of evolutional history production has slightly exceeded decomposition so that a highly oxygenic atmosphere has replaced the original reducing atmosphere of the earth. Man, of course, is tending to reverse this trend by increasing, decomposition (burning of fuels, etc.) at the expense of production. The most immediate problem is created by the increase in atmospheric CO_2 since relatively small changes in concentration can have large effects on the heat budget of the earth.
7) Ecological studies indicate that diversity is directly correlated with stability and perhaps inversely correlated with productivity, at least in many situations. It could well be that the preservation of diversity in the ecosystem is important for man since variety may be a necessity, not just the spice of life!
8) At the population level it is now clear that the growth form of the human population will not conform to the simple sigmoid or logistic model since there will always be a long time lag in the effects of crowding, pollution and overexploitation of resources. Growth will not "automatically" level off as do populations of yeasts in a confined vessel where individuals are *immediately* affected by their waste products. Instead, the human population will clearly overshoot some vital resource, unless man can "anticipate" the effects of overpopulation and reduce growth rates *before* the deleterious effects of crowding are actually felt. Intelligent reasoning behavior seems now to be

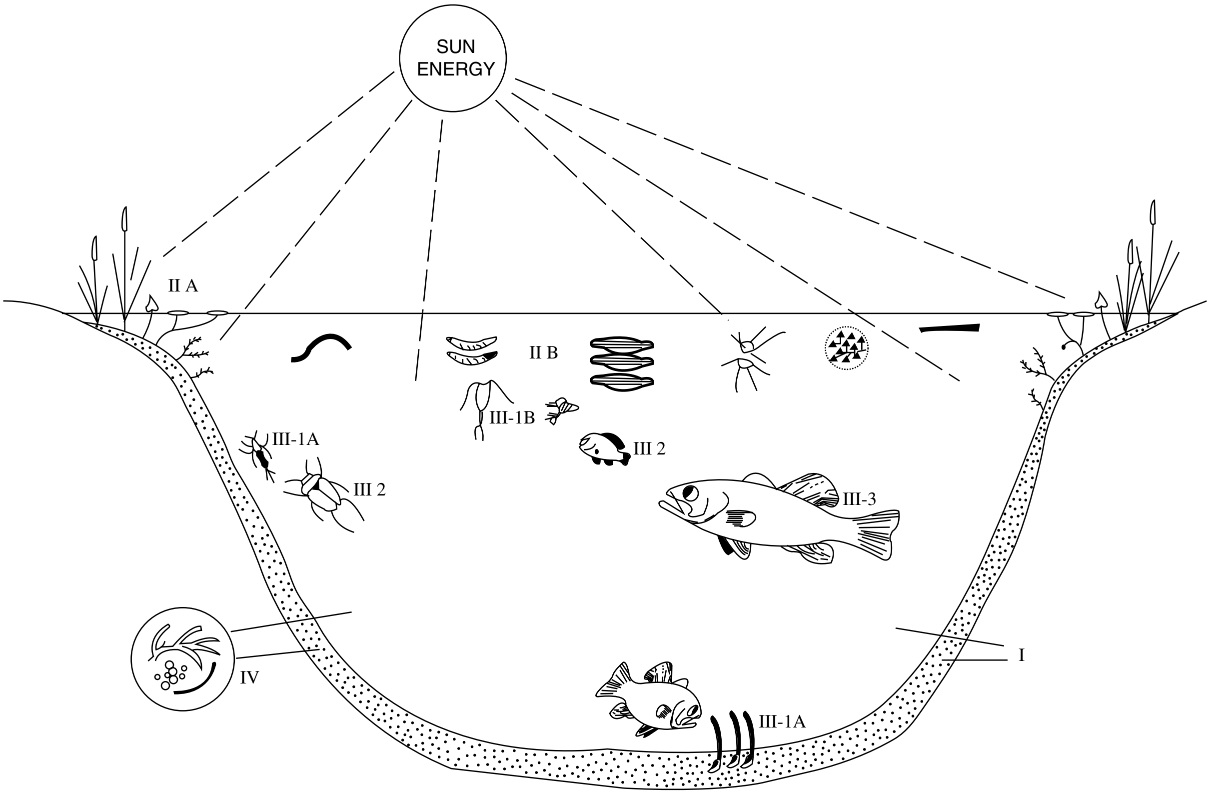


FIGURE 3 Diagram of the pond ecosystem. Basic units are as follows? I, abiotic substances—basic inorganic and organic compounds; IIA, producers—rooted vegetation; IIB, producers-phytoplankton; III-1A, primary consumers (herbivores)—bottom forms; III-1B, primary consumers (herbivores)—zooplankton; III-2, secondary consumers (carnivores); III-3, tertiary consumers (secondary carnivores); IV, saprotrophs—bacteria and fungi of decay. The metabolism of the system runs on sun energy, while the rate of metabolism and relative stability of the pond depend on the rate of inflow of materials from rain and from the drainage basin in which the pond is located. (From Odum, 1971.)

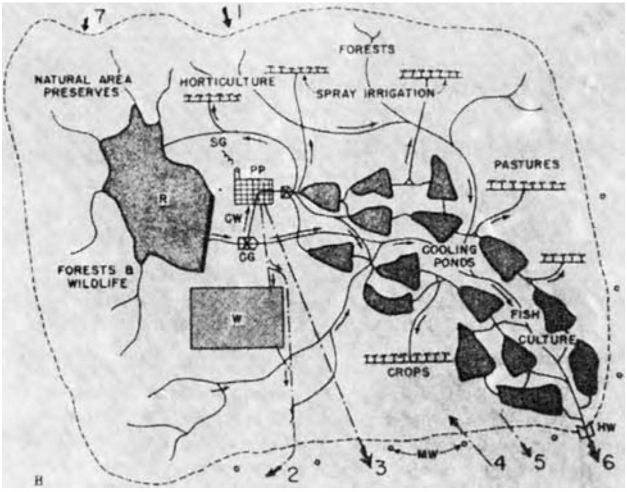


FIGURE 4 Schematic design for a waste management park for the atomic power plant (PP) of the future which is located in a natural watershed basic (outlined by the dashed lines). Waste heat (i.e., thermal pollution) in the reactor cooling water (CW) flowing from a large storage reservoir (R) is completely dissipated by evaporative cooling from the network of shallow ponds and spray irrigation systems. The warm ponds may be used for fish culture, sport fishing, or other recreation purposes. Irrigation of portions of the terrestrial watershed increases the yield of useful forest or agricultural products, while at the same time water recycles through the "living filter" of the land back into stream, ponds, and ground-water. Low-level nuclear wastes and solid wastes are contained within carefully managed land fill area (W); high-level nuclear wastes in spent fuel elements are exported to a special nuclear burial ground located outside of the management part. Stream flow, ground water, and stack gases are continuously monitored by hydrological weirs (HW), monitoring wells (MW), and stack gas control systems (SG) in order to make certain that no air or water pollution leaves the controlled area. The chief inputs and outputs for this environmental system include (see numbered marginal arrows): 1, input of sunlight and rainfall; 2, export of nuclear wastes to burial grounds; 3, electric power to cities, etc.; 4, input of nuclear and other fuels; 5, output of food, fibres, clean air, etc.; 6. downstream flow of clean water for agriculture, industry, and cities; 7, public and professional use for recreation, education, and environmental research. The size of such a complete waste management park will depend on regional climates and topography, and on the amount of electrical or other energy diverted to power cooling, but something on the order of 10,000 acres for a 2500 megawatt power plant would be the minimum needed to insure 100% pollution control and allow for accidents and mechanical malfunctions. However, such waste treatment capacity could also support a certain amount of light industry within the park. Heavy industry should be located within its own waste management park. (From Odum, 1971).

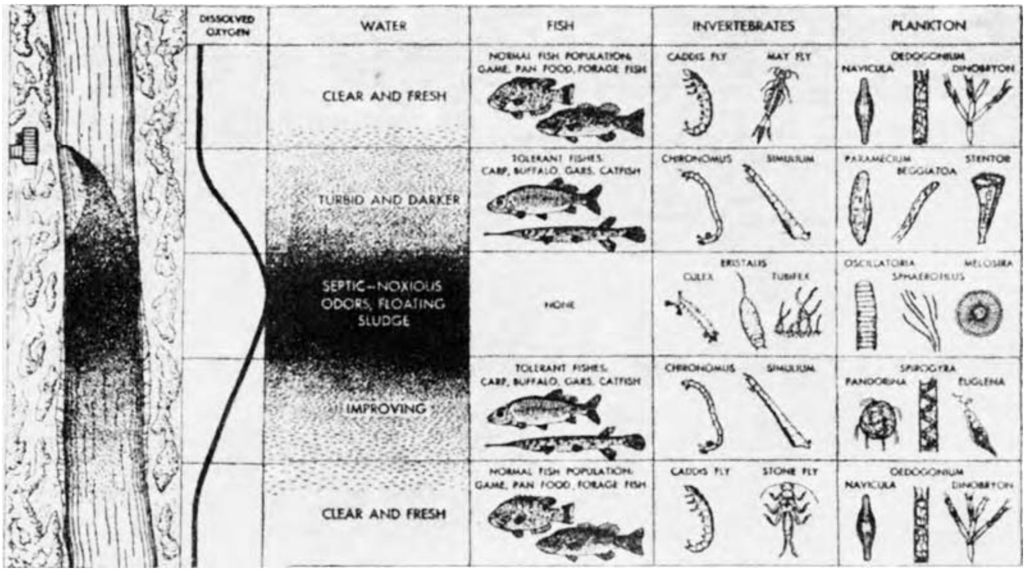


FIGURE 5 Pollution of a stream with untreated sewage and the subsequent recovery as reflected in changes in the biotic community. As the oxygen dissolved in the water decrease (curve to the left), fishes disappear and only organisms able to obtain oxygen from the surface (as in Culex mosquito larvae) or those which are tolerant of low oxygen concentration are found in zone of maximum organic decomposition. When bacteria have reduced all of the discharged material the stream returns to normal. (After Eliassen, *Scientific American,* Vol. 186, No. 3, March, 1952.)

the only means to accomplish this, as emphasized at the beginning of the article.

9) Some of the most important "breakthroughs" in ecology are in the area of biogeochemical cycling. Since "recycle" of water and minerals must become a major goal of human society the recycle pathways in nature are of great interest; there seem to be at least four major ones which vary in importance in different kinds of ecosystems: (1) recycle via microbial decomposition of detritus, (2) recycle via animal excretion, (3) direct recycle from plant back to plant via symbiotic microorganism such as mycorrhizae associated with roots, and (4) autolysis, on chemical recycle, with no organism involved. Pathway 3 seems to be especially important in the humid tropics which suggests tropical agriculture might be redesigned to include plant foods with mycorrhizae.
10) The principles inherent in limiting factor analysis and in human ecology can be combined to formulate the following tentative overview: In an industrialized society energy (power, food) is not likely to be limiting, *but the pollution consequences of the use of energy and exploitation of resources is limiting.* Thus, pollution can be considered the limiting factor for industrialized man—which may be fortunate since pollution is so "visible" that it can force us to use that reasoning power which is supposed to be a special attribute of man.

ECOSYSTEM DEVELOPMENT

Principles having to do with the development of ecosystems, that is ecological succession, are among the most relevant in view of man's present situation. I have recently reviewed this subject (Odum, 1970); accordingly a brief summary will suffice for this piece.

In broad view ecosystems develop through a rapid growth stage that leads to some kind of maturity or steady state (climax), usually an oscillating steady state. The early successional growth stage is characterized by a high Production/respiration (P/R) ratio, high yields (net production), short food chains, low diversity, small size of organism, open nutrient cycles and a lack of stability. In contrast, mature stages have a high biomass/respiration (B/R) ratio, food web, low net production, high diversity and stability. In other words, major energy flow shifts from production to maintenance (respiration).

The general relevance of the development sequence to land use planning can be emphasized by the following "mini-model" that contrasts in very general terms young and mature ecosystems:

Young	Mature
Production	Protection
Growth	Stability
Quantity	Quality

It is mathematically impossible to obtain a maximum for more than one thing at a time, so one can not have both extremes at the same time and place. Since all six characteristics are desirable in the aggregate, two possible solutions to the dilemma suggest themselves. We can compromise so as to provide moderate quality and moderate yield on all the landscape, or we can plan to compartmentalize the landscape so as to simultaneously maintain highly productive and predominantly protective types as separate units subjected to different management strategies. If ecosystem development theory is valid and applicable to land-use planning (total zoning), then the so-called multiple-use strategy, about which we hear so much will work only through one or both of these cases, because in most cases, projected multiple uses conflict with one another. Some examples and suggestions for implementing compartmental plans are considered in the above mentioned paper.

REFERENCES

Birge, E.A. (1915), The heat budgets of American and European lakes, *Trans. Wis. Acad. Arts. Lett.* **18,** 166–213.

Cairns, J. *Rehabilitating damaged ecosystems,* CRC Press, Boca Raton, 1988.

Dokuchaev, V.V., *Teaching about the zones of nature,* (in Russian). Reprinted 1948, Moscow.

Forbes, S.A. (1887), The lake as a microcosm, *Bull. Sc. H. Peoria,* reprinted in: *Ill. Nat. Hist. Surv. Bull.* **15,** 537–550 (1925).

Hardin, G. (1968), The Tragedy of the commons, *Science* **162,** 1243–1248.

Hutchinson, G.E. (1964), The lacustrine microcosm reconsidered, *Amer. Sci.* **52,** 331–341.

Jerrers, J.N.R., *Practitioners handbook in the modelling of dynamic changes in ecosystems,* Wiley, Chichester, 1988.

Jorgensøn, S.E., *Modelling the fate and effect of toxic substances in the environment,* Elsevier, Amsterdam, 1984.

Leopold, Aldo (1933), The conservation ethic, *Jour. Forestry* **31,** 634–643.

MacArthur, R.H., *Geographical ecology, Patterns in the distribution of species,* Princeton University Press, 1984.

Marsh, G.P. (1864), *Man and Nature:* or *Physical Geography as Modifed by Human Action,* reprinted 1965 by Harvard Univ. Press, Cambridge.

Mellos, K., *Perspectives on ecology,* Macmillan, Basingstoke, 1989.

Möbius, K. (1877), Die Auster und die Austernwirtschart. Berlin, Transl. 1880. Rept. U.S. Fish Comm., 1880, 683 to 751.

Odum, H.T. (1970), *Environment, Power and Society,* John Wiley and Sons, N.Y., p. 331.

Odum, H.T. (1967), Biological circuits and the marine systems of Texas, in *Pollution and Marine Ecology* (Olson and Burgess, Eds.), John Wiley and Sons, N.Y., pp. 99–157.

Odum, H.T., B.J. Copeland, and E.A. McMahan (1969), Coastal Ecological Systems of the United States. A report to the Federal Water Pollution Control Administration, Marine Sci. Inst., Univ. Texas, Mimeograph.

Odum, Eugene P. (1969), The attitude lag, *BioScience* **19,** 403.

Odum, Eugene P. (1970a), The strategy of ecosystem development, *Science,* **164,** 262–270.

Odum, E.P. (1970b), Optimum population and environment: A Georgia microcosm, *Current History* **58,** 355–359.

Odum, E.P. (1970c), The attitude revolution, in *The Crisis of Survival,* Scott, Foresman Co., Glenview, Ill., pp. 9–15.

Odum, E.P. (1971), *Fundamentals of Ecology,* 3rd Edition, W.B. Saunders, Philadelphia.

Polunin, N.V., *Ecosystem theory and practice,* Wiley, Chichester, 1986.

Sukachev, V.N. (1944), On principle of genetic classification in bioeoenology, translated and condensed by F. Raney and R. Daubenmire. *Ecol.* **39,** 364–367. (See also Silva Fennica, **105,** 94, (1960)).

Tansley, A.G. (1935), The use and abuse of vegetational concepts and terms, *Ecol.* **16,** 284–307.

Taylor, L.R. (Ed.), *The Optimum Population for Britain,* Academic Press, N.Y., p. 182, 1970.

Walter, H. and S.W. Brecke, Ecological systems of the biosphere; Ecological principles in global perspective (Ökologie der Erde), Springer Verlag, Heidelberg, 1985.

White, L. (1967), The historical roots of our ecological crisis, *Science* **155,** 1203–1207.

EUGENE P. ODUM
University of Georgia

EFFECTS OF AIR POLLUTANTS: see AIR POLLUTANT EFFECTS

EFFECTS OF CHEMICALS

Almost daily one reads reports in the press about new dangers found or suspected in foods, medicines and other products. Public awareness of possible perils in everyday products has increased greatly. Smoking, asbestos, toxic and hazardous wastes (once called industrial wastes) are widely discussed.

Problems resulting from technological advances are being uncovered at an alarming rate. While it is possible to predict, with some accuracy, benefits to be expected from improvements and new approaches in applied science, seldom have serious efforts been made to determine adverse effects resulting from these. The general public is increasingly aware of potential hazards from new products and processes and the trend in industrialized countries is toward greater control and regulation of hazardous undertakings. The question as to the more effective approach, gentle persuasion to gain voluntary compliance or strong legislation for strict regulation, has not been satisfactorily answered. Probably there never will be agreement.

In the United States, the Office of Technology Assessment was dissolved. The stated reason was that this Office did long-term studies which were not immediately useful to a legislator. In remarks at the Conference on Technical Expertise and Public Decisions at Princeton University, the then Chairman of the House Science Committee said that it was more desirable to depend on the views of those most interested in the topic—the lobbyists.

The relationship among the air, water, and soil is not a static one. Effects of a pollutant may be demonstrated progressively in the various compartments into which the environment is divided. A substance may be initially present, without apparent ill effects, in one compartment. Later, the same substance, or a demonstrable derivative, may appear in a different part of the environment in a most undesirable way. An excellent example is the high mercury level found in fish. Initially, it was felt that disposal into the marine environment would conveniently remove a bothersome waste. However, by previously unsuspected paths the metal found its way into the systems of game and food fish. Food, thus, has become a secondary distributor of pollutant material.

Chemical pollutants may be divided into four categories:

1) Natural chemicals in excess.
2) Naturally occurring toxins.
3) Mixtures of air and water pollutants which produce adverse effects but with only partially defined or undefined components.
4) Synthetic chemicals.

The first group includes chemicals such as nitrates and nitrites. Nitrates, for example, can cause methemoglobinemia in infants by reduction of the capacity of the blood to carry oxygen. An intermediate reduction is involved. A famous case in New York involved accidental introduction of sodium nitrite into oatmeal and the resulting problems of "Eleven Blue Men." The victims, all heavy users of alcohol, apparently tried instinctively to compensate for low salt in the body. Through an accident, sodium nitrite instead of sodium chloride, was placed in salt shakers in a public eating place. The eleven victims all became cyanotic, with the characteristic blue color giving the name to the episode. Nitrites may also react with secondary amines to form nitrosamines, some of the which have been shown to be carcinogenic, teratogenic, and mutagenic, all in microgram doses. Oxides of nitrogen, thought to be significant in smog production, can also form nitrosamines.

The second group, natural fungal and plant toxins, usually are introduced into the human ecosystem through accident or carelessness. Conditions of harvesting, storage, and processing have been shown to be of possible importance.

An excellent example of the third group has already been mentioned. Mercury was discharged to the receiving water as an apparent ultimate solution to a waste disposal problem. The two components of the system, mercury and water, were assumed to be non-reactive. Unfortunately, the two-component system was, in fact, a multicomponent system and no endeavor was made to determine the complete mechanism. The problem was further complicated by the introduction of edible fish into the chain.

Minimata Disease is named for the city in which it occurred. Inorganic mercury was discharged in the effluent of a local industrial plant. Through action of marine organisms, the mercury was converted to lipid-soluble methyl mercury, which was taken up in the food chain to fish, the staple of the local diet. 43 deaths and about 700 serious illnesses were acknowledged by local authorities in the 1950s. Some unofficial estimates have put the death toll as high as 800. In 1989 two former company officials were given prison sentences for the 1950s pollution. Lawsuits resulting from the pollution were finally settled in 1996.

In Japan in 1968 about 1800 people developed a malady similar to chloracne after ingesting rice oil contaminated with a chlorinated biphenyl. Known as Yusho or Rice Oil Disease, the rice oil used in cooking had been contaminated by a PCB which had leaked from a faulty heat exchanger. In 1978 an outbreak similar to Yusho Disease occurred in Taiwan and is known as the Yu-Cheng Incident. The cause was the same, a faulty heat exchanger.

The last group, synthetic chemicals, includes pesticides and fertilizers used in agriculture, food additives, compounds containing heavy metals, plasticizers, fuel additives, household chemicals, industrial chemicals, therapeutic and prophylactic drugs, and drugs of abuse. Food additives may be intentional or accidental. Anti-oxidants and dyes are added routinely to many foods. However, almost any of the aforementioned may be accidentally introduced into food, often with most unpleasant results. There is much controversy concerning synthetic agriculture chemicals. Advantages and disadvantages are numerous and no definite decision has been reached concerning continued use of many substances. NTA (Nitrilotriacetic Acid) as a substitute for phosphates in detergents, is another excellent example of conflicting use. The problem is far from restricted to simple direct physiological effects.

Food additives may be classified as to function. They find use as coloring material, flavor enhancers, shelf life extenders, and in protection of food nutritional value. While valuable, color additives are not always essential. However, many of the foods now enjoyed by modern western society would not be possible, in their present form, without food additives.

It is estimated by the World Health Organization that about one fifth of the food produced in the world is lost by spoilage. Preservation, or retardation of spoilage, can be accomplished by addition of chemical preservatives, or by physical means such as freezing, drying, souring, fermenting, curing or ionizing radiation. There is some concern that irradiation of food may have adverse effects and leave unwanted residues. Little is known about possible chemical chain reactions.

Food additives have been classified by Kermode into five broad groups:

1) Flavors
2) Colors
3) Preservatives
4) Texture agents
5) Miscellaneous

In Table 1 are displayed food additives declared by the United States Food and Drug Administration to be "generally recognized as safe." Not included in Table 1 is a large group of natural flavors and oils. To be on the list an additive must have been in use before 1958 and meet specifications for safety. Materials introduced after 1958 must be tested individually in order to quality for inclusion of the FDA list. Examples of materials formerly listed but now removed are cyclamate sweetners and saccharin. A ban on cyclamates ordered by the US government after tests revealed development of bladder cancer in laboratory rats fed on a diet containing cyclamates. Further testing, after the ban, found cancer development in the same test species at dosage rates one sixth as large as those which brought about the ban.

TABLE 1

Additives listed by the U.S. Food and Drug Administration as "generally recognized as safe." (Courtesy, *Scientific American*)

ANTICAKING AGENTS
Aluminum calcium silicate
Calcium silicate
Magnesium silicate
Sodium aluminosilicate
Sodium calcium aluminosilicate
Tricalcium silicate
CHEMICAL PRESERVATIVES
Ascorbic acid
Ascorbyl palmitate
Benzoic acid
Butylated hydroxyanisole
Butylated hydroxytoluene
Calcium ascorbate
Calcium propionate
Calcium sorbate
Caprylic acid
Dilauryl thiodipropionate
Erythorbic acid
Gum guaiac
Methylparaben
Potassium bisulfite
Potassium metabisulfite
Potassium sorbate
Propionic acid
Propyl gallate
Propylparaben
Sodium ascorbate
Sodium benzoate
Sodium bisulfite
Sodium metabisulfite
Sodium proponate
Sodium sorbate
Sodium sulfite
Sorbic acid
Stannous chloride
Sulfur dioxide

TABLE 1*(continued)*

Additives listed by the U.S. Food and Drug Administration as “generally recognized as safe.” (Courtesy, *Scientific American*)

Thiodipropionic acid
Tocopherols
EMULSIFYING AGENTS
Cholic acid
Desoxycholic acid
Diacetyl tartaric acid esters of mono- and diglycerides
Glycocholic acid
Mono- and diglycerides
Monosodium phosphate
derivatives of above
Propylene glycol
Ox bile extract
Taurocholic acid
NUTRIENTS AND DIETARY SUPPLEMENTS
Alanine
Arginine
Ascorbic acid
Aspartic acid
Biotin
Calcium carbonate
Calcium citrate
Calcium glycerophosphate
Calcium oxide
Calcium pantothenate
Calcium phosphate
Calcium pyrophosphate
Calcium sulfate
Carotene
Choline bitartrate
Choline chloride
Chopper gluconate
Cuprous iodide
Cysteine
Cystine
Ferric phosphate
Ferric pyrophosphate
Ferric sodium pyrophosphate
Ferrous gluconate
Ferrous lactate
Ferrous sulfate
Glycine
Histidine
Inositol
Iron, reduced
Isoleucine
Leucine
Linoleic acid
Lysine
Magnesium oxide
Magnesium phosphate
Magnesium sulfate
Manganese chloride
Manganese citrate
Manganese gluconate
Manganese glycerophosphate
Manganese hypophosphite
Manganese sulfate
Manganous oxide
Mannitol
Methionine
Methionine hydroxy analogue
Niacin
Niacinamide
D-pantothenyl alcohol
Phenylalanine
Potassium chloride
Potassium glycerophosphate
Potassium iodide
Proline
Pyridoxine hydrochloride
Riboflavin
Riboflavin-5-phosphate
Serine
Sodium pantothenate
Sodium phosphate
Sorbitol
Thiamine hydrochloride
Thiamine mononitrate
Threonine
Tocopherols
Tocopherol acetate
Tryptophane
Tyrosine
Valine

(continued)

TABLE 1 *(continued)*

Additives listed by the U.S. Food and Drug Administration as “generally recognized as safe.” (Courtesy, *Scientific American*)

Vitamin A
Vitamin A acetate
Vitamin A palmitate
Vitamin B_{12}
Vitamin D_2
Vitamin D_3
Zinc sulfate
Zinc gluconate
Zinc chloride
Zinc oxide
Zinc stearate
SEQUESTRANTS
Calcium acetate
Calcium chloride
Calcium citrate
Calcium diacetate
Calcium gluconate
Calcium hexametaphosphate
Calcium phosphate monobasic
Calcium phytate
Citric acid
Dipotassium phosphate
Disodium phosphate
Isopropyl citrate
Monoisopropyl citrate
Potassium citrate
Sodium acid phosphate
Sodium citrate
Sodium diacetate
Sodium gluconate
Sodium hexametaphosphate
Sodium metaphosphate
Sodium phosphate
Sodium potassium tartrate
Sodium pyrophosphate
Sodium pyrophosphate, tetra
Sodium thiosulfate
Sodium tripolyphosphate
Stearyl citrate
Tartaric acid
STABILIZERS
Acacia (gum arabic)
Agar-agar
Ammonium alginate
Calcium alginate
Carob bean gum
Chrondrus extract
Ghatti gum
Guar gum
Potassium alginate
Sodium alginate
Sterculoia (or karaya) gum
Tragacanth
MISCELLANEOUS ADDITIVES
Acetic acid
Adipic acid
Aluminum ammonium sulfate
Aluminum potassium sulfate
Aluminum sodium sulfate
Aluminum sulfate
Ammonium bicarbonate
Ammonium carbonate
Ammonium hydroxide
Ammonium phosphate
Ammonium sulfate
Beeswax
Bentonite
Butane
Caffeine
Calcium carbonate
Calcium chloride
Calcium citrate
Calcium gluconate
Calcium hydroxide
Calcium lactate
Calcium oxide
Calcium phosphate
Caramel
Carbon dioxide
Carnauba wax
Citric acid
Dextrans
Ethyl formate
Glutamic acid

(continued)

TABLE 1*(continued)*

Additives listed by the U.S. Food and Drug Administration as “generally recognized as safe.” (Courtesy, *Scientific American*)

Glutamic acid hydrochloride
Glycerin
Glyceryl monostearate
Helium
Hydrochloric acid
Hydrogen peroxide
Lactic acid
Lecithin
Magnesium carbonate
Magnesium hydroxide
Magnesium oxide
Magnesium stearate
Malic acid
Methylcellulose
Monoammonium glutamate
Monopotassium glutamate
Nitrogen
Nitrous oxide
Papain
Phosphoric acid
Potassium acid tartrate
Potassium bicarbonate
Potassium carbonate
Potassium citrate
Potassium hydroxide
Potassium sulfate
Propane
Propylene glycol
Rennet
Silica aerogel
Sodium acetate
Sodium acid pyrophosphate
Sodium aluminum phosphate
Sodium bicarbonate
Sodium carbonate
Sodium citrate
Sodium carboxymethylcellulose
Sodium caseinate
Sodium citrate
Sodium hydroxide
Sodium pectinate
Sodium phosphate
Sodium potassium tartrate
Sodium sesquicarbonate
Sodium tripolyphosphate
Succinic acid
Sulfuric acid
Tartaric acid
Triacetin
Triethyl citrate
SYNTHETIC FLAVORING SUBSTANCES
Acetaldehyde
Acetoin
Aconitic acid
Anethole
Benzaldehyde
N-butyric acid
d- or *l*-carvone
Cinnamaldehyde
Citral
Decanal
Diacetyl
Ethyl acetate
Ethyl butyrate
Ethyl vanillin
Eugenol
Geraniol
Geranyl acetate
Glycerol tributyrate
Limonene
Linalool
Linalyl acetate
1-malic acid
Methyl anthranilate
3-Methyl-3-phenyl glycidic acid ethyl ester
Piperonal
Vanillin

The Delaney Amendment, dealing with food additives, has been in place since 1958. A substance which produced cancer in laboratory animals could not be used as a food additive. It was overhauled significantly in 1996. More foods are now covered. However, states cannot set standards which are more strict than the federal rules. The standard for pesticides in raw and processed food is based on the amount which may be expected to produce cancer in one person in a million.

It is estimated that there are as many as 1400 natural and synthetic flavors available today. Increased use of flavoring has paralleled the introduction of new food processing and distribution techniques.

Flavor enhancers are closely related to flavors. The best known of these is monosodium glutamate (MSG). There is evidence linking excessive intake of MSG to Kwok's disease, sometimes known as "Chinese restaurant syndrome." Symptoms include a tightening of the neck and face muscles, at times accompanied by nausea, headache, and giddiness. Some countries have placed limits on the amount of MSG in foods and require that the presence of this substance be prominently noted.

Color additives are utilized primarily for the purpose of giving an appetizing appearance to food. Synthetic dyestuffs are used much more often than natural materials. It can be expected that the synthetic dyes will be produced in a high state of purity. In general, small amounts of dye will suffice to give the desired food color. A material that gives a color that is durable is needed by the food processor. Regulations governing the colors that may be added to foods vary markedly from country to country and this causes some difficulties for food processors engaged in international operations.

Some commonly used preservatives are benzoic acid, sodium benzoate, sorbic acid, monosodium phosphate, sodium propionate, and sulfur dioxide. Of special interest is the use of sulfur dioxide for inhibition of mold and discoloration in wine. Sulfur dioxide is, of course, a prominent air pollutant. There is anxiety that the acceptable daily intake of sulfur dioxide will be exceeded in countries where wine intake is heavy. Experimental evidence indicates that sulfur dioxide inhibits the growth rate, probably due to destruction of Vitamin B_1 by sulfite.

Antioxidants, added to prevent rancidity in fatty foods, can be put in the classification of preservatives. A second use for antioxidants is developing with the growing practice of display of food in transparent containers and wrapping. Attendant exposure to light causes discoloration, not necessarily detrimental to nutritional value in itself, but detracting from attractiveness. The most commonly used antioxidants are butylated hydroxyanisole, butylated hydroxytoluene, propyl, octyl and dodecyl gallates, and tocopherols. Antioxidant effect of a substance can often be increased by addition of a second material, producing a synergistic reaction and allowing more effective product control.

Texture agents include emulsifiers, stabilizers, and thickeners. These are the largest single class in terms of total quantity consumed in food. Use of these agents has contributed greatly to the development of the new convenience foods.

The miscellaneous group includes acids, alkalies, buffers, neutralizing agents, anti-glazers, release agents, anticaking materials, clarifying agents, and foaming agents. All of these substances are primarily manufacturing and processing aids. Indeed, without many of these materials the range of modern foods would be much more limited.

Testing of food additives as with most other chemicals, is not usually done with human subjects. Exposure studies of both limited duration and long term are made on mice, rats, dogs, and other laboratory animals. When deviation from normal response is noted, the largest dosage that does not produce the change is used as the base for acceptable intake for humans. Normally, this dosage is reduced by a factor of about 100 in most countries. The allowable, or acceptable, dosage is expressed as milligrams of the substance per kilogram of receptor body weight. It must be emphasized that extrapolation from animals to humans is always a difficult undertaking.

International guidelines have been issued by the World Health Organization and the Joint Expert Committee on Food Additives of the Food and Agriculture Organization. The Codex Alimentarius Commission of the Food and Agriculture Organization has published six principles concerning the use of food additives.

1) The use of an additive is justified only when it has the purpose of maintaining a food's nutritional quality, enhancing keeping quality or stability, making the food attractive, providing aid in processing, packing, transporting or storing food or providing essential components for food in special diets. An additive is not justified if the proposed level of use constitutes a hazard to the consumer's health, if the additive causes a substantial reduction in the nutritive value of the food, if it disguises faulty quality or the use of processing and handling techniques which are not allowed, it deceives the customer or if the desired effect can be achieved by other manufacturing processes that are economically and technologically satisfactory.
2) The amount of additive should not exceed the level reasonably required to attain the desired effect under good manufacturing conditions.
3) Additives should conform to an approved standard of purity.
4) All additives in use, or proposed, should be subjected to adequate toxicological evaluation and permitted additives should be kept under observation for deleterious effects.
5) Approval of additives should be limited to specific foods or specific purposes and under specific conditions.
6) Use of additives in foods consumed mainly by special groups within the community should be determined by the food intake of that group.

Toxicity is the general term applied to adverse biological effects in man resulting from pollutants. Effects may range from lessened health levels to death. Chronic toxicity is of major interest here. Included under effects of chornic toxicity are carcinogenicity, teratogenicity, and mutagenicity. It has been suggested that some behavioral disorders are seated in the effect of toxic substances. This area has not been extensively investigated.

One or more of these types of toxicity may be induced by pollutants in the environment. Pollutants may cause adverse effects alone or may interact synergistically with ordinarily harmless substances to give unexpected results. The mechanism of smog formation is an excellent example, although Man is but indirectly affected.

It has been estimated that, over the expected life span of one malformed child, the total costs for care are in the neighborhood of a million dollars. This figure, which does not include loss of earnings, is a high price for society to pay.

Widespread, long-term genetic effects due to environmental pollution cannot be accurately predicted but the cost is certainly great. Such cost estimates have not, until recently, been applied in matching benefits to cost in evaluation of a potentially hazardous substance. Hazards from a particular substance need not be accepted when another substance of equal worth is available. Mandatory testing is usually looked upon as an unnecessarily bothersome expense by producers of synthetic substances. In fairness, however, it must be said that some chemicals, in use for years, have been unexpectedly indicated as potential hazards and the economic loss has been heavy. Cyclamates are a good example. It is of interest to speculate if such a situation could have been avoided. The basic question concerns the adequacy of existing legislation and difficulty of implementing regulations.

It has become increasingly clear that many diseases formerly regarded as spontaneous are caused by environmental pollutants. These diseases include cancer, birth defects, and mutations. The problem is compounded by greatly increased exposure of the population to new synthetic chemicals and their degradation products. Environmental effects of these substances are not usually adequately evaluated.

There is now overwhelming evidence that many human cancers are due to carcinogenic substances in the environment. These are, then, preventable. Studies of epidemiological factors have indicated strongly local environmental factors are of significance in cancer incidence.

There is a demonstrable link between cigarette smoking and lung cancer and other cancers have been shown to be related to smoking. Cigarette, pipe and cigar smoking have become socially unacceptable. Ayers has described the cigarette as a private air pollution source. There is strong evidence that effects of smoking are experienced by persons in the vicinity of the smoker. Environmental tobacco smoke can be defined as a mixture of sidestream smoke from the cigarette and mainstream smoke exhaled by the smoker. Smoking in schools, public buildings and businesses is now widely banned.

In 1604 James I described smoking as “A custom loathsome to the eye, hateful to the nose, harmful to the brain, dangerous to the lungs.” The first major paper suggesting a link between smoking and lung cancer was published in 1939. In 1950 Wynder and Graham reported that, in a group of 650 men with ling cancer, 95% had smoked for at least 25 years. Doll and Bradford, in a 1951 report of interviews with 1357 patients with lung cancer, found that 99.5% were smokers. There was a marked decrease of cigratte smoking after the 1963 Report of the Surgeon General of the US Public Health Service. In the same year, an internal document of a major tobacco company stated that the company was in the business of selling nicotine, an addictive drug effective in release of stress mechanisms. The British government banned television advertising of cigarettes in 1965. The first health warnings appeared on American cigarette packs in 1966 and on British packs in 1971. A 1984 review article in the *Journal of Epidemiology* confirmed the link between smoking and cervical cancer. A 1990 study related lung cancer in nonsmokers to passive childhood smoking. A former Prime Minister accepted a contract in 1992 with a tobacco company as an adviser on strategy for selling cigarettes in Eastern Europe and Developing Nations. In 1993 the US Environmental Protection Agency classified environmental tobacco smoke as a Class A carcinogen. The State of Texas, in a 1996 suit against a cigarette maker, quotes a company executive as saying “We don’t smoke the—, we just sell it. We reserve that for the young, the black, the poor and the stupid.” President Clinton, the first US President to engage in open conflict with the tobacco industry, declared tobacco to be an addictive drug.

While Legal Aid to fund claims against tobacco companies by former smokers was refused in the UK in 1996, lawsuits against tobacco companies in the United States have been filed by individuals and as class actions. 22 states have sued tobacco companies to regain vast sums of public money expended on care for smoking related maladies. Previously confidential documents from company files and information supplied by whisle-blowers have shown that the companies were aware of adverse health effects of smoking as early as the 1930s.

The tobacco industry, faced with decreasing usage at home, resorted to a creative marketing approach. It was decided that a new market must be developed. One large tobacco company targeted black people and young, blue-collar females. There was a strong public outcry when this was reported in the news media and the company withdrew the campaign. It was a public relations disaster.

There is increased effort to market American cigarettes in the Third World. American tobacco products enjoy a good reputation as to taste and are popular overseas. Officials in some developing countries have accused the US of following a double standard with regard to addictive drugs at home while promoting cigarette use abroad.

Cancer of the oral cavity in Asia is linked with chewing of tobacco leaves and betel nuts. High incidence of gastric cancer in Chile, Iceland, and Japan has been associated epidemiologically with a diet high in fish. It has been suggested that nitrosamines, formed by reaction between nitrites and secondary amines of the fish, may be a significant factor. Nitrites are employed as a preservative of fish. Dietary contamination with aflatoxin is thought to be a causitive effect of liver cancer in the Bantu. Aflatoxin is a fungal carcinogen.

Ingestion of azoglucoside carcinogens with Cycad plants is probably the chief cause of liver cancer in Guam. Cancer of the esophagus in Zambians may be related to high nitrosamine content of Kachasu spirits.

There is well documented evidence of occupationally related cancers. These include bladder cancer in the rubber and aniline dye industries, lung cancer in uranium workers, nasal sinus cancer in wood workers, skin cancer in shale oil workers, and lung cancer in asbestos workers.

The question of asbestos assumed great importance in the 1980s. Once widely used in building, it was found to be a serious health hazard. It is now felt that no exposure to airborne particles is safe.

The Environmental Protection Agency estimates that 30,000,000 tons were used between 1900 and 1980. Asbestos abatement has become big business. One estimate places the value as high as $200 billion per year. Asbestos has been used in more than 3000 products over the years. These include duct work, exterior shingles, floor and ceiling tiles, plaster, pipe lagging, cement, drywalls, theater curtains, brake linings, clutch facings and baby powder.

Asbestos occurs naturally as chrysotile, crocidolite, amosite, anthophyllite, actinolite tremolite. The thin, tiny fibers are not dangerous until disturbed. Then particles of a certain size can lodge in the lungs. When the substance can be crumbled under hand pressure—friable—it is considered extremely dangerous. Building renovation and demolition can release particles into the air. Water damage can loosen the binding matrix.

In the 1930s asbestos inhalation was linked to asbestosis and lung cancer. It was later shown that mesothelioma, cancer of the lining of the lung, is caused by asbestos. Gastrointestinal and larynx cancers have also been associated with asbestos. In the 1970s the Department of Health, Education and Welfare estimated that 8 1/2 to 11 million workers have been exposed occupationally in the last 40 years. Asbestos related illnesses can take 20 to 40 years to develop. World War II shipyard workers and prewar insulation workers comprise the majority of afflicted persons. In one study a premature death rate of 48% was found among insulation workers of the 1940s. Custodial personnel are also thought to be at high risk. Studies have found that wives of asbestos workers may well be in danger. It is thought that fibers on the husbands' clothes are the reason.

Asbestos fibers suspended in water are not thought to pose a hazard. However, steam pipe explosions have put fibers into the air, causing whole blocks to be evacuated and necessitating expensive cleanup operations.

Some claims that chrysotile, the asbestos form most commonly used in the US, may pose less danger than other varieties. This position has not yet been validated scientifically. The variety of asbestos versus fiber size as more important in causing disease is the current debate topic. Some European nations and Canada make distinctions among asbestos varieties but, in the US, the EPA and OSHA treat all forms the same for rule making purpose. The UK and Scandinavia follow closely the US approach.

Considerable uncertainty exists in risk assessment for nonoccupational and environmental settings.

Action on asbestos abatement began in the schools. Since the period for appearance of asbestos related diseases can be as long as 20 to 40 years, it was felt that school-age children were at particular risk. In 1982 the EPA Asbestos-in-Schools required all public and private schools to inspect for friable asbestos and report to parents and employees if any were found. The rule did not require abatement. In 1984 the Asbestos Hazard Abatement Act gave funds to assist in abatement. The Asbestos Hazard Emergency Response Act (AHERA) of 1986 established rules and regulations concerning identification, evaluation and control of asbestos containing materials in schools. It further required schools to identify friable and nonfriable asbestos found and submit management plans to state governors. Provisions were included for periodic reinspections, even though inspections were already made as a result of the notification rule of 1982. Of great importance was that portion of the Act dealing with removal situations and certification of workers. In the early days of removal, when guidelines were not yet available, there were too many "rip and skip" operators who performed such work without regard for proper procedures or air quality monitoring. In many cases there was probably more asbestos in the air after the operation than there was before.

It is felt that asbestos abatement in private homes will not be covered by formal regulations. However, it is estimated that about 75,000 commercial and industrial buildings contain friable asbestos.

There are four basic abatement approaches.

1) Removal
2) Encapsulation, in which friable asbestos is bound in a matrix.
3) Isolation of the asbestos containing area
4) Repair

All of these means have a place in dealing with the asbestos problem and it is necessary to seek advice of a qualified professional before undertaking any action. It should be noted that, in the early days of abatement efforts, some education departments allowed only removal. Encapsulation was forbidden. In retrospect, this inflexible position might be criticized.

The National Institute of Building Sciences (NIBS) has taken the position that there is no single correct answer for all situations.

Environmental carcinogens can be classed into two categories, potent and weak. Potent carcinogens, such as nitrosamines and aflatoxins, can induce cancers in animals in very low concentrations. Isolation of these substances in food has given rise to endeavors to relate food distribution patterns with local cancer incidence.

Weak carcinogens, such as atmospheric pollutants, some pesticides, and food additives, have effects much more difficult to evaluate and thus may pose as great a threat as the potent carcinogens because they are less likely to be recognized as a significant epidemiological factor. Often, evaluation must be by indirect means. A causal relationship has been demonstrated to exist between urban air pollution and lung cancer. Lung cancer mortality patterns differ in various sections of the United States and Great Britain. It has been clearly shown that increased mortality due to lung cancer is related to increased urbanization. A survey in the late 1950s found lung cancer rates in the United States to be 39 per one hundred thousand in rural areas and 52 per one hundred thousand in urban areas. Similar data were gathered in Great Britain and the added effect of smoking was evaluated. The conclusion of significance of urban air pollution is inescapable.

The first demonstration that environmental pollutants can cause genetic damage came with the discovery that high energy radiation induces mutations. Later, the development of the nuclear industry focused increased attention on the dangers of unwanted genetic change. Associated with possible genetic damage in this industry was the danger of radiation induced carcinoma. Accordingly, safeguards were developed and limits were set on radiation levels. These were functions of exposure time. Limits were changed periodically in the light of fresh evidence. It is of interest to note that changes were usually downward. Radiation from X-Ray machines and fluoroscopes was treated in the same way. A widely available hazard was the fluoroscope for fitting of shoes. It is impossible to determine possible genetic damage from this source but it was probably considerable. Such installations have all but disappeared. Medical and dental radiation sources are firmly regulated. There was suspicion that some chemicals might induce mutations but only about thirty years ago was the mutagenic effect of mustard gas on fruit flies shown. There is concern in some circles that strongly mutagenic chemicals, their effects not yet recognized, may already be in wide use.

A mutation can be a chemical transformation of a single gene or a rearrangement of a chromosome. The former, called a gene or point mutation, can cause an alteration in function. The latter may be microscopically visible and is known as chromosome aberration. In studies of human subjects the aforementioned changes are not always easily determined.

When genetic function of a cell is changed while reproductive capacity is unaltered the genetic change is transmitted to descendant cells. Mutations in germ cells are most serious in long-term effects because changes are transmitted to future generations.

Mutagenic effects are of many types. These range from lethal effects to changes so slight as to remain unnoticed. It is thought that many inherited diseases are based in mutagenic effects. At this time there is no known way to adequately evaluate long-term mutagenic effects. It is a cause for serious concern for future generations. Genetic effects over a long term are statistical rather than discrete. As natural selection has had its effect over a long period, so will externally induced mutations make their effect felt over a long period. Dominant mutations appear in the next generation while recessive mutations require contributions from both parents. This characteristic may not appear for many generations.

Most mutations are harmful or neutral. Many mutants which in the past would have produced death or lessened fertility now remain. This is due to higher standards of health care. The former equilibrium, in which old mutations disappeared about as rapidly as new mutants appeared, has been upset in favor of new mutants. Natural selection does not apply as strongly as before. It has been suggested that medical problems of the future will be more and more due to genetic origin. One estimate places our present health burden as being 25% of genetic nature.

If a mutation causes a gene to have a sterilizing or lethal effect, only one generation is involved. Less severe effects, however, may involve many generations. The less severe the effect, the more people will be exposed. Mild mutational effects, affecting many people, will have a greater public health impact than one severe, or fatal, case. Unfortunately, many of these milder effects may be difficult to detect.

Many chemical mutagens in the environment pose threats which have not yet been adequately evaluated. Mammalians test systems are available but extrapolation from smaller animals to man is indeed difficult and liable to error.

Trimethylphosphate, until recently, was added to gasoline for control of surface ignition and spark plug fouling. This substance has been implicated in chromosome damage in rats. The dosage in rats at which damage occurred was at subtoxic level. However, extrapolation of dosage levels to man is difficult and man's exposure to the substance in the environment is almost impossible to estimate.

Congenital malformations are those abnormalities which can be recognized at birth or shortly thereafter. The study of these abnormalities, which can cause serious disability or death, is known as teratology. A broader definition can include microscopial, biochemical, and functional abnormalities of prenatal origin. It is estimated that about 3% of live births involve congenital malformations, but this is only an estimate. Lack of adequate data at the national level makes such figures rather rough and probably inaccurate.

Three categories of human teratogens have been identified. These are: (1) viral infections, (2) ionizing radiation, and (3) chemicals. Thalidomide is the best known of the third category. Experimental evidence of teratogenicity of some compounds had been in existence for many years but requirements of three generation reproductive tests in animals were established only after the sensational thalidomide disaster of 1962. Some questions concerning the validity of such tests have been raised.

Pollutant materials to which humans are exposed must be examined for toxicity. These examinations must also include specific tests for carcinogenicity, mutagenicity, and teratogenicity. These effects have been studied in the past by separate disciplines and there has been inadequate interchange of results and ideas. It is to be hoped that the emerging profession of environmental scientist will aid in breaking down these historical and somewhat artifical barriers among disciplines.

There is need for more sensitive and reliable methods of testing effects of single chemicals, degradation products and mixtures. External environmental effects must also be included as system parameters in evaluation of pollutant effects. It is possible that human tests may become necessary, but moral and legal considerations pose serious questions.

Agents under test must be administered at subtoxic, toxic, and chronic levels. Effects of possibly significant but normally non-toxic agents on the system containing the toxic agent under investigation must also be evaluated.

Testing procedures must be realistic and reflect the pathways of human exposure. However, the exception to this rule is in sensitivity. Evaluation of carcinogenic effect may require administration to test animals by different means. Normal ingestion in humans might be by inhalation while adequate dosage in rats might require addition of the substance in question to daily food. This is particularly true when dealing with weak environmental carcinogens.

Metabolic compatibility between a test species and humans is important for extrapolations of test results. There is seldom a one to one correspondence and most authorities feel that at least two test species should be examined. There is much testing data on rodents, pigs, and subhuman primates. Thus, further utilization of these animals is indicated. In special cases, however, a less common species may be necessary in order to gain the desired data.

It is necessary to test levels much higher than human exposure for carcinogenic, teratogenic, and mutagenic substances. It must be noted that, even when large numbers of mice are tested, the number is still small compared to the millions of humans that might be exposed in the everyday environment. Not all humans exposed would respond, in any event, to the dangerous substance. As an example, meclizine, an antihistamine used for treatment of morning sickness, has been found to be teratogenic in the rat but not so in a relatively small number of women tested. The question as to effects on a larger population is unanswered. For thalidomide, humans are found to be 60 times more sensitive than mice, 100 times more sensitive than rats, 200 times more sensitive than dogs, and 700 times more sensitive than hamsters. It is obvious, therefore, that predictions as to teratogenic doses of thalidomide on the basis of animal testing would mean very high and dangerous exposure levels.

Complicating the testing picture is the increased effect of substances when a second, supposedly innocuous, substance greatly increases the undesirable end effect. Such reactions were the basis of the Delaney Amendment of 1968 to the Food, Drug, and Cosmetic Act (US). The Amendment required that no additive be considered safe if found, under appropriate testing conditions, to induce cancer in man or animal.

Recent advances in molecular biology have made it possible to give weight to a broad range of evidence, including details as how toxic agents affect human cells and on genetic material which controls cell reproduction. The relatively new concept of virtual risk may find application. It should be noted, however, that use of cost/benefit may not be desirable and should be treated with caution, if not suspicion. At times this concept has been used to evade existing environmental regulations. Elected officials are finding that efforts to weaken public health and pollution control legislation are unpopular.

Often, non-carcinogenic alternatives are available as replacements for questionable materials. There is considerable controversy about the pesticide DDT. Some authorities question the continued utility of DDT for control of cotton insects due to development of resistant strains. This controversy will not soon die down. Cyclamates, banned in recent years, were an intentional food additive and of no value nutritionally. Only after their widespread use for a number of years was there shown a carcinogenic danger.

The argument of economic loss to special interest groups is of no validity. In view of the uncertainty of long-term effects of carcinogenic materials, zero tolerance levels must be imposed for these substances.

It is important that chemicals, and their derivatives, suspected or implicated as toxic, carcinogenic, teratogenic, and mutagenic be detected and monitored in the environment. Epidemiological studies may show effects but it is necessary to quantitatively establish occurrence of these substances. It has been possible to demonstrate the relationship between cigarette smoking and lung cancer although some special interest groups seriously disputed this. However, in this case were two dissimilar populations, smokers and non-smokers. In the case of exposure of the general population to a suspect substance, evaluation and epidemiological treatment of data are more difficult. Apart from rubella virus, no known teratogens, such as ionizing radiation, mercury, etc. have been positively identified epidemiologically in the highly industrialized countries. This indicates the great need for better and more comprehensive gathering of data relating to birth defects. It is to be hoped that environmental effects should be segregated from natural and spontaneous mutation effects. It would be necessary to monitor special indicator traits. Such traits would have to be chosen with great care and followed for a considerable period.

Recently it has been suggested that mutation rates could be monitored by means of data on spontaneous abortions. Teratogens may cause chromasome aberrations but these act after fertilization and would be against a normal background.

The question of data on occupational hazards, long the major field of endeavor of the industrial hygienist, must be raised in connection with legislation thought to be inadequate and outdated. The correctness of much of the information on which maximum allowable concentrations are based is in doubt at this time. Probably a critical reexamination will soon come about as a result of increased environmental awareness.

Present toxicological techniques are not sufficiently sensitive to monitor adequately many weak carcinogens. Compounding the problem is the difficulty of recognizing effects of many substances in the general population. Special situations as to sample population at times help to simplify the problem, such as cigarette smoking. But there is certainly a great area in which problems are strongly suspected to exist but statistical validation is not possible at this time. Yet tolerance and allowable concentration limits must be set. There is a most fruitful area for research here.

Regulatory agencies are often understaffed and underfunded. Attention of existing resources is usually allocated on the "squeaky wheel" principle. They move from current crisis to current crisis and are frequently unable to devote enough time to prevention of new crises. This is contrary to good public health practice.

At times competition among various interest groups can lead to inconsistent positions by responsible officials. In the Report of the Surgeon General (1981) it was stressed that smoking, particularly cigarette smoking, can cause many types of cancers. However, the Surgeon General declined to take issue with the continued granting of subsidies to tobacco growers by noting that price supports are considered an economic and agricultural issue and not an issue concerning public health.

The most visible of Surgeon Generals, Dr. C. Everett Koop, was also the most outspoken about the dangers of tobacco use. He became something of a folk hero and was highly respected.

There have been some instances where persons not in sympathy with environmental control and public health have been appointed to high positions in environmental protection. The need for environmental protection does not change with political climate. Efforts to regress are a serious economic error. Damage from deferred action can cost many times the sums initially saved and the health effects can be cumulative. Public health practice is the basis of the great majority of environmental efforts.

REFERENCES

Adams, J., Virtual Risk and the Management of Uncertainty. Paper presented at the *Royal Society Conference on Science, Policy and Risk.* London. March 18, 1997.

Anon., Minimata Case Is Settled Decades Later. *Chemical Engineering,* p. 46, June 1996.

Anon., New Standards in the Works for Peticide Regulation, *Chemical Engineering,* p. 25, August 1996.

Ayers, S.M., *Conference on Metropolitan Air Pollution*, Wagner College, New York, 1967.

Barth, D.S. *et al., J.A.P.C.A.,* **21**, 544, 1971.

Brandt, A.D., *Industrial Health Engineering*, John Wiley and Sons, New York, 1947.

Chiaramonte, J. *et al.*, *N. Y. S. J. of Medicine*, p. 394, Feb. 1970.

Epstein, S.S., *Nature,* **228**, 816, 1969.

Faith, W.L. and A.A. Atkisson, *Air Pollution*, 2nd Ed., Wiley-Interscience, New York, 1972.

Ferrand, E.F., Effects of Air Pollutants, *The Encyclopedia of Environmental Science and Engineering*, 4th Ed., Gordon and Breach Science Publishers, New York. 1998.

Glasser, M., L. Greenburg and F. Field, *Arch. Environ. Health,* **15**, 684, 1971.

Heuss *et al., J. A. P. C. A.*, 541, 1971.

Hopkins, E.S. and W. Schulze, *The Practice of Sanitation*, Williams and Wilkins, Baltimore, 1958.

Junge, C.E., *Air Chemistry and Radioactivity*, Academic Press, New York, 1963.

Karaffa, M.A. *et al.*, Evaluation of Asbestos Levels in Two Schools Before and After Asbestos Removal, EPA/600/52-89/010, Environmental Protection Agency, Washington, D.C., 1989.

Kendall, D.M., A Summary of Panel Recommendations: Report of a Panel on Food Safety to the White House Conference on Food, Nutrition and Health, Washington, D.C., 1969.

Kermode, G.O., *Scientific American*, **3**, 15, 1972.

MacNeil–Lehrer News Hour, September 20th, 1989.

Maxcy, K., Ed., Rosenau, *Preventive Medicine and Public Health,* 8th Ed., Appleton-Century-Crofts, New York, 1956.

Needleman, H.L. and P.J. Landrigan, Toxins at the Pump, *N.Y. Times,* March 13th, 1996.

Oberle, M.W., *Science*, **165**, 991, 1969.

Perlman, M.E. *et al., Pediatrics*, **47**, 391, 1971.

Policy Making, Department of Trade and Industry, London, March 18, 1997.

Proctor, R.N., *Cancer Wars*, Basic Books, New York, 1995.

Report of the Surgeon General, 1981.

Report of the Surgeon General, 1989.

Rouche, B. *Eleven Blue Men*, Little, Brown and Co., New York, 1953.

Salvato, J., Environmental Health, *The Encyclopedia of Environmental Science and Engineering*, 4th Ed., Gordon and Breach Science Publishers, N.Y., 1998.

Walker, R.S., The Quest for Knowledge versus the Quest for Votes, *IEEE Technology and Society Magazine*, Spring 1997.

PAUL BAHAM
U.S. Merchant Marine Academy

JAMES R. PFAFFLIN
Gillette, New Jersey

ELECTROSTATIC PRECIPITATION

This article deals with high efficiency (99.5%) particulate removal techniques often required of modern central station power plants. The reader is also referred to the article "Particulate Removal" for a discussion of control methods including those used when more moderate conditions apply.

Electric power companies are required to analyze proposals for, and subsequently to purchase, electrostatic precipitators based on cost and performance.

The basic design factors which determine the collection efficiency are the collecting plate area, the velocity of the gas, the time that the gases are in contact with the discharge wires and collecting plates, and the electrical system supplying the useful power to the flue gas. It is the differences in these factors in the manufacturers' proposals that give the engineer the most trouble in choosing the precipitator that will continually produce the required efficiency. The amount of useful power, and therefore the collection efficiency, is primarily determined by the number of active high tension electrical bus sections into which the precipitator is divided (see Figure 1).

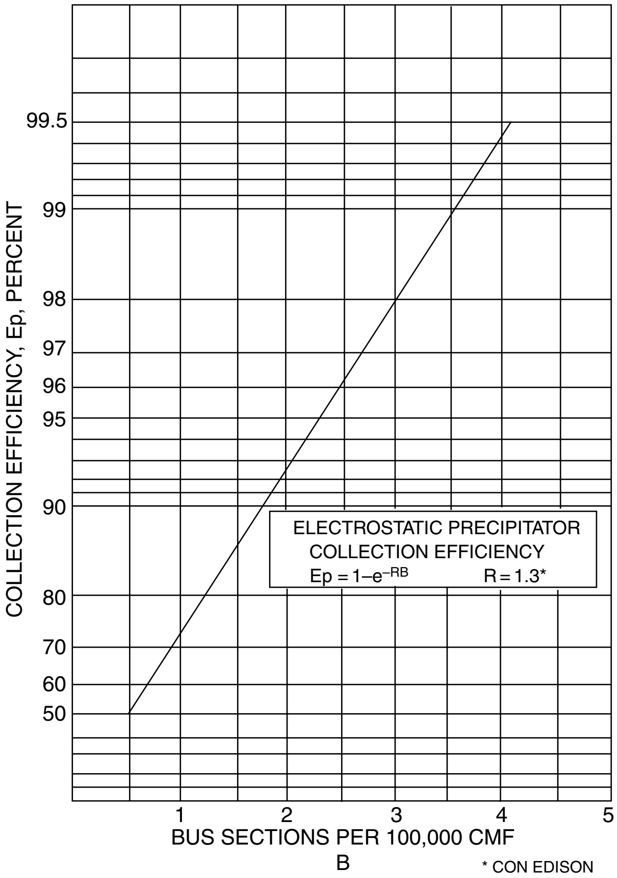
The collection efficiency of a precipitator is closely related to the useful amount of electrical power than can be supplied to the precipitator, the greater the useful power, the higher the efficiency. If we imagine a precipitator with all the discharge wires being supplied by one power source through a single cable, the highest voltage that could be maintained between the wires and the collecting plates would be limited by the first wire to spark excessively. The reason that one wire may spark excessively before another is due to many factors including uneven distribution of the gas and dust as they enter the precipitator, uneven build up of ash on the wires and plates, mechanical misalignment of the wires or plates and the fact that the collection process produces a different amount of ash in the gas at the entrance and discharge end of the precipitator. Even if all the wires spark, at the same voltage, there is an appreciable loss in efficiency due to lowered voltage in the wires operating in parallel because the excessive sparks from one wire affect all the others.

From this it is evident that the ideal precipitator would be one in which each wire has its own stabilization control and power source, but this, of course, would not be economically feasible. Somewhere between these extremes is the practical number of power sources or electrical bus sections that will continually produce the desired efficiency.

Figure 1 shows the efficiency curve which may be used in preparing specification and predicting actual operating efficiency. This relationship between efficiency and active bus sections has been referred to by White[2] as the "Ramsdell Equation."

An active bus section refers to a separately energized precipitator section where a transient electrical disturbance in a given section is not reflected in any other section. This condition exists when either one section is energized by a single rectifier or when two sections are energized by a double half wave rectifier.

A design criterion or an equation for the physical sizing of precipitation is required. A curve based on Con Edison's own experience and that of other utilities is presented in Figure 2.



* CON EDISON
1.0% SULPHUR
300°F

FIGURE 1

The fundamental efficiency formula for an electrostatic precipitator is

$$E = 1 - e^{-w/30.48\times AV}$$

where

E = Collection Efficiency
A = Collection Area
V = Rate of Gas Flow
w = migration velocity, factor which is related to useful electrical power

Low sulphur coal ash is not easy to collect with electrostatic precipitators because of its high resistivity. This was inferred from Figure 3 which shows by tests the effect of lowering the sulphur in the coal on collection efficiency.

The Ramsdell Equation based on bus sections is a variation of the fundamental formula and is expressed as

$$E = 1 - e^{-RB}$$

where

E = Collection Efficiency
B = Number of active bus sections which is related to useful electrical power
R = Performance Constant

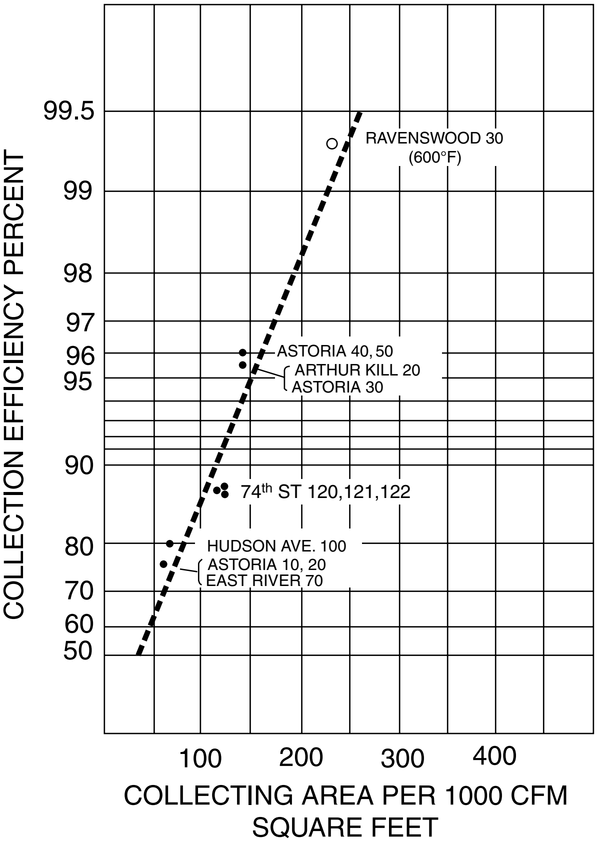


FIGURE 2 Collecting efficiency vs Collecting area per 1000 CFM. Using this efficiency test data and precipitator collecting area we were able to plot the equation or the dotted curve shown in Figure 2.

Using tests data with medium and low sulphur coals and information volunteered by other utilities for high sulphur coals, a family of curves may be drawn (see Figures 4 and 5).

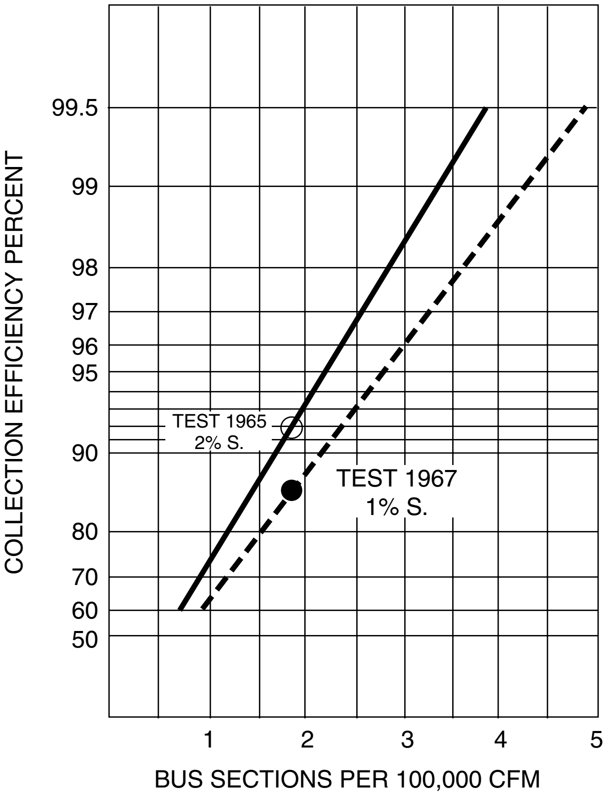


FIGURE 3

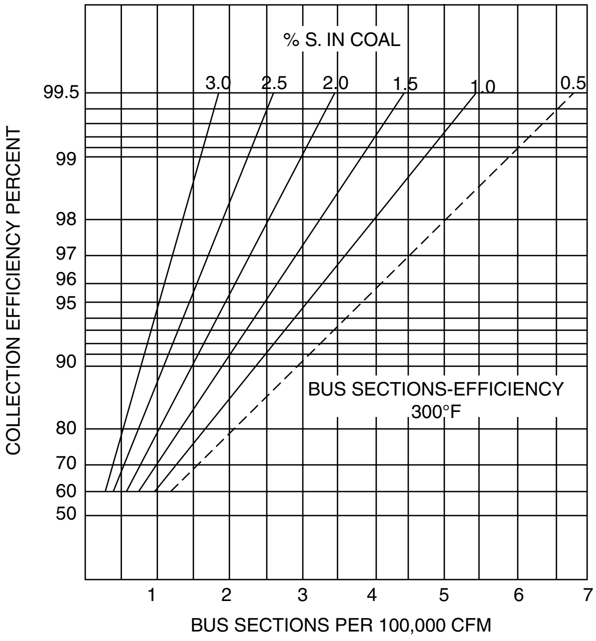
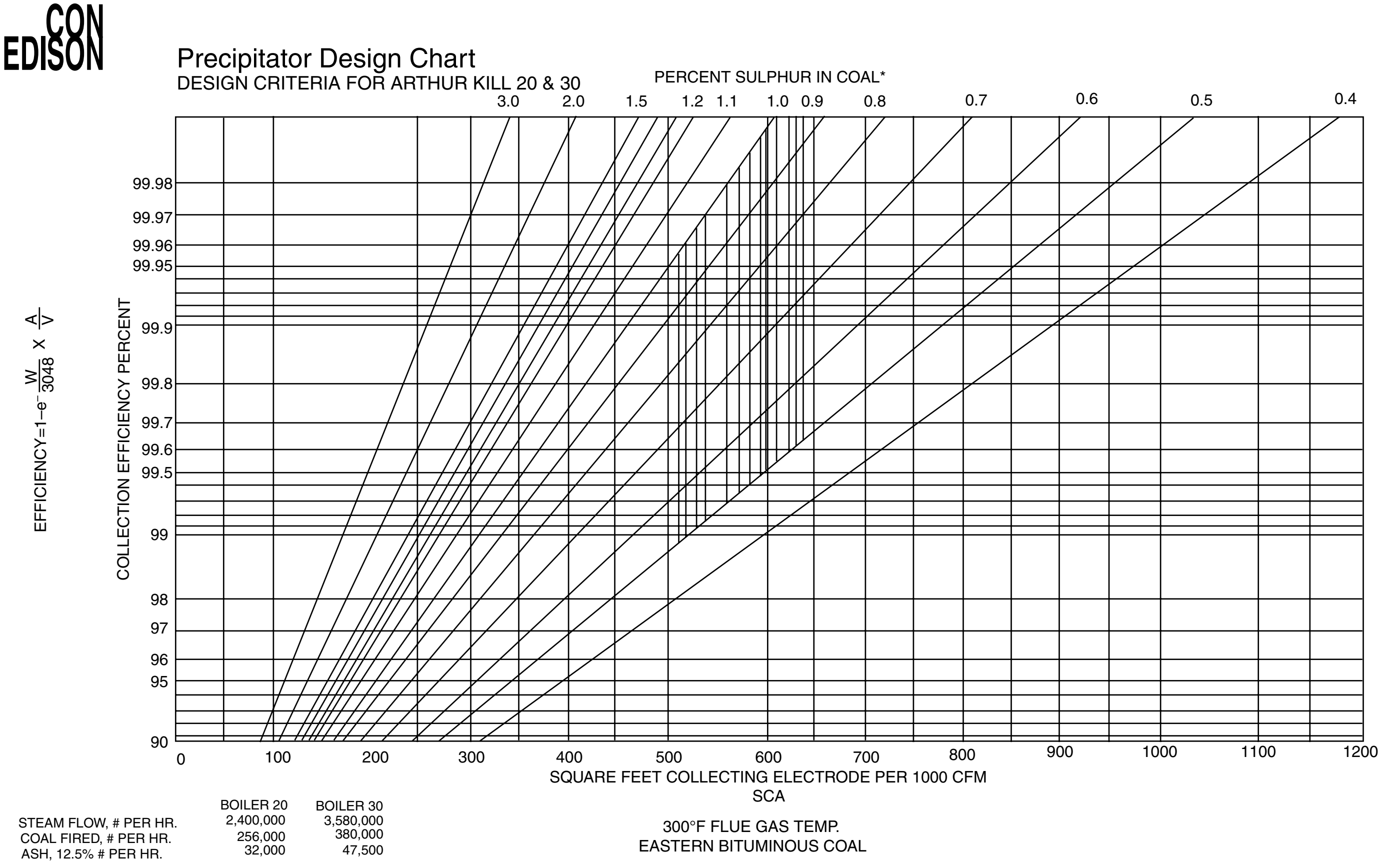


FIGURE 4



	BOILER 20	BOILER 30
STEAM FLOW, # PER HR.	2,400,000	3,580,000
COAL FIRED, # PER HR.	256,000	380,000
ASH, 12.5% # PER HR.	32,000	47,500
FLY ASH, # PER HR.	25,600	38,000
GAS FLOW, ACFM	1,350,000	1,850,000
AVG PRECIP.EFF %	99.5	99.5
EMISSION #/10^4 CTU	.033	.033
EMISSION GRAINS/CU FT	0.01	0.01

R.G.Ramsdell Jc
DEC. 16, 1980

FIGURE 5

There is another feature of precipitator design that is of vital importance to collection efficiency, namely the velocity of the ash laden gas flowing through the collector. The lower the velocity, the greater treatment time available to thoroughly charge the flue gas and the lower the velocity, the less chance there is for reentraining or sweeping off the fly ash accumulated on the collecting plates. The lower the sulphur content of the coal, the higher the resistivity of the ash. Since, under these conditions the collected ash has difficulty sticking to the collecting plates, lower sulphur coals require lower gas velocity. Figure 6 indicates the maximum gas velocities required to insure the required collection efficiencies when burning 1.0% sulphur coals.

It is known that the resistivity of the fly ash is related to sulfur content of the coal burned and also to the temperature of the flue gas.

We have already seen that the lower the sulphur content, the higher the resistivity of the fly ash but as the flue gas temperature drops below 300°F, the high resistivity effect of lower sulfur is substantially reduced. A 30°F decrease in gas temperature under certain conditions will offset the effect of a 1.0% decrease in the sulfur content of the coal.

All the data we have previously discussed in this paper was at a nominal operating temperature of 300°F.

In order to insure that adequate electric power is available for charging our precipitators we again reviewed the latest units on our system. Two sets of curves were developed from this study. Figure 7 shows the total rectifier capacity in milliamps and Figure 8 shows the total transformer capacity in kilovolt-amperes.

These curves are based on 2.0% sulfur coal. Having more electric capacity is of little value with the low sulfur coals with high resistivity ash because the determining factor is how much power the ash laden flue gas will absorb, not how much of a charge can be imposed on the gas. On the other

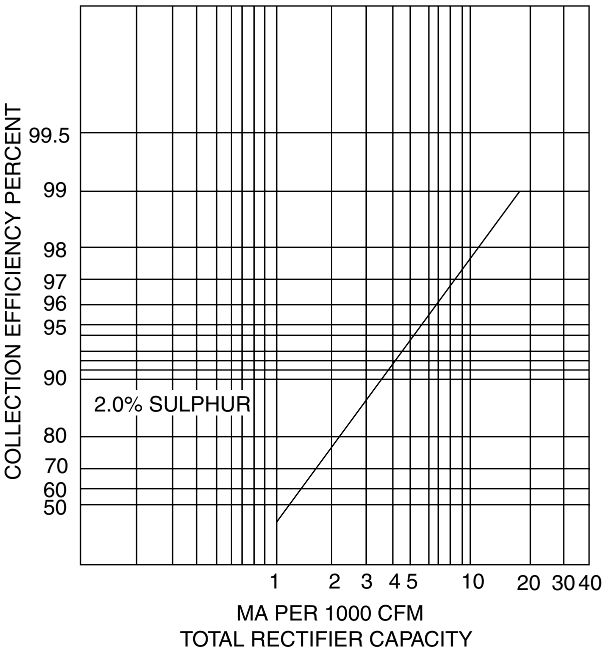


FIGURE 7 Precipitator design curve.

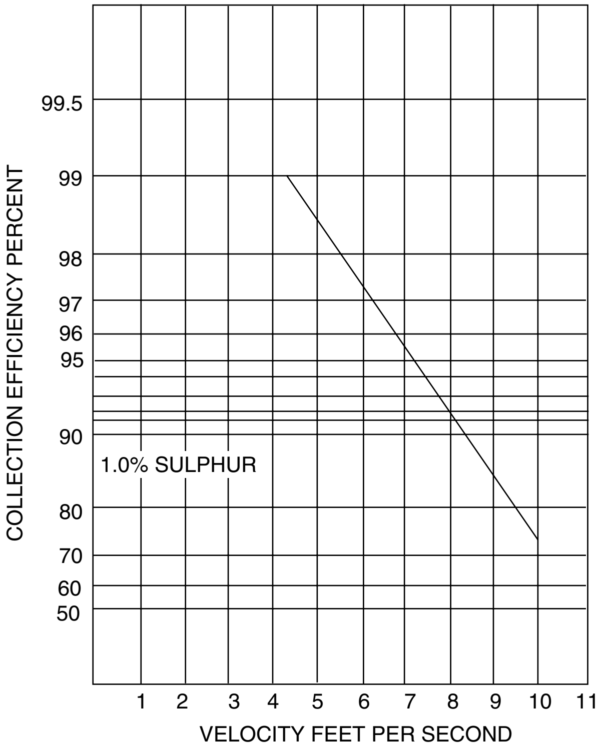


FIGURE 6 Precipitator design curve.

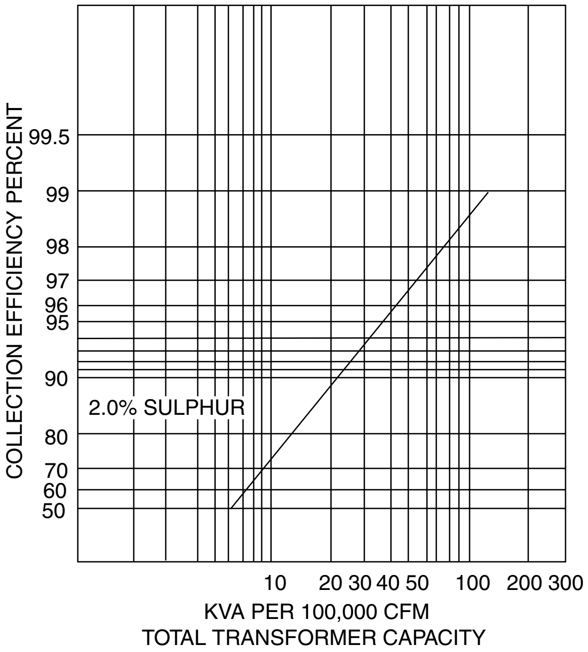


FIGURE 8 Precipitator design curve.

hand high sulfur coal ash does require much more electric capacity than shown on these curves. Precipitators operating under these conditions are sometimes referred to as power "Hogs."

Certain basic criteria should be met in order to attain high collection efficiency at a modern power plant installation. These include:

1) Low velocity of the gas passing through the collection zone.
2) High time of contact of the gas in the precipitator.
3) Reasonably large collecting surfaces.
4) High corona power.

The collection efficiency of a precipitator is closely related to the useful amount of high corona power that can be supplied to the precipitator: the greater the useful power, the higher the efficiency. For better efficiency this power should be distributed among many energized collecting sections, each having an individual control and power supply.

Burning low sulfur coals requires the upgrading of dust collection equipment. This leaves a company with the following alternatives if physical space exists:

1) Add collecting surface and electrical sets to existing 99.75% collectors for example to maintain the necessary efficiencies when burning low sulfur coals.
2) Add new precipitators in series with the existing precipitators to increase the efficiency for example from 99% to 99.8% when burning low sulfur coal.
3) SO_2 conditioning and pulsed energization.[7]

"HOT" PRECIPITATOR CASE STUDY

Hot gas temperature (>600°F) precipitation offers a feasible alternative for the low sulfur eastern coal situation.

This approach was used for the newest operating coal fired unit, at Con Edison's Ravenswood No. 30. The location of this precipitator, between the economizer outlet of the boiler and the air heater inlet, is shown in Figure 9.

The location of this "hot" precipitator was predicated on three considerations:

1) Anticipated reduced air heater fouling by locating the precipitator ahead of the air heater.
2) Ability to burn low sulfur coals without affecting the collection efficiency (see Figure 10).
3) The boiler was designed to burn oil as an alternate fuel and it was desired to be able to operate the precipitator when burning this oil. After extensive tests on a pilot installation at Ravenswood No. 10 while burning oil, it was determined that the precipitated oil ash caught in a "hot" condition could be handled. It has been demonstrated that such a precipitator is effective in collecting oil ash.

The collector is made up of four separate combination units, two double-decked for the north boiler and two double-decked for the south boiler. Extensive model study was required to attain the most efficient flue design that would result in proper gas and dust distribution entering each precipitator. The height of this precipitator for the 1000 MW unit is over 15 stories. The latest performance test on this Ravenswood No. 30 collector when burning coal has met the most optimistic expectations.

As previously mentioned, the design of the precipitator units for Boiler 30 at the Ravenswood Station featured the "hot gas" concept primarily because of its more efficient characteristics in collecting the particulate matter from flue gas while firing either fuel oil or low sulfur coal. Proposed air-pollution control legislation at the time confirmed the need for such characteristics in new precipitator equipment. An interesting structural problem arose in the design of the supporting steel for these precipitator units which we feel was resolved in a rather unique fashion.

Because the gas to be handled at the higher temperatures is much greater than would be required at the normal" cold precipitator" temperature levels, the equipment itself must be bigger and, therefore, heavier. The four precipitator units at Ravenswood—Blr 30 required a building volume approximately 90′ × 243′ × 167′ high which included space to install the large associated flue sections. The decision was made to enclose the building on all its exposed sides with uninsulated metal siding (and to provide a roof) for the following reasons:

1) To eliminate the need to weatherproof the flue and equipment insulaton.
2) To eliminate wind loading on the large exposed surfaces of the flues and precipitator units.
3) To improve the appearance of the installation which is only 150′ from a public street.
4) To reduce the external sound levels around the installation.
5) To reduce future maintenance costs.

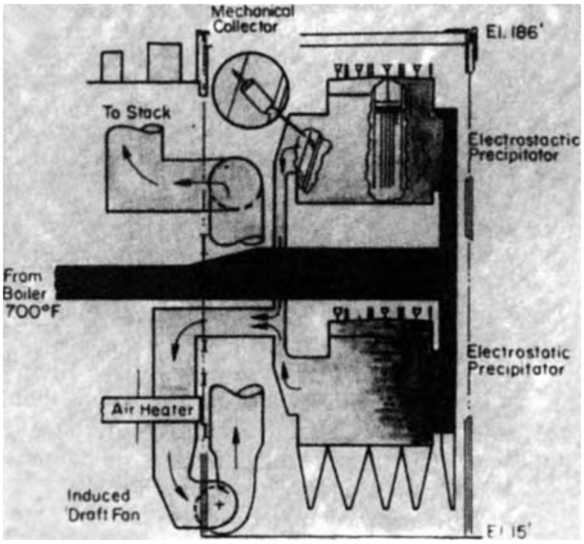


FIGURE 9

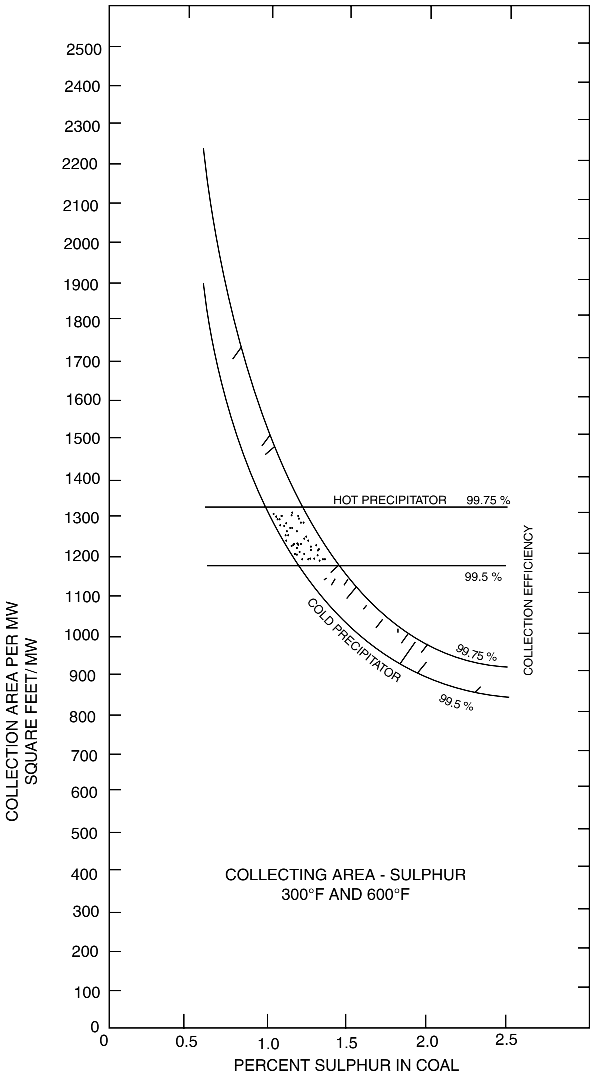


FIGURE 10

Initially, there was no reason to believe that these units would create an excessive expansion problem for the supporting steel, even with the high operating temperature expected. Calculations indicated that only a 20° temperature rise would result overall within the enclosed building with the possibility of a few local hot spots developing similar to what you would normally find in the boiler house.

Boiler 30 started out in construction as an oil fired unit with provisions included to convert to coal fired at some future date. It was decided during the initial construction to convert immediately even if it would not be ready to burn coal until two years after its initial operation burning oil.

Due to the close erection schedule which the coal conversion work on the unit was to follow, it was necessary to start the equipment foundations and supporting steel in the field as quickly as possible. Thus the steel was designed before the equipment design was completed.

The four units were designed to be installed on two elevations in a double deck fashion as shown in Figure 9. The lower units were supported from steel erected on the +42′−8″ elevation (Grade elevation is +15′ − 0″) which allowed for a convenient column arrangement below. The upper units, however, were to be supported from the +126′−2″ level and presented a more difficult column design. Since the interior columns that supported the lower units could not be carried up through the equipment to the upper supports, deep girders had to be utilized to span over the lower units to provide the required support. The maximum span required was 90′–0″ which resulted in a girder depth of 9′−1½″ for a total girder weight of approximately 110 tons. A high yield strength steel was utilized for these girders (ASTM A440 – f_y = 46 KSI). Bracing the building was also a challenge because of the equipment space requirements, and could only be provided around the periphery of the units except below the 42′−8″ elevation.

When the hopper detail drawings were received for the precipitator sections, it was noted that the upper hopper plate stiffeners were located very close to the steel support girders (see Section 1–1 in Figure 11). It was obvious that the hopper insulation above these upper stiffeners would be almost impossible to filed install with the stiffeners in place. But more important, the supporting steel member was now in a “heat pocket” which would not afford much air movement for cooling. Since the supporting members were designed for the same top elevations in both directions, this greatly reduced the possibility of air movement longitudinally along the interior members. This condition existed at all hopper locations at both levels. The possibility of over-heating the supporting steel resulting in an excessive outward movement of the support columns now had to be reckoned with for the safety of the structure.

Schemes were immediately proposed to provide some type of forced ventilation or cooling system which were dismissed because of lack of space as well as for economy reasons. Conditions in the field at the time were such that erection of the lower steel supports was nearing completion, and the lower precipitator shell plates were being delivered to the job site. A solution had to be found which would not delay the precipitator erection, and yet result in a stable structure under significant expansion movement.

The answer was to reduce the relative expansion of support framing in any one direction by providing expansion points at certain key connections. The centerlines for the north and south units were located on the 390 and 320 Column Lines respectively (see Figure 11). It was felt that the building should move north and south of these lines symmetrically as the building heated up during operation. This could be done by stiffening the support steel on these centerlines and by cutting loose the connections on the 36 Line and replace the fixed connections with movable ones. This meant that the maximum expansions would take place by moving the 29 and 430 Column Lines outward and at the same time allowing the center of the building (36 Line) to absorb the inward expansions on sliding connections. This approach had the advantages of having to cut loose only one column line instead of several, and also it reduced individual relative expansions to

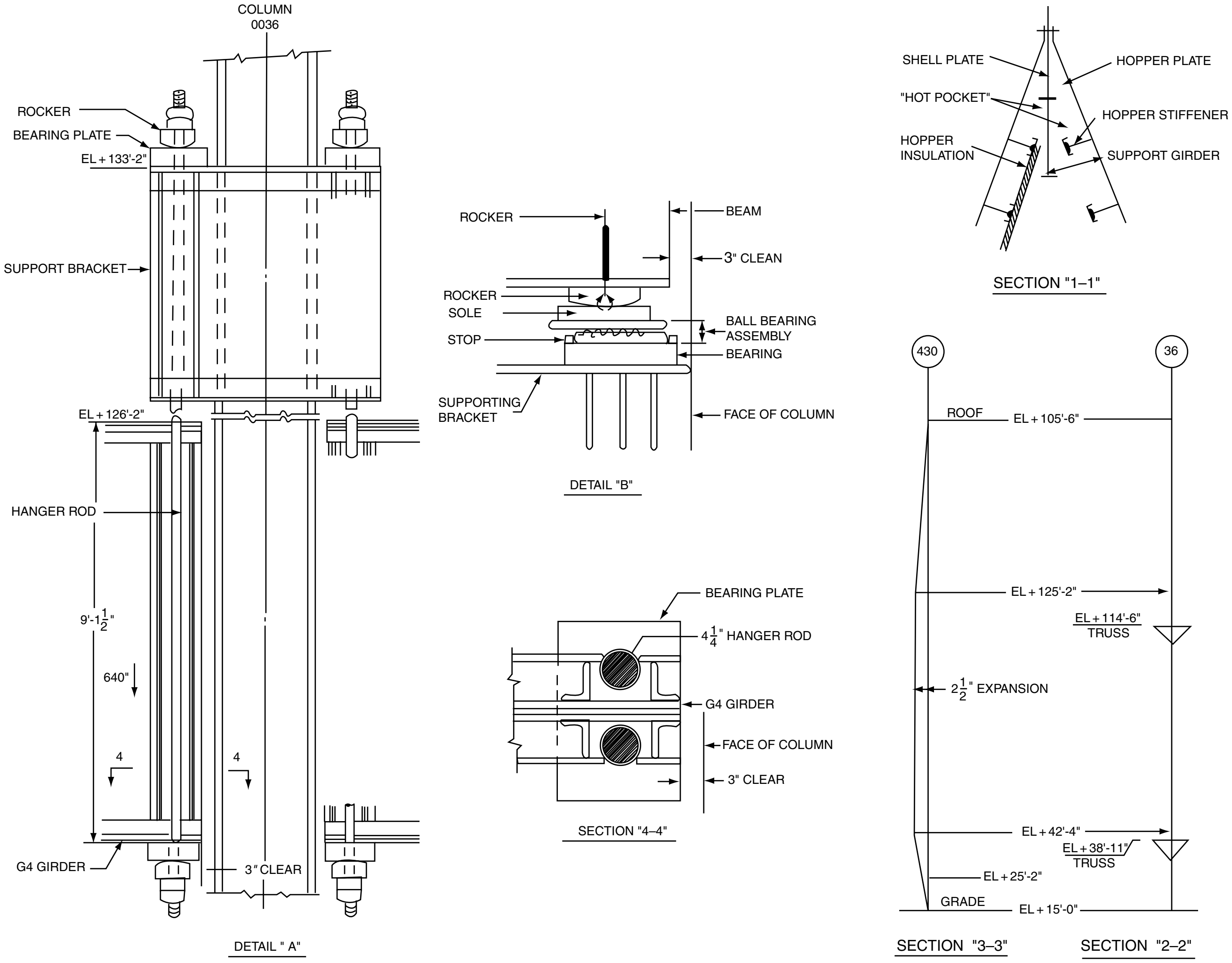


FIGURE 11

movement over one quarter of the building length rather than over one half, at least in the north–south direction.

Design Phase

To accomplish the above design changes, four major obstacles had to be overcome:

1) The centerline support girders for each precipitator unit had to be fixed, or at least stiffened, so that they could come as close to being a zero expansion line (N–S) as possible.
2) Expansion points had to be designed for the 36 Line girder connections.
3) Column Line 36 had to be braced in order to stabilize the building after the connections were cut free.
4) Provisions had to be made for the outward movement of the peripheral columns around the building.

For the normal operating temperature of the units (670°F), it was decided to assume the support steel temperature would rise to an average of 500°F during operation, and provide for the large movements involved. This temperature rise would be applied to all steel directly below the plan dimensions of each unit with the framing beyond these dimensions only assumed to rise to 200° F. Under this assumed combination of temperature rises, the maximum differential movement expected was approximately 2½″ in the north–south direction as well as in the easterly direction. The heavily braced LL Column Line in the boiler house would act as the zero movement line in the east–west direction.

The center support girders for the lower units were restrained by shifting adjacent vertical diagonal bracing to fix the PP320, PP390, M320, and MM390 columns at the + 42′−8″ elevation. The G5 girders on the upper units were stiffened with horizontal members and bracing as indicated in Figure 10.

The expansion joint design for the 36 Line connections utilized ball bearing assemblies except for the G4 connections at Column QQ36 at the +126′−2″ elevation. Detail B in Figure 11 shows the ball bearing type that was selected. The bearings (50 per assembly) were 7/8″ in diameter, and were made of M-50 steel as were the top and bottom assembly plates. The ball bearing retainer was made of stainless steel. The reaction of the G4 girder at Column QQ36 was 640 kips, which ruled out a ball bearing design as the size that would be required was too large for the space available. It was decided to hang the huge girders in a pendulum fashion from a bracket above, and allow them to expand by swinging. The connection finally used is shown in Detail A in Figure 11. For the hanger rod material, a high temperature service chrome-moly steel was selected (ASTM 193-Grade B7). The bar material was heat treated and stress relieved. Mill tests on the material used indicated yield points of 98 KSI and higher. The total required length of only 4¼″ diameter rod was 19′−10″

Horizontal bracing has to be added to the 36 Column Line to carry interior lateral loads out to the vertically braced RR and MM column lines. This was accomplished by adding two horizontal trusses, one at elevation 38′-11″ and the other at elevation 114′−11″ (see Section 2–2 in Figure 11). It was possible to place the lower truss just below the expansion connections for the precipitator support steel at elevation +42′−8″ so the "vertical" members of the horizontal truss could also act as supports for the ball bearing assemblies required for the lower units. This was not the case for the upper units as the existence of other steel (including the G4 girders) made it impossible to install the upper truss any closer to the +126′−2″ elevation. New support brackets were added to the 36 line columns to support the assemblies for the upper precipitator units.

The last major design obstacle to be overcome was the exterior column movement that would result from the 2½″ outward expansion of the framing at elevations +42′−8″ and +126′−2″. It was accomplished by reinforcing the column base billets (and in a few cases by extending the footings as well) to transfer the eccentric column loads which would be carried down predominantly by the exterior flanges. Section 3–3 in Figure 11 indicates the type movement expected. It was also necessary to cut free the support steel of some platforms at elevation +25′−2″ so as not to restrain the exterior columns close to their bases which would induce high moment forces in the columns. This platform steel was resupported on lubrite plate bearings before it was cut free to expand. The column splice plates were found to be flexible enough to carry the moments through the joints at the expanding levels.

Construction Stage

Once the necessary alterations were designed and approved, the difficult task of implementing the changes in the field still remained to be done. One major limitation in this phase of the job was that the precipitator erector was not anxious to have any building connections cut free until his erection work was essentially completed. He had placed a crawler crane on a runway structure atop the 126′−2″ support steel to erect the top units, and he felt that the vibrations already being experienced with crane movements were large enough with the original fixed connections in place, without increasing them by adding expansion joints. This meant that the expansion connections would have to be effected while they were carrying their full design dead loadings. This loading amounted to approximately half the total design load with the remaining half consisting primarily of fly ash loadings.

The erection work was done in three separate stages. First, the column base plates were reinforced as this material was easy to obtain and required little fabrication. Second, all the remaining work except for cutting free the expansion connections was erected. That is, the bracing changes required to fix the lower support columns, the stiffening of the upper centerline support girders, the two horizontal trusses along the 36 Column Line, and the brackets to support the ball bearing assemblies. During this stage also, the ball bearing assemblies were jacked up under the support girders, and the hanger rod connections at column QQ36 were installed. The assemblies were positioned by jacking up the support brackets. The G4 girder connections required the hanger rods to be lowered down through the upper bracket into position. After the top and bottom bearing and rocker plates were in place, the upper nuts were then turned until a snug fit was obtained. The third and final stage included the actual burning free of the original bolted connections in a very careful manner. Less than a sixteenth of an inch drop was observed in the G4 girder elevation after the rods were completely loaded at this point. The entire cutting operation at all elevations was done in less than four weeks.

Operating Experience

All precipitator units have been operating successfully. The tie-in to the boilers being done in two stages. The two southerly units being connected first and then the northerly ones directly after. This was accomplished by taking out the south and north boilers alternately, thus reducing the rate of the unit during this period to 500 MW. The complete tie-in period was just less than six months.

Position readings of several exterior columns were taken before any precipitator units were energized. Readings were taken at the +42′ and +126′ elevations of each column measured. These "cold" readings were taken to obtain base measurements from which to determine the outward movements of the columns once in operation. "Hot" readings were taken last summer with all the units energized and showed that the north columns had moved a maximum of 1½″ to the north, and the east columns had moved a maximum of 1 11/16″ to the east. Both north and east maximum movements were recorded at the +126′−2″ elevation. Comparing this to the design movement of 2½″ it appears that the assumed steel temperature rises were ample. The fact that the +126′−2″ framing is actually expanding more than the +42′−8″ framing is not surprising. It was expected that the upper reaches of the building would be hotter by convection, and because the elevation +126′−2″ framing has an operating unit both above and below it. The lower elevation steel is more readily cooled by outside air being pulled through louvers at grade and across the basement floor by the forced draft fans in the boiler house.

A visit to the precipitator building with all units energized indicated that the building temperature rises gradually as you ascend with the upper areas being as much as 50° warmer than in the basement. The "hot spots" that were originally expected were found to exist, especially where little clearance was left between the precipitator units and the entrance flues on the east side of the building above the +126′−2″ elevation. Air temperatures as high as 220°F were recorded in such areas. Support steel temperatures were taken where access was possible and found to be as high as 240°F. It is felt that the 500°F design temperature probably exists in the middle regions of the support steel at the +126′−2″ elevation but only thermo couples could confirm this. A temperature differential of 115°F was recorded between the top and bottom flanges of the north G4 girder at column QQ36.

In all, the precipitator building at Ravenswood Station—Boiler 30 houses some very unique equipment which created special temperature problems to support it and to enclose it. Operating experience indicates that the designs undertaken to satisfy these conditions are working well.

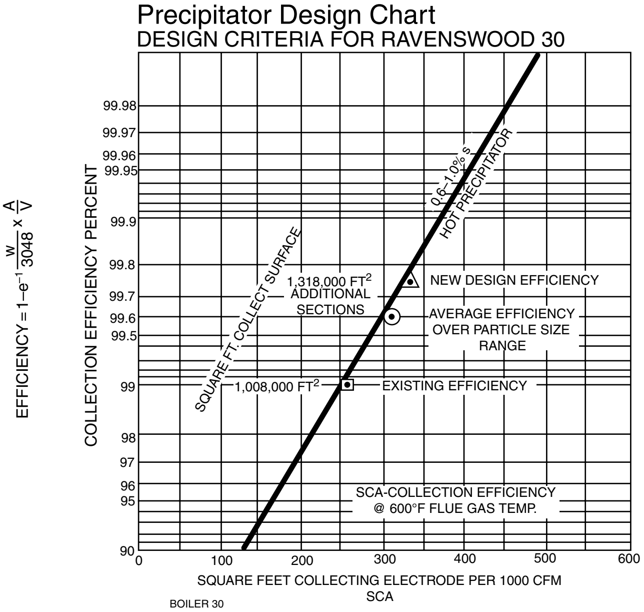
PARTICULATE CONTROL ANALYSIS OF DEIS

The precipitators proposed in 1981 will upgrade the controlled particulate emissions to below 0.033 lb/million Btu.[5] This rate assumes that coal with a heating value of 12,200 Btu/lb and 12.5% ash will be burned, that 80% of the ash if fly ash, and electrostatic precipitators have a design efficiency of 99.75% will be employed. This rate is equivalent to that from burning 0.3% sulfur oil.

Particulate emission control with the original precipitator was adequate to meet plume opacity standards. After the precipitator is upgraded, plume opacity is expected to be below 10%, or less than half of the opacity standard. This is comparable to the opacity when burning 0.3% sulfur oil and is well within the State and City limit.

Con Edison's plan was to commence coal burning using the existing hotside electrostatic precipitator at Ravenswood Unit 3 and to increase the design collection efficiency from 99.0% to 99.75%. This will control the total particulate emission rate to less than 0.033 lb/million Btu, which is equivalent to that from 0.3% sulfur oil. The existing Ravenswood Unit 3 precipitator was tested at 99.2% to 99.6% efficiency when burning 1% sulfur coal. The upgraded precipitator design includes the following to insure that the high collection efficiency is maintained:

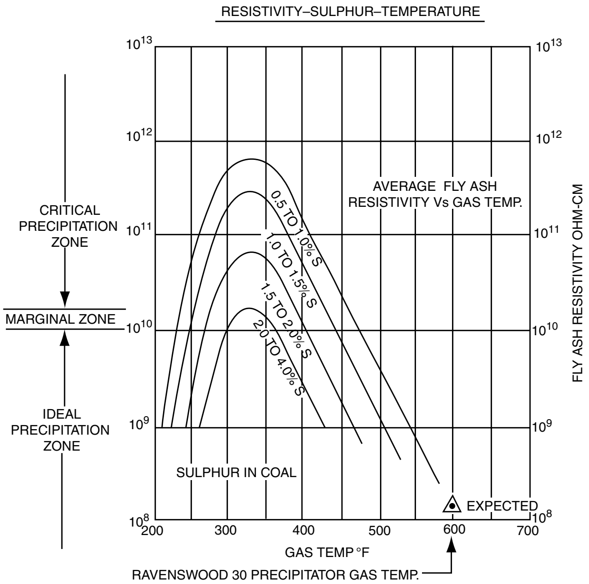
> The mechanical dust collectors will be replaced with an additional 310,000 sq ft. of electrical collecting surface area. This will result in a specific collection area (SCA) of 329 (hot side). Figure 12 indicates that this will provide a design collection efficiency of 99.75% while burning coal with a sulfur content of 0.6–1%.
>
> Electrical sections will be isolated so that failure of one section will not affect performance of other sections. As many as 7% of the electrical sections could be out of service without degrading precipitator efficiency below 99.6%.



	BOILER 30
STEAM FLOW, #PER HR.	6,500,000
COAL FIRED, #PER HR.	685,000
ASH, 12.5% #PER HR.	85,600
FLY ASH, #PER HR.	68,600
GAS FLOW, CFM ∂ 600°F	4,007,000
AVG. PRECIP. EFF. %	99.6
EMISSION GRAINS/CUFT ∂ 600°F	0.010
GAS FLOW, CFM ∂ 300°F	2,920,000
EMISSION GRAINS/CUFT ∂ 300°F	0.01
EMISSION #/10^4 BTU	.033

FIGURE 12A

EASTERN BITUMINOUS COAL



R.G. RAMSDELL JR.
DEC 16, 1980

FIGURE 12B

Electrostatic Precipitation

Existing electrostatic precipitator performance may be improved by the use of wide plate spacing (replacing weighted wire discharge electrodes with rigid type), by using intermittent energization (i.e., blocking selected half-cycles of power to the transformer–rectifier sets powering the ESP) and by applying flue gas conditioning.[7] The importance of modeling ESP performance via 3 different types of computer models enables utilities to optimize their upgrade taking into account hot-to-cold side conversion, fuel switching and sizing for bidding and licensing purposes.[8]

The effects of sulfur and sodium on fly ash resistivity and performance have been discussed in the literature.[9,10]

Daub[11] discusses the effects of computer controlled energization of the tranformer–rectifier sets of a precipitator. Typical results in various European situations are presented in Table 1. Greico and Wedig[12] discuss the performance and economics of ESP's for removal of heavy metals in coal, oil, and orimulsion fired units.

Good design provides for a mass flow 12% above that anticipated, and for a 5% variation in flow distribution between the precipitator boxes.

For particulate control application at Ravenswood Unit 3, the reasons for choosing to upgrade the hotside precipitator included.

1) Only minimal modification work is required for upgrading the existing precipitator.
2) Electrostatic precipitators will meet performance criteria on either coal or oil firing.
3) The low pressure drop across the system eliminate the need for additional booster fans or ID fan modifications.
4) Electrostatic precipitators are a comparatively low maintenance system.
5) Electrostatic precipitators performance is not adversely affected by a rapidly changing gas flow or boiler load.
6) Performance of the precipitator is not affected by changing ash characteristics or coal sulfur content since ash resistivity is low due to high gas temperatures (Figure 12b, 13).
7) Power requirements are comparable to a bag-house filter.
8) Ash removal is more reliable due to the higher temperatures.
9) Air heater performance and maintenance is improve because of the cleaner flue gas.

Fabric Filters: Alternative to Proposed Precipitator

Fabric filters (baghouses) can be installed at Ravenswood Unit 3 if the existing precipitators are removed. They offer no advantage in collection efficiency and opacity over the proposed precipitation upgrading (Figures 14 and 15), but will increase the reconversion costs by about $90 million more than the proposed precipitator upgrading.

Fabric filters, as applied in the utility industry, operate by drawing dust-laden flue gas through a porous fabric bag woven of multifilament glass yarn. During operation, a fly-ash cake is formed over the cloth pores (with the glass filaments forming a

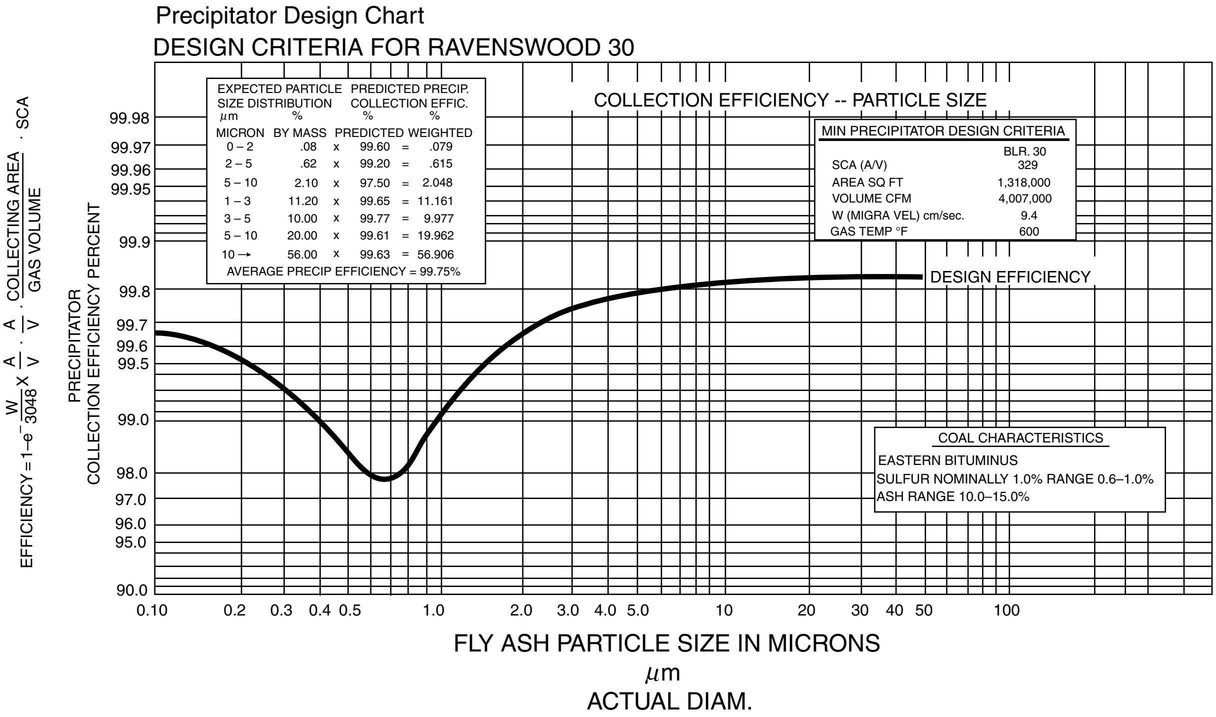


FIGURE 13 Precipitator collection efficiency as a function of particle size.

TABLE 1

	Power plant Coal-fired boiler	**Power plant** Coal-fired boiler	**Power plant** Coal-fired boiler	**Paper mill** Lime reburning	**Steam boiler** Stoker with lime stone	**Cement Plant** Kiln	**Cement Plant** Kiln
Esp make	Walther	FLS miljϕ	FLS miljϕ	FLS miljϕ	FLS miljϕ	FLS miljϕ	FLS miljϕ
Resistivity level	High	Medium	Medium	Low	High	High	High
Energy savings	90%	91%	64%	84%	78%	70%	65%
Emission reduction	25%	33%	7%	0%	65%	57%	47%

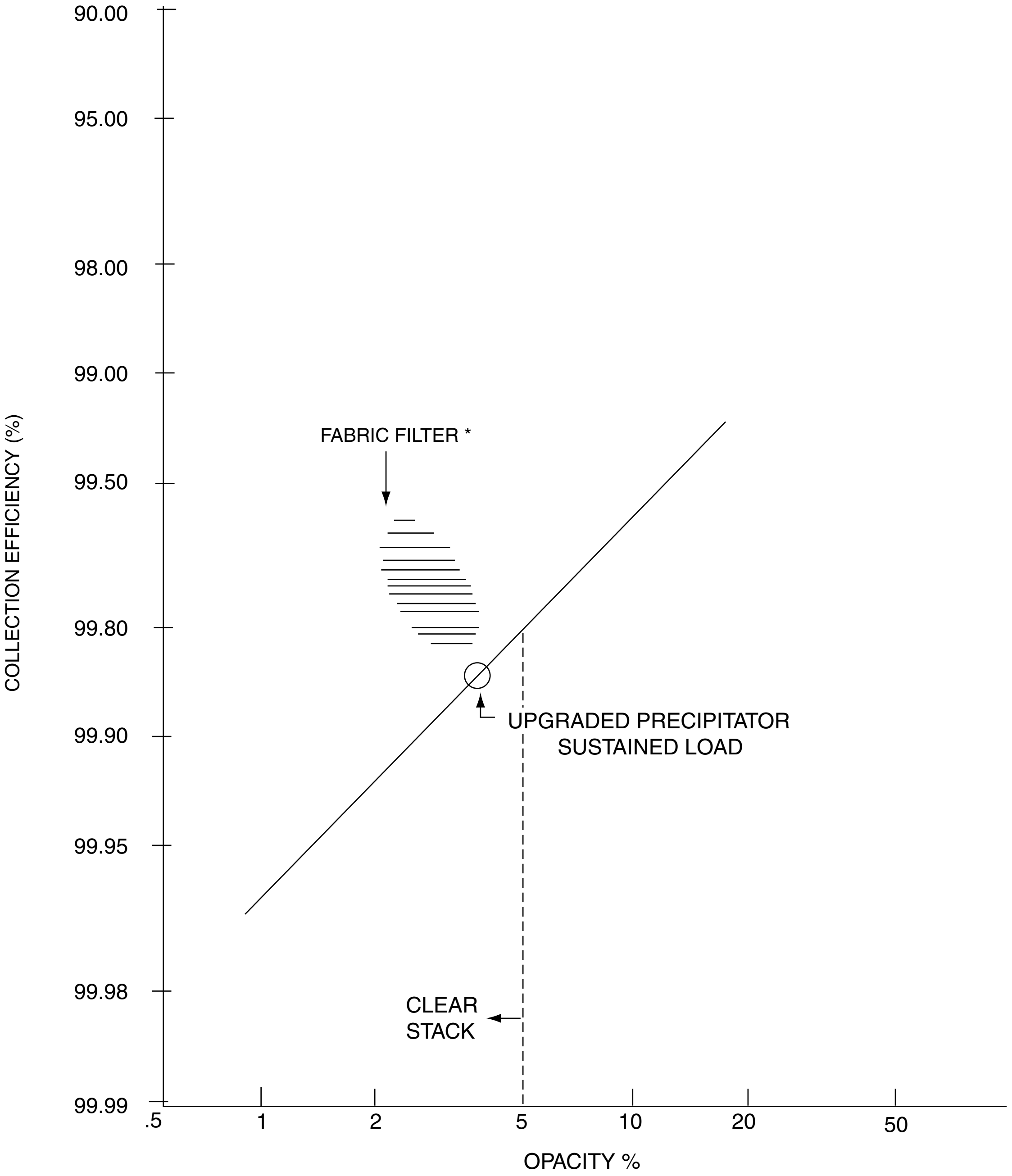


FIGURE 14 Ravenswood Unit 3 plume opacity as a function of particulate collection system efficiency.[5]

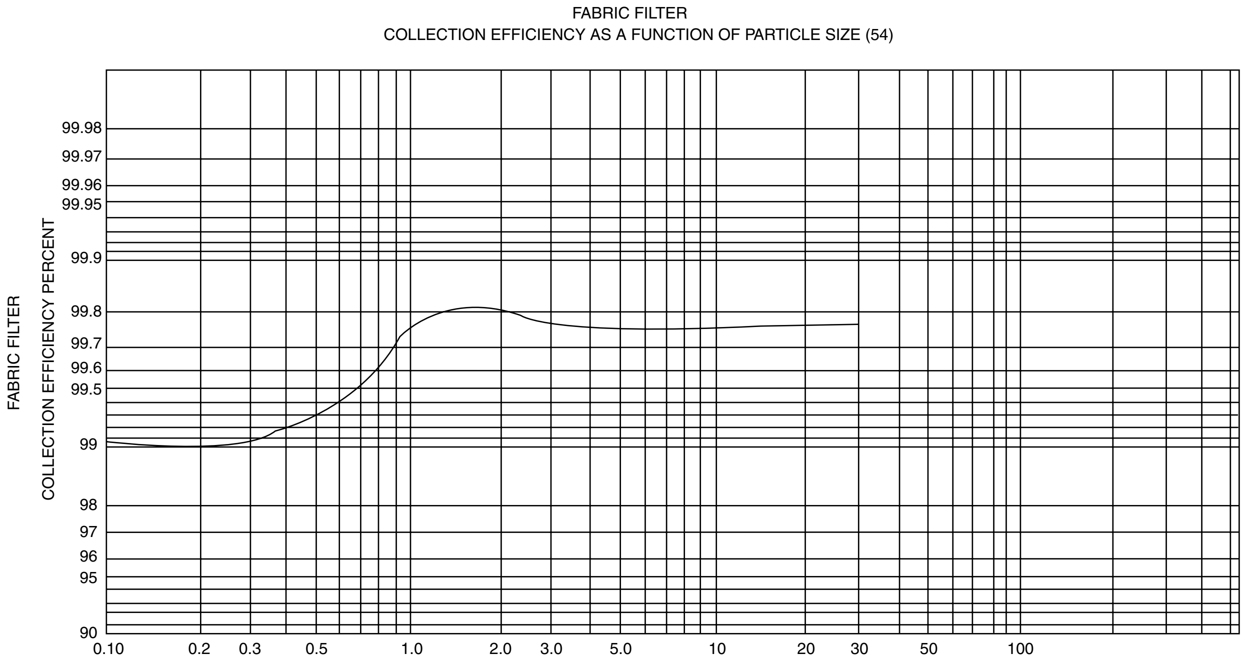


FIGURE 15 Fly ash particle size in microns, actual diameter.

substrate), with subsequent ash collection achieved by impaction of suspended particles on previously collected particles, as the gas flows through the cake. The collected dust cake is periodically dislodged into a bottom hopper by back flushing the fabric with clean flue gas or by shaking. This type of operation necessitates a multi-compartment approach with a typical installation consisting of 10 to 30 separately isolated compartments, each containing about 500 1 ft × 30 ft cylindrical bags.

The filtering medium (cloth) selected for use in a fabric filter must be compatible with the temperature and pH of the effluent. For coal-fired power plants, the present-day filtering medium for a fabric filter is glass fiber bags treated with either teflon or silicon-graphite for lubrication to avoid abrasion between the fabric fibers. The coated fiber fabric is resistant to the chemical attack of flue gas constituents and is capable of withstanding operating temperatures up to approximately 500°F. However, glass bags require precautions in handling and cleaning, due to their potential for self-abrasion. If a spray dryer is used for SO_2 removal, acrylic fiber bags are a less expensive alternative and are less vulnerable to self-abrasion. Their operating temperature limit is 280°F, which while above the typical operating temperature for gas leaving a spray dryer, is below the temperature of unscrubbed flue gas. Advances in filter media technology are continuing, with emphasis on increasing the thermal and chemical resistance of the bags and extending bag life.

As ash accumulates on the filtering media, the pressure loss across the fabric filter increases. In general bags are cleaned automatically upon reaching a specific pressure drop across the fabric filter or by a timing cycle. The compartmentalization of fabric filters on coal-fired boilers allows for one compartment to be isolated for cleaning while the remaining compartments treat the flue gas. To date, the cleaning of fabric filters on coal-fired boilers has generally been accomplished by reversing gas flow against the bag surface, by shaking the bags, or by a combination of the two approaches. The fly ash is then released from the bags and falls into a hopper for removal.

Prior to mid-1978, eight coal fired electric utility stations were equipped with fabric filters. The best documented of this group include Sunbury Units 1 & 2 for Pennsylvania Power and Light and Nucla Units 1, 2 and 3 for Colorado UTE. For the most part, these units were small (less than 50 MW) stoker-fired units with high air heater exit temperatures, (350°F), and fired with low sulfur coal.

EPA assessed the data from these units in mid-1978 and concluded that this technology was capable of reducing outlet emissions to less than 0.03 lb per million Btu (99.8% efficiency). This analysis formed the basis in a large part for the new NSPS for electric utilities of June, 1979.

An additional 1000–1500 MW of fabric filter capacity came on line during the 1978–1979 period and approximately 4000 MW of capacity has been installed in the 1980–1983 period, including numerous large units in the 350–750 MW size class. The fabric filter commitments are generally for use on western coals that have proven very difficult to control by standard electrostatic precipitator technology. This is because filter performance is not sensitive to fly ash composition. Another factor, which contributes to the rapid growth of fabric filter applications in the West, is that filters tend to achieve a lower stack opacity than comparable ESPs operating on western coal. Plume opacity and atmospheric visibility have historically been of much greater concern in the Western states.

The fabric filter has the following advantages for Ravenswood Unit 3:

1) Gas distribution is not as critical as in the precipitator.
2) The fabric filter can be compartmentalized to allow for maintenance while the fabric filter is onstream. This design allows for high availability to the power boiler.

The fabric filter has the following disadvantages for Ravenswood Unit 3:

1) Replacement of the existing precipitator with a baghouse will cost $120 million (1991) but will not improve the collection efficiency of the particular control system beyond that resulting from the proposed precipitator upgrading.
2) Fuel switching may cause permanent fabric blinding because of the sticky characteristics of oil ash.
3) New induced draft fans will be required because of a higher system pressure drop.
4) The furnaces were not designed for baghouses; therefore, the pressure drops that occur during a malfunction of the forced draft fans may exceed design limits.
5) Relatively limited field experience is available on fabric filter use on large units burning eastern coal.
6) Higher maintenance will be required than for electrostatic precipitators.
7) The maximum fabric filter operating temperature is approximately 500°F; therefore, complicated duct work would be required in order to install the baghouse downstream of the air preheater, which would require more plant site area.

REFERENCES

1. Ramsdell, R.G. Jr. (1961) *Prediction of Fly Ash Precipitation* A.I.E.E. Paper CP 61–399.
2. White, Harry J., (1963) *Industrial Electrostatic Precipitation,* Addison-Wesley Publishing Co.
3. Ramsdell, R.G. Jr., *Design Criteria for Precipitators for Modern Central Station Power Plant,* paper presented at American Power Conference April 23–25, 1968.
4. Ramsdell, R.G. Jr. and C.F. Soutar, *Anti-Pollution Program of Consolidated Edison Company of New York,* paper presented at ASCE Conference Environmental Engineering, Chattanooga, Tenn. May 13–17, 1968.
5. Consolidated Edison Company of New York, October 1981. Environmental impact statement—Ravenswood Generation Station Unit 3, Reconversion from Oil to Coal.
6. Orange and Rockland Utilities, Inc., May 1981, Environmental impact statement—Lovett Generating Station Units 4 and 5, Reconversion from Oil to Coal.
7. Altman, R., March 1989, Improving Electrostatic Precipitator Performance, EPRI Jnl., p. 42.
8. Grieco, G., April 1988, Proper ESP Evaluation Demands Extensive Computer Modelling, Power, p. 73.
9. Han, H.P. and E.N. Ziegler, The Performance of High Efficiency Electrostatic Precipitators, *Envir. Progress* **3,** 201 (1984).
10. Harrison, W.A. *et al.,* Medium Sulfur Coal and Fly Ash Resistivity, *Jnl. Air Poll. Contr. Assoc.,* **38,** 209 (1988).
11. Daub, Kare, Low cost improvement of existing electrostatic precipitators, Power Tech. International, Spring 1996, pp. 37–40.
12. Grieco, G.J. and C. Wedig, Controlling heavy metals with electrostatic precipitators, Power Tech. International, Spring 1996, p. 18.

ROGER G. RAMSDELL, Jr.
Consolidated Edison Company, Inc.

ENERGY SOURCES—ALTERNATIVES

INTRODUCTION

Abundant and dependable sources of energy are vital to the social and economic welfare of an industrial nation. Energy resources, including fossil and nuclear fuels as well as solar, water, tidal, and geothermal energy, may be captured or recovered and converted into other energy forms for a variety of household, commercial, transportation, and industrial applications that enrich the human environment. The economic benefits of energy utilization are generally described by the relationship between the per capita GNP of a nation and its per capita energy use shown in Figure 1, which is based on United Nations data.[1] On the other hand, economic indicators such as GNP are not the sole measure of the quality-of-life and use of energy, particularly in the energy conversion step where resources are consumed to produce power, have been identified as major contributors to air and water pollution and to the despoilment of land; environmental factors that threaten the integrity of the biosphere and possibly endanger the life support system on earth.

Energy sources have effects on the environment that occur both naturally and as a result of man's efforts to exploit them. Natural effects range from the life sustaining action of solar radiation to the destructive manifestations of geothermal energy, volcanic eruption, which cause extensive damage and spew particulate matter and gases into the atmosphere at a rate rivaled by conventional man-made processes. These and other natural phenomena have not been controlled and should not be considered further because it can be presumed that these natural effects are now widely surpassed by human faults and advanced technologies which not yet consider in the due extent secondary environmental effects. These environmental effects encountered in the harnessing of various energy sources depend on the forms of energy in demand and on the resources and systems that are utilized to satisfy that demand. The impacts of this activity are controllable to varying degrees by the use of alternate fuels or resources, and by the proper design and application of the systems that are used to supply and utilize energy. Figure 2 is a simplified flowsheet indicating all of the activities required to deliver a useful form of energy to the consumer. In mining and drilling to recover energy resources, the environmental problems encountered include the despoilment of land, safety of the workers involved, and water pollution. The strip mining of coal and uranium is especially damaging to land areas, and underground coal mining is notoriously hazardous. Acid mine drainage is a significant source of pollution of streams and waterways, and oil drilling in offshore areas carries the possibility of uncontrolled discharges of crude oil that are damaging to the marine environment. Each fuel or resource has unique potential for environmental damage in its shipment and storage and in refining and treatment operations where waste materials are generated and must be disposed of. The major fraction of emissions and wastes, however, are generated in the energy conversion processes. A distinction is made in this step between centralized energy conversion, where fuels and other resources are consumed and converted to other energy forms, usually electrical, for distribution and utilization, and a decentralized operation, where fuels are distributed to the location of the ultimate utilization of energy. Although the centralized conversion system generally has economies of scale which may facilitate an increased investment in environmental protection systems, the concentrated nature of the activity aggravates emission and waste disposal problems. In certain cases, such as in the disposal of waste heat, dilution of emissions or discharges may be a valid control technique and is feasible with numerous distributed sources while it may not be as practical with large concentrated sources.

The release of wastes and pollutants from energy activities is at a level where significant external costs are being sustained in terms of hazards to health and damage to materials and property. From the economic viewpoint, these wastes and pollutants are residuals which have heretofore been disposed of to the environment by the producing activity with little or no economic penalty. The social and other external costs attributable to these residuals were external to the producing activity and did not affect decisions regarding the utilization and allocation of resources in that activity. Ayres and Kneese[2] have proposed a mathematical framework for tracing the flow of residuals and charging the external costs attributable to them to the producing activity. The materials flow diagram that they have developed is shown in Figure 3 and addresses the energy conversion operation in relation to industrial activity in the production of goods. The inputs are energy resources and other raw materials which are converted by this industrial activity into final goods and residuals. The residuals, or waste products, identified with the energy conversion operation include carbon dioxide, carbon monoxide, sulfur dioxide, oxides of nitrogen, fly ash, slag, radioactive materials, and waste energy in the form of thermal energy and noise. Control of these residuals may be

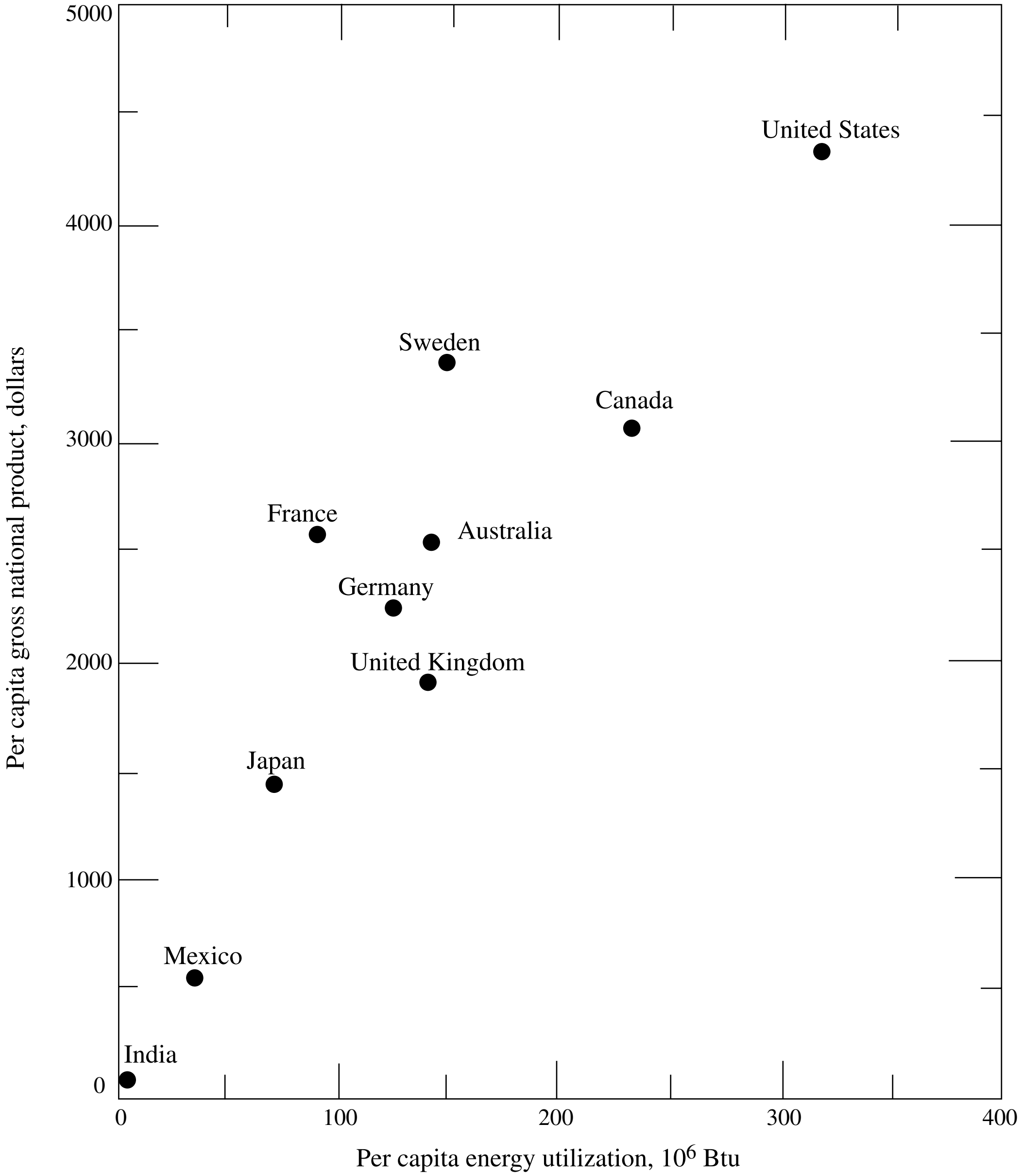


FIGURE 1 Relationship between per capita GNP and energy utilization for selected countries in 1968.[1,3]

stimulated by economic measures such as taxation, that lead to the internalization of all of the costs imposed on society by their discharge, or by regulatory action. Regardless of how controls are stimulated, however, the quality and form of residuals and the cost of their control will depend ultimately on the technological alternatives that are available. Because of the complexity of both the energy system and its impacts on the environment, it has been useful and will be of more and more importance for the humanity to view the subject from four reference points: *ENERGY DEMAND, ENVIRONMENTAL FACTORS, ENERGY RESOURCES,* and *ENERGY CONVERSION.* Consideration of the existing and projected patterns of *ENERGY DEMAND,* and a review of all of the *ENVIRONMENTAL FACTORS* involved in satisfying this demand, will serve to define the extent of the conflict between energy and the environment as well as to indicate trends that can aggravate or help to resolve the problem. Measures that can be taken to insure an adequate supply of energy with minimum environmental damage are centered on the availability of alternative *ENERGY RESOURCES* and on *ENERGY CONVERSION* systems which help to protect the environment or to harm it less by advanced or new conversion technologies. In the short term, fuels may be treated, modified, or substituted, and emission controls can be provided in fuel handling and in energy conversion operations, while in the longer term the development of new resources and an advanced conversion system may also be dictated. Both near- and long-term measures must be discussed to illustrate the range of selection available among alternatives. Considering the conventional energy sources, the increasing environmental problems and all energy sources alternatives, there can be predicted five essential developments:

1) the conventional energy sources will build also in the future the hard core of energy activities, substituting nevertheless as much as possible petroleum by other energies to strengthen the availability of petroleum into a longer future
2) the new energy sources—alternatives as weak energies will substitute partly the short-term demand on conventional energy sources and take over in the long-term future an increasing but not substantial portion of the total energy demand
3) the transportation and traffic systems of the world, which make a major contribution to environmental degradation, will be converted to new, or alternative energy sources, thus fundamentally changing their nature
4) the protection of environment necessitates for both, the conventional and the new energy technologies, high investment costs which will increase the cost of final energy for the consumer
5) to find out the best combination of conventional energies and energy sources—alternatives of old and new energy technologies and necessities as well as economic possibilities to protect the environment, the countries—and it can be only a national way of solution—must elaborate a long-term planning system like masterplan or energy strategies with computerized instrumentation, a typical task for the national energy commissions.

The energy strategies must not be understood as a kind of legislation, but they will give all industry connected with energy production, conversion and application the best path into the future.

ENERGY DEMAND

Energy is utilized in various forms by the demand sectors of the economy. Household and commercial consumers require thermal energy for comfort and for the preparation of food, and electrical energy for lighting and for the operation of electronic and motor-driven appliances that provide entertainment, communication, and labor-saving functions. Energy in mechanical form is needed for both public and private transportation. Industry has similar needs with added requirements involving thermal and chemical energy for materials processing, refining, and the manufacture of synthetic materials and consumer goods. The use of specific energy resources to satisfy these demands depends on their relative economics and convenience as well as on the systems that are available for conversion into the form of energy that is ultimately required.

In principle, especially in the northern countries, the heat demand is much higher than the demand on mechanical force. The demand on public and private light varies between one and fifteen percent. Highly industrialized and densely populated countries are the biggest consumer of light. The energies heat, mechanical force and light are called effective, final or util energies. While in the sixties still all of the energy demand in the United States was satisfied by consumption of the fossil fuel resources; coal, oil, and natural gas, the parts of the individual primary energy agents in the meantime have changed. Small amounts of power were supplied by hydroelectric and nuclear power plants and even less by solar, geothermal, tidal, and other energy resources. On the other hand the part of nuclear energy was increased.

Table 1 shows the direct consumption of energy resources by demand sector in the United States for the year 1968 with a projection to the year 2000. The consumption of energy resources in the production of electricity represents a centralized conversion, and the ultimate disposition of electrical energy to the various demand sectors is given in Table 2, again for the year 1968, with a projection to the year 2000. Since the data presented represent resource consumption, they imply an efficiency of energy conversion and utilization. Comparing the fuel consumption and electrical output of utilities in 1968, an average efficiency of about 32% is indicated. Fuels burned for thermal energy in decentralized household, commercial, and industrial applications

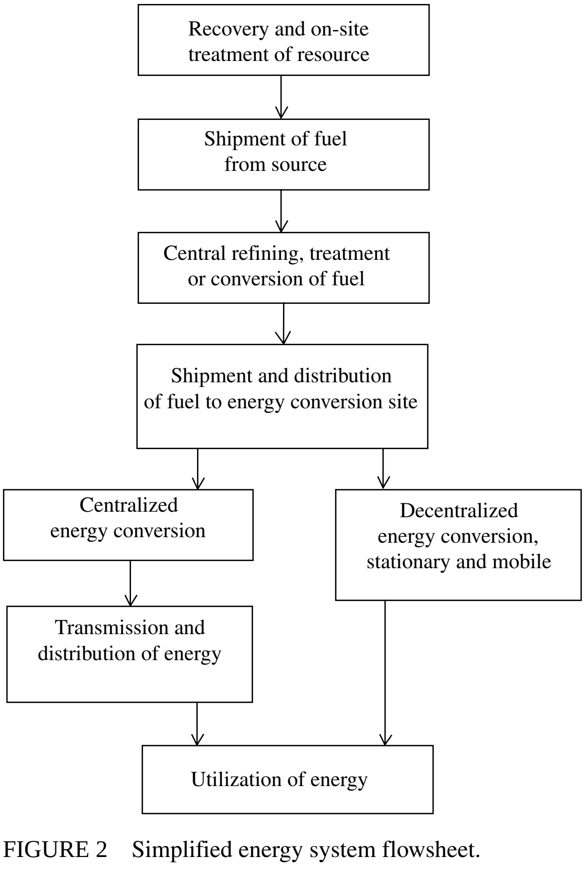


FIGURE 2 Simplified energy system flowsheet.

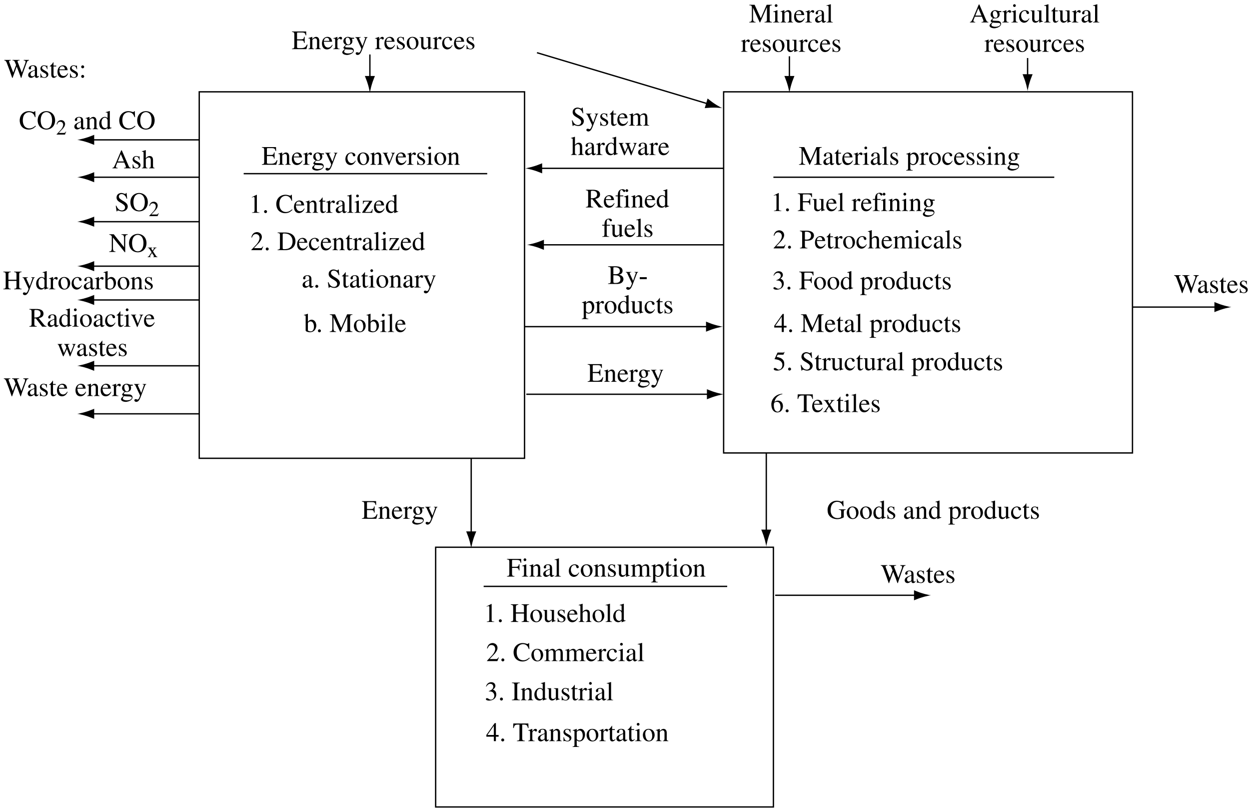


FIGURE 3 Materials flow diagram in economic activity.[2]

are used at efficiencies of 60% to 75%, while fuels used for transportation are converted first to thermal and then to mechanical energy at an average over-all efficiency of about 25%. The source of the 1968 data is the Minerals Yearbook[3] while the projections are those of Morrison and Readling,[4] who developed a loglinear relationship between the major subcomponents of energy demand and several independent variables that are assumed to influence demand, such as; population, economic indicators, and technological developments. Coefficients of the independent variables were to be obtained by regression analysis using data for the period 1947 to 1965.

The year 2000 projection of a total energy resource demand of 168.6×10^{15} Btu is a medium range case, and assumes a normal rate of improvement in energy conversion efficiency and the continuation of present trends of competition and substitution in energy markets. Other projections were made by Morrison and Readling on the basis of different assumptions, including high- and low-range cases that reflect changes in energy demand patterns and technological developments anticipated in energy conversion systems. The high-range case forecasts a demand of 209.4×10^{15} Btu in the year 2000 and assumes the establishment of an all-electric economy, where electricity is generated in fossil, hydroelectric, and nuclear power plants at an average efficiency of 43% and is consumed in household, commercial, and industrial sectors at 10% efficiency and in transportation, with battery-driven vehicles, at an efficiency of 42%. The low-range forecast is 106.5×10^{15} Btu and assumes that energy inputs to the household, commercial, and industrial sectors are satisfied with electricity from fuel cells operating at an efficiency of 75% with 50% recovery of waste heat, and energy requirements in the transportation sector are also provided by fuel cells but at an efficiency of 50%. This again, is an all-electric economy but no central station power generation is needed under these circumstances. Natural gas was the fuel considered for all sectors in the low-range case, however, alternative fuels suitable for fuel cells could also be provided by hydrogen or methane produced by the gasification of coal, or by hydrogen made by the electrolysis of water. Of course such forecasts are highly speculative since it is not certain that fuel cells will be economical in such application in the next 30 years. Numerous other forecasts of energy demand have been made and are summarized in a review[5] performed by Batelle-Northwest for the Energy Policy staff of the Office of Science and Technology (OST). All of these forecasts neglected supply restraints and thus do not reflect the effect of possible energy cost variations due to shortages of resources, increased requirements for environmental protection, and technological developments. None indicate any significant exploitation of solar, geothermal or tidal energy resources.

TABLE 1
Total US energy consumption by fuel and consuming sector
(trillions of Btu's in direct fuels used)

		1968[3] (actual)	2000[4] (projected)
Household and commercial:			
Gas		6,451	19,066
Petroleum		6,581	2,000
Coal		568	
	Total	13,600	21,066
Industrial:			
Gas		9,258	17,504
Petroleum		4,474	13,090
Coal		5,616	2,000
	Total	19,348	32,594
Transportation:			
Gas		610	1,000
Petroleum		14,513	41,649
Coal		12	—
	Total	15,135	42,649
Electric utilities:			
Gas		3,245	4,128
Petroleum		1,181	861
Coal		7,130	18,720
Hydro (coal equiv.)		2,359	5,056
Nuclear		130	43,526
	Total	14,045	72,291
Miscellaneous:			
Gas		—	—
Petroleum		295	—
Coal		1	—
	Total	296	—
Total Consumption		62,424	168,600

TABLE 2
Utility electricity distribution by consumer sector
(trillions of Btu's in electrical form)

		1968[3]	2000[4]
Household and commercial		2,469	20,045
Industrial		2,043	10,793
Transportation		18	100
	Total	4,530	30,938

The forecast methodology with its multiple "demand scenarios" has led to the development of computerized demand models with more or less comfort on the input side. Such models are useful for the analysis of different "scenarios" and consequences for the energy application. The uncertainty of some factors of the consumer market recommends to limit the comfort of full particulars of demand models.

The medium-range forecast for the year 2000, summarized in Table 1 indicates a shift in energy demand to electricity generated in centralized plants with its share of direct resource usage increasing from about 22% in 1968 to 43% in 2000. This shift is apparent in current trends with the demand for total energy growing at about 4% per year while the demand for electricity is growing at around 7% per year not considering the question of future acceptance. Nuclear power is projected to account for over half of the electrical generation in 2000. Such changes in energy demand patterns are seen to have profound effects on the amount of resources consumed and, consequently, on environmental effects. For example, in the case of space heating which constitutes a growing market for electrical energy, fossil resources may be used directly for their thermal effect at an average efficiency about twice that involved in the conversion to electricity in central station power plants. This results in a doubling of the amount of fossil fuel that must be burned to provide a given amount of space heat to the consumer. Although this substitution would be expected to result in the production of more wastes and emissions, it may be justifiable on the basis of the economies of scale that are at work in the centralized operation, which permit better control of emissions, and the opportunity it provides to utilize the more abundant nuclear fuels which cannot be applied economically in small decentralized conversion systems.

In a different example, for space heating and probably also some demand of process heating thermal power plants (coal, lignite, and nuclear) could be used by the double purpose electricity generating and district heating system, that would increase substantially the efficiency of such power plants, involving however the high investment costs for the take-out of heat in the power plants and the long distance tube circulating system to the main demand centers.

The demand indicated in Table 1 for petroleum in 1968 in the various sectors requires more detailed attention since this energy resource is consumed in a variety of product forms. The total demand for petroleum products was 4.9×10^9 barrels and Table 3 shows how this total breaks down among the various products. The end use of most of these is self-evident and it may be seen that less than 2% of the product was consumed as petrochemical feedstock. Diesel fuel for mobile and stationary power plants is included along with the lighter fuel oils in the distillate fuel oil category. Naphtha-type jet fuel is generally used in military aircraft while the kerosine-type fuel is used in civil aircraft.

The non-energy demand for natural gas as petrochemical feedstock and in the production of carbon black for the rubber industry accounted for about 2½% of the total demand for that resource. In the case of coal, about half of the resources used in the industrial sector were consumed as coke, carbonized from coal. About 1% of the total coal demand is consumed as industrial raw materials for the production of crude light oil and coal tars. There is much concern over the depletion of petroleum and natural gas for energy purposes when the demand for these materials, as organic molecular building blocks in the petrochemical industry, is growing rapidly and will expand further as natural mineral resources are depleted and must be replaced in their markets

TABLE 3
Domestic demand for crude oil products, 1968[3] (millions of barrels)

Gasoline—motor gasoline	1925
aviation gasoline	30
Jet fuel—naphtha type	127
kerosene type	222
Special naphthas	27
Kerosine	103
Distillate fuel oil	863
Residual fuel oil	680
Lubricants, wax, coke, asphalt, road oil, and still gas	427
Liquefied gases for fuel and chemicals	386
Petrochemical feedstock	93
Miscellaneous	17
Total domestic product demand	4900

with synthetic materials. This dilemma suggests, for the long term, the implementation of an all-electric economy based on power plants consuming the more abundant nuclear fuels and thereby conserving fossil fuel reserves.

A more detailed breakdown of energy utilization for specific applications within the household, commercial, industrial, and transportation sectors is presented by Schurr.[6]

ENVIRONMENTAL FACTORS

All of the activities involved in supplying energy to the various demand sectors have some impact on the environment. The benefits to the human environment obtained in the utilization of energy have a social and economic basis. These benefits must be acknowledged and are also discussed in some detail by Schurr.[6] Other environmental factors involved in energy related activities may be measured in terms of land use, employment and working conditions in energy industries, air and water pollution, radiation hazards, solid wastes, and noise. These environmental factors are attributable to either the characteristics of the energy resource utilized, or to the technology used in the particular activity that is the causative agent.

Land Use

Physical space is required for all energy-related activities, with mining, electric power production, and electric power transmission being the largest users. The strip mining of coal is especially damaging to the landscape and represents an increasing method of recovery of coal, with about 34% of the coal produced in 1968 mined by this technique. In 1965 about 1.3×10^6 acres were involved in strip mining and the reclamation of this land was estimated to cost an average of \$230/acre for complete restoration and \$149/acre for partial restoration.[7] Strip mining areas are under study as possible solid waste disposal sites to serve urban areas that are nearby, where the cost of transporting the wastes is not excessive. Uranium mining poses potential problems similar to those of coal mining but on a reduced scale. The effects of oil and gas drilling, barring accidental discharges, are less severe.

Fuel shipments by rail and truck consume space with attendant noise, hazards, and general inconvenience. In 1968 there were 209,478 miles of petroleum pipeline and 861,000 miles of natural gas pipeline in service, including distribution lines.

The electric utilities require substantial areas of land for power generation and transmission. About 400 acres of land are required for a 3000 MW(t) nuclear power plant and 1200 acres are needed for a coal-fired plant of the same size, including space for on-site coal and ash storage. These sites are usually preferred near a source of cooling water, such as a river, lake, or other body of water. There are also over 300,000 miles of high-voltage overhead electric transmission lines in service with an average right-of-way width of 110 feet, for a total land requirement of over 4 million acres. This latter figure does not include local distribution lines although there is a growing tendency to put these, as well as high voltage transmission lines, underground. There is about 2400 miles of underground transmission line in use in the 69 to 345 kV range. Electrical energy cannot be conveniently stored in large quantities and some utilities use pumped storage facilities to satisfy peak electrical demands. Elevated water reservoirs are required at pumped storage sites and, like hydroelectric power plants, they are usually located in mountainous areas where there are natural sources of water. The electric utilities are facing severe problems in the siting of new facilities. Recognizing these difficulties and the growing demand for electric power, the Energy Policy staff of the OST has recommended[8] more long-range integrated planning by utilities, participation in this planning by environmental agencies, and notice to the public of new plant sites at least five years in advance of construction. In addition to calling for an expanded research and development program on pollution controls and improved generation techniques, the report asks for more attention to underground power transmission and advanced siting approaches such as offshore locations and underground nuclear installations. The availability of a transmission grid that could move very large blocks of power over some distance would relieve some siting constraints and could also improve service reliability.

In spite of numerous studies and analyses of locations and transmission systems, the above described experiments, tendencies, and intentions differ from the actual energy reality. The high tension up to 1100 kV (a.c.) serves less for the integration of long distance nuclear power plant capacities but formation of strong and dependable grids of interconnection which partly already pass over the border of neighborhood countries. Such transmission lines are planned mainly for the safety of supply of electricity in big areas of demand and for the compensation of peak loads respectively for the support of transmission systems when power plants and transmission lines drop out of operation. Regarding the field of production, the energy reality leads to realize the inclination to build the thermal power plants, that means in the future mainly the nuclear power plants, in case this technology

TABLE 4
Employment and safety record in selected energy industries

Industry	No. of employees, 1969 (thousands)	No. of injuries, 1968		Injury rate per 10^6 man-hours, 1968	
		fatal	non-fatal	fatal	non-fatal
Coal mining	136	311	9460	1.33	40.36
Petroleum and natural gas extraction	282	102	9069	0.10	9.19
Petroleum refining	146	—	—	—	8.6[c]
Electric, gas, and sanitary services	664	—	—	—	6.2[c]
Automotive service stations	575[a]	—	—	—	14.5[c]
Fuel dealers	101[a]	—	—	—	—
Federal government services[b]	3006	—	—	—	6.9[c]

[a] 1967 data.
[b] Excluding military.
[c] Includes fatal injuries.
Source: Statistical Abstracts of the US, 1970.

remains or will be again accepted by political authorities and population, proceed nearer to the demand centers. This could be of double advantage, should the district heating technology be extended, what is recommendable. Whatever kind of energy source may be used, and in this connection the energy sources—alternatives are not excluded, the psychological distance to the energy in general will be shorter at the end of this century, and this fact may contribute to resolve some technical and economic problems, and regarding heat it could be advantageous to reduce urban air pollution.

Employment and Safety

The direct employment of labor in the various energy industries is a small fraction of the total labor force, however, such statistics do not reflect the vital need for energy in manufacturing and industrial processes and to provide comfort, light, and labor-saving functions in all working activities. The employment of labor in some energy industries is listed in Table 4 along with the injury rate in each industry. Additional data on the fatal injury rate and the number of injuries, both fatal and nonfatal, are given for mining activities. Statistics on injuries in federal government service, excluding military personnel, are also given for comparison purposes. It is apparent that the underground coal mining activity has the poorest safety record of the energy industries listed, the most significant problems involving collapse of mine tunnels and explosions due to the methane released from the coal bed during mining operations. Research and development is in progress in the coal mine safety area, including the study of methods of stabilizing mine tunnels as the coal is extracted and other methods of protecting the workers in the event of a fire or mine failure. An example of one novel scheme to guard against fire in high-methane mines is the provision of an oxygen-free atmosphere in the mine with the workers using portable breathing apparatus.[9]

Air Pollutants

A survey[10] performed by the National Air Pollution Control Administration (NAPCA) of the emission of air pollutants in the United States for the year 1968 identifies energy conversion operations in stationary and mobile power plants as the major source of carbon monoxide, particulates sulfur dioxide, hydrocarbons, and oxides of nitrogen. The total emissions of these pollutants are indicated by source in Table 5. The Office of Air Programs in the Environmental Protection Agency (EPA), formerly the NAPCA, has also published a compilation of air pollutant emission factors[11] which give the average quantity of pollutants released in the operation of various fuel-energy conversion system combinations. The emissions are given for uncontrolled sources unless otherwise noted. Control measures that may be used to reduce these emission factors involve treatment of fuels, modification of the energy conversion system, or treatment of effluents depending upon the particular case.

Carbon Monoxide

This poisonous gas is produced by the incomplete combustion of carbon in fuels. When combustion conditions are well controlled and excess air can be provided, combustion is generally complete and CO_2 is produced with only trace amounts of CO. In mobile power plants, such as internal-combustion engines which account for almost two-thirds of the total emissions of this pollutant, operating conditions are quite variable and CO may be in equilibrium in high concentrations with other combustion products at the high temperatures involved. These hot gases are cooled rapidly and there is not sufficient time to convert the CO to CO_2 before they are expelled from the cylinder.

Emission factors for CO from uncontrolled sources at average operating conditions are given in Table 6. Control of these emissions is accomplished, when possible, by proper

TABLE 5
Estimated nationwide emissions, 1968[10] (in millions of tons per year)

Source	Carbon monoxide	Particulates	Sulfur oxides	Hydrocarbons	Nitrogen oxides	Total
Transportation	63.8	1.2	0.8	16.6	8.1	90.5
Fuel combustion in stationary sources	1.9	8.9	24.4	0.7	10.0	45.9
Industrial processes	9.7	7.5	7.3	4.6	0.2	29.3
Solid waste disposal	7.8	1.1	0.1	1.6	0.6	11.2
Miscellaneous[†]	16.9	9.6	0.6	8.5	1.7	37.3
Total	100.1	28.3	33.2	32.0	20.6	214.2

[†] Primarily forest fires, agricultural burning, coal waste fires.

design of the combustion chamber to prevent regions of insufficient air and by providing residence time and temperature conditions that will allow conversion of all of the CO to CO_2. In cases where combustion conditions cannot be adequately controlled to permit complete conversion, the exhaust gases may be reacted with air either thermally or catalytically to complete the reaction to CO_2.

Particulates[12, 13]

This category of pollutants includes any matter, solid or liquid, in which individual particles are larger than molecular size but smaller than 500 microns. The burning of coal in central station power plants accounts for about 64% of all particulate emissions even with the widespread use of electrostatic precipitators with efficiencies of up to 99.5%. This particulate release is mostly in the form of fly ash, which is predominantly carbon, silica, alumina, and iron oxide, and when recovered from stack gases, poses a solid waste disposal problem. Markets for fly ash involve construction uses and are growing. Oil has a lower ash content than coal but is also a source of particulate emissions. The particulate emissions from mobile power plants include lead compounds, carbon, and metallic oxides. Lead alkyl additives used in gasoline to raise its octane number are present in concentrations of 2 to 4 g/gal and this lead content is ultimately discharged to the environment. The automotive and petroleum industries are moving toward reduction or elimination of these additives.

Emission factors for particulates from uncontrolled sources are given in Table 7. The emissions from the combustion of coal are given as a function of the per cent ash content, which depends on the grade of coal but averages about 10%. Wet bottom and cyclone-type burners capture more ash in the furnace in the form of bottom ash, which is a solid waste, and thereby reduce the fly ash particulate matter released with the flue gas. Control of particulate emissions from power plants is provided by cyclone separation, filtering of the flue gases, and electrostatic precipitation, and these techniques are used extensively.

Particulate matter in the atmosphere has adverse effects on materials and health and will have long-term effects on the thermal energy balance of the earth. It provides condensation nuclei which can increase cloud cover and thereby reflect more of the solar energy input to the earth back into space. It has been estimated[14] that a 1% increase in the earth's cloud cover can reduce the average temperature of the earth by 1.4°F and an increase of 5% could lead to the return of an ice age. Volcanic eruptions are believed to account for most of the particulate contamination of the upper atmosphere and in the past have been a source of meteorological changes. Jet aircraft exhausts add water vapor and particulates to the upper atmosphere and can also contribute to meteorological effects.

TABLE 6
Carbon monoxide emission factors[11]

	Average emissions per unit of fuel burned
Stationary Sources	
Coal:	
Household and commercial	10 lb/ton
Industry	2 lb/ton
Utility	1 lb/ton
Fuel oil:	
Household	5 lb/1000 gal
Commercial and industrial	0.2 lb/1000 gal
Utility	0.04 lb/1000 gal
Natural gas:	
Household and commercial	20 lb/million ft^3
Utility and industrial	0.4 lb/million ft^3
Mobile Sources	
Gasoline-powered vehicle urban 1970	2620 lb/1000 gal
Diesel-powered bus and truck	225 lb/1000 gal
Aircraft, long range turbofan	26 lb/engine-flight[a]
Jumbo jet	28 lb/engine-flight[a]

[a] Emissions during portion of flight at altitudes less than 3500 ft, including takeoff and landing.

Sulfur Dioxides[15,16]

SO_2 is produced in the combustion of sulfur-bearing fuels such as coal and residual oil and is released to the atmosphere in the flue gases from power plants using these fuels. Approximately 60% of the SO_2 released to the atmosphere in 1968 was produced in the burning of coal while the combustion of oil accounted for another 14%.

TABLE 7
Particulate emission factors[11]

	Average emissions per unit of fuel burned
General	16 A[a] lb/ton
Dry bottom	17 A lb/ton
Wet bottom	13 A lb/ton
Cyclone	2 A lb/ton
Spreader stoker	
Greater than 10×10^6 btu/hr	13 A lb/ton
Less than 10×10^6 btu/hr	2 A lb/ton
Hand-fired equipment	20 lb/ton
Utility	8 lb/1000 gal
Industrial and commercial	23 lb/1000 gal
Industrial and commercial	15 lb/1000 gal
Household	10 lb/1000 gal
Utility	15 lb/million ft^3
Industrial	18 lb/million ft^3
Household and commercial	19 lb/million ft^3
Gasoline-powered vehicle, urban 1970	8 lb/1000 gal
Diesel-powered bus and truck	13 lb/1000 gal
Aircraft, long range turbofan	8 lb/engine-flight[b]
Jumbo jet	10 lb/engine-flight[b]

[a] Where letter A is shown, multiply number given by the per cent ash in coal.

[b] Emission during portion of flight at altitudes less than 3500 ft, including takeoff and landing.

TABLE 8
Sulfur dioxide emission factors[11]

	Average emissions per unit of fuel burned
Coal	38 S[a] lb/ton
Fuel oil	157 S[b] lb/1000 gal
Natural gas	
Gasoline-powered vehicle, urban 1970	5 lb/1000 gal
Diesel-powered bus and truck	27 lb/1000 gal
Aircraft, long range turbofan and jumbo jet	2 lb/engine-flight[c]

[a] S indicates % sulfur in fuel.

[b] This factor is for residual oil. For distillate oil use factor of 142 S.

[c] Emission during portion of flight at altitudes less than 3500 ft, including takeoff and landing.

This ratio may still change for the worse, since:

1) the combustion of oil will decrease, especially in power plants;
2) the burning of coal will increase until nuclear energy can substantially replace the coal in new power plants—supposed nuclear power plants will be accepted by population also in the larger future.

Therefore new coal burning power plants will be equipped with desulfurization equipment that may eliminate in increasing measure the emission of sulfur dioxide.

Emission factors are given for various uncontrolled sources in Table 8 as a function of the per cent sulfur content of the fuel. The average sulfur content of coal is about 2.5% by weight while that of crude oil is closer to 0.3%, but these levels can vary considerably depending on the source. In petroleum refining, the sulfur generally remains in the heavier fractions and reaches much higher concentrations in residual fuel oil, for example. Sulfur dioxide emissions may be controlled by fuel selection using either low sulfur fuels or by removing the sulfur from the fuel before burning it. At the beginning of the seventies other techniques have been proposed and developed to capture sulfur compounds during combustion or to remove them from the flue gas before it is discharged to the atmosphere. Flue gas scrubbing processes developed, under test and further improvement include the use of MgO, alkalyzed alumina, or limestone to neutralize oxides of sulfur, forming precipitates of sulfites and sulfates that can be collected as solids. A catalytic oxidation process, based on the technology of sulfuric acid production, that captures sulfur emissions as SO_3 in a weak acid solution is also quite promising.

Oxides of Nitrogen,[17] Hydrocarbons,[18] and Photochemical oxidants[19]

NO_x and reactive organic substances such as hydrocarbons (HC), when exposed in the atmosphere to solar ultraviolet radiation, form photochemical smog consisting of aldehydes and a group of oxidants including ozone and peroxyacyl nitrates.

Oxides of nitrogen are produced in all combustion processes that use air as the oxidant at a rate that is dependent on both the flame temperature and air supply. Nitric oxide (NO) is formed at high combustion temperatures and normally would convert to the more stable nitrogen dioxide (NO_2) as the gas temperature is reduced. The kinetics of that reaction are slow, however, and the gases cool to near ambient conditions before very much NO_2 can be formed. In the combustion of methane, the peak equilibrium NO_x emission concentration is developed with about 15% excess air while the peak flame temperature occurs at a slightly deficient air fuel ratio.

With different fuels the NO_x emissions will vary, depending on the theoretical flame temperatures and the combustion process used. Table 9 gives average uncontrolled emission factors for NO_x. These emissions may be reduced, but not eliminated, by the adjustment of the air supply to a fuel-rich condition and operation at as low a combustion temperature as possible by means of two stage combustion, exhaust gas recirculation, or water injection. Some of the flue gas scrubbing techniques proposed to control SO_2 emissions in power plants may also be effective for the removal of NO_x from the gas stream.

The main sources of hydrocarbon emissions are evaporation of fuel from vehicles and in fuel handling operations,

TABLE 9

Emission factors for nitrogen oxides[11]

	Average emissions per unit of fuel burned
Coal	
Household and commercial	6 lb/ton
Industry	18 lb/ton
Utility	18 lb/ton
Fuel oil	
Household and commercial	12 lb/1000 gal
Industry, tangentially fired unit	40 lb/1000 gal
Utility	105 lb/1000 gal
Natural gas	
Household	50 lb/million ft^3
Commercial	100 lb/million ft^3
Industry	230 lb/million ft^3
Utility	390 lb/million ft^3
Gasoline-powered vehicle, urban 1970	183 lb/1000 gal
Diesel-powered bus and truck	370 lb/1000 gal
Aircraft, long range turbofan	5 lb/engine-flight[a]
Jumbo jet	6 lb/engine-flight[a]

[a] Emission during portion of flight at altitudes less than 3500 ft, including takeoff and landing.

TABLE 10

Hydrocarbon emission factors[11]

	Average emissions per unit of fuel burned
Coal	
Household and commercial	3 lb/ton
Industry	1 lb/ton
Utility	0.3 lb/ton
Fuel oil	
Household	3 lb/1000 gal
Industry and commercial	3 lb/1000 gal
Utility	2 lb/1000 gal
Natural gas	
Household and commercial	8 lb/million ft^3
Utility and industry	40 lb/million ft^3
Gasoline-powered vehicle urban 1970	330 lb/1000 gal
Diesel-powered bus and truck	37 lb/1000 gal
Aircraft, long range turbofan	17 lb/engine-flight[a]
Jumbo jet	3 lb/engine-flight[a]
Vehicle gas tank and carburetor	75 lb/1000 gal
Filling[b] of automobile tanks	12 lb/1000 gal
Filling[b] of service station tanks	12 lb/1000 gal

[a] Emission during portion of flight at altitudes less than 3500 ft, including takeoff and landing.

[b] Assumes splash fill.

and unburnt fuel in vehicular exhausts. Emission factors for various sources of hydrocarbons are given in Table 10. Blowby of gasoline vapors from the cylinder into the crankcase was a large automotive emission source but has been largely brought under control by positive crankcase ventilation techniques that recycle these vapors back to the combustion chamber. Control of other sources may be provided by better fuel handling techniques, improved combustor design, treatment of exhaust gases and reduction of fuel volatility. The treatment of automotive exhaust gases involves control of both CO and hydrocarbon emissions using thermal oxidation in the 1600°F to 1800°F temperature range, or catalytic oxidation with transition metal oxide materials. These catalysts are generally poisoned by lead additives and offer extra incentive to eliminate lead additives in fuels.

Carbon Dioxide

This gas is inevitably produced in the complete combustion of carbon-bearing fuels. It is not normally thought of as a pollutant, however, there is some concern over its long-term effect on the thermal equilibrium of the earth because of the so-called "greenhouse effect." Carbon dioxide is a strong absorber in the long wavelength infrared region of the electromagnetic spectrum, a region where much of the energy that the earth loses by radiation to space is concentrated. The CO_2 absorbs this energy, reradiating it in turn to space at a lower source temperature. This mechanism effectively reduces the heat loss from the earth and causes its temperature to increase.

Bolin[20] has reviewed the carbon cycle mechanisms that are at work in the biosphere. The current annual release to the atmosphere of carbon, as CO_2, attributable to fossil fuel combustion is about 5×10^9 tons while a total of some 200×10^9 tons have been released to the atmosphere by combustion processes during the past century. During this same time period the average concentration of CO_2 in the atmosphere has increased from 290 to 320 ppm, a change that accounts for only about ⅓ of the total CO_2 emissions. Apparently, removal processes are effective and include photosynthesis, solution in ocean waters, and deposition in carbonate sediments. One analysis[14] indicates that an increase in atmospheric CO_2 concentration to 400 ppm, which will be attained in 1990 if fossil fuel consumption increases at an annual rate of 5% and if 60% of the CO_2 emissions remain in the atmosphere, would increase the temperature of the earth by 1.4°F. The growth of vegetation is stimulated by high CO_2 levels but, of course, carbon fixed in organic compounds by photosynthesis is ultimately released again either by combustion or natural decay processes. Also, to be considered are the increase of world population and the energy consumption per capita connected with a considerable decrease of surface available for natural vegetation.

Water Quality[14]

The energy activities have an impact on water quality in several ways. Severe damage is done to the marine environment in the accidental discharge of oil either from offshore drilling operations or from oil tankers and other vessels that carry petroleum products. In 1969 there were 1007 reported spills of oil in quantities over 100 bbl each. Vessels were identified as the source of 53 of the seaspills, 331 were from shore facilities, and 144 were unidentified. Individual discharges can be quite large however. In 1967 the tanker Torrey Canyon ran aground and discharged approximately 400,000 bbl of crude oil into the sea. It has been estimated[21] that the oil well blowout in the Santa Barbara channel in 1969 released about 75,000 bbl of oil by the 100th day of discharge. Smith[22] has outlined methods of dealing with spills after they have occurred that have been applied in the United Kingdom, including skimming, absorption, jelling, sinking, and emulsification.

Acid mine drainage, consisting of sulfuric acid and iron compounds formed by the reaction of water and air with sulfur-bearing minerals, contaminates local streams in coal mining areas. Martin,[23] in 1968, estimated that over 4 million tons of polluted water was released annually in the United States to about 11,000 miles of streams and other receiving waters from both abandoned and active mines. Control measures include eliminating the source of polluted water drainage by sealing off underground mines and restoring the surface of strip mines, and by collection and treatment of the drainage.

The petroleum- and coal-producing industries, in 1964, discharged about 7.3×10^{12} gal of industrial waste, or about 10% of the total discharge of industrial waste water. This waste involved 500×10^6 lb of biochemical oxygen demand (BOD), only about 2% of the total industrial waste water BOD. These industries also required 1.2×10^{12} gal of cooling water in the same year which was 2.4% of the total industrial cooling water demand. Electrical utilities accounted for 81.3% of the total demand by using 40.7×10^{12} gal of cooling water. This water is returned to the environment in a heated condition, which reduces the amount of dissolved oxygen that can be retained and thus has an effect similar to oxygen demanding wastes, but it is not otherwise contaminated. The need for cooling water is discussed in a more general context, along with alternate means of heat rejection, in the following section.

Thermal Effects

Heat must be rejected from all thermodynamic power cycles in quantities that depend on the efficiency of the energy conversion operation. The primary sources of waste heat are central station power plants, which can reject heat from their condensers to water or to the atmosphere through cooling towers, and automotive power plants, which discharge their waste energy to the atmosphere in exhaust gases. Atmospheric discharges of thermal energy can lead to local meteorological effects while discharges to water can have effects on the marine ecology.

In a 1000 MW (electrical output) fossil fuel power plant of 40% thermal efficiency, about 1275 MW of thermal energy must be rejected to the condenser coolant while about 225 MW is discharged directly to the atmosphere in the flue gas. A nuclear power plant of the same electrical capacity operating at 33% efficiency must reject all of its waste heat, 2000 MW, to the condenser coolant.

Discharges of waste heat can be controlled by improving the thermal efficiency of power plants, and the severity of local effects may be reduced by distributing the waste energy in large quantities of air or water. Many regulatory agencies have placed restrictions on the allowable temperature difference between inlet water and the water returned to the marine environment. Cooling ponds may be employed in such instances to hold up the effluent water until it has cooled to the allowable discharge temperature. Applications for this "waste" energy have been proposed,[24] including space heating and refrigeration in urban areas, thawing of ice-bound seaways, agricultural use to stimulate growth and to extend the growing season, and in aquaculture to stimulate the growth of algae, shellfish, and other potential marine food sources.

The use of waste heat for district heating systems may have a good chance to be applied on a large scale, although this technology needs a relatively high initial investment.

Radiation

There has been little addition to natural radiation levels in the environment as a result of energy related activities, excluding nuclear weapons testing as a source.

In the nuclear industry radiation problems begin with the mining of uranium-bearing ores. The alpha particle emitters, particularly radon gas, in the uranium-238 decay chain that ends with stable lead-206 pose the major health problem. Radon is chemically inert and diffuses from rock surfaces into the mine atmosphere where it and its daughter products can be inhaled by workers. Exposure standards and control measures have been published by the Bureau of Mines.[25] Mine tailings also contain these radiation sources which complicate the disposal of these solid wastes.

The largest source of radioactive materials are the products produced by the fissioning of uranium and plutonium fuel in nuclear reactors. These fission products are formed in the encapsulated fuel elements and must be handled in the reactor waste treatment system only in small amounts that may leak from the elements during reactor operation. Other radioactive materials found in the reactor system are isotopes produced by the irradiation of materials of construction, corrosion products, and coolant additives. The fission products must be handled in the fuel processing plant where uranium and plutonium that has not fissioned is recovered for use in new fuel elements. Table 11 indicates the activity of the long-lived radioactive materials that are generated in the fuel elements of a nuclear reactor and that pose a long-term disposal problem. Isotopes with a short half-life may be simply held-up until they decay to a harmless product that may be released to the environment, but the others must be chemically reacted and concentrated for disposal by some other technique. The long half-life materials must be sequestered in a manner such that they cannot be released to the environment for centuries. A most promising approach to long-term disposal under study by the Atomic Energy Commission (AEC) involves the concentration of radioactive wastes and their conversion into an inert solid form by calcination or fixation in glassy materials, which would be stored underground in isolated geological formations such as salt domes, shale, or rock.[26] Two long-lived radioactive isotopes that cannot be easily captured and concentrated for such disposal are krypton-85 and tritium. Krypton-85 is a chemically inert fission product with a 10.76-year half-life. Tritium, with a half-life of 12.26 years, is produced both by fission and by neutron irradiation of boron and lithium contained in the reactor coolant and ends up in aqueous form in waste treatment systems. Both of these isotopes are discharged to the environment in liquid and gaseous effluents but are first diluted to permissible concentration levels.

Fossil-fuelled power plants release small quantities of radioactive material that are attributable to trace amounts of uranium and thorium in coal and oil. A study performed by Department of Health, Education, and Welfare personnel[27] compared potential exposures to the population in the vicinity of coal fired and nuclear power plants. The only radiation sources attributable to either type of plant that could be detected above background levels were emissions to the atmosphere. A plant fired with a specific coal containing 9% ash and with 97.5% effective fly ash removal could lead to an exposure of 3.08×40^{-4} millirem of radiation annually per MW of electrical power produced. The exposure potential from a typical pressurized water reactor (PWR) plant was 1.05×10^{-5} millirem on the same basis. A boiling water reactor (BWR) included in the study had a higher exposure value but is not representative of plants of this type that are now being built. In any case, the maximum off-site exposure attributable to any nuclear plant was in the 5 to 15 millirem/yr range. The difference between the BWR and PWR in this study is attributed to the fact that the PWR had the ability to store gaseous wastes for about 60 days prior to release to the atmosphere, while the BWR had only a 30 minute hold-up. Larger volumes of waste gases must be handled in the BWR even though the total quantity of radioactive materials produced are comparable in the two reactor types.

TABLE 11

Long-lived radioactive materials generated in typical nuclear reactor fuel

Material	Half-life (in years)	Activity[a] (megacuries[b]/10^{15} Btu of thermal energy produced by fission)
Tritium	12.26	0.146
Krypton-85	10.76	3.4
Iodine-129	1.6×10^{7}	1.03×10^{-5}
Other fission products		
Cesium-137	30	40.2
Strontium-90	28	28.5
Technetium-99	2.1×10^{5}	4.2×10^{-3}
Transuranium isotopes		
Americium-241	458	0.091
Curium-244	18.1	1.34
Plutonium-238	89	1.27
Plutonium-239	2.4×10^{4}	0.116
Plutonium-240	6760	0.212
Plutonium-241	13	55

[a] After 90 days of cooling time following operation of a typical batch of 2.84% U-235, non-recycle, PWR fuel at a specific power of 40 kW/kg U to a burnup of 35,000 MW-days/long ton.

[b] A curie is the amount of a given radioactive substance which undergoes 3.7×10^{10} radioactive transformation/sec (equivalent to 1 g of radium).

Derived from data presented in: Safety Analysis Report, Barnwell Nuclear Fuel Plant, Allied Chemical Nuclear Products, Inc., Amendment 2, Docket No 50–332–4, March 21, 1969.

Solid Wastes[28]

In 1969 over 30 million tons of fly ash were recovered from power plants burning coal and lignite. About 20% of this is used commercially as a soil conditioner and as an additive to concrete and asphalt mixes. Bottom ash, which accounts for about one-fifth of the ash produced in coal-fired plants, is used in the manufacture of cinder blocks, as road ballast, and on highways in icy weather. Unused portions of these materials represent a solid waste disposal problem and are dumped in spoil areas which occupy land are unsightly.

Large amounts of solid wastes are generated in coal mining, uranium mining, and shale oil recovery. The residual wastes from the processing of coal are put in culm banks and coal waste piles. These contain a high content of carbonaceous material and besides their unsightly appearance and the landslide hazard that they represent, they can ignite spontaneously and emit air pollutants. A survey has located waste piles from 470 bituminous coal cleaning plants and 800 culm banks each containing 10,000 tons or more of waste material from the cleaning of anthracite coal. In 1964 about 495 refuse deposits located in 15 states and containing about 500×10^{6} yds^3 of material were found to be burning.

Overburden material from strip mines is piled in spoil banks which may be thousands of feet long and over 100 feet high. These are a source of acid water, will generally not support vegetation, and render the land unsuitable for recreation or industrial use. A thickness of overburden material averaging about eleven times the coal thickness was handled in US strip mining activities during 1970.

Control of the solid waste problem in these activities will require increase utilization of ash from coal burning power plants and of mill tailings from coal and uranium ore treatment. An obvious disposal method for the portions of these wastes that cannot be used and for spoil bank material is to replace them in depleted strip mining areas and thereby assist in restoration of the land whenever this is feasible.

Noise[14]

Energy conversion devices, particularly those used for mobile applications, are noisy in operation. Noise can impair hearing and has physiological and psychological effects on people.

The average level of sound at normal conversation amounts to 55 db (A) in a distance of 3 feet, the sound of elocution amounts to 70 db (A). For the noise in a living room a level of 35–40 db (A) is acceptable, the critical level of noise begins with 40 db (A), above 40 db the noise can reduce the attention and concentration of working people. For working places they are required by law to limit the noise level to:

1) mainly intellectual work 55 db (A).
2) simple or mainly mechanical office work 70 db (A).
3) all other work in office or workshop 85 db (A).

From 90 db (A) on it is required to wear ear protection.

Traffic noise, mainly from vehicular power plants, can reach 90 db (A) in urban areas and on heavily traveled highways, while a four-engine jet aircraft produces up to 120 db (A) on take-off at a distance of 200 feet. In internal combustion engines the sources of noise are the rapid combustion of fuels, high-velocity gas discharges through exhaust valves, and engine vibrations. The noise output is held down by muffling of the exhaust gases and by acoustic lining of the engine compartment. Turbine noises from both aircraft engines and stationary power plants arise from high-speed rotating components and high-velocity gas flows, and can be controlled by reducing the fan blade tip speed, the use of acoustic linings, and varying the geometry of the exhaust nozzle.

The necessity to build power plants also in the vicinity of towns and villages leads to expensive noise protection equipment.

ENERGY RESOURCES

Modification of fuel form and the substitution of alternate energy resources are useful procedures for controlling environmental effects. The selection among alternative resources depends on their relative abundance, ease of recovery, convenience in handling, and economics. Some resources, such as fossil and nuclear fuels, are present on earth in only finite amounts and are not renewed to any extent, although in the case of nuclear fuels the energy content of known reserves may be quite large. The other major category of energy resources involve energy that is supplied continually to the biosphere over a geological time span, including geothermal and tidal energy as well as solar energy and its derivatives, wind, waves, and hydropower.

In 1980 the average cost of coal, oil, and natural gas at their source was US $55–60 per ton, US $30–35 per bbl, and $1.6–1.8 per 1000 ft^3, respectively. Within the last years these costs have drastically escalated and further adjustments may be attributed to the cost of providing environmental protection systems and to governmental policy actions in specific areas such as mine safety, oil import quotas, and natural gas price regulation. Current prices for nuclear fuel supplied to fission reactors indicate a cost for this form of energy of about US $1.8–2 per 10^6 Btu.

The cost of treating, refining, and transporting fuels adds considerably to the price paid by the consumer for energy. Figure 4 shows the portion of the cost of electrical energy, produced in central station power plants, that is attributable to the transportation of coal, oil, and natural gas fuels and to high-voltage electric transmission as a function of distance and transport technique.[29] The spatial separation between the location of resources and the centers of energy demand poses the problem of optimal siting of fuel treatment, refining, and energy conversion activities and has been studied in a geographical format by Manners.[30] Consideration of environmental factors imposes additional constraints on such siting problems.

A number of estimates of the abundance of energy resources are available and have been assembled and refined by Hubbert[31] of the US Geological Survey. The exploitation of an exhaustible energy resource generally begins at a low level and increases to a period of maximum utilization, ultimately returning again to a low level as the resource is depleted or as recovery costs become prohibitive. Of interest with respect to an energy resource is the estimated time of peak annual consumption and the time at which 90% of the resource is expected to be consumed. The summation of annual recovery over the entire recovery time span is the ultimate recovery for that resource, Q_∞. The cost of obtaining a resource increases as that resource is depleted. In general, it is necessary to qualify any statement concerning resource abundance with a definition of the cost constraints that are assumed. In view of such difficulties, estimates of resource availability vary widely.

The estimated supply of coal and nuclear fuels can provide the projected US needs over several centuries, while Hubbert's analyses indicate that over 90% of the oil and natural gas resources may be consumed by the year 2015. Others are much more optimistic about new discoveries of oil and gas. Homan and Lovejoy[32] have reviewed the methods used in arriving at oil and gas supply estimates and compare the results arrived at by several investigators.

Any estimate of the abundance of a certain energy resource must however consider that the different process of substitution in the meantime under way may rapidly change the global picture of the quantity of energy resources. The process of substitution makes it necessary to analyze alternative energy strategies and to formulate a selected long-term strategy for each country or even multinational areas.[33]

Another pertinent characteristic of a specific fuel or energy resource is its versatility. Some fuels can be modified or treated with relative ease to provide a convenient form and can be utilized in a wide spectrum of energy conversion system while others may be limited in their periods of availability or in their adaptability. This characteristic accounts more than any other for the great demand for natural gas or oil and its derivative products. Coal, lignite, and their products coals are burnable products of decomposition of former vegetable substance in different phases of preservation. Coals consist of more than 50% (weight) and more than 70% (volume) of carbon and contain at maximum 30% unburnable parts (ash). Coals have brown up to black color. The vegetable substances produce first peat and then progressing through lignite, subbituminous, bituminous, and anthracite stages. It has been estimated that there is about

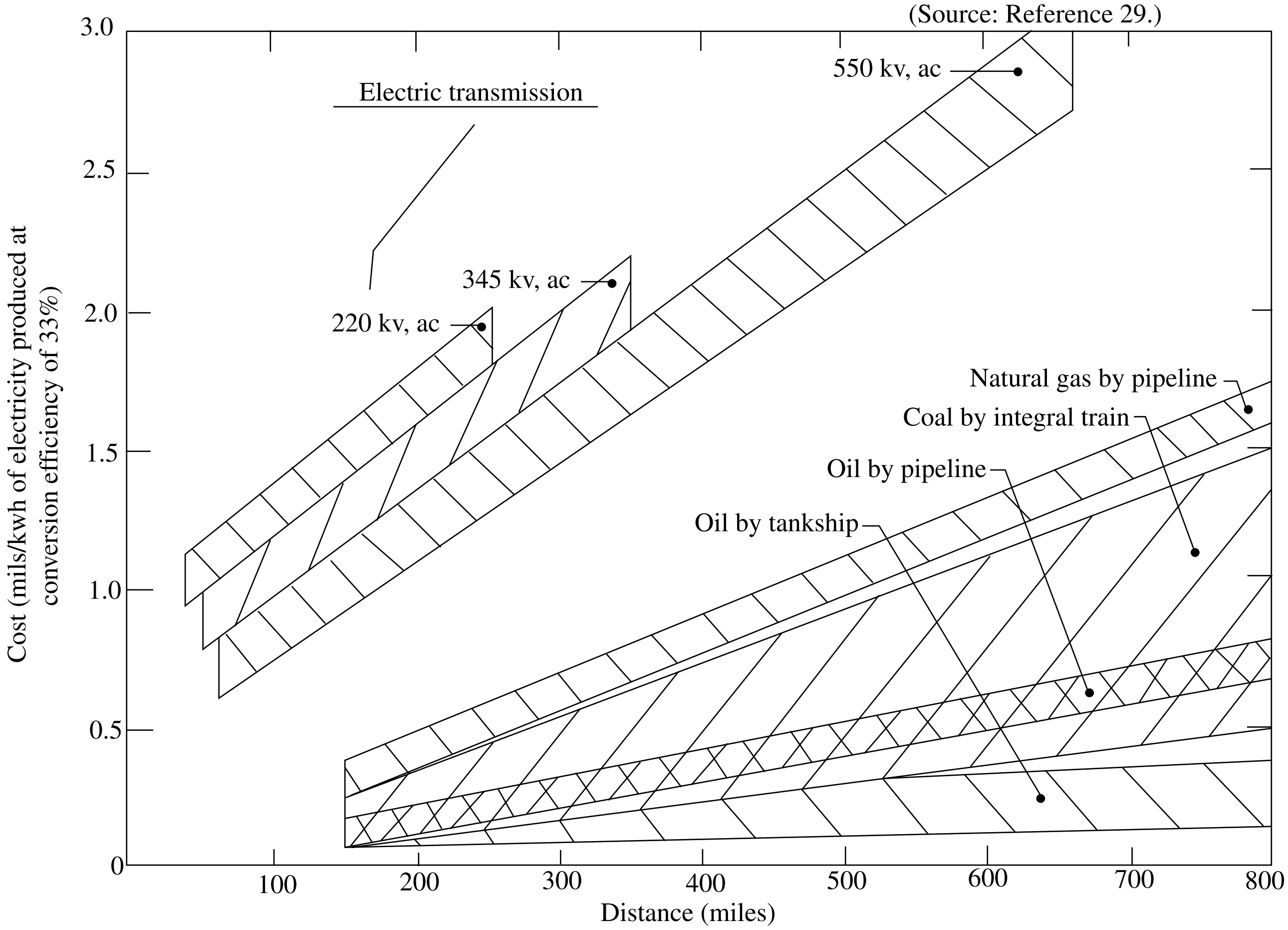


FIGURE 4 Portion of central station electric cost attributable to the transportation of energy.

1.58×10^{12} tons of coal in the US in seams at least 14 inches thick and at depths of less than 3000 feet, and about half of this is considered recoverable at a cost near current levels. These deposits are described by grade and sulfur content in Table 12. On the same basis, the world's coal resources are 8.3×10^{12} tons. The time of peak annual recovery and the time at which about 90% of this resource will be consumed are projected to be the year 2150 and 2440, respectively, for the US and 2150 and 2380 for the world. The annual extraction of bituminous coal in the world was in 1974 2245 M of tons, the extraction of lignite 847 M of tons (Data source: Friedensburg/Dorstewitz, Enzyklopaedie Naturwissenschaft und Technik, Muenchen, 1980).

Sulfur is present in coal as pyrite (FeS_2) and organic sulfur in roughly equal amounts. The organic sulfur is difficult to remove, while only 30% of the sulfur present as pyrite is removed by the standard coal-cleaning technique of crushing and flotation. Processes are available for the liquefaction and gasification of coal and lignite, the process for the liquefaction and gasification of coal had worked before the petroleum era and is working with improved technologies in semi-industrial installations, the process for the liquefaction and gasification of lignite may go in the near future into the phase of pilot plants. These processes will provide a sulfur-free fuel in more convenient form. Hydrogenation plants are foreseen that will produce 100,000 bbl/day of synthetic crude oil from 3.24×10^5 tons of coal[34] at a cost after sufficient development of about the double of the world marked price of natural crude oil, that means, all gasification and liquefaction plants are not yet competitive with the conventional process of refining natural crude oil. But the engagement of several industrial countries demonstrates that the technical processes of substitution exist and are available for a certain period to overcome the problems of modification of the world energy economy until alternative energy sources of long-term significance and necessary quantities are available for the then modified applications in the human life.

The liquid of synthetic crude oil has a high gasoline content and the process yields a high-sulfur char residue. In the gasification of coal, the sulfur is recovered as H_2S. Carbonization, or pyrolysis, yields a gas, mostly hydrogen,

TABLE 12
Estimated remaining coal reserves of the United States by rank and sulfur content, 1965 (in billions of tons[a])

			Coal rank		
Sulfur content %	Anthracite	Bituminous	Subbituminous	Lignite	Total
< 0.7	14	104	257	345	720
0.8–1.0	—	111	131	61	303
1.1–1.5	—	49	—	41	90
1.6–2.0	—	43	1	—	44
2.1–2.5	—	48	—	—	48
2.6–3.0	—	52	—	—	52
3.1–3.5	—	90	—	—	90
3.6–4.0	—	128	—	—	128
< 4.0	—	105	—	—	105
Total	14	730	389	447	1580

[a] Coal in seams at least 14 in thick and less than 3000 ft deep in explored areas. Approximately one-half of these reserves are considered recoverable.

Data source: Decarlo, J.A., E.T.T. Sheridan, and Z.E. Murphy, Sulfur Content of US Coals, Dept of Interior. Bureau of Mines, Information Circular 8312, 1966, Washington, DC.

with 15 to 30% of the original energy content of the coal with the balance in the coke that is produced. Hydrogasification processes can produce methane, a pipeline quality gas. The waste products produced in a coal conversion include particulates, CO_2 and sulfur compounds but they are generally easier to handle in this step which may also be carried out at a site more remote from population centers than where the fuel is consumed in an energy conversion operation.

Hydropower

Hydropower is the most important of the regeneratable energy sources because of its highest efficiency at the energy conversion. But its share on the total energy production lies worldwide at only 3%. The possibilities of extension are limited and it is expected that in the year 2000 hydropower will bring only 5% of the worldwide needed energy. That means in absolute figures: throughout the world installed electrical capacity of hydropower was in 1962 2.857 GW, while in 1978 it amounted to 3.960 GW. Some countries have almost used their potential on available and usable hydropowers:

Switzerland	near 100%
Germany	about 75%
Austria	about 60%
Europe	more than 50%
USA	about 22%
Brazil	about 20%

There are two types of hydroelectric power plants:

1) Run-of-river power plants for the use of affluent water;
2) Storage power plants (power stations with reservoir) where the influx can be regulated with the help of a reservoir.

Mostly greater differences in altitudes are being used, like mountain creeks. Power stations with reservoirs are generally marked by barrages with earth fill dam or concrete dams. A special kind of storage power stations is the pump storage power station which makes it possible to use surplus energy of the grid to pump the water to the reservoir and to provide kinetic energy.

Though hydropower generally can be called environmentally acceptable, there exist also some problems:

1) Change of ground water level and fill up of the river bed with rubble.
2) Risk of dam breaks.
3) Great demand for land space for the reservoir (the Kariba Dam in Central Africa produced a reservoir surface of 1930 square miles and made necessary the resettling of 50,000 people).
4) Diminution, but partly also increase of value of recreation areas.

As the hydropowers of the world are limited, the world energy demand however is rising, finally the share of hydropower will decrease.

Crude Oil

Hubbert estimates that the ultimate recovery, *QF* of crude oil in the US, including Alaska, will be about 200×10^9 bbl.

Other estimates are higher than this by a factor of 2 or 3. The world resources are felt to be between 1350 and 2100 × 10^9 bbl. Peak annual recovery and the 90% depletion point are projected to be attained by the years 1995 and 2000, respectively, for the US and 2000 and 2050 for the world, considering that the processes of saving and substitution have begun. Petroleum resources may be extended considerably by the recovery of this fuel from shale oil. About 80 × 10^9 bbl of oil can be recovered in the US under present conditions, using shale with a content of oil of more than 10 gallons per ton. The total US resources of shale oil in place with more than 10 gallons of oil per ton is figured at 1430 × 10^9 bbl. Before a substantial production could be achieved, however, high investments must be affected—and prototype plants of semi-industrial and industrial size would have to be constructed.

A wide range of control of product yield is available in the oil refining process. Separation of light and heavy fractions are generally performed by distillation into liquefied petroleum gas (mostly propane and butane), gasoline, kerosine, fuel oil, and residual crude. The normal yield of gasoline from a typical crude is about 15% but this can be increased substantially by thermal and catalytic cracking techniques. In 1968 the average gasoline yield from US refineries was 45% of the crude oil consumption. Petroleum can also be used to produce methane, hydrogen, ammonia, methanol, and other clean fuels that are being studied for various applications including mobile power plants.

The sulfur content of crude oil is generally concentrated by refining operations in the heavier fractions, such as the residual oil used to fuel central station power plants. Table 13 gives the 1966 crude oil production in the US by area and sulfur content of the original crude. Since much crude oil and petroleum product is imported, similar information on foreign production is given in Table 14. Sulfur may be removed from oil by several techniques including delayed coking, solvent deasphalting, and hydrogen treatment.

Natural Gas and Natural Gas Liquids

The principal component of natural gas delivered to the consumer is methane. As recovered in the field the gas contains hydrocarbons of the composition C_nH_{2n+2} (methane, ethane, propane, butane, and pentane), natural gas liquids, and other gases such as hydrogen sulfide (H_2S) and, in some cases, helium. The H_2S content varies from field to field and is removed to recover and sell the sulfur. Natural gas resources are classified into nonassociated, associated, and dissolved categrories depending on whether the gas exists alone, in a gas cap with oil beneath it, or dissolved in oil. About 75% of the gas reserves are in the nonassociated category. Gas associated with oil generally provides pressurization of the oil pool and stimulates its recovery, so the gas is usually left in place until the oil is depleted. Natural gas liquids (NGL) are recoverable from the gas and supplement other liquid fuel resources.

The ultimate recovery, Q_∞, of natural gas in the US is estimated by Hubbert at about 1200 × 10^{12} ft^3 and for the world at 8000 to 12,000 ×10^{12} ft^3. Cumulative production through 1969 was about 400 × 10^{12} ft^3 in the US. On the basis of these figures, the year of peak annual recovery will be about 2000 and 90% of the resources will be consumed by the year 2025.

Uranium and Thorium

Natural uranium has an isotopic makeup of 0.711% uranium-235, 0.006% uranium-234, and the balance is uranium-238. Only uranium-235 among the naturally occurring isotopes is fissionable to any extent and natural uranium must be enriched in this isotope to between 2 and 3% to provide fuel for current nuclear reactors. If only the naturally occurring uranium-235 were utilized in fission reactors, the nuclear industry could be limited, for economic reasons, to those US resources of U_3O_8 that are recoverable at a suitable cost. These are estimated at 660,000 tons and would furnish only about 232 × 10^{15} Btu of energy. Fortunately, it is possible to extend the degree of utilization of our uranium resources by converting the abundant uranium-238 isotope into fissionable plutonium-239. Thorium, which is available in large quantities, can also be converted into a fissionable isotope, in this case uranium-233. These conversion processes, involving the capture of neutrons by the fertile isotopes, may be carried out in power producing reactors, which are called breeders if the amount of new fissionable generated is greater than that consumed, or converted if the amount generated is less than that consumed. Breeding is required in order to fully utilize uranium-238 and thorium-232 resources and makes the economics of nuclear power less dependent on the cost of fuel resources. There are estimated to be about 25 × 10^6 tons of U_3O_8 in the US recoverable at reasonable costs and eventual utilization of all of this natural uranium, including the uranium-238 isotope after conversion into plutonium-239, would develop some 1.25 × 10^{21} Btu of energy.

Fusion Resources

There are several fusion processes that are currently under study or in development as a source of energy but none have been attained in a controlled reactor. These reactions depend on the hydrogen isotope deuterium, D, in seawater as a primary fuel. One process, the D–D reaction involves the feeding of deuterium into a magnetically confined high temperature plasma and has the net result of converting five deuterium atoms into one atom each of helium-4, helium-3, and hydrogen, in addition to two neutrons, releasing 24.9 MeV of energy in the process. An alternate and, in theory, easier reaction to obtain is the D–T reaction which involves the fusion of one atom of deuterium and one atom of tritium to produce an atom of helium-4, one neutron, and 17.6 MeV of energy. The neutrons generated in the D–T reactor must be captured in a lithium-6 blanket

$$\begin{aligned} D + T &\rightarrow {}^4He + n + 17.58\ \text{MeV} \\ D + T &\rightarrow {}^3He + n + 3.27\ \text{MeV} \\ D + D &\rightarrow T + p + 4.04\ \text{MeV} \end{aligned}$$

TABLE 13
United States crude oil production by area and sulfur content category—1966

Area	Crude oil production, % of US	Annual crude oil production, 10^6 bbl sulfur, weight %					
		0.00–0.25	0.26–0.50	0.51–1.00	1.01–2.00	2.00	Total
Alaska	0.47	14.3	—	—	—	—	14.3
Appalachia	0.68	20.4	—	0.1	—	—	20.5
California	11.34	25.7	40.7	110.1	122.5	45.4	344.4
Gulf Coast	29.52	568.4	328.4	0.4	—	—	897.2
Illinois, Indiana, Kentucky	3.02	67.3	24.2	—	—	—	91.5
Michigan	0.47	12.0	0.9	1.2	0.1	—	14.2
Mid-continent	23.09	215.2	227.9	145.1	79.6	33.8	701.6
Rocky Mountain	9.08	140.4	23.1	47.7	14.0	50.7	275.9
Southeastern US	2.14	13.8	20.0	1.5	7.1	22.5	64.9
West Texas and Southeast New Mexico	20.19	150.2	107.4	90.2	171.3	94.4	613.5
Totals	100.00	1227.7	772.6	396.3	394.6	246.8	3038.0

Source: McKinney, C.M., and E.M. Shelton, Sulfur Content of Crude Oils of the Free World, US Dept. of the Interior, Bureau of Mines, Report of Investigations 7059, 1967, Washington, DC.

producing another atom of helium-4, an additional 4.8 MeV of energy, and most importantly, an atom of tritium which is recovered and fed to the fusion device, along with fresh deuterium to fuel the reactor. Thus, the D–T concept produces a net 22.4 MeV for each deuterium atom supplied, but also requires a continuous supply of lithium-6 and seems to be limited by the availability of this material rather than deuterium.

There are 1.028×10^{25} atoms of deuterium per cubic meter of seawater or about 1.5×10^{43} atoms in all of the oceans. Recovery and utilization of 1% of these in a D–D Reaction would provide 1.14×10^{26} Btu of energy.

The measured and probable resources of lithium in the US amount to about 21×10^6 tons containing 6.54×10^{34} atoms of lithium-6. Since the number of lithium-6 atoms recoverable appear to be less than the deuterium resources, the energy available through the D–T reaction is limited by the availability of lithium-6 to about 2.2×10^2 Btu, an amount comparable to the energy content of the world's coal resources.

In spite of all research and development work for fusion reactors no reliable statement can be given whether it will be possible to reach the aim to develop an economically working fusion reactor. On the other hand the successful development of such a fusion reactor would be an enormous gain and worth all efforts.[36]

Geothermal Sources

The temperature of earth increases with depth at the rate of about 2°F every 100 feet in rock and at about 3 times the rate near natural steam sources. The thermal energy in deep underground zones can be recovered either by tapping natural steam or hot water sources, or by fracturing dry hot rock formations and flooding them with water to generate steam for use in a conventional steam-electric power plant. As this energy is withdrawn it is replaced by the normal conduction of heat from below. The installed electrical capacity in 1980 of such power plants was about 1350 MW in the world, the capacity of heat plants about 5500 MW. Recent surveys indicate a potential electrical output of 20,000 MW from natural steam sources in California's Imperial Valley, with a like amount available in Oregon. The use of nuclear explosives to fracture hot rock formations can extend the availability of this resource considerably and is under study by the AEC.

In comparison with other energy sources the presently used geothermal energy is extremely low, even if it may locally be of great importance. There is no hope that in the near future its contribution to the total energy consumption will be more than 1–2%.

Tidal and Wave Power

The energy resource of tidal power results from gravitational interaction between the earth and the moon. It may be harnessed by constructing dams in tidal areas to create a difference in elevation between two bodies of water. The potential energy in the higher body of water can then be recovered to produce electrical power by the same techniques used in hydroelectric plants. The maximum energy recoverable is proportional to the area of the basin and to the square of the tidal range. In North

TABLE 14

Foreign crude oil production by area and sulfur content category

Area and sulfur content range, weight %	1996 production 10^6 bbl	%
Africa:		
0.00–0.25	637	63.7
0.26–0.50	144	14.4
0.51–1.00	216	21.6
1.01–2.00	—	—
>2.00	3	0.3
Canada:		
0.00–0.25	112	35.3
0.26–0.50	18	5.7
0.51–1.00	107	33.8
1.01–2.00	40	12.6
>2.00	40	12.6
Middle East:		
0.00–0.25	—	—
0.26–0.50	—	—
0.51–1.00	—	—
1.01–2.00	1509	44.8
>2.00	1862	55.2
South America:		
0.00–0.25	24	1.6
0.26–0.50	19	1.3
0.51–1.00	52	3.5
1.01–2.00	225	15.2
>2.00	1161	78.4

Source: McKinney, C. M., and E. M. Shelton, Sulfur Content of Crude Oils of the Free World, US Dept. of the Interior, Report of Investigations 7059, 1967, Washington, DC.

America various sites on the Bay of Fundy have a total potential for the production of about 29,000 MW of electrical power, although little of this can be developed economically.

The use of wave power of the oceans was proposed repeatedly and in different systems, firmly installed at the beach or near the beach swimming. Theoretically it is possible to use the wave power and this can be demonstrated on models. But in great quantities it will not be economical, and as the aim is the production of electrical energy, it would be necessary to install enormous conventional stand-by plants because of its irregular turn.

Biomass

In the ecology the entire living components of ecological systems are called biomass. Biomass is subject to dynamic changes because phases of production follow phases of consumption and decomposition. As this runs by law of nature but principally is controllable by suitable production terms, there arose already before the productional ecological research out of practice the agricultural production oriented economy which is, under several aspects, only a synonym for productional ecology. The productional ecology integrates the biological energy flow. As with each productional process a portion of the absorbed energy at living organisms is used to maintain metabolism, there remains from the gross production only a fraction as usable net production. We must consider the turnover which makes the production, and this turnover is not comparably high competed with the hard energy.

Another problem is that the use of biomass as industrial initial position for energy production at equal conditions for the protection of the environment can easily become uneconomical. The conclusion is that biomass will reach only a small percentage of the total energy volume, should be used in small plants of conversion only, brings principally advantages only for decentralized small consumers, and has no economic basis should the regulations for the protection of the environment be valid also for big biomass systems.

Sun[37,38,39,40,41]

The rate at which solar energy intercepts the diametral plane of the earth is about 17.7×10^{10} MW. At the surface of the earth the average solar power is about 1820 Btu/ft^2-day but this figure can vary widely depending upon the location and, of course, is only available during sunny periods.

The principle difficulty in utilizing solar thermal energy for endothermic reactions is thus not so much matching the technology of the power plant to the solar-specific conditions, but to tie in the processes for which the power demand is usually constant to the fluctuating conditions of insolation. The problem of intermittent energy supply originating in the day/night cycle has therefore to be countered by the development of storage devices, hybrid strategies or by selecting interruptable processes, before large-scale utilization of solar energy can be considered.

Possibilities which show promise for the future are the development of technically efficient devices suitable for commercial exploitation, and a similarly attractive proposition appears to be a method of storing the fluctuating solar energy in processes which are interruptable, or at least can be modified over time. Conversion of such processes to solar power appears conceivable, and likewise simultaneous utilization of thermal and photochemical energies could result in advantages that up to now have not been fully exploited. Possibly hydrogen produced by solar power will play a part in the storage and transportation of energy from solar plants. In all solar thermal power plants, the low radiation flux density of the incoming solar radiation is collected and concentrated by a field of mirrors, referred to as collectors or heliostats. As much as possible of the radiation energy is absorbed and converted to energy. The balance of plant elements correspond to the conventional components of a normal power plant, but they have to be matched to the solar-specific mode of operation.

Figure 5 shows the general arrangement of a central receiver system power plant, which shall be briefly

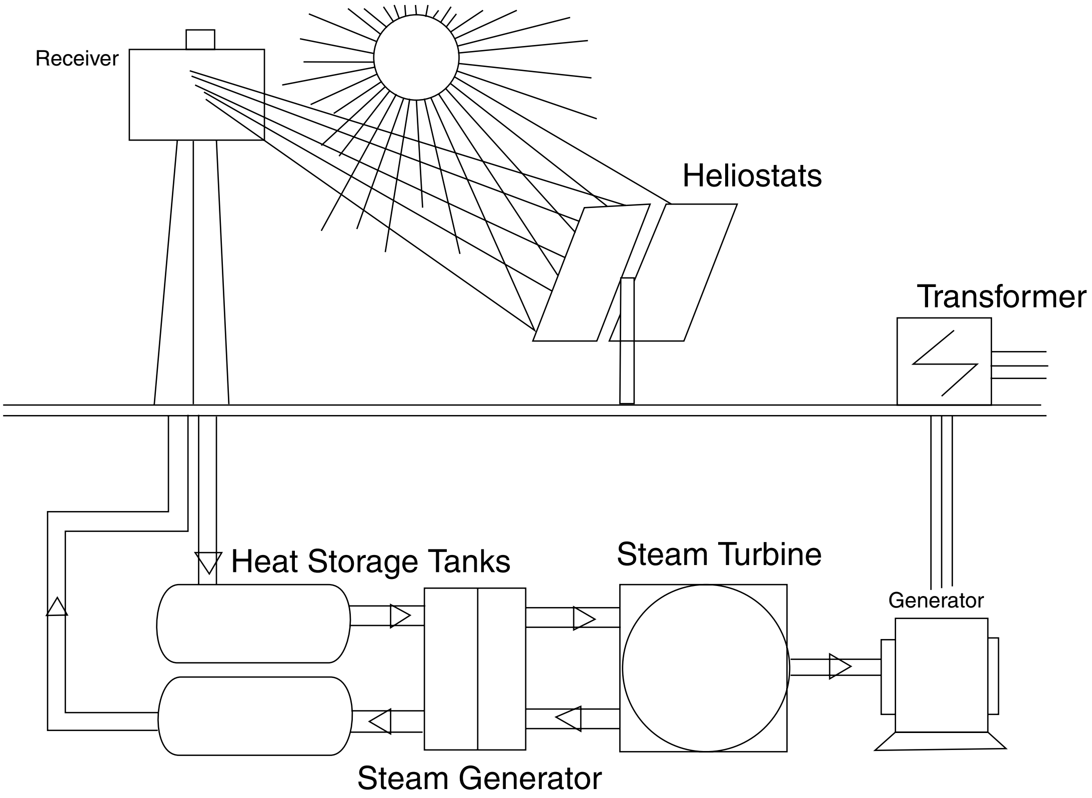


FIGURE 5 General arrangement of Central Receiver Systems (CRS).

described in the following as a example of a solar power plant. In this CRS (Central Receiver System) tower concept, heliostats reflect all incident radiation to a central receiver at the top of a tower. Within this receiver, temperatures ranging from 500 to 1000°C can be obtained. Table 15 lists the tower plants constructed up to now, of which solar 1 in Barstow in California, USA, Table 16, is the largest at 10 MW. For some 10 years, such prototype power plants have been designed, constructed and tested in different parts of the world. In their key characteristics, such as technology, conceptual design, size, heat carrying medium and power may differ from each other. For the tower power may differ from each other. For the tower power plants Solar 1 (Barstow, USA), CESA 1 (Almeria, Spain), Sunshine (Nio, Japan) and Eurelios (Adrano, Italy), steam was employed as the heat-carrying medium. The SSPS plant in Almeria, Spain, was operated with liquid sodium; the French power plant Themis used a molten salt for this purpose.

Since the beginning of 1987, the plants have been brought together under one organization under the leadership of Spain to make up the "Platform Solar de Almeria" (PSA) and now represent the European centre for solar thermal energy research.

Research work in the future will result from consultations between West Germany and Spain, as well as the participation of other nations in the SSPS project (participants: Switzerland, West Germany, United Kingdom, Italy, Sweden, USA). The activities are directed to the three following areas:

- maintenance and upgrading of the plants
- high-temperature technology of the components (in particular concentrator, receiver, storage device)
- utilization of the radiation and thermal energy during endothermic reactions in chemical processes for the production of transportable fuels and chemicals

In the farm concept, which is under investigation not only in Almeria but also in the USA and Australia, solar radiation is focused by parabolic troughs onto pipelines under glass functioning as receivers. The maximum temperatures attainable with such systems are in the range 300 to 400°C depending on the oils used as heat-transfer media.

Direct conversion of solar energy into electrical energy is, environmentally, the most acceptable way of electrical power generation. Solar cells are already supplying spaceships with electrical energy. For the construction of solar power plants, concept and feasibility studies are in hand. Even if electricity generation in this type of plant does not fulfill conventional ideas of "profitability" great importance must be attached to attempts to introduce environmentally-neutral systems of power generation into the world. These endeavours are

TABLE 15
Data for solar power plants

Name	SSPS	Eurelios	Sunshine	CESA 1	Themis	Solar One
Location	Almeria	Adrano	Nio	Almeria	Targasonne	Barstow
Country	Spain	Italy	Japan	Spain	France	USA
Completion year	1981	1981	1981	1983	1982	1982
El. output (MW)	0.5	0.7	0.8	1.0	2.5	10
Heat transfer fluid	sodium	steam	steam	steam	molten salt	steam

TABLE 16
Principal data of Solar One in Barstow, California, USA

Reflecting heliostat area	71,447 m^2
Heat transfer medium in receiver	steam
Storage capacity equivalent to	28 M Whe
Net electrical output	10 M We

worthy of generous support by governments. In any case, the studies will be continued, and progress is considerable.

Refuse

Solid waste products including paper, rags, cardboard, and plastics may be considered as an energy resource. Combustible wastes may be used as a fuel for electric power production and process heat and have a heating value of about 5000 Btu/lb. If fully utilized, these wastes could furnish about 10% of the fuel needs of central station power plants. Environmental problems are encountered in burning this material just as in the incineration of waste, and emissions depend upon the chemical characteristics of the refuse.

An alternate use of refuse, studied by the Bureau of Mines,[42] involves its conversion by pyrolysis into fuels such as combustible gases and light oils as well as other useful by-products. A ton of urban refuse containing mainly plastics can be converted, by pyrolysis at 900°C, into 328 lb of solid residue, 1.5 gal of tar, 0.5 gal of light oil (mostly benzine), 97.4 gal of aqueous liquor, 31.5 lb of ammonium sulfate, and 18,058 ft^3 of gas consisting of hydrogen, methane, CO, and ethylene. Other processes may be used to produce predominantly liquid fuels rather than a gaseous product.

Agricultural wastes, from animals and vegetation, are more abundant than urban wastes and constitute a large, though difficult to exploit, energy resource.

ENERGY CONVERSION

Most of the air pollutant and waste heat discharges attributable to energy-related activities are released in the energy conversion step. Here, energy resources are consumed to provide other energy forms either for direct utilization or for transmission to centers of energy demand.

The range of technological alternatives available in energy conversion is illustrated in Figure 6. The columns and rows represent input and output energy forms, respectively, with specific devices or power plants capable of making the transformation indicated in the appropriate block. The most widely used systems involve chemical (fossil fuel) and nuclear energy inputs and the following conversions or series of conversions:

1) For space heat (non-electric);
   chemical → thermal.
2) For the generation of electricity;
   Chemical or nuclear → thermal → kinetic → electrical.
3) For the propulsion of a vehicle (non-electric);
   chemical → thermal → kinetic.

In the conversion from thermal to kinetic energy a power cycle is used and its efficiency is limited by the second law of thermodynamics to a value less than that of the theoretical Carnot cycle. The Carnot cycle efficiency is equal to the difference between the temperatures at which heat is supplied to and rejected from the cycle working fluid, divided by the absolute temperature at which the heat is added. This theoretical limit on conversion efficiency increases for a given heat rejection temperature as the temperature at which heat is supplied to the cycle working fluid is raised.

A 1968 inventory of prime movers that are involved in the kinetic energy portion of various conversion systems is given in Table 17. Although automotive engines make up over 90% of the total installed horsepower, this figure is deceptive because of differences in the annual usage of the various power plants. A better indication of their relative use can be obtained from fuel consumption data.

The state-of-the-art in energy conversion systems and associated operations has been reviewed in a study performed for the OST[43] and by the FPC in both its 1964 and most recent 1970 National Power Survey.[29] It is convenient in reviewing alternative power devices that might be used to reduce harmful environmental effects, to maintain the distinction between centralized and decentralized power plants, where the latter may be either stationary or mobile.

From To	Electromagnetic	Chemical	Nuclear	Thermal	Kinetic (mechanical)	Electrical	Gravitational
Electromagnetic		Chemiluminescence (fireflies)	Gamma reactions (Co^{60} source) A-bomb	Thermal radiation (hot iron)	Accelerating charge (cyclotron) Phosphor	Electromagnetic radiation Electroluminescence	Unknown
Chemical	Photosynthesis (plants) Photochemistry (photographic film)		Radiation catalysis (hydrazine plant) Ionization (cloud chamber)	Boiling (water/steam) Dissociation	Dissociation by radiolysis	Electrolysis (production of aluminum)	Unknown
Nuclear	Gamma-neutron reactions ($Be^9 + \rightarrow Be^8 + n$)	Unknown					
Thermal	Solar absorber (hot sidewalk)	Combustion (fire)	Fission (fuel element) Fusion		Friction (brake shoes)	Resistance-heating (electric stove)	Unknown
Kinetic	Radiometer Solar cell	Muscle	Radioactivity (alpha particles) A-bomb	Power cycles (thermal expansion turbines, combustion engines)		Motors Electrostriction (sonar transmitter)	Falling object (water)
Electrical	Radio antenna Solar cell	Fuel cell Batteries	Nuclear battery	Thermoelectricity, Thermionics, Thermomagnetism, Ferroelectricity	MHD Conventional generator Piezoelectricity		Unknown
Gravitational	Unknown	Unknown	Unknown	Unknown	Rising object	Unknown	

Source: Adapted from USAEC matrix.

FIGURE 6 Energy conversion matrix.

Centralized Power Plants

To take advantage of economies of scale, electrical energy is generated in large centralized facilities for distribution to consuming activities. The demand for electrical energy varies on a daily, weekly, and seasonal cycle. Since this form of energy cannot be conveniently stored, the installed generating capacity must match the peak demand and some of this capacity must be idle during periods of low demand. The more economic plants in the system, which may be fossil or nuclear facilities operating on a steam power cycle or hydroelectric plants, are usually operated as much as possible in a so-called baseloaded condition, while the older and less efficient plants and special gas-turbine installations are utilized only during periods of peak demand. The use of pumped storage can serve to flatten out the demand curve by using excess power produced during periods of low demand to pump water uphill where it can be used during later periods of high demand by recovering its potential energy. Transmission ties between electric utility systems in different regions of the country can be used to take advantage of load diversity since the timing of seasonal and daily peak demands varies depending on location.

Electric generating plants with an output of over 1000 MW are not uncommon and stations of up to 10,000 MW capacity are anticipated by the year 2000, perhaps involving an energy center concept where the power source is integrated on the same site with industrial, agricultural, and/or desalination operations. The concentrated environmental effects of large central station power plants are receiving increased attention and have been the subject of congressional hearings.[40] Natural gas and low-sulfur residual oil may be used in steam-electric plants with minimal air pollution but these fuels are limited in their availability. The provision

TABLE 17

Total horsepower of all prime movers, 1968 (10^6 horsepower)

Item	Horsepower
Automotive[a]	16,854
Nonautomotive	
Factories	52
Mines	43
Railroads	58
Merchant ships and vessels	20
Farms[b]	292
Electric central stations	372
Aircraft	137
Total	17,828

[a] Includes cars, buses, trucks, and motorcycles.

[b] Includes about 1.3 × 10^6 hp in work animals and 22,000 hp in windmills.

Source: Statistical Abstracts of the US, 1970.

of control systems for SO_2 and NO_x emissions from steam plants fired with coal or high sulfur oil will increase the cost of power generation. Baron[41] estimates that flue gas scrubbing would add about 7% to the cost of power generation in a 1000 MW output plant using oil with a 1% sulfur content. The provision of dry cooling towers that reject waste heat to the atmosphere with no waste usage would increase the cost in the same plant by another 30%. The EPA has recently proposed[46] emission standards for coal-fired steam generators that would limit particulate releases to 0.2 lb per 10^6 Btu, SO_2 releases to 1.2 lb per 10^6 Btu, and NO_x to 0.7 lb per 10^6 Btu heat input.

The necessity to substitute oil by other energy agents as well as the little acceptance of nuclear power plants by the population of industrialized countries may lead for a certain time to a renaissance of conventional coal and lignite fired power plants but with a more sophisticated equipment for protection of environment. This protection will not include dry cooling towers, but washing equipment of flue gas to wash out partly SO_2 and NO_x as well as precipitators for dust. This environment protecting equipment increased the cost of the last built 750 MW coal fired power plant in Germany up to 27%. With full desulfurization and wet cooling tower the provision of environment protection will cost about 35%. It seems that there are two tendencies:

1) Big power plants of coal and lignite fired units with full SO_2 protection, and wet cooling tower (350 and 750 MW).
2) In the second place nuclear power plants with two and more units of 1000 and 1300 MW considering the difficulty to find new locations accepted by the population.

Improvements in the efficiency of energy conversion reduce the fuel consumption and attendant environmental effects directly for a given electrical output. Steam-cycle power plants operating with fossil fuels attain a thermal efficiency of about 40% with a maximum steam temperature in the power cycle of 1000° to 1050°F, which is well below the combustion temperature of these fuels. Current nuclear power plants operate with a maximum steam temperature of about 550°F and have a lower thermal efficiency. The cycle efficiency increases as the steam temperature is raised but the more severe temperature and pressure conditions would lead, in both fossil and nuclear systems to materials problems and higher plant capital costs that would more than offset the economic gains of increased efficiency. Topping cycles, which are added to the steam cycle and use other thermodynamic working fluids such as mercury or potassium in higher temperature ranges, therefore providing higher efficiency, have been studied for fossil-fueled systems and actually operated in the case of mercury but the gains appear to be marginal. Thermionic topping devices[47] that convert thermal energy directly to electrical energy have also been proposed and promise to increase the over-all efficiency of conventional plants to about 50% performance must be improved before they can be used to advantage. A combined cycle, involving a gas turbine topping unit that would operate at high temperature on gas produced from coal and would reject its heat to a conventional steam cycle, is another promising system and could have an overall efficiency in the 42 to 43% range.

Some promising alternatives for centralized power production that can utilize fossil resources are fuel cells and magnetohydrodynamic (MHD) generators. The number of nuclear reactor plants being installed by utilities is expanding and further developments in breeder reactors and fusion systems are also anticipated. Solar power plants are of great interest in the longer term as a clean energy source.

Fuel Cells

This direct conversion device is not limited in efficiency by thermodynamic considerations and can convert chemical energy directly into electrical energy without the intermediate conversion to thermal and kinetic energy. Over-all efficiencies as high as 75% have been attained in special purpose applications. With hydrogen and oxygen fuel fed to electrodes submerged in a suitable electrolyte, this device produces 0.7 to 0.85 volts (dc)/cell and many cells must be connected in series to produce useful voltages for transmission. Development work is in progress on cells that could operate on gasified coal or other gaseous fuels with air as the oxidizer and a conceptual design has been completed for such a plant.[48]

Air pollution and waste heat discharges would be reduced in a fuel cell power plant because of the improved efficiency of fuel utilization. If coal were used as the source of a gaseous fuel, the impurities in the coal and particulates would have to be controlled in the gasification plant, but such control would be easier to provide in this operation than in a conventional power plant. In the course of the development of fuel cells, it is more likely that they will be applied first to smaller decentralized power systems.

Magnetohydrodynamics

MHD generators can utilize either a fossil or nuclear thermal energy source. In a fossilfueled system, hot combustion gases, seeded with potassium or cesium to make the gas conductive, are expanded at high velocity through a magnetic field. The dc current produced in the moving conductive gas is picked up at electrodes embedded in the walls of the gas channel in various geometries depending upon the particular type of generator. The advantage of this system is its high thermal efficiency of 50 to 60% when operated at a gas temperature in the 4000° to 5000°F range. A panel[49] has conducted a study of MHD for the OST and concluded that additional research and development should be performed before an MHD plant is constructed. Much of this development effort must be directed toward materials research to permit extended operation at high temperatures.

The environmental effects of MHD generation would be similar to those in any fossil-fueled combustion system but would be mitigated because of the improved efficiency of conversion. The seed material in the gases could not be released in any case, for economic as well as environmental reasons, and must be recovered. It is anticipated that recovery of the other pollutants can be accomplished in the same operation.

Fission Reactors

Current light-water cooled nuclear reactor systems used for electric power production operate with fuel slightly enriched in the uranium-235 isotope and have a conversion ratio of about 0.6, i.e., they generate about 60% as much new fuel as they consume by converting fertile uranium-238 into fissionable plutonium-239. These plants have a thermal efficiency of about 33% because of limitations on the temperatures at which they can operate. Alternate reactor systems that are under study, including the molten salt reactor (MSR) and the high temperature gas-cooled reactor (HTGCR), can generate steam at conditions comparable to those of fossil-fueled plants and the molten salt system, when fueled with uranium-233 fuel that is produced from thorium by neutron capture in a reactor, can breed about 5% more fuel than it consumes. All of these systems operate with a thermal (or slow) neutron spectrum in the reactor. By operating with a higher energy neutron spectrum in a fast reactor, better neutron economy can be achieved and more neutrons are available for the conversion of fertile material. Such reactors may be cooled with helium, steam, or liquid sodium. These fast breeder reactors, when developed and integrated into the centralized power system, will greatly extend the nuclear energy resources by fully utilizing the more abundant uranium-238 isotope to produce additional fissionable fuel. An analysis of the various reactor concepts and their potential roles in a nuclear power economy has been performed by the AEC[50] and led to the selection of the liquid metal cooled fast breeder reactor (LMFBR) as the high priority development task.

The various topping and MHD systems discussed previously are also applicable to nuclear power plants and may be useful additions if higher reactor operating temperatures can be reached.

All reactor systems generate similar quantities of radioactive materials for a given thermal energy output and the handling of these materials is closely controlled during all fuel cycle and reactor operations. The amount of such materials produced, and waste heat discharged, for a given electrical output depends, of course, on the plant efficiency. It is important to recognize that in all reactor systems the cooling water that carries away the waste heat is physically separated from the reactor primary coolant whether it be gas, water, or liquid sodium. Because of the large quantities of radioactive materials present in the fuel region of a reactor during operation, there is concern over incidents that might accidentally discharge this material to the environment and no nuclear plants have been constructed in regions of high population density. Protection against such discharge is provided by safety systems that minimize the possibility of an accident and by incorporating several physical barriers within the plant to contain the radioactive products in the unlikely event that they are released from the fuel region.

Fusion Reactors

The successful development of a controlled and economic fusion device would make available the tremendously large energy reserves of deuterium in sea water. Current research is directed toward devices in which the combination of plasma temperature, density, and confinement time required for a sustained reaction may be attained. The status of this research and the outlook for fusion power has been summarized by Rose.[51] The deuterium-tritium (D−T) reaction, wherein these materials are confined as a plasma at a temperature of about 40×10^{6}°C in a high strength magnetic field while they react, releases most of its energy in the form of neutrons that must be captured in a lithium blanket to generate new tritium fuel. The blanket fluid may then be circulated to generate steam for a conventional steam-cycle system. Topping cycles may also be used depending on the operating temperatures and the over-all efficiencies will be competitive with fossil-fired power plants.

The deuterium-deuterium (D−D) reaction appears to be more difficult to attain, requiring confinement at a temperature of some 350×10^{6}°C, but releases much of its energy in the form of charged particles which might be used to generate electricity directly without using a power cycle. Higher operating efficiencies may be achieved in such a system. There are still other fusion reactions that are of interest because of their special characteristics.

It now seems that fusion power will be more economical in very large plants of greater than 5000 MW capacity and such plants will pose substantial waste heat disposal problems unless very high efficiencies are attained. In the D−T concept, tritium must be recovered from the blanket and handled during the fuel cycle. Another source of radioactive products the structural material that is activated by the neutrons produced in either reaction. In any case, the problems of disposing of radioactive materials appear to be less severe than those in fission reactors and, because of the small amounts of fuel material present in the device at any given time, the accident hazard is apparently minimal.

Solar Power Plants In-Orbit and Terrestrial-Based Systems

Several concepts for utilizing solar energy in a centralized power plant have been proposed. Glaser[52] outlined a conceptual system that uses a satellite with a large array of solar cells placed in a synchronous orbit. The solar cells would convert the solar energy to electricity and this energy would be beamed by microwave to a receiving station on earth. To supply the annual electrical energy demand of the northeast region of the US in 1980, estimated at 1.7×10^{15} Btu, a 105 square mile array of silicon cells would be required on the satellite and the receiving antenna on earth would be of the order of 20 square miles. Another scheme proposed by Meinel[53] involves a terrestrial-based plant that would capture thermal energy from the sun by means of specially coated surfaces to provide a "greenhouse" effect. A working fluid would be heated to a temperature of about 1000°F for use in a power cycle to produce electricity. A terrestrial-based system of 1000 MW electric capacity would require a land area of about 5.5 square miles. The satellite concept, like the utilization of fossil and nuclear resources, does add to the thermal burden that must be dissipated to the environment, while the terrestrial-based solar plant merely implies a redistribution of energy that impinges on the earth in any event.

Single and Dual Purpose Power Plants: Advantages and Problems

The more simple service of electricity generating units is the operation of single purpose power plants, on the other hand a second purpose can help to increase the efficiency of the plant, using for instance, the waste heat for district heating systems or desalination of sea water. Naturally the structure of the regional or national energy economy must allow the technical and economical interconnection of two aims. The increasing cost of fuel oil and natural gas may promote the use of fossil-fueled and nuclear power plants for the district heating in high populated towns and regions. When fully used for the two purposes the efficiency of a coal-fired power plant can increase up to 75 to 80%. But the high investment costs of big district heating systems and the take out of heat in the power plant brings the advantage of higher efficiency only during the winter period, in the summer time there will be needed in parallel the normal wet cooling tower. This is only one argument, the other is, that the coal, fired in the plant boiler, substitutes fuel oil unburdening the national balance of payment from the import cost for petroleum. If suitable a dual purpose power plant may be taken into consideration in the prefeasibility study period of new projects and here, in particular, also under the aspect that district heating systems could result in a dramatic reduction of air pollution in towns.

Decentralized Stationary Power Plants

Energy conversion systems currently utilized in this category involve mainly combustion devices for space heat and industrial process heat applications. For economic reasons they are generally not as amenable to as high a level of emission control as larger centralized systems, save for conversion to cleaner fuels such as natural gas and perhaps methane or hydrogen produced from coal. These applications are also candidates for conversion to electrical energy thereby shifting the emission problem from decentralized plants to centralized power generating facility.

Decentralized electric power generating facilities of up to 5 MW output are sometimes installed in shopping centers and housing complexes. These now use diesel generators or gas turbines but could also use fuel cells when these are developed. Waste heat from these power plants is used for space heating and air-conditioning, and it is claimed[54] that up to 85% of the energy in the fuel may be effectively utilized in these systems as compared with only about 40% in central station power plants.

The use of solar energy for individual home space conditioning and water heating will further on increase but involves a substantial capital investment in these applications. The payback time for the investment costs, however, will become shorter, because the natural gas, fuel oil, and electric costs will rise in future overportional to the cost of primary energy.

Nuclear reactors are not competitive in small decentralized applications for economic and technical reasons partially attributable to the requirement for shielding and other auxiliary systems. They are used, like other new devices such as fuel cells, in special military and space applications and may have potential in very large aircraft. Certain isotopes, produced by neutron irradiation in a nuclear reactor, have also been used as energy sources for small power systems in space and remote terrestrial applications.

Mobile Power Plants

Energy conversion devices used in motor vehicle, rail, and aircraft applications must provide acceptable performance characteristics in terms of horsepower–weight ratio and specific fuel consumption. The gasoline-fueled, internal-combustion (I-C) engine has dominated the motor vehicle field and the gas turbine has achieved the same level of preeminence in aircraft. An appraisal of automotive transportation in relation to the problem of air pollutant was performed by a Department of Commerce panel.[55] Alternate systems were reviewed and a five-year development program was recommended to support innovative developments on energy sources, propulsion systems, and emission control devices. Alternative energy sources were considered further in system studies performed on electric vehicles by A.D. Little[56] and on the whole spectrum of unconventional low pollution-potential power sources by Battelle Memorial Institute.[57] It also is possible to use alternate fuels such as propane, methanol, natural gas, and hydrogen in I-C engines to reduce the level of emissions. Propulsion devices that may have some future potential include gas turbines, external combustion systems operating with a steam of organic Rankine cycle or with a Stirling cycle, fuel cells, and battery driven electric motors. The characteristics of some

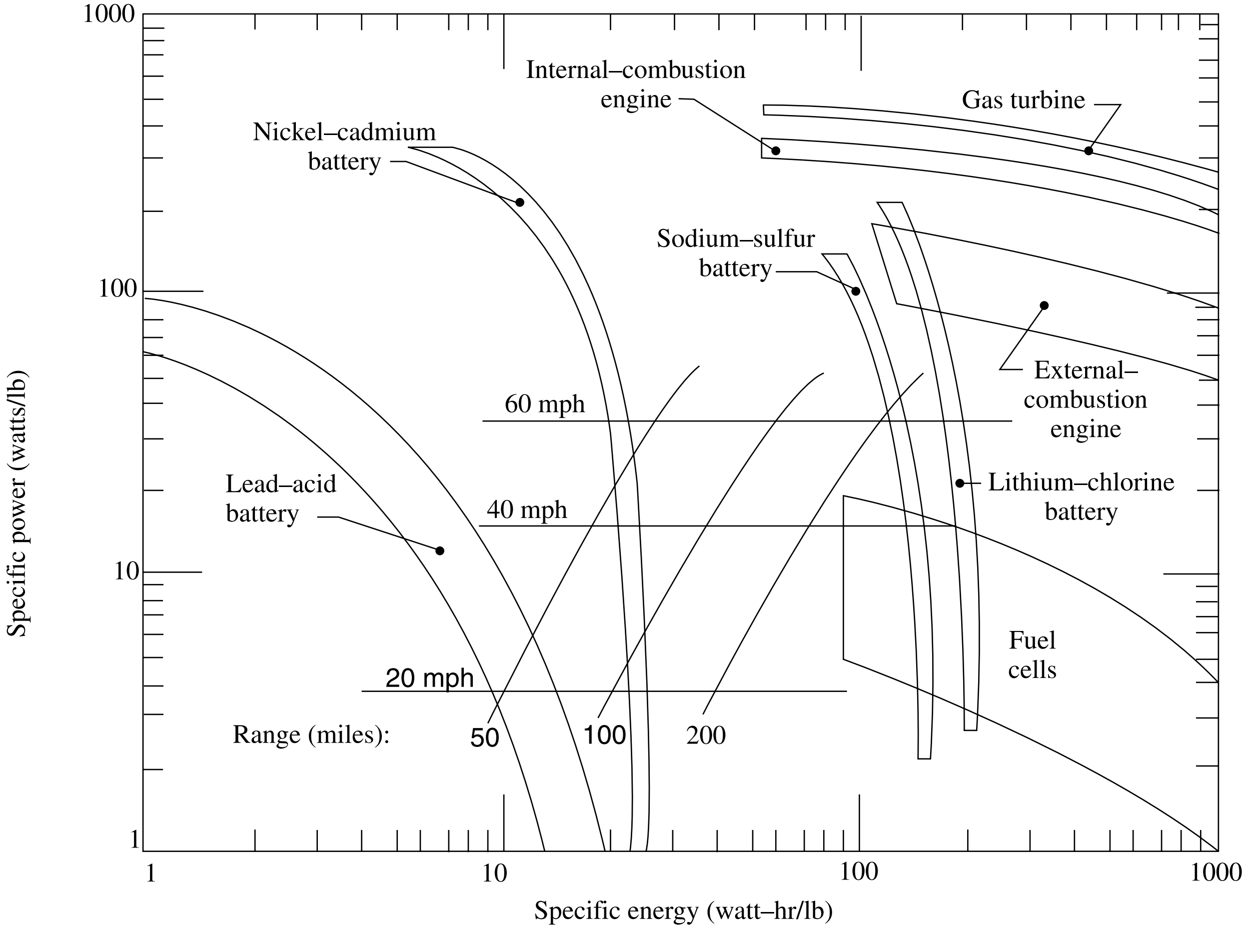


FIGURE 7 Vehicle requirements, and characteristics of mobile power plants for a 200 lb vehicle with a 500 lb power source.

alternative power plants are illustrated in Figure 7 which indicates the range of specific power in watts/lb and specific energy in watt-hr/lb available with these devices. The propulsion system requirements for a 2000 lb vehicle propelled with a 500 lb power source at speeds of 20, 40, and 60 mph, and for vehicle ranges of 50, 100, and 200 miles are also indicated. From these characteristics it may be seen that the gas turbine and external-combustion engine are the systems most competitive with the current I-C engine. Various types of battery systems are shown and these as well as fuel cell systems may be applicable to limited-purpose vehicles. The sodium–sulfur and lithium–chlorine systems may approach the performance of combustion engines but are still in the development stages. Introduction of a battery driven vehicle will shift the emission problem from the automobile to a central station power plant where the electrical energy required to recharge the battery would be generated (see section: The Problem of Transport).

Emissions from automotive power plants are variable and depend upon engine, load, level of maintenance, and individual driving habits. To allow for variable engine loads, emission standards are often related to a given driving cycle involving specified warm-up, acceleration, deceleration, cruise, and idel periods. The hydrocarbon, CO, and NO_x emission standards now in effect in California and those established by the EPA for 1974 concurrent are shown in Table 18 along with emissions from a representative uncontrolled gasoline fueled I-C engine as well as alternative power plants. The pollutants from current I-C engines have been reduced from uncontrolled levels by control of gasoline evaporation sources and by crankcase ventilation systems that recirculate cyclinder blow-by gases back to the combustion chamber. Further reductions of CO and hydrocarbon releases, especially NO_x, will involve high temperature or catalytic reaction of exhaust gases in unregulated or regulated catalysts, which are more and more applied in Gasoline I-C engines. The advantages of the gas turbine and external combustion engines, which can operate on distillate fuels such as kerosine, over the I-C engine lie in the better control of combustion conditions that can be achieved by separating the fuel combustion chamber from the working parts of the engine. The emission figures for a battery driven electric vehicle indicated in Table 18 were developed by Agarwal[58] on the basis that the electrical energy is provided from a central station

TABLE 18
Emissions from alternate vehicular power plants

	Emissions (g/mile)		
	HC	CO	NO_x
California standard, 1971[a]	2.2	23	4
EPA standard for 1975[b]	0.41	3.4	3
EPA standard for 1976[b]	0.41	3.4	0.4
Gasoline I-C engine, no controls	11.6	64	5.6
Natural gas I-C engine[a]	1.5	6	1.5
Diesel engine[c]	3.5	5	4
Regenerative gas turbine[c]	0.32	3.5	1.9
Stirling engine[c]	0.1	1.0	2.6
Steam engine[a]	0.2	1.0	0.4
Electric battery vehicle[d]	0.04	—	4.3

[a] Congressional Record H 903, February 17, 1970.

[b] Federal Register, July 2, 1971, pp. 12658.

[c] Personal communication, R. Tom Sawyner.

[d] Reference 58; emissions in this case are from a coal-fired central power plant and also include about 16.25 g of SO_2/mile and 2.1 g particulates/mile.

power plant. It was assumed that 0.95 lb of coal was burned at the power plant for each mile of vehicle travel. In this case the emissions would also include an SO_2 release of 16.25 g/mile if the coal contained 2% sulfur and no control were provided. This actual situation however will change when coal fired centralized power plants are equipped with desulfurization installations or when in the future the percentage of electricity produced in nuclear power plants will be much higher.

Diesel engines used in trucks, buses, and trains are accountable for a small fraction of transportation vehicle emissions but have local objectionable effects such as smoke and odor. Aircraft engines have similar effects but in the long run may be able to effectively utilize alternate fuels such as methane and hydrogen that will reduce emissions.

The expansion of mass transportation facilities to reduce automobile usage, especially in urban areas, is an effective way to utilize energy more efficiently in moving people from place to place. A bus carrying 50 people and traveling at a speed of 60 mph may consume about 0.003 gallons of fuel/passenger mile, while an automobile carrying three people will consume about 0.023 gallons of fuel/passenger mile traveling at the same speed with a corresponding increase in emissions and traffic congestion.

Disintegrated Methods

Analyzing alternative energy sources and alternative energy economies the centralized social order ruling presently was compared with disassociated way of life. For this alternative social order disintegrated methods of energy supply were suggested. These suggestions start from the point of view that the effective (final) energies heat, mechanical work and light can also be produced in small closed systems. As they must not be available steadily, the temporal not-necessity of these models get a favorable valuation in the ranking list. Investment for connection, transmission and interconnection become inapplicable as well as measures for environment. Such considerations lead to self-operating island stations on the basis of a suitable primary energy, for example, crude oil or diesel fuel which through conversion may produce mechanical work, heat, and light. Also biomass or bioenergies are part of such computer models. These disintegrated methods are possibly suitable for peripheral agricultural structures and may relieve the world energy supply unessentially. In industrialized countries they cannot be applied.

The Problem of Transport

The demand on kinds of effective energies, especially in industrial countries with cool climate are generally divided as follows:

	percentage of total supply	possibilities of improvement of efficiency of conversion
heat	60–69%	10–15%
mechanical work	25–35%	3–4%
light	1–15%	less than 1%

whereby chemical energy as effective energy, which is hard to inquire statistically and without important share, shall be neglected here.

In Germany the effective energies were supplied to 43.6% to private households, 36.8% to the industry and 19.6% to the field of transport and traffic (all figures valid for 1976).[59] If in the sector transport and traffic the above mentioned 19.6% (48.6 Mio tons SKE) would be again 100%, the following figures refer to

road	85.8%
railroad	5.1%
airlines	6.4%
inland navigation	2.7%

The main part of energy (and that mainly converted from crude oil) flew into the road traffic—8% of the total primary energy of the FRG. Railroad traffic, i.e., “Deutsche Bundesbahn” and private railroad companies, consumed only 5.1% or 0.8% of the total primary energy consumption. Though road traffic consumed tenfold of primary energies, the output of transport amounted only to 2.3-fold of the output of railroad (1976). Hence it follows that:

1) Railroad is with regard to energy consumption more economical than road.
2) The electrification of the railroad has reduced within 10 years the primary energy consumption to round about 50%.
3) Road traffic offers the greatest possibilities for saving energy and for the substitution of crude oil products by electricity.

A favourable combination for the energy economies of industrial countries would be:

1) A far-going electrification of railroad.
2) An increased portion of transport and traffic on electrified railroad.
3) Electrification of road traffic by electric-driven cars (battery cars).

As an electric-driven car with battery at present can drive about 200 miles, it can be suggested that for long-distance traffic the railroad should be used and for city traffic the electric car. For trips which by all means cannot be made by electrified railroad nor by electric cars, the conventional motor car should be available (as far as possible by rented cars). This combination would extend the world stock of crude oil, but for the transport sector it would mean a great change. Also the vehicle industry should help to carry this energy alternative. To bring "The Problem of Transport" into a favorable relationship to the above-mentioned 3-cornered problem energy/economy/ecology, it is unavoidable that, to a great extent, overland air traffic must be replaced by electrically-driven high speed trains travelling at 250 to 300 miles per hour. Environmental problems will make this development unavoidable.

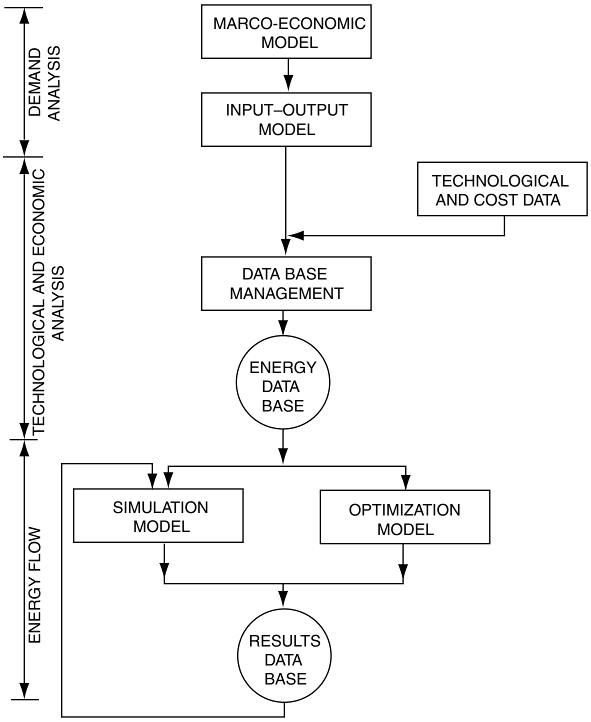


FIGURE 8 General flow-chart of the energy model.

Energy Strategies[33]

State and regional energy authorities in many countries bear the responsibility to bring about fundamental alterations of the energy economy, the ultimate object of which is to assure an adequate supply of clean, low cost, safe, and dependable energy. In general it can be assumed that population and national economies will continue to expand but that the extent and nature of this growth is as difficult to forecast as it is to accurately determine energy demands over a long period.

There are, as explained before, numerous alternative technical, and ecological economic solutions to these problems, comprising different priorities in the research and development of new energy systems, the choice of various primary energy sources, and the application of particular energy conservation and transport technologies. Further, non-technical elements of planning and shaping the energy economy lie in the enactment of tax, subsidy policy, and pricing.

The three main aspects of energy planning are as follows:

1) The problems presented by energy economies should not be considered in isolation. Rather, because of their central importance to the whole of a country's economy, these problems must be treated as sub-systems, interacting with other components in the economic nexus, such as technological concepts and environmental factors.
2) The energy system is mapped in a model. Depending upon the particular questions the model is to answer, three broad methods may be correspondingly applied: optimization models (linear programming, dynamic programming); simulation models; input/output analysis.
3) The model is mathematically designed for computer analysis. Interdisciplinary working groups, representing several scientific fields, are required to develop the model because of the complexity of the systems to be mapped and the many different areas to which the model's elements and subsystems belong.

This modelling method serves to furnish the basis for rational choices about energy supply as an integral part of a national (or regional) economy. After thorough investigation of the interrelation between the energy economy, the national economy and the environment, the knowledge of all technical and economic facts and influences as well as alternative possibilities will be given into formulated computer programs and become the foundation for long-term energy planning. The programs can simulate alternative strategies, check the effects of measures adopted, and make quantitative statements and forecasts about the energy economy. Figure 8 shows a simplified flow-chart of an energy model used for national energy studies.

In Figure 9 are shown the different steps from the inventory of energy resources with own sources, trade, and sources provided by trans-border energy transfer.

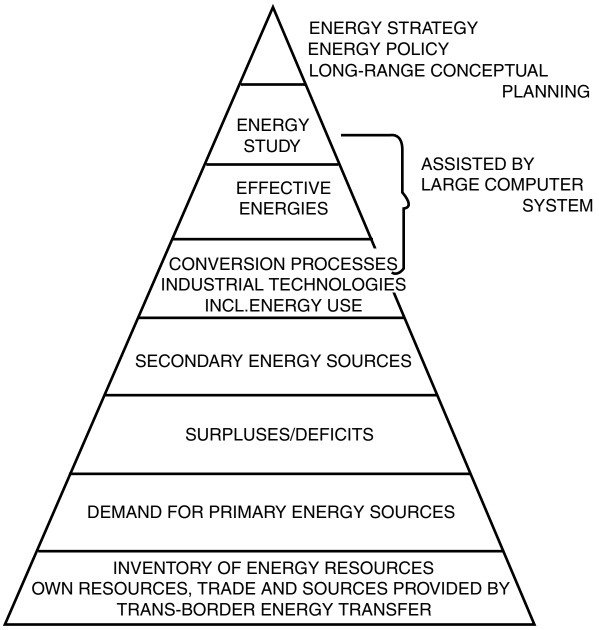


FIGURE 9 Energy pyramid analysis and strategy.

The implementation of a national or regional energy study yields the following results:

1) A long-range development plan for the energy system in terms of exploitation and extraction capacities, energy conversion and production capacities, total demand in relation to capital investment and population growth.
2) Annual consumption of primary energy and its structure.
3) An energy flow diagram.
4) Annual costs for each operation in the energy system.
5) Equipment requirements for each process and the additional capacity investment.
6) Ratio of indigenous to imported energy.
7) Sites for new plants.
8) Effects of new technologies on the energy system.

CONCLUSION

The role of energy in the well being and progress of a nation is obviously quite complex, involving many conflicting social, economic, and environmental factors. It is generally recognized that a national energy policy in many countries is urgently needed. The governments of especially the industrial countries have understood the dependence of their economy from a healthy energy economy and initiated steps to secure the basis for a reorganization of governmental and semi-governmental institutions engaged with energy and to be able to give reliable recommendations to the industry. The endeavours of governments and industries become more difficult because of the lack of healthy bases and the uncertainty how the numerous energy technologies will be developed and which availabilities and costs in the long-term future can be expected for the different primary energy agents. The energy program of the USA includes a broad program to accelerate the development of fast breeder reactors, SO_2 control techniques, coal gasification and liquefaction, and several other energy conversion techniques up to the bio-energy. Furthermore the program includes the development of leasing provisions for geothermal fields and shale oil resources and the acceleration of gas lease sales on the Outer Continental Shelf, steps that are intended to increase the availability of fuels which are or will be in short supply. The balance for the program outlined in the energy messages of the Government includes programs to utilize energy more wisely and specifies increased insulation requirements in Federal Housing Administration standards for federally insured homes. Increasing attention is being directed toward individual energy utilizing activities, such as space heating and cooling, to determine how more efficient systems and practices can serve to reduce per capita energy demands. However, the crucial revelation of the past decades is that, put simply, there is a relationship between energy and the environment fundamental to the survival of mankind—energy demand cannot be met by employing technological and economic means if over the long term the finite resources of the environment are consumed in the process. The solution of this problem in which energy, ecology and economy are, as it were, represented as being at the apexes of a triangle requires that human and artificial intelligence be deployed which are available but needs governments to channel their endeavours before it is too late, and to hinder vested interest from deflecting them from their course. In this respect, the USA has assumed a leading role. Subsequent efforts are being undertaken by other governments in industrialized or semi-industrialized countries and it may be of advantage to interchange any experience and know-how in the problematical fields of energy and environment.

A tremendous increase has been seen in renewable energy sources such as wind and solar power throughout the world.[61] The reduction in air pollution by these renewables results directly from the reduction in fossil fuel combustion (see www.refocus.net). Renewables are approaching 2% of the utility mix in some states of the United States. Wind parks, both on- and offshore can generate many megawatts with large wind turbines producing power on the order of one MW each.[62]

REFERENCES

1. United Nations 1969 Statistical Yearbook, Statistical Office of the U.N., Department of Economic and Social Affairs, New York 1970.
2. Ayres, R.U. and A.V. Kneese, Production, Consumption and Externalities, *American Economic Review,* June 1969.
3. *Minerals Year-Book,* 1968, Vol. Vol. I–II, Metals, Minerals, and Fuels, USGPO, Washington, D.C., 1969.
4. Morrison, W.E. and C.L. Reading, An Energy Model for the United States, Featuring Energy Balances for the Years 1947 to 1965 and Projections and Forecasts to the Year 1980 and 2000, U.S. Dept. of the Interior, Bureau of Mines Information Circular 8384, USGPO, Washington, D.C., 1968.

5. A Review and Comparison of Selected United States Energy Forecasts prepared for the Office of Science and Technology by the Battelle Memorial Institute, December 1969, USGPO, Washington, D.C., 1970.
6. Schurr, S.H. and B.C. Netschert, *Energy in the American Economy, 1850–1975, An Economic Study of Its History and Prospects,* The Johns Hopkins Press, Baltimore, Md., 1960.
7. *The Economy, Energy, and the Environment,* A Background Study prepared for the use of the Joint Economic Committee. Congress of the United States, September 1970, USGPO, Washington, D.C., 1970.
8. *Electric Power and the Environment,* A report sponsored by the Energy Policy Staff Office of Science and Technology, August 1970, USGPO, Washington, D.C., 1970.
9. Rice, J.K., Coal Mining in an Oxygen-Free Atmosphere, American Society of Mechanical Engineers Winter Annual Meeting, Nov. 1970, New York, Paper No. 70-WA/PID-4.
10. *Nationwide Inventory of Air Pollutant Emissions 1968,* August 1970, U.S. Dept. of HEW, NAPCA Publication No. AP-73, USGPO, Washington, D.C., 1970.
11. *Compilation of Air Pollutant Emission Factors* (revised), U.S. Environmental Protection Agency, Office of Air Programs, Publication No. AP-42, 1972, USGPO, Washington, D.C., 1972.
12. *Control Techniques for Particulate Air Pollutants,* January 1969, U.S Dept of HEW, NAPCA Publication No. AP-51, USGPO, Washington, D.C., 1969.
13. *Air Quality Criteria for Particulate Matter, January 1969,* U.S. Dept. of HEW, NAPCA Publication, No. AP-49, USGPO, Washington, D.C., 1969.
14. *Environmental Quality,* the First Annual Report of the Council on Environmental Quality, August 1970, USGPO, Washington, D.C., 1970.
15. *Control Techniques for Sulfur Oxide Air Pollutants,* January 1969 U.S. Dept. of HEW, NAPCA Publication No. AP-52, USGPO, Washington, D.C., 1969.
16. *Air Quality Criteria for Sulfur Oxides,* January 1969, U.S. Dept. of HEW, NAPCA Publication No. AP-50, USGPO, Washington, D.C., 1969.
17. Bartok, W. *et al., Systems Study of Nitrogen Oxide Control Methods for Stationary Sources.* Vol. II, November 1969, PB 192, 789.
18. *Air Quality Criteria for Hydrocarbons,* March 1970, U.S. Dept. of HEW, NAPCA Publication No. AP-64, USGPO, Washington, D.C., 1970.
19. *Air Quality Criteria for Photochemical Oxidants,* March 1970, U.S. Dept. of HEW, NAPCA Publication No. AP-63, USGPO, Washington, D.C., 1970.
20. Bolin, B., The Carbon Cycle, *Scientific American,* Vol. 223, No. 3, September 1970.
21. McCaull, J., The Black Tide, *Environment,* Vol. 11, No. 9, November 1969.
22. Smith, J.W., Problems in Dealing with Oil Pollution on Sea and Land, *Journal of the Institute of Petroleum,* Vol. 54, No. 539, November 1968.
23. Martin, E.J. and R.D. Hill, Mine Drainage Research Program of the Federal Water Pollution Control Administration, Preprints of Papers presented before the Second Symposium on Coal Mine Drainage Research, Mellon Institute, May 14–15, 1968, Pittsburgh, Pa.
24. Miller, A. *et al.,* Use of Steam Electric Power Plants to Provide Thermal Energy to Urban Areas, ORNL-HUD-14, January, 1971 National Technical Information Services, U.S. Department of Commerce, Springfield, Va.
25. Rock, R.L. and D.K. Walker, Controlling Employees Exposure to Alpha Radiation in Underground Uranium Mines, U.S. Dept. of the Interior, Bureau of Mines, USGPO, Washington, D.C., 1970.
26. Schneider, K.J., Solidification and Disposal of High-Level Radioactive Wastes in the United States, *Reactor Technology,* Vol. 13, No. 4, pp. 387–415.
27. Martin, J.E., E.D. Harward, and D.T. Oakley, Comparison of Radioactivity from Fossil Fuel and Nuclear Plants, Nov. 1969, Appendix 14. Hearings before the JCAE, Congress of the United States, Environmental Effects of Producing Electric Power, October and November 1969, USGPO, Washington, D.C., 1969.
28. Solid Waste Management, Prepared by the Ad Hoc Group for the Office of Science and Technology, May 1969, USGPO, Washington, D.C., 1969.
29. National Power Survey, Federal Power Commission, 1964, USGPO, Washington, D.C., 1964.
30. Manners, G., *The Geography of Energy,* Aldine Publishing Co., Chicago, 1967.
31. Hubbert, M.K., *Resources and Man,* Chapter 8, Energy Resources, W.H. Freeman and Co., San Francisco, 1969.
32. Lovejoy, W.F. and P.T. Homan, *Methods of Estimating Reserves of Crude Oil, Natural Gas, and Natural Gas Liquids,* The Johns Hopkins Press, Baltimore, Md., 1965.
33. Möllenkamp, F.W. and E.C. Zoller, *Formulating Energy Strategies for Long-Term Energy Supply,* Energy Development, Muenchen-Graefelfing, 1980.
34. De Carlo, J.A., *An Alternative to Oil Imports: Crude Synthetic from Coal, Oil Import Controls,* hearings before the Sub-committee on Mines and Mining of the Committee on Insular Affairs, House of Representatives, 1970, USGPO, Washington, D.C., 1970.
35. Cap, F. and K. Schiepf, *Energieversorgung,* Probleme und Ressourcen Stuttgart, 1981.
36. Stegemann, D., Kernkraftwerk, *Enzyklopaedie Naturwissenschaft und Technik,* München, 1980.
37. Kesselring, P. and M. Becker, Further development of solar thermal energy technology, VDI Reports No. 704, 1988.
38. Hillesland, T. and P. De Laquil Results of the U.S. Solar Central Receiver Utility Studies, VDI Berichte Nr. 704, 1988.
39. Grasse, W., Design principles of solar thermal power plants—results and experiences from operating experimental plants, VDI Reports No. 704, 1988.
40. Fricker, H.W., PHOEBUS—a 30 MWe demonstration power plant, VDI Reports No. 704, 1988.
41. Brakmann, G., Experience with Small Solar Power Systems Project (SSPS) in Almeria, lecture New Delhi, April 1988.
42. Sanner, W.S. *et al.,* Conversion of Municipal and Industrial Refuse into Useful Materials by Pyrolysis, August 1970, Dept. of the Interior, Bureau of Mines, report of investigations 7428, USGPO, Washington, D.C., 1970.
43. Energy R & D and National Progress, Prepared for the Interdepartmental Energy Study by the Energy Study Group under the direction of Ali Bulent Cambel, USGPO, Washington, D.C., 1965.
44. Environmental Effects of Producing Electric Power. Hearings before the JCAE, Congress of the United States, October and November 1969, USGPO, Washington, D.C., 1969.
45. Baron, S., Cut Pollution at What Price? *Electrical World,* January 19, 1970.
46. *Federal Register,* August 17, 1971, p. 15706.
47. Thermionic Topping Converter for a Coal-Fired Power Plant, Office of Coal Research R & D Report No. 52, U.S. Dept. of the Interior, USGPO, Washington, D.C.
48. Final Report, Project Fuel Cell, Office of Coal Research R & D Report No. 57, U.S. Department of the Interior. USGPO, Washington, D.C., 1970.
49. MHD for Central Station Power Generation: A Plan for Action prepared for the Office of Science and Technology, by the Panel on Magnetohydrodynamics, June 1969, USGPO, Washington, D.C., 1969.
50. U.S. Atomic Energy Commission, Potential Nuclear Power Growth Patterns, WASH-1098, December 1970, USGPO, Washington, D.C.
51. Rose, D.J., Controlled Nuclear Fusion: Status and Outlook, *Science,* Vol. 172, No. 3985, 21 May 1971, pp. 797–808.
52. Glaser, P.E., Power from the Sun: Its Future, *Science,* Vol. 162, No. 3856, 22 Nov. 1968, pp. 857–861.
53. Meinel, A.B. and M.P. Meinel, *Bulletin of the Atomic Scientists,* October 1971, pp. 32–37.
54. Clark, J.M. Jr., Total Energy, The State-of-the-Art, National Petroleum Refiners Association, National Fuels and Lubricants Meeting, Sept. 17–18, 1969, New York, Paper No. F & L-69–65.
55. The Automobile and Air Pollution, A Program for Progress, U.S. Dept. of Commerce, October 1967, USGPO, Washington, D.C., 1967.

56. George, J.H.B., L.J. Stratton, and R.G. Acton, *Prospects for Electric Vehicles, A Study of Low-Pollution-Potential Vehicles-Electric,* May 15, 1968, Arthur D. Little, Inc., Cambridge, Mass.
57. Hoess, J.A. and R.C. Stahman, Unconventional Thermal, Mechanical, and Nuclear Low-Pollution-Potential Power Sources for Urban Vehicles, International Automotive Engineering Congress, Society of Automotive Engineers, Detroit, Mich., January 13–17, 1969. SAE Paper No. 690231.
58. Agarwal, P.D., Electric Car and Air Pollution, Automotive Engineering Congress, Society of Automotive Engineers, Detroit, Mich., January 11–15, 1971. SAE Paper No. 710190.
59. Steinmetz, P. Energieverbrauch (Verkehr), *Enzyklopaedie Naturwissenschaft und Technik,* München, 1980.
60. Grasse, W., Solar Thermal Power Applications, Present State and Trends, Report on Solar World Congress 1987.
61. Wachsmann, U. and M.T. Tolmasquim, Wind Power in Brazil, *Refocus,* Jan/Feb 2003.
62. Rice, J., Renewable Energy, *Power Engineering,* April 2004.

FRIEDRICH-WERNER MÖLLENKAMP
Fichtner Beratende Ingenieure

KENNETH C. HOFFMAN
Mathtech, Inc., Centralized Power Plants

ENVIRONMENTAL ASSESSMENTS AND RELATED IMPACTS

BACKGROUND

This author has previously written on the subject of environmental assessment by outlining its typical content requirements, as well as the types of expertise that are required in the preparation of an Environmental Impact Statement(EIS). The author has also discussed some of the problems associated with the EIS process and has provided some suggestions in improving the quality of EIS documents. This paper examines the role of the EIS at the project level, where ultimately, most decisions of project approvals or denials are made in the 1990s by municipal reviewers rather than by state or federal agencies.

INTRODUCTION

With the advent of the National Environmental Policy Act (i.e., NEPA) in 1969, the requirements for environmental impact reporting was originally restricted by the Act to Federal projects, or projects subsidized in whole or in part by Federal funding. Since NEPA, many states and municipalities have developed their own environmental impact requirements to aid in the review of projects under their jurisdiction. As a result of the above evaluation, which has coincided with a general decentralization of power under the Reagan administration, the approval or denial of proposals often rest with local review of the environmental impacts of the specific project regardless of its funding source (i.e., public or private). The evolution of ultimate decision-making at the local level is inevitable when one considers that the most potentially noxious sitings (e.g., nuclear power plants, resource recovery facilities, airports, etc.) would be accepted more critically by municipalities serving as potential host communities than by state or federal entities. Since environmental standards promulgated at the federal and state levels can be adopted or be made more restrictive at the municipal level, the opportunity presents itself for municipalities to utilize local environmental impact ordinances to tightly regulate land-use development within their boundaries. Even in cases where states have the power to "force" sitings in municipalities in order to insure the health and well being of all its constituents, most have tried, for political reasons, to justify their positions through an environmental review process, which, in many cases, has caused extensive delays and/or total abandonment of such sitings. As such, the preparation and presentation of environmental impact statements at the municipal level have often become the most critical element in the consideration and fate of many development projects.

VARIATION IN EIS REPORTING AT THE MUNICIPAL LEVEL

The quality of EIS reporting at the municipal level often exhibits tremendous variations for one or more of the following reasons:

1) The expectations of municipalities relating to EIS documentation varies from it merely being a formality with the application process to the document being a pivotal component in the approval or denial of the project. In municipalities where growth is encouraged and variances to the zone plane are given consideration with regularity, the environmental impact statement requirements may be minimal. In contrast, in municipalities which practice "no-growth" policies and/or rigidly protect their zone plan, the review and critique of the EIS often is used as a weapon for denial or delay of applications.
2) The assessment ordinances, as promulgated, often allow too much subjective interpretation insofar as EIS preparation is concerned. Examples are as follows:

   a. Normally there are no regulations specified for the preparers of component portions of the EIS to be identified or to provide their credentials pertinent to the sections they have prepared.
   b. There generally are no guidelines to provide the applicant with information as to the extent of documentation required to deem an EIS "complete" for submission purposes.
   c. Specific environmental quality standards to be maintained are often either not specified in the ordinances, or are written in qualitative rather than quantitative terms. This encourages assessors to respond to environmental issues in qualitative terms rather than to conduct proper monitoring programs to establish baseline data, project thereon the added impact of the

proposal, and assess resultant values versus existing quantitative standards. The latter approach is clearly a meaningful measure of determining environmental conditions.

d. Most municipalities do not have detailed baseline data available in such environmental categories as ambient noise, air and water quality levels. This condition makes it difficult and expensive for the applicant to obtain a reliable data base, and helps create the huge variation in real data presentation found in impact statements.

e. Often distinction is not provided in the assessment requirements for projects of vastly varying magnitudes of scale and sensitivity. This generally results in unnecessary detail provided in small projects, and insufficient detail provided in more complex proposals.

Because of the subjectivity noted above in attempting to "adequately" respond to many municipal environmental ordinances, many applicants are reluctant to commit sufficient funds to insure the preparation of a comprehensive assessment. As such, impact statement preparers may indirectly be encouraged to prepare statements hastily in order to accelerate the application process. Unfortunately, and perhaps unfairly, this can appear to be a reflection on the assessment preparer rather than on the process itself, which may help to foster inadequate responses.

SUGGESTED METHODOLOGIES FOR PROVIDING APPROPRIATE EIS DOCUMENTATION AT THE MUNICIPAL HEARINGS

In order to resolve some of the aforementioned difficulties encountered when attempting to prepare an appropriate EIS, the following suggestions are offered:

1) The statement preparer should confer early in the process with the client and all the other professionals involved to develop a scope of services needed to prepare an adequate assessment *independent* of the ordinance requirements. The scope should include the professionals needed, the developed and generated data required, the respective environmental quality standards which exist, and the general adversities that will have to be mitigated if the proposal is to be approved.
2) The scope of effort should clearly reflect the magnitude of the project, and should concentrate on critical environmental issues associated with the project.
3) The assessment specialist should meet early in the application process with the individuals, e.g., Township Engineer, Environmental Consultant, and Environmental Commission members who will be directly involved with the review of the assessment. These individuals should be questioned as to their specific areas of concern, which, in turn, will generate more comprehensive analyses in the EIS. The reviewers may also be helpful in citing other documents, reports, etc., known to the Township, which may be useful reference materials for the EIS preparer.
4) After the EIS is completed, the reviewers should be provided copies well in advance of the formal hearing on the application. Hopefully, any questions, differences, etc., can be resolved prior to the hearing. Although this approach won't always be agreed to by the reviewers, it will establish a good faith effort by the applicant to communicate and resolve differences with the assessment reviewers.

It should be appreciated that not all applications are approved, and further, that assessments can be used (properly or improperly) as the main reason to reject an application. Because of this fact, it is most important that the EIS be well written and well documented such that it can prevail, if needed, in an appeal situation where more objective review may be involved. In essence, a good report should ultimately stand the test of objective critical review even if that situation never occurs.

GENERAL GUIDELINES FOR PREPARING AND PRESENTING ENVIRONMENTAL IMPACT STATEMENTS AT THE PROJECT (MUNICIPAL) LEVEL

Practice in the preparation of EISs at the project level has generally focused on addressing the inventory, impacts, mitigations provided, and potential alternatives to the project including the so-called "no build" alternative. In addition, most EISs do not include a traffic or planning analysis as these documents are normally prepared separately by traffic and planning consultants. Generally, the concept of "no build" in an EIS prepared for a project is not a realistic consideration when one is hired specifically to defend a particular application. Furthermore, applicants desirous of receiving approvals for a specific development plan on a particular site normally are not seriously interested in any other permitted or conditional use alternatives allowed in the zoning regulations of the affected jurisdiction. As such, EISs are usually prepared and reviewed basically as a go or no-go situation for a specified development plan.

Regarding the levels of sophistication required in the preparation of an EIS, it generally is a function of two factors, namely, the scale (i.e., magnitude) and sensitivity of the project and the anticipated formal opposition to the project. While theoretically, the level of effort required in an EIS should be independent of the extent and nature of the opposition, one must recognize that additional care in the preparation is crucial when the statement can be expected to stand the test of extreme scrutiny by individuals dedicated to defeating the project by attacking and/or discrediting portions of the EIS.

Lastly, it should be appreciated that even if no expected opposition to a project may arise, one should prepare a document to the extent that the preparer can feel professionally comfortable with the report findings and could testify with confidence on same under the potential of cross-examination.

PACKAGING AND DELIVERY OF THE EIS

The packaging of the EIS generally resides with one individual who may have written the entire EIS or may have prepared it in concert with other professionals. The packager must be well versed in all discipline areas involved in the EIS to be able to edit the entire report and blend it into a cohesive document for presentation purposes. The document should contain an executive summary at the beginning of the report to provide an overview of the scope of work and the pertinent findings in the EIS.

PERSONNEL INVOLVED IN EIS PRESENTATION

Project level EIS preparation and presentation will require at the minimum a site engineer, a traffic engineer, an architect, a planner, an attorney, and an environmental engineer well versed in the management and development of EISs. Depending upon the specific project and its relative complexities, it may be wise to call upon other professionals with varied backgrounds in such disciplines as zoology, botany, archeology, hydrology, noise and air quality assessment, fiscal impact analysis, socioeconomics, etc., who may have to prepare sections of the EIS and defend and/or support the project in the above environmental areas.

Since these experts and their findings may well determine the viability of the project, it is important that the credentials of these experts be recognized and respected at least in the regions in which the development is proposed.

PRESENTATION OF THE EIS TO THE PUBLIC

The EIS should be viewed as a mechanism for "selling" the proposal in question, and the capability of the presentors involved in the public hearing process can greatly affect the acceptance or rejection of the project.

The presentation should include the purpose of the EIS, the nature of the study undertaken, basic findings from the study including any unavoidable adverse impacts found, and methods to be employed to mitigate the impacts. Often, the overall findings of the EIS can be communicated to the municipal reviewers by the person responsible in charge of managing and packaging the EIS.

In an adversarial situation where opposing expert witnesses are anticipated as well as cross-examination by an attorney(s), it may be prudent for the applicant to have the individual experts describe their specific contributions in the EIS document to establish a proper record. In such instances, it is wise for the attorney of the applicant to review the testimony of the respective witnesses well before the public hearing so that the presentation can be cohesive and effective. Each witness should describe the inventory study conducted, the projections of changes resulting from the project, comparisons of same (where applicable) to related environmental standards, resultant beneficial and/or adverse impacts generated if the proposal is constructed and operated, and mitigations suggested and/or incorporated to alleviate or minimize adverse impacts to acceptable levels.

Because EIS documents normally must be defended at public hearings, it is important that the witnesses have good written and oral skills, and are well versed in expert testimony proceedings.

The use of visual or summary materials is effective for public presentations in that audiences generally can follow visual material more closely than solely listening to speakers. At times, hand-outs may be useful to highlight findings. The visual material often aids speakers in the "flow" of information they will be presenting through the presentation of key headings and/or issues. It is helpful to have another individual objectively critique the visual material prior to presentation at the formal hearings for purposes of interpreting the effectiveness of the material.

THE NOMINAL GROUP TECHNIQUE

The Nominal Group Technique is a methodology used for site selections which incorporates a quantitative means for differentiating between alternative sites considered based upon a weighting and scoring system of environmental factors associated with each alternative. Although quantitative decision-making models have been utilized in environmental assessment reporting since the advent of NEPA, the Nominal Group Technique is unique in that it places the decision-making in site selection on a nominal group of citizens, usually appointed by municipal officials, who reside in the areas that may be ultimately impacted by the site selection process. After the consultants have removed potential sites from consideration based upon generally recognized exclusionary criteria (such as wetlands, floodplains, archeologically significant sites, conservation areas, farmland preservation districts, etc.), the nominal group, with guidance from the environmental consulting firm preparing the environmental impact report will generally perform the following functions in the process:

1. Develop a list of environmental factors that they collectively deem pertinent in the site selection process.
2. Develop a relative weighting of importance of each of the environmental factors noted above. The weighting is normally based upon some arbitrary scale (e.g., 0 to 100 with 100 being of greatest significance and zero indicating no significance). Each member provides a weighting figure for each environmental factor considered, and a weighted average value is determined for

each of the factors involved. Typical environmental impact studies may include of the order of 25 to 35 different environmental factors for consideration.

3. Once the environmental factors are determined by the nominal group, the consulting group generally develops criteria for the nominal group to aid them in weighting the impact of each environmental factor involved for the sites in question so that the nominal group can "score" each of the sites involved.

For example, assume that the nominal group considered noise as a factor in the siting of an airport, and further, they judged that noise impacts should have an importance weighting of 87 (out of 100).

In order to assess the impacts of sound generation from a proposed airport on the potential sites to be considered, the consulting group would review the demographics and housing characteristics of the sites under consideration and develop scoring criteria which is applied to all sites such as in the following example:

SCORING LEVEL:	SOUND LEVEL CONDITIONS
0 points	—no dwelling unit within 2 miles of the proposed airport boundaries.
1 point	—less than 5 dwelling units within 2 miles of the proposed airport boundaries.
2 points	—between 6 and 25 dwelling units within 2 miles of the proposed airport boundaries.
3 points	—between 26 and 75 dwelling units within 2 miles of the proposed airport boundaries.
4 points	—greater than 75 dwelling units within 2 miles of the proposed airport boundaries.

As such, if the area surrounding a particular site in question had between 6 and 25 dwelling units within 2 miles of the airport boundaries, it would achieve a score of 87 × 2, or 174 points for the factor of noise, whereas, if it had no dwelling units within 2 miles, it would obtain a score of 87 × 0, or zero points for the factor of noise. For this system scoring approach, the higher the point value accrued for the sites would indicate the sites most environmentally sensitive to impacts resulting from a proposed airport.

Each environmental factor would be scored by the nominal group in a fashion as noted above, and the cumulative score for each site would be tallied and the sites ranked accordingly in terms of the least to the greatest sensitivity to the proposal.

In order to insure the integrity of the nominal group (i.e., to avoid potential conflicts of interest in their voting behavior), the sites they would be scoring would be "masked" so that they would not be able to identify the sites in question from the data provided to them by the consulting groups.

Normally, professionals from the consulting group are available to the nominal group to respond to any technical questions the group might have in formulating their numerical evaluations in the process. Also, in general, a series of rounds of voting would be utilized until the nominal group felt that they had reached a consensus.

Use of the Nominal Group Technique methodology for site selection purposes has the following recognized benefits:

- It provides a resultant site or sites for ultimate selection purposes which is determined on a quantitative basis.
- The nominal group members are unbiased in that they do not know the sites they are evaluating during the site selection process.
- The process is democratic, and interactive sessions between members allows for interchange and stability of the process.
- The group members are the decision makers rather than the consultants and, as such, they (i.e., the public) govern their own destiny in the site selection process.

The Nominal Group process is an excellent tool for consultants and/or County or State Agencies, in applications which face strong opposition, to provide the burden of decision-making on the Nominal Group Committee.

The effectiveness of the Nominal Group Technique Method in the site selection process is related to the following factors:

- The size of the nominal group.
- The credentials of the nominal group to conduct assessments.
- The method of selection of the nominal group members.
- The charge given to the group before and during the site selection processes.
- The role of the group and their involvement (if any) in selecting exclusionary criteria.
- The basis utilized for weighting the environmental issues considered by the group members.
- The basis utilized for scoring each of the issues considered by the group members.
- The analysis and interpretation of the data received from the process and its related statistical significance.

Depending upon how the above issues are handled in a particular study, the results can range from excellent to ridiculous, and the Nominal Group Members can play roles ranging from enlightened decision-makers to manipulated individuals who serve solely to justify a site selection at the conclusion of the process.

POTENTIAL FLAWS IN IMPLEMENTATION OF THE NOMINAL GROUP TECHNIQUE APPROACH

One must be careful in the implementation of the Nominal Group Technique not to allow for flaws in the process which can ultimately be utilized by astute objectors to defeat the proposal. Once the validity of the process can be critically questioned, the Nominal Group Technique, in itself, can prove to be the death knell of the study.

Some of the flaws noted in studies previously reviewed and critiqued by this author which utilized the Nominal Group Technique are as follows:

- The nominal group members may not be asked to participate in developing exclusionary criteria to initially exclude some sites.
- The nominal group size varied from five (5) members to twelve (12) members for three major studies reviewed by this author. One can seriously question the statistical significance of the findings of such a sample of respondents.
- Environmental criteria and nominal group members were changed in one study as it proceeded to reduce the number of feasible sites in the staged process. The problem which occurs is that with inconsistent evaluation criteria and evaluators, different final conclusions regarding site selection may be drawn depending upon when the various criteria and evaluators entered into the process.
- As nominal group members vote through a series of rounds in order to arrive at a "consensus" in the process, a reasonable consistency in the voting pattern through each round for each individual group member should be fairly evident. Major shifting in the voting posture of group members suggests a lack of understanding and knowledge of the issues in evaluating the sites, which places resultant findings into serious question.
- When weighing each site for each environmental factor considered, the point system used to differentiate between sites must be sensitive enough for evaluators to recognize distinct differences between the sites. If such sensitivity cannot be established, the quality of the resultant scores can be questioned. Generally, sensitivity can be achieved if and only if the consultants have thoroughly reviewed each site in question to establish meaningful criteria for nominal group members to vote upon.

Because of common flaws as noted above, it would be relatively easy for people well versed in the impact assessment process to either stop or seriously delay a project.

POSSIBLE STEPS TO ENHANCE THE SUCCESS OF UTILIZING THE NOMINAL GROUP TECHNIQUE IN ASSESSMENT REPORTING

Steps which may be taken to increase the probability of successful outcomes (i.e., sitings) utilizing the Nominal Group Technique approach are as follows:

1. Utilize a statistically significant sample of nominal group members to insure that potential errors in site selection can't be attributed solely to sample size.
2. Use a consistent nominal group with consistent criteria throughout the entire evaluation process.
3. Screen candidates to ensure that they have sufficient knowledge of the proposal and related impacts associated therewith to make credible value judgments.
4. Nominal group members should participate in the selection of exclusionary criteria which will be used to eliminate sites at the beginning of the process.
5. Provide the environmental consultants with responsibility for the project with sufficient time and budget to reasonably develop criteria for each environmental factor considered by the nominal group for the group to differentiate between the subject sites.

Although the above steps will not insure the acceptance of the proposals in question, it will enhance the probability of success and it will not detract from the potential benefits of utilizing the Nominal Group Technique approach in environmental impact analyses.

Since the early 1970s, the concept of the wetlands and the resultant need for wetlands delineation in the United States has become a paramount issue in site development and related assessment reporting. Parcels which are characterized as wetlands are normally deemed so on the basis of their soil strata, their relatively shallow depth to seasonal high groundwater table and by the nature of the vegetation that they will be able to sustain. If a parcel is deemed a wetland (which is a somewhat subjective approach usually negotiated in the field by engineers and botanists representing the applicants and the review agencies), buffers must also be established beyond the delineated wetlands based on the "quality" of the wetlands.

In densely populated areas, where prime developable land is often scarce, the potential for encountering wetlands on a site is a distinct possibility. From a site development standpoint, it is incumbent that a wetlands delineation survey be conducted initially by competent professionals to assess the potential loss of site acreage. This approach should also be considered by potential developers prior to purchasing or taking an option to purchase a site. Too often, this author has witnessed applications well into the review process which are either withdrawn or are no longer economically feasible to construct because of subsequent findings of wetlands on the site. This often leads to time consuming and costly litigation between applicants and their design professionals.

The analysis of traffic-induced noise and air quality impacts in environmental assessment reporting has been conducted in most reports since the passage of NEPA. Normally, many attendees at public hearings are confused about the logarithmic nature of decibel levels (the standard noise descriptor) and how multiple sources of sound are added in terms of decibels. In addition, they generally do not have the technical background to assess results generated from noise and air pollution mathematical models often employed by professionals preparing assessments. However, the public, over the last thirty years, have demonstrated an increasing tendency to voice more concerns at hearings related to noise and air quality issues, and quality of life issues as a whole.

In addition, because traffic noise generation by vehicles on public highways, and rail and aviation noise generation are not normally regulated by municipalities, and further, because the agencies which regulate rail and aircraft noise have less stringent or no standards of compliance at residential property lines (or properties in the case of aircraft) compared to typical municipal noise ordinances, noise generation from these sources is being challenged more often by the public.

In populated and well developed communities, the remaining parcels to be developed are often situated in close proximity to rail lines, airports and heavily traveled highways. Lastly, society is trending toward greater longevity, and noise and air pollution more adversely affect senior citizens than the rest of the general population. As such, credible assessment reporting in the disciplines of noise and air quality, as well as providing meaningful mitigation techniques in site design and orientation for residential applications proposed in proximity to major potential noise and air quality generators are critical for one to be successful in convincing an often cynical public.

Lastly, the air quality standards promulgated for mobile source (i.e., traffic) pollutants in the Clean Air Act of 1970, with the exception of ozone (which is generated about equally by mobile and stationary air pollution sources) have remained unchanged, yet most urban areas, where air pollution generation is normally at the highest levels, have remained in compliance over the period in question. This has occurred despite a major increase in driver registration and vehicular miles traveled on US highways since 1970, due in major parts to technological advances (e.g., catalytic mufflers since the mid 70s, which have reduced vehicular emissions to offset population growth. However, in the United States in the near future, we will be challenged with the need to find alternative fuel sources for vehicles which generate less air pollutants if we are to remain in compliance with the Ambient Air Quality Standards promulgated in 1970 for mobile source pollutants.

In summary, issues such as wetlands delineation, air pollution and traffic-induced noise must be given serious attention in assessment reporting for applications to be approved by reviewing bodies. Increased public awareness on how effectively to critique decision-making methodologies and basic assumptions associated with mathematical models employed in conjunction with assessment preparation is placing an increased burden to defend properly their application in the reporting process.

RISK ASSESSMENT REPORTING AND ASSOCIATED STRATEGIES

If an application involves a project which can (or may) be deemed as a potential threat to life, limb or property resulting from its construction and/or operation, reviewing agencies will normally require a risk assessment to be prepared in addition to an environmental assessment. The function of a risk assessment is to quantify (where possible) the risk of death, injury and potential loss of personal property associated with the worst-case scenarios that can be perceived occurring on the subject site. Typically, risk assessors are to prepare their assessment utilizing the principle of Murphy's Law that "all that can go wrong, will!" All aspects of the operation on the site, including transportation, handling, storing, processing of the product involved and related by-product generation and disposal must be considered, and probabilities of risk quantified (based upon industry practice and history) for each phase of the operation on an annual basis.

For all aspects of risk considered, one assesses the so-called "safe separation distance" (SSD) from the site should a particular failure mode occur either on the site or by transmission from the site (e.g., vapor dispersion of a product or by-product). If potential sensitive receptors are located within the SSD, then those cumulative risks from all potential modes associated with the site in question must be quantified. For parcels located beyond the SSD, the risks are considered negligible.

Once a potential risk is quantified for a particular site, mitigations to the risk can be investigated, based on the proposed implementation of mitigating factors for the subject site and the historical value based on industrial experience related to their effectiveness in reducing risk. Examples of mitigating factors may include the following: availability of fire-fighting capability on the site or within a short distance of the site, accident history of the particular company involved, training of personnel on site, etc. Generally, a value between 0 and 1 is assessed for each mitigating factor, which is then multiplied by the cumulative risk probability for the proposal. This, in essence, produces a reduced probability of risk.

The risks associated with the proposal are typically compared to the "normal risk" to which individuals are subjected in conducting their daily lives. To develop such a perspective, published annual fatality statistics due to accidents (vehicular, aircraft, drowning, lightning, electrocution, falls, etc.) are available from all industrialized nations which, based on total population statistics, can be normalized to a probability of fatality per year basis for the general population. Utilizing the above information, a risk assessment can then compare the risks associated with functioning daily as opposed to the risk associated with living in proximity to a proposed development with a perceived and quantified risk.

The purpose of the above comparisons is to provide decision-makers with a quantitative means to evaluate the risks associated with a proposal with its potential for impact on the quality of life of individuals who would be living in close proximity to the site. Although any application will provide some increased risk to the general public (which will normally create a negative response), the *relative* risk compared to the daily risk to which one is normally subjected provides reviewers with an added tool upon which to render an informed decision. Lastly, it is not the role of a risk assessment preparer to try to convince the public to accept a particular project! Rather, the burden is to provide sufficient quantitative information for decision makers to render an informed decision on a difficult application which will generally be viewed with anxiety by neighboring residents.

FORENSIC ENGINEERING REPORTING AND ASSOCIATED STRATEGIES

In certain situations, environmental engineers normally engaged in environmental and/or risk assessment reporting may be approached to conduct forensic engineering studies.

Forensic engineers are generally involved in providing technical support and, perhaps, expert testimony for attorneys representing plaintiffs or defendants in civil claims wherein a plaintiff has been killed or injured allegedly due, in part, to one or more defendants. Attorneys representing plaintiffs or their estate in cases of the death of the plaintiff generally accept cases on the basis of contingency (i.e., they collect a fee for service only either through a settlement out of court or by winning the court case), whereas defendants are represented by attorneys they retained directly or attorneys representing insurance companies who insure the defendant (which is typical for small businesses and/or corporations).

Environmental engineers operating in the forensics area for a plaintiff will generally be asked to prepare a report which supports the claims of an attorney representing a plaintiff versus one or more defendants. Engineers representing a defendant will be asked to review the plaintiff's expert(s) reports and to prepare a report which can either totally or partially discount claims by the opposing expert(s) against the defendant.

Forensic engineering studies are unique compared to environmental or risk assessments in that claims are often introduced after the occurrence of the incident which precipitated the case. As such, forensic engineers normally must digest a volume of paperwork regarding the incident itself, as well as reports and interrogatories that may have been responded to by parties with knowledge of the incident, or other professionals (i.e., physicians, psychologists, etc.) who have been asked to present their expert opinions on the case. Forensic engineers additionally will normally perform site visits, speak to parties of interest, and perform technical analyses (where appropriate) in support of preparation of their expert report.

Environmental and risk impact report preparers will invariably be required to also provide expert testimony before municipal, state or federal hearing Officers to defend their document and related findings as a part of a process to secure approvals to construct and operate a structure(s). In forensic engineering investigations, at least in the United States, it has been the author's experience (as well as others in the discipline with whom he has consulted) that only about 10% of the cases ever go to trial. The remaining 90% of the cases are settled out of court. The reasons for this may be explained by some or all of the following.

The plaintiff's attorney is paid only on contingency and will often try to avoid the risk of a lengthy trial by accepting a settlement out of court. In major cases, defendants are usually defended by insurers who would prefer to make a settlement payout as a "nuisance" value rather than pay attorneys to prepare and be involved in a lengthy court case.

Because technical experts are generally required to "reconstruct" events which occurred years before they were retained and because many of the parties of interest in the case cannot be contacted due to death, illness, change of address, etc., the resultant expert reports are more subject to scrutiny and question than when one is reporting on current situations. As such, attorneys for both plaintiffs and defendants may be more concerned about the abilities of their experts to defend their positions under intense cross examinations by opposing lawyers. Civil cases are generally heard and decisions rendered by a jury, and both sides consider this to be a difficult body "to read" and select on the jury, particularly, when cases involve issues and arguments which are highly technical in nature. As a result, the technical reports prepared by the respective forensic engineers representing the plaintiff and defendant are often the key determinant in the amount of settlement award that will ultimately be agreed upon by both sides.

Environmental engineers who wish to pursue employment in the field of forensic engineering can best accomplish this by informing injury litigators and/or insurance companies of their interest.

REFERENCES

Canter, L., *Environmental Impact Assessment.* 2nd Edition. McGraw-Hill Publishers. 1996.

Delbecq *et al., Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes.* Scott Foresman and Company. 1975.

Dresnack, R., Environmental Assessments and Related Impacts. *Advances in Environmental Science and Engineering,* **2**, Gordon and Breach Science Publishers. 1979.

Dresnack, R., Environmental Impact. *Advances in Environmental Science and Engineering,* **4**. Gordon and Breach Science Publishers. 1981.

Dresnack, R., Environmental Assessments and Related Impacts. *The Encyclopedia of Environmental Science and Engineering.* 4th Edition. Gordon and Breach Science Publishers. 1998.

Goldfarb, W., Environmental Law. *The Encyclopedia of Environmental Science and Engineering.* 4th Edition. Gordon and Breach Science Publishers. 1998.

Leopold, L.B. *et al.,* Procedure for Evaluating Environmental Impact. Geological Survey Circular 645. 1971.

Pfafflin, J., Environmental Impact, in *Clean Production—Environmental and Economic Perspectives,* pp. 285–291, K. B. Misra, Editor. Springer-Verlag. Heidelberg. 1996.

Plater, G. Z. B., Abrams, R., Goldfarb, W. and Wisch, D. L., *Environmental Law and Policy: Nature, Law and Society,* 3rd Edition. Aspen Publishers. 2004.

Van de Ven and Delbecq, Nominal versus Interacting Group Process for Committee Decision-Making Effectiveness. *Academy of Management Journal.* June 1971.

ROBERT DRESNACK
New Jersey Institute of Technology

ENVIRONMENTAL EDUCATION

The broad field of environmental science and engineering is truly interdisciplinary in nature and any discussion of environmental education must address the many facets and levels of this topic. Is the effort to increase understanding and respect for one's surroundings? This might well begin in the primary school and extend to the post secondary level. Is the intent to produce a practicing professional in one or more of the fields which comprise environmental science and engineering? This involves undergraduate and, possibly, graduate work. The discussion presented here will deal with practicioners. Practicioners will mean scientists, engineers, managers, planners, policy makers, environmental lawyers, etc. 'Etc.' is used because almost every human action will affect the environment and it is not possible to restrict the disciplines. The fundamental breakdown of environmental effort might be public health, conservation and resource management but even this classification is too simplistic.

The education must be quite broad. It has been said that a technician knows everything about the job except its place in the universe. One cannot be certain that the first position after graduation will last until retirement. A broad background gives greater flexibility. Although a topic may be 'hot' (newsworthy) at the moment, one may be certain that equally serious and sensational problems will arise during the working life of the student. While concentration can be in one or more of many areas, an environmental scientist or engineer must have more than superficial knowledge of physics, chemistry, biology and microbiology. (Odum defines ecology as 'environmental biology'). Meteorology, oceanography, geology, geography and hydraulics will be important as will be fundamentals of non-ideological economics. Mathematics should be carried to a fairly high level, particularly statistics. Knowledge of local, national and international statutes and regulations will be applied constantly. An awareness of politics will be *very* important. Public health practice and the fundamentals of conservation should be studied. Computer literacy will be necessary. It is very important that written and oral communications skills be developed. The world is becoming ever smaller and competence in one or more foreign languages will be valuable. In addition to the scientific and cultural subjects just mentioned, an engineering student will study the specialized parts of that particular field of applied science. One is, in essence, becoming a well-rounded person in a broad discipline. It will be concluded that there is simply too much material to compress into a four-year course of study. This is correct and graduate study will be necessary in the majority of cases.

The preceding brief discussion is meant as an introduction to the reader of the many disciplines which might well be involved in education leading to careers in environmental science or environmental engineering. One final observation should be offered. When entering upon one's post secondary education there will be pressure pushing the student toward currently fashionable fields. Such a course should be resisted. One can expect to work on the order of forty years after the terminal degree and it is wise to follow a field in which one has developed an interest and can expect to be satisfied.

REFERENCES

Nicholas Murray Butler, *Introduction to "Engineers and Engineering in the Renaissance"*. W. B. Parsons. Williams and Wilkins, Baltimore, 1939.

Eugene B. Golub, *Environmental Education.* Encylopedia of Environmental Science and Engineering. Second Edition. Gordon and Breach Science Publishers, New York, 1983.

Eugene B. Golub, *Environmental Education.* Encyclopedia of Environmental Science and Engineering. Third Edition. Gordon and Breach Science Publishers, New York, 1992.

P. H. Jones, *Environmental Education.* Encyclopedia of Environmental Science and Engineering. First Edition. Gordon and Breach Science Publishers, New York, 1976.

J. R. Pfafflin, P. Baham and F. S. Gill, *Dictionary of Environmental Science and Engineering.* Gordon and Breach Publishers, Reading, 1996.

Andrew Porteous, *Dictionary of Environmental Science and Technology.* Revised Edition. John Wiley and Sons Ltd., Chichester, 1992.

Abel Wolman, *Water, Health, and Society.* Selected Papers by Abel Wolman. Gilbert F. White, Editor. Indiana University Press, Bloomington, 1969.

EUGENE B. GOLUB
New Jersey Institute of Technology

ENVIRONMENTAL HEALTH

INTRODUCTION[†]

Definitions

"Environmental Health" encompasses what is also known as environmental engineering and sanitation, public health engineering, and sanitary engineering. It is concerned with "the control of all those factors in man's physical environment which exericise or may exercise a deleterious effect on his physical development, health and survival,"[1] with consideration of the physical, economic and social[‡] impact of the control measures applied. Included is the application of engineering principles to the control, modification or adaptation of the physical, chemical, and biological factors of the environment in the interest of man's health, comfort, and social well-being. The concern is not merely with simple survival and prevention of disease and poisoning, although even these are not entirely under control. In addition, environmental health involves the maintenance of an environment that is suited to man's efficient performance, and to the preservation of comfort and enjoyment of living today and in the future.[2,3]

The environment is defined as the sum of all external influences and conditions affecting life and the development of an organism. The Committee on Environment of the American Public Health Association goes a step further and says that "The Environment is considered the surroundings in which man lives, works and plays. It encompasses the air he breathes, the water he drinks, the food he consumes and the shelter he provides for his protection against the elements. It also includes the pollutants, waste materials, and other deterimental environmental factors which adversely affect his life and health."[4]

Scope

It is apparent from the above that the field of environmental health covers an extremely broad area of man's universe. The World Health Organization, the American Public Health Association, and others have suggested what the scope should be and how the programs should be administered. In all instances, a totality is envisioned with consideration of the impact of man on the environment and the impact of the environment on man, with a balanced appraisal and allocation of available resources to both.

A report of a WHO expert Committee[5] considers the scope of environmental health to include or relate to the following:

1) Water supplies, with special reference to the provision of adequate quantities of safe water that are readily accessible to the user, and to the planning, design, management, and sanitary surveillance of community water supplies, giving due consideration to other essential uses of water resources.
2) Wastewater treatment and water-pollution control, including the collection, treatment, and disposal of domestic sewage and other waterborne wastes, and the control of the quality of surface water (including the sea) and ground water.
3) Solid-waste management, including sanitary handling and disposal.
4) Vector control, including the control of arthropods, molluscs, rodents, and other alternative hosts of disease.
5) Prevention or control of soil pollution by human excreta and by substances detrimental to human, animal, or plant life.
6) Food hygiene, including milk hygiene.
7) Control of air pollution.
8) Radiation control.
9) Occupational health, in particular the control of physical, chemical, and biological hazards.
10) Noise control.
11) Housing and its immediate environment, in particular the public health aspects of residential, public, and institutional buildings.
12) Urban and regional planning.
13) Environmental health aspects of air, sea, or land transport.
14) Accident prevention.
15) Public recreation and tourism, in particular the environmental health aspects of public beaches, swimming pools, camping sites, etc.
16) Sanitation measures associated with epidemics, emergencies, disasters, and migrations of populations.
17) Preventive measures required to ensure that the general environment is free from risk to health.

[†] This chapter is based on material which appears in a book by the author entitled *Environmental Engineering and Sanitation*, John Wiley & Sons, Inc., New York, NY, 1972.

[‡] Includes political, cultural, educational, biological, medical, and public health.

The APHA Committee on Environment[4] proposed the following program areas and also the planning considerations, and methods to implement programs.

Environmental Program Area

Wastes

1) Air
2) Sewage and liquid
3) Solid

Water Supply
Housing and Residential Environment
Food and Drugs
Radiation
Noise
Accidents
Occupational and Institutional Hazards
Vectors
Recreation

Planning Considerations

Health
Economic
Demographic and Land Use
Social
Esthetic
Resource Conservation
(Also manpower, facilities, and services.)

Methods and Technics

Research
Demonstration
Education
Standards
Legislation
Inspection
Enforcement
Planning
Evaluation
Incentives
Systems Analysis

The discussions that follow will briefly review and introduce some of the major program areas listed above. Detailed treatment of some of the subject matter can be found in other chapters of this Encyclopedia.

ENVIRONMENTAL HEALTH AND ENGINEERING PLANNING

The Principle of the Total Environment

The central community or city is dependent upon the surrounding areas for its human resources and its economic survival. The surrounding areas are dependent in whole or in part on the central community or city for employment, cultural and other needs. Increasing urbanization is erasing the boundary between the two except for the governmental jurisdictions involved and even these are being blended. The health, welfare, library, police, transportation, recreation, drainage, water, sewerage, and superimposed school systems are examples of combined services now provided in some areas. It is self-evident that regional problems require regional solutions even though operationally individual control may be temporarily required as a political expedient until mutual confidence is established between adjoining governmental units. This emphasizes however the necessity of interim comprehensive planning so that adjoining governmental units may capitalize on improvements and services provided to the long term mutual benefit of their residents. For no one lives in a vacuum and the interdependency of governmental units must be recognized as a fact of life.

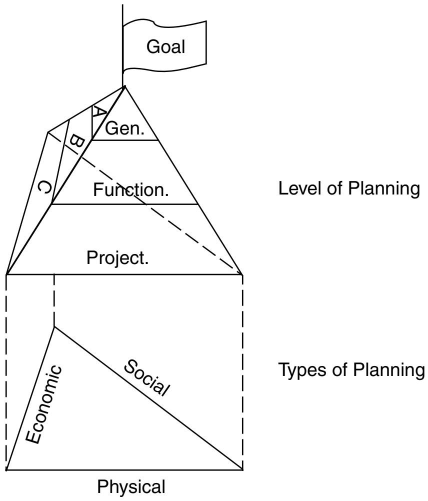


FIGURE 1 Types and levels of planning. (Applicable to National, State, Regional and Local planning.) From Joseph A. Salvato, Jr., *Environmental Engineering and Sanitation,* John Wiley & Sons, Inc., New York, NY, 1972.

Planning and the Planning Process

It is important that environmental health and engineering planning take place within the context of comprehensive regional or area wide planning. It is equally important however that comprehensive community planning fully recognize the environmental health and engineering functions and needs of the area. As used here, the term planning means the systematic process by which goals (policies) are established, facts are gathered and analyzed, alternative proposals and programs are considered and compared, resources are measured, priorities are established, and recommendations are made for the deployment of resources designed to achieve the established goals.[6]

There are many types and levels of planning for the future ranging from family planning to national planning for survival. Figure 1 shows three different types of planning and three different levels of planning. These are all interdependent.

A. *General, Overall Policy Planning*—Identification of goals, aspirations and realistic objectives. Establishment of functional priorities.
B. *Functional Planning*—Such as for transportation, water supply, wastewater, recreation, air pollution, solid wastes, or medical care facilities in which *alternative functional solutions are presented,* including consideration of the economic, social and ecological factors, advantages and disadvantages.
C. *Project Planning*—Detailed engineering and architectural specific project plans, specifications, drawing and contracts for bidding. Plan of action.

Construction, Operation and Maintenance—Plan adjustment as constructed; updating and planning for alterations and new construction.

Figure 2 gives an example of the *general, overall policy planning process.* It recognizes in the process the environmental health and engineering goals and objectives, considerations, plans, and the implementation programs and devices needed. The next level of planning, depending on the functional priorities established, would be specific functional planning such as for recreation, transportation, sewerage, solid wastes, or environmental health.

Environmental Health and Engineering Planning

Since this chapter deals with environmental health, a report outline for that purpose follows.

1) Letter of transmittal to the contracting agency.
2) Acknowledgments.
3) Table of contents.
    a) List of tables.
    b) List of figures.
4) Findings, conclusions, and recommendations.
5) Purpose and scope.
6) Background data and analysis, as applicable, including base maps, reports, and special studies.
    a) Geography, hydrology, meteorology, geology, and ground water levels.
    b) Population and characteristics, past, present, future, and density.
    c) Soils characteristics; flora and fauna.
    d) Transportation and mobility. Adequacy and effects produced, present and future.
    e) Residential, industrial, commercial, recreational, agricultural, and institutional development and redevelopment.
    f) Land use, present and future; spread of blight and obsolescence; inefficient and desirable land uses.
    g) Drainage, water pollution control, and flood control management.
    h) Water resources, multi-use planning and development with priority to water supply, environmental impact.
    i) Air and water pollution, sewerage, and solid wastes management.
    j) Public utilities and their adequacy—electricity, gas, oil, heat.
    k) Educational and cultural facilities, size, location, effects.
    l) Economic studies—present sources of income, future economic base and balance, labor force, markets, industrial opportunities, retail facilities, stability.
    m) Sociological factors, characteristics, knowledge, attitudes; behavior of the people and their expectations.
    n) Local government and laws, codes ordinances.
    o) Special problems, previous studies and findings, background data.
7) Supplementary background environmental health and engineering information.
    a) Epidemiological survey including mortality, morbidity, births and deaths, and specific incidence of diseases; social, economic, and environmental relationships; also water-, insect-, and food-borne diseases; animal- and animal- related diseases, airborne and air-related diseases, pesticide and chemical poisonings; adequacy of data and programs.
    b) Public water supply, treatment, and distribution including population served, adequacy, operation, quality control, cross-connection control, storage and distribution protection, operator qualifications. For individual systems—population served, special problems, treatment and costs, adequacy, control of well construction.
    c) Waste-water collection, treatment and disposal, adequacy of treatment and collection system, population served, operator qualifications; sewer connection control. For individual systems—population served, special problems, control of installations. Water pollution control.
    d) Solid wastes management—storage, collection, transportation, processing, and disposal; adequacy.
    e) Air resources management and air pollution control including sources, air quality, emission standards, problems and effects, regulation, and control program.
    f) Housing and the residential environment—control of new construction, house conservation

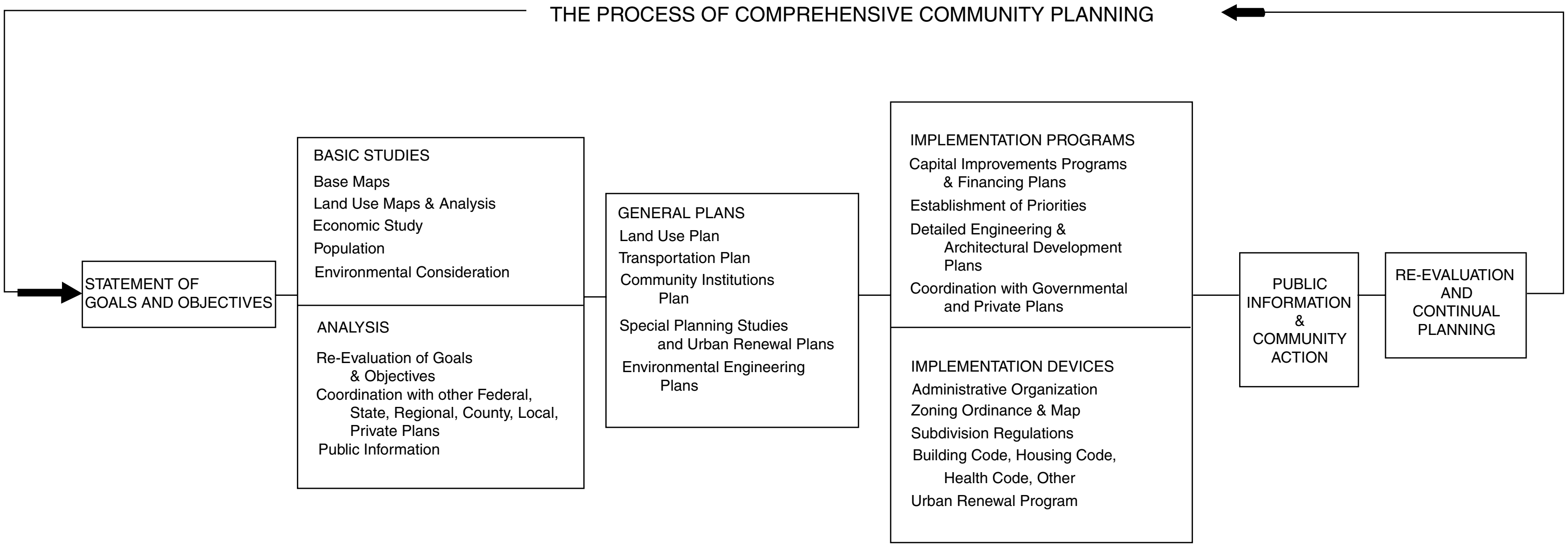


FIGURE 2 An example of the planning process. From Joseph A. Salvato, Jr., “Environmental Health and Community Planning,” *Journal of the Urban Planning and Development Division,* ASCE, Vol. 94, No. UPI, Proc. Paper 6084, August 1968, pp. 22–30.

and rehabilitation, enforcement of housing occupancy and maintenance code, effectiveness of zoning controls, and urban renewal. Realty subdivision and mobile home park development and control, also effect of development on the regional surroundings and effect of the region on the subdivision, including the environmental impact of the subdivision.
   - g) Recreation facilities and open space planning, including suitability of water quality and adequacy of sewerage, solid waste disposal, water supply, food service, rest rooms, safety, and other facilities.
   - h) Food protection program—adequacy from source to point of consumption.
   - i) Nuclear energy development, radioisotope and radiation environmental control including fallout, air, water, food, and land contamination; thermal energy utilization or dissipation, and waste disposal; naturally occurring radioactive materials; air, water, plant, and animal surveillance; federal and state control programs, standards; site selection and environmental impact, plant design and operation control; emergency plans.
   - j) Planning for drainage, flood control, and land use management.
   - k) Public health institutions and adequacy of medical care facilities such as hospitals, nursing homes, public health centers, clinics, mental health centers, rehabilitation centers, service agencies.
   - l) Noise and vibration abatement and control.
   - m) Noxious weed, insect, rodent, and other vermin control, including disease vectors and nuisance arthropods; regulation, control, and surveillance including pesticide use for control of, aquatic and terrestrial plants, and vectors; federal, state, and local programs; and effects of water, recreation, housing, and other land resource development.
   - n) Natural and manmade hazards including safety, slides, earthquakes, brush and forest fires, reservoirs, tides, sand storms, hurricanes, tornadoes, high rainfall, fog and dampness, high winds, gas and high tension transmission lines, storage and disposal of explosive and flammable substances and other hazardous materials.
   - o) Aesthetic considerations, also wooded and scenic areas, prevailing winds, and sunshine.
   - p) Laws, codes, ordinances, rules, and regulations.
   - q) Environmental health and quality protection; adequacy of organization and administration.

8) The comprehensive regional plan.
   - a) Alternative solutions and plans.
   - b) Economic, social, and ecologic evaluation of alternatives.
   - c) The recommended regional plan.
   - d) Site development and reuse plans.

9) Administration and financing.
   - a) Public information.
   - b) Administrative arrangements, management, and costs.
   - c) Financing methods—general obligation bonds, revenue bonds, special assessment bonds; taxes, grants, incentives, federal, and state acid.
   - d) Cost distribution, service charges, and rates. Capital costs—property, equipment, structures, engineering, and legal services; annual costs to repay capital costs, principal and interest, taxes. Regular and special charges and rates.
   - e) Legislation, standards, inspection, and enforcement.
   - f) Evaluation, research, and re-planning.

10) Appendices
   - a) Applicable laws.
   - b) Special data.
   - c) Charts, tables, illustrations.

11) Glossary.
12) References.

HEALTH ASPECTS

Life Expectancy

The life expectancy at birth has varied with time, geography, and with the extent to which available knowledge concerning disease prevention and control could be applied. Table 1 shows the trend in life expectancy with time. The gains in life expectancy between 1900 and 1968 shown in Table 1a have occurred mostly in the early years, 21 years at birth and 15.5 years at age 5, reducing to 4.8 years at age 45 and 2.4 years at age 70. The life expectancy gains are due to better sanitation and nutrition and to the conquest of the major epidemic and infectious diseases including immunization and chemotherapy.

The vital statistics in Table 2 are of interest in that they show the changes in major causes of death in 1900 related to 1960 and the net reduction in total death rate. Table 2a shows the leading causes of death as of 1967. The leveling off that is apparent in the United States is due to our inability thus far to identify the causes and to control the chronic, non-infectious diseases such as heart disease and cancer.

Communicable Disease Control

Although the communicable diseases as causes of death have been largely brought under control in the more advanced countries, this is not the case in the undeveloped areas of the world. Even in the so-called advanced countries, illnesses associated with contaminated drinking water and food are not uncommon.

TABLE 1
Life expectancy at birth

Period or year	Life expectancy
Neanderthal man (50,000 BC–35,000 BC)	29.4[a]
Upper paleolithic (600,000 BC–15,000 BC)	32.4[a]
Mesolithic	31.5[a]
Neolithic anatolia (12,000 BC–10,000 BC)	38.2[a]
Bronze age—Austria	38[a]
Greece classical (700 BC–460 BC)	35[a]
Classical Roman (700 BC–200 AD)	32[a]
Roman empire (27 BC–395 AD)	24
1000	32
England (1276)	48[a]
England (1376–1400)	38[a]
1690	33.5
1800	35
1850	40
1870	40
1880	45
1900	48
1910	50
1920	54
1930	59
1940	63
1950	66
1960	68
1970	71

Note: The 1970 life expectancy reported by the United Nations for Sweden was 71.9 for males and 76.5 for females and for the United States 67.0 for males and 74.2 for females.

[a] E.S. Deevy, Jr., "The Human Population," *Scientific American,* Vol. 203, No. 3, September 1960, p. 200.

Life expectancy figures from 1690 to 1970 are for the United States.

TABLE 1a
Increase in life expectancy between 1900 and 1968 at selected ages[a]

Age	Years added	Age	Years added
0	21	45	4.8
1	15.5	55	3.6
5	12.0	60	3.1
15	10.4	65	2.7
25	8.8	70	2.4
35	6.7	75	2.0

[a] *Reference:* The New York Encyclopedic Almanac 1971, New York, NY, p. 496.

TABLE 2
Selected vital statistics—deaths per 100,000 in the United States

	Death rate for year	
Cause	1900	1950–60
Pneumonia and influenza	202	32
Tuberculosis	195	7
Diphtheria	40	0+
Measles	14	0.2
Diarrhea	110	—
Typhoid	31	0+
Malignant neoplasms	64	150
Cardiovascular and renal diseases	345	520
All causes	1719	934

TABLE 2b
Leading causes of death—deaths per 100,000 population in the United States

Cause	Death rate
Diseases of the heart	364.5
Cancer	157.2
Cerebral hemorrhage (stroke)	102.2
Accidents	57.2
Motor vehicle	26.8
All others	30.4
Pneumonia and influenza	28.8
Certain diseases of early infancy	24.4
Arteriosclerosis	19.0
Diabetes mellitus	17.7
Other diseases of the circulatory system	15.1
Other bronchopulmonic diseases	14.8
Cirrhosis of liver	14.1
Suicide	10.8
Congenital malformations	8.8
Homicide	6.8
Other hypertensive disease	5.6
Other and ill-defined	88.7
Total of all causes	9.537

From U.S. Public Health Service, 1967.

Between 1946 and 1960 a total of 228 waterborne outbreaks with 25,984 cases were reported in the United States.[7] An outbreak in California involving *Salmonella typhimurium* affected an estimated 18,000 persons in a population of 130,000.[8] An explosive epidemic of infectious hepatitis in India with about 29,300 cases of jaundice was attributed to inadequate chlorination and poor operation control.[9]

Sewage normally contains organisms causing various types of diarrhea, dysentery, infectious hepatitis, salmonella infections, and many other illnesses. It becomes obvious that all sewage should be considered presumptively contaminated, beyond any reasonable doubt, with disease producing organisms. The mere exposure of sewage on the surface of the ground, such as from an overflowing cesspool or septic tank system, or its improper treatment and disposal into a stream or lake, immediately sets the stage for possible disease transmission. The means may be a child's ball, the house fly, or ingestion of contaminated water or food. Sanitary safeguards, including adequate water treatment, must therefore be always maintained if the water- and filthborne diseases are to be held in check.

It may appear inconceivable, but there are still many urban areas, as well as suburban areas, in the United States and abroad where the discharge of raw or inadequately treated sewage to roadside ditches and streams is commonplace. Although the disease hazard is ever present, the public pressure for sewage treatment and water pollution abatement stems more from aesthetic, recreational, and related economic considerations rather than from actual disease hazard and transmission. As a matter of fact, a critical reappraisal may be in order to review current expenditures and perhaps achieve a better balance in the allocation of public funds for the public good.

Whereas a safe and adequate water supply is taken for granted by most people in the United States, for about 2 billion people, about two thirds of the world's population, this is still a dream. The availability of any reasonably clean water in less developed areas of the world just to wash and bathe would go a long way toward the reduction of such scourges as scabies and other skin diseases, yaws and trachoma, and the high infant mortality. The lack of a safe water makes commonplace high incidences of shigellosis, amebiasis, schistosomiasis, leptospirosis, infectious hepatitis, typhoid, and paratyphoid.[10]

Between 1923 and 1937 there were an average of 43 milkborne outbreaks with 1724 cases and 47 deaths reported each year in the United States. Between 1938 and 1956 an average of 24 milkborne outbreaks per year with 980 cases and 5 deaths were reported to the U.S. Public Health Service. Between 1957 and 1960 the outbreaks averaged 9 and the cases 151 per year. There were no deaths reported since 1949. The success achieved in the control of milkborne illnesses can be attributed to the practical elimination of the sale of raw milk, greatly improved equipment and to effective control over the pasteurization of milk and milk products.

Whereas milkborne diseases have been brought under control, foodborne illnesses remain unnecessarily high. Between 1938 and 1956 there were reported 4647 outbreaks, 179,773 cases and 439 deaths. In 1967, there were still 273 outbreaks reported with 22,171 cases and 15 deaths. The major bacteria related to foodborne illnesses in recent years (1967–68) are Salmonella, *C. perfringens,* and Staphylococcus. Banquets accounted for over half of the illnesses reported with schools and restaurants making up most of the rest. The largest number of outbreaks occurred in the home.[11]

In addition to waterborne and foodborne diseases, consideration should be given to the environmental related respiratory diseases, the insectborne diseases and zoonoses, and the many miscellaneous diseases. These too must be controlled to the extent possible where indicated. The common cold, encephalitides, malaria, rabies, and disabilities related to air pollution are examples.

More effect must be directed to the total environmental pressures and insults to which the body is subjected. The cumulative body burden of all deleterious substances gaining access to the body through the air, food, drink, and skin must be examined both individually and in combination. The synergistic effects and neutralizing effects must be learned in order that proper preventive measures may be applied. It has been difficult to determine the effects of the presence or absence of certain trace elements in air, water, and food on human health. Some elements such as fluorine for the control of tooth decay, iodine to control goiter, and iron to control iron-deficiency anemia have been recognized as being beneficial in proper amounts. But the action of trace amounts ingested individually and in combination of lead, cadmium zinc, hexavalent chromium, nickel, mercury, manganese, and other chemicals are often insidious. The effects are extended in time to the point where direct relationship with morbidity and mortality is difficult to prove in view of the many possible intervening and confusing factors. Some may even prove to be beneficial.

Mercury

A brief discussion of one of these metals, namely mercury, is of interest. The poisoning associated with (a) the consumption of mercury contaminated fish in Japan between 1953 and 1964; (b) bread made from mercury-contaminated wheat seed in West Pakistan in 1961, in Central Iraq in 1960 and 1965, and in Panorama, Guatemala in 1963 and 1964; (c) pork from hogs fed methylmercury treated seed in Alamogordo, New Mexico; and (d) methylmercury treated seed eaten by birds in Sweden, directed worldwide attention to this problem. The discovery of moderate amounts of mercury in tuna and most fresh water fish, and relatively large amounts in swordfish, by many investigators in 1969 and 1970 tended to further dramatize the problem.[12,13,14]

The organic methylmercury forms of mercury are highly toxic. Depending on the concentration and intake, it can cause unusual weakness, fatigue and apathy followed by neurological disorders. Numbness around the mouth, loss of side vision, poor coordination in speech and gait, tremors of hands, irritability and depression are additional symptoms leading possibly to blindness, paralysis, and death. The mercury also attacks vital organs such as the liver and kidney. It concentrates in the fetus and can cause birth defects.

Mercury has an estimated biological half-life of 70 to 74 days in man, depending on such factors as age, size, and metabolism, and is excreted mostly in the feces at the rate of about one percent per day. Mercury persists in large fish such as pike one to two years.

Mercury is ubiquitous in the environment. The sources are both natural and manmade. Natural sources are leachings and volatilization from mercury containing geological formations. Manmade sources are waste discharges from chloralkali and pulp manufacturing plants, mining, chemical manufacture and formulation, the manufacture of mercury seals and controls, treated seeds, combustion of fossil fuels, fallout, and surface runoff. The mercury ends up in lakes, streams and tidal waters, and in the bottom mud and sludge deposits.

Microorganisms and macroorganisms in water and bottom deposits can transform metallic mercury, inorganic divalent mercury, phenylmercury, and alkoxialkylmercury into methylmercury. The methylmercury thus formed, and perhaps other types, in addition to that discharged in wastewaters, are assimilated and accumulated by aquatic and marine life such as plankton, small fish, and large fish. Alkaline waters tend to favor production of the more volatile dimethylmercury, but acid waters are believed to favor retention of the dimethyl form in the bottom deposits. Under anaerobic conditions, the inorganic mercury ions are precipitated to insoluble mercury sulfide in the presence of hydrogen sulfide. The process of methylation will continue as long as organisms are present and they have access to mercury. The form of mercury in fish has been found to be practically all methylmercury and there is indication that a significant part of the mercury found in eggs and meat is in the form of methylmercury.

The amount of mercury in canned tuna fish has averaged 0.32 ppm, in fresh swordfish 0.93 ppm, in freshwater fish 0.42 ppm (up to 1.4 and 2.0 ppm in a few large fish such as Walleyed Pike), and as high as 8 to 23 ppm in fish taken from heavily contaminated waters. The mercury in urban air has been in the range of 0.02 to 0.2 μg per cu m, in drinking water less than 0.001 ppm, and in rain water about 0.2 to 0.5 ppb (μg/1). Reports from Sweden and Denmark (1967–69) indicate a mercury concentration of 3 to 8 ppb (ng/gm) in pork chops, 9 to 21 ppb in pig's liver, 2 to 5 ppb in beef fillet, 9 to 14 ppb in hen's eggs, and 0.40 to 8.4 ppm in pike.

In view of the potential hazards involved, steps have been taken to provide standards or guidelines for mercury. The maximum allowable concentration for 8 hour occupational exposure has been set at 0.1 mg metallic vapor and inorganic compounds of mercury per cubic meter of air. For organic mercury the threshold limit is 0.01 mg per cu m of air. A maximum allowable steady intake (ADI) of 0.03 mg for a 70 kg man is proposed, which would provide a safety factor of ten. If fish containing 0.5 ppm mercury (the actionable level) were eaten daily, the limit of 0.03 mg would be reached by the daily consumption of 60 gm (about 2 ounces) of fish.[13] The proposed standard for drinking water is 0.002 ppm. A standard for food has not yet been established; 0.05 ppm has been mentioned.

There is no evidence to show that the mercury in the current daily dietary intake has caused any harm, although apparent health does not indicate possible non-detectable effects on brain cells or other tissues. Nevertheless, from a conservative health standpoint, it has been recommended that pregnant women not eat any canned tuna or swordfish. Also implied is caution against the steady eating of large fresh water fish or other large saltwater fish. The general population should probably not eat more than one fish meal per week.

The identification of mercury as coming from manmade and natural sources requires that every effort be made to eliminate mercury discharges to the environment. At the same time the air, drinking water, food, fish, and other wildlife, aquatic plants and animals, surface runoff and leachates, precipitation, surface waters, and man himself should be monitored. This should be done not only for mercury, but also for other potentially toxic and deleterious chemicals. Further research and studies are needed to determine the subtle and actual effects of mercury and other metals, as normally found, on man and his environment.

WATER SUPPLY

A primary requisite for good health is an adequate supply of water that is of satisfactory sanitary quality. It is also important that the water be attractive and palatable to induce its use; for otherwise, water of doubtful quality from some nearby unprotected stream, well or spring may be used. Where a municipal water supply as available, it should be used as such supplies are usually under competent supervision, ample in quantity and also provide fire protection. However, this is not always the case. Because of the excellent water service generally available in the United States, the people and public officials have tended to become complacent and take for granted their water supply. As a result, in some instances, funds have been diverted to other more popular causes rather than to maintenance, opeation, and upgrading of the water supply system.

Status of Water Supply

A survey[15] made by the Public Health Service in 1962 is of interest in pointing out the number and type of public water supplies in the United States and populations served. It was reported that there were 19,236 public water supplies serving approximately 150,000,000 people; 75% were ground water supplies, 18% were surface water supplies, and 7% were a combination. Of significance is the finding that 75,000,000 people in communities under 100,000 population were served by 18,873 public water supplies, and 77,000,000 people in communities over 100,000 were served by 399 public water supplies. Also, of the 19,236 supplies, 85% served communities of 5,000 or less. The information emphasizes the need for giving at least equal attention to small public water supplies as is given to large supplies. In addition, millions of people on vacation in relatively uncontrolled rural environments depend on small water supplies which often are not under close surveillance.

The Public Health Service completed a study in 1970 covering 969 small to large public surface and ground water supply systems serving 18,200,000 persons (12% of the total United States population served by public water supplies) and 84 special systems serving trailer and mobile home parks, institutions, and tourist accommodations.[16] Although the drinking water supplies in the United States rank among the best in the world, the study showed the need for improvements. Based on the 1962 USPHS Drinking Water Standards it was found that in 16% of the 1969 communities surveyed the water quality exceeded one or more of the *mandatory* limits established for coliform organisms (120), fluoride (24), lead (14). It is of interest to note that of the 120 systems that exceeded the coliform standard, 108 served populations of 5,000 or less and that 63 of these were located in a state where disinfection was not frequently practiced or was inadequate. An additional 25% of the systems exceeded the *recommended* limits for iron (96), total dissolved solids (95),

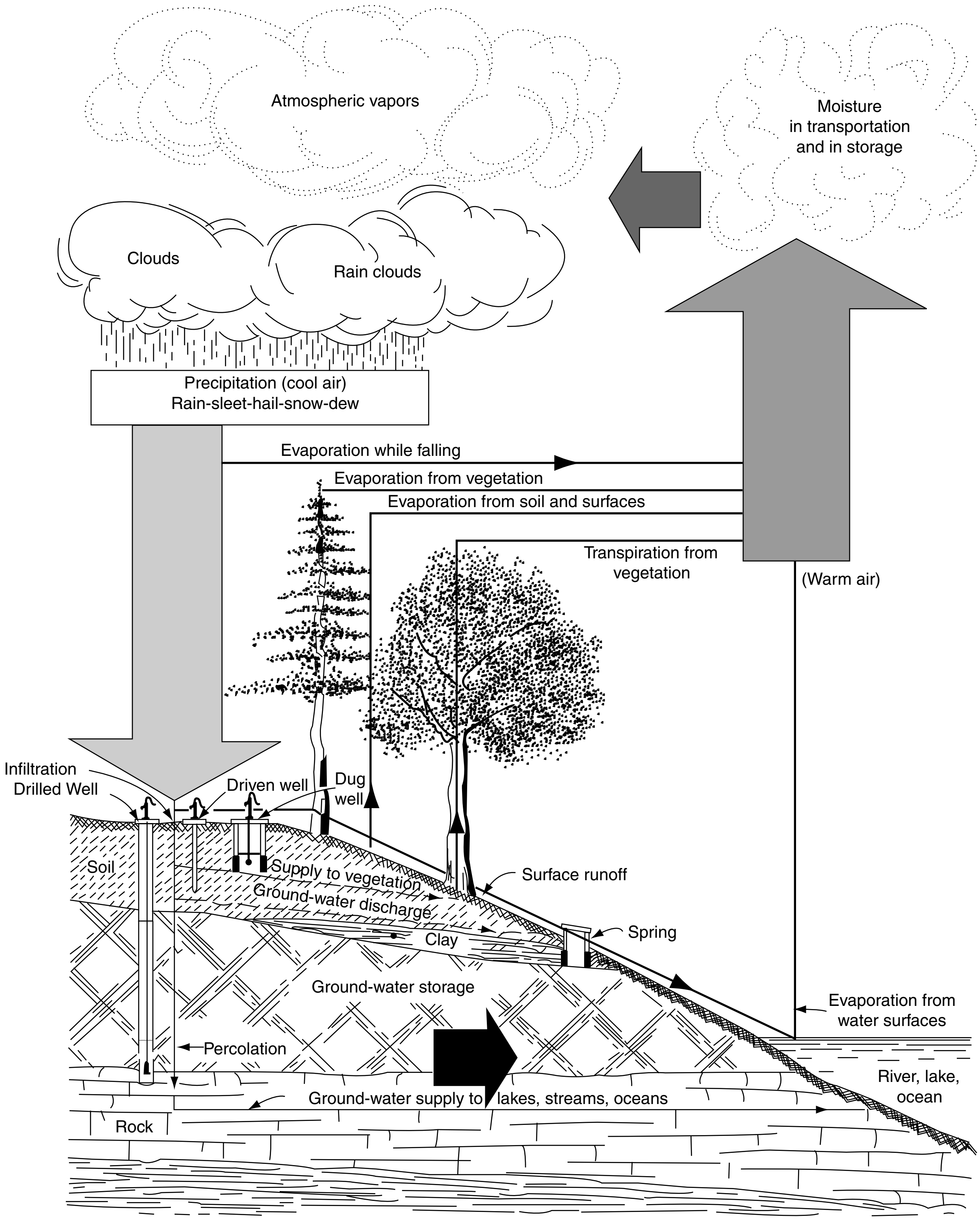


FIGURE 3 The hydrologic or water cycle. From Joseph A. Salvato, Jr., *Environmental Engineering and Sanitation,* John Wiley & Sons, Inc., New York, NY, 1972.

TABLE 3
Surface water criteria for public water supplies

Constituent or characteristic	Permissible criteria	Desirable criteria
Organic chemicals:		
Color (color units)	75	<10
Odor	+	Virtually absent
Temperature[a]	do	+
Turbidity	do	Virtually absent
Microbiological:		
Coliform organisms	10,000/100 mlb	<100/100 mlb
Fecal coliforms	2000/100 mlb	<20/100 mlb
Inorganic chemicals:	(mg/l)	(mg/l)
Alkalinity	+	+
Ammonia	0.5 (as N)	<0.01
Arsenic[a]	0.05	Absent
Barium[a]	1.0	do
Boron[a]	1.0	do
Cadmium[a]	0.01	do
Chloride[a]	250	<25
Chromium,[a] hexavalent	0.05	Absent
Coppera	1.0	Virtually absent
Dissolved oxygen	^{3}4 (monthly mean) ^{3}3 (individual sample)	Near saturation
Fluoride[a]	+	+
Hardness[a]	do	do
Iron (filterable)	0.3	Virtually absent
Lead[a]	0.05	Absent
Manganese[a] (filterable)	0.05	do
Nitrates plus nitrites[a]	10 (as N)	Virtually absent
pH (range)	6.0-8.5	+
Phosphorus[a]	+	do
Selenium[a]	0.01	Absent
Silver[a]	0.05	do
Sulfate[a]	250	>50
Total dissolved solids[a] (filterable residue)	500	<200
Uranyl ion[a]	5	Absent
Zinc[a]	5	Virtually absent
Organic chemicals:		
Carbon chloroform extract[a] (CCE)	0.15	<0.04
Cyanide[a]	0.20	Absent
Methylene blue active substances[a]	0.5	Virtually absent
Oil and grease[a]	Virtually absent	Absent
Pesticides:		
Aldrin[a]	0.017	do
Chlordane[a]	0.003	do
DDT[a]	0.042	do
Dieldrin[a]	0.017	do
Endrin[a]	0.001	do
Heptachlor[a]	0.018	do
Heptachlor epoxide[a]	0.018	do
Lindane[a]	0.056	do
Methoxychlor[a]	0.035	do
Organic phosphates plus carbamates[a]	0.1c	do
Toxaphene[a]	0.005	do
Herbicides:		
2, 4-D plus 2, 4, 5-T, plus 2, 4, 5-TP[a]	0.1	do
Phenols[a]	0.001	do
Radio activity:	(pc/1)	(pc/1)
Gross beta[a]	1000	<100
Radium-226[a]	3	<1
Strontium-90[a]	10	<2

(*continued*)

[a] The defined treatment process has little effect on this constituent. (Coagulation, sedimentation, paid rapid filtration and chlorination.)

[+] No concensus on a single numerical value which is applicable throughout the country. See reference.

[b] Microbiological limits are monthly arithmetic averages based upon an adequate number of samples. Total coliform limit may be relaxed if fecal coliform concentration does not exceed the specified limit.

[c] As parathion in cholinesterase inhibition. It may be necessary to resort to even lower concentrations for some compounds or mixtures. (Permissible levels are based upon the recommendations of the Public Health Service Advisory Committee on Use of the PHS Drinking Water Standards.)

From *Water Quality Criteria,* Report of the National Advisory Committee to the Secretary of the Interior, April 1, 1968, Washington, DC, p. 20.

manganese (90), fluoride (52), sulfate (25), nitrate (19). The study also showed that 56% of the systems were deficient in one or more of the following: source protection; disinfection or control of disinfection; clarification (removal of suspended matter) or control of clarification; and pressure in the distribution system. It was also reported that 90% of the systems did not have sufficient samples collected for bacteriological surveillance; 56% of the systems had not been surveyed by the state or local health department within the last three years; in 54%, cross-connection prevention ordinances were lacking; in 89%, reinspection of existing construction was lacking; in 61%, the operators had not received any water treatment training; in 77%, the operators were deficient in training for microbiological work and 46% of those who needed chemistry training did not have any. Finally, the smaller communities had more water quality problems and deficiencies than the larger ones, showing the advisability of consolidation and regionalization where this is feasible.

Water Cycle

The movement of water from the atmosphere to the earth and back again to the atmosphere can be best illustrated by the

TABLE 4
Guides for water use

Type of establishment	Gallons per day[a]
Residential:	
Dwellings and apartments (per bedroom)	150
Temporary quarters:	
Boarding houses	65
Additional (or non-resident boarders)	10
Camp sites (per site)	100
Cottages, seasonal	50
Day camps	15–20
Hotels	65–75
Mobile home parks (per unit)	125–150
Motels	50–75
Restaurants (toilets and kitchens)	7–10
Without public toilet facilities	2 –3
With bar or cocktail lounge, additional	2
Summer camps	40–50
Public establishment:	
Boarding schools	75–100
Day schools	15–20
Hospitals (per bed)	250–500
Institutions other than hospitals (per bed)	75–125
Places of public assembly	3–10
Turnpike rest areas	5
Turnpike service areas (10% of cars passing)	15–20
Arusement and commercial:	
Airports (per passenger)	3–5
Country clubs	25
Day workers (per shift)	15–35
Drive-in theaters (per car space)	5
Gas station (per vehicle serviced)	10
Milk plant, pasteurization (per 100 lb of milk)	11–25
Movie theaters (per seat)	3
Picnic parks with flush toilets	5–10
Self-service laundries (per machine)	400
Shopping center (per 100 sq ft floor area)	250
Stores (per toilet room)	400
Swimming pools and beaches with bath-houses	10
Farming:	
Livestock (per animal)	
Cattle	12
Dairy	35
Goat	2
Hog	4
Horse	12
Mule	12
Sheep	2
Steer	12
Poultry (per 100)	
Chickens	5–10
Turkeys	10–18

[a] Per person unless otherwise stated.

From *Rural Water Supply,* New York State Department of Health, Albany, NY, 1966.

hydrologic or water cycle shown in Figure 3. The average annual precipitation in the United States is about 30 inches of which 72% evaporates from water and land surfaces and transpires from plants, and 28% contributes to the ground water recharge and stream flow.[17]

Water Quality

Water quality is usually determined by the water source, the treatment it receives and its method of distribution. Standards of water quality take into consideration the physical, chemical, and microbiological characteristics, the radioactivity, and compliance with the *Public Health Service Drinking Water Standards* 1962.[18] Interpretation of water quality analyses should be based on a sanitary survey. This is defined as "(1) An investigation of any condition that may affect public health. (2) A study of conditions related to the collection, treatment, and disposal of liquid, solid, or airborne wastes to determine the potential hazard contributed from these sources to the environment. (3) A study of wastewater discharges on sources of water supply, on bathing or other recreational waters, on shellfish culture, and other related environments."[19]

Surface waters by their very nature are subject to natural and manmade pollution which in many instances is not readily controllable. Hence the selection of a surface water supply source must take into consideration the treatment that may be needed now and in the future to ensure that the finished water will at all times meet the *Drinking Water Standards.* It is obvious that this objective can be met with greater certainty if the source of water is from a protected watershed and reservoir rather than from a stream passing through urban areas which may also be used for wastewater disposal and navigation.

Raw water quality criteria for public water supplies have been developed which recognize the great variability in surface water quality. The criteria are shown in Table 3. The *permissible criteria* are for raw waters which can by conventional rapid sand filtration treatment or less meet the *Drinking Water Standards.* The *desirable criteria* are for raw waters which can meet the *Drinking Water Standards* at less cost and with greater factors of safety than is possible with waters meeting the desirable criteria. It is important to point out that the constituents marked with an asterisk in Table 3 are not removed by conventional rapid sand filtration treatment. In some instances additional treatment with coagulant aids or activated carbon may suffice, and in others elimination of the source of the contaminant may be the only practical answers.

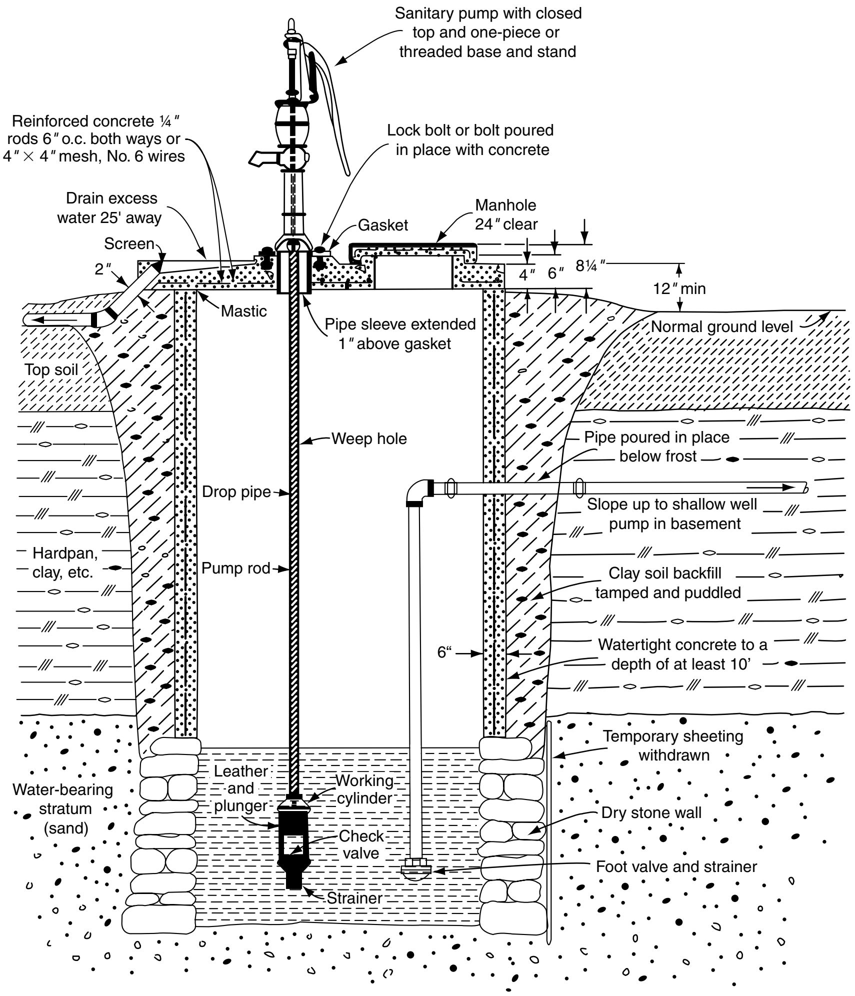


FIGURE 4 A properly developed dug well. Reprinted, with permission, from *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., copyright, 1958.

Water Use

The quantity of water used for domestic purposes will, in general, vary directly with the availability of the water, the habits of the people, the number and type of plumbing fixtures provided, the water pressure, the air temperature, the newness of a community, the type of establishments, and other factors. Wherever possible, the actual water consumption under existing or similar circumstances and the number of persons served should be the basis for the design of a water and sewerage system. Special adjustment must be made for industrial use. Municipal water consumption varies widely dependent on location, industrial usage, metering, size, economic, social and other factors. An average is 155 gpcd.

Table 4 gives estimates of water consumption at different types of places. Additions should be made for car washing, lawn sprinkling, and miscellaneous uses. If provision is made for firefighting requirements, then the quantity of water provided for this purpose to meet fire underwriters' standards will be in addition to that required for normal domestic needs in small communities.

Sources of Water

The sources of water supply are divided into two major classifications; namely, ground water and surface water. To these should be added for completeness rain water and demineralized water. Ground water supply sources include dug, bored, driven, and drilled wells; rock and sand or earth springs, and infiltration galleries. Figure 4 shows details of well and spring construction and sanitary protection. Standards for well construction are given in *AWWA Standard for Deep Wells,* AWWA A 100–66 published by the American Water Works Association and in *Recommended State Legislation and Regulations,* Department of Health, Education, and Welfare, Public Health Service, Washington, D.C., July 1965. Surface water supply sources include lakes, reservoirs, streams, ponds, rivers, and creeks. The development of surface water supplies is covered in standard sanitary engineering texts and state publications.

Note: 1. Springs should be located at least 100 ft and preferably 200 ft from privies, cesspools, barnyards, leaching pits, tile fields, etc. Springs should not be located downgrade

TABLE 5

A classification of waters by concentration of coliform bacteria and treatment required to render the water of safe sanitary quality

Group No.	Maximum permissible average MPN coliform bacteria per month[b]	Treatment required[a]
1	Not more than 10% of all 10-ml or 60% of 100 ml portions positive; not more than 1.0 coliform bacteria per 100 ml.	No treatment required of underground water, but a minimum of chlorination required to surface water, advised for ground water.
2	Not more than 50 per 100 ml.	Simple chlorination or equivalent.
3	Not more than 5000 per 100 ml and this MPN exceeding in not more than 20% of samples.	Rapid sand filtration (including coagulation) or its equivalent plus continuous chlorination.
4	MPN greater than 5,000 per 100 ml in more than 20% or samples and not exceeding 20,000 per 100 ml in more than 5% of the samples.	Auxiliary treatment such as 30–90 days storage, pre settling, pre-chlorination, or equivalent plus complete filtration and chlorination.
5	MPN exceeds group 4.	Prolonged storage or equivalent to bring within groups 1–4.

[a] Physical, inorganic and organic chemicals, and radioactivity concentrations in the raw water and ease of removal by the proposed treatment must also be taken into consideration. See Table 3, *Water Quality Criteria,* Report of the National Technical Advisory Commitee, to the Secretary of the Interior. April 1, 1968, Washington, DC. And *Manual for Evaluating Public Drinking Supplies* U.S. Department of Health, Education, and Welfare, Public Health Service, Environmental Control Administration, Cincinnati, OH 1969.

[b] Fecal coliforms not to exceed 20% or total coliform organisms. The monthly geometric mean of the MPN for group 2 may be less than 100 and for group 3 and 4 less than 20,000 per 100 ml with the indicated treatment.

Adapted from *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., New York, 1958.

so as to be in the direct line of drainage from sources or pollution. 2. If a handpump is provided it should be of sanitary design, installed with a watertight connection.

Note: 1. *Locations.* Wells should be located at least 100 ft and preferably 200 ft from privies, barnyards, leaching pits, cesspools, tile fields, and other sources of pollution. Wells should not be located downgrade so as to be in the direct line of drainage from sources of pollution. *Concrete.* Mix one bag of cement, 2 cu ft of sand, and 3 cu ft of gravel. Then add 5 gal of water, for moist sand, and mix again.

Water Treatment

As an aid in determining the treatment that should be given water to make it safe to drink, the United States Public Health Service has classified waters into several groups.[20] The treatment required by this classification is based upon the most probable number (MPN) of coliform bacteria per 100 milliliters (ml) ot sample. The classification is summarized in Table 5. It needs to be supplemented by chemical, physical, and microscopic examinations. For water to be generally acceptable, other treatment may be required in addition to that necessary for the elimination of disease-producing organisms. People expect the water to be safe to drink, attractive to the senses, sort, nonstaining, and neither scale-forming nor corrosive to the water system. The various treatment processes employed to accomplish these results are shown in the flow diagram in Figure 8. The untrained individual should not attempt to design a water treatment plant for life and health will be jeopardized. This is a job for a competent sanitary engineer.

Plant capacity—50 to 100% greater than average daily demand, with clear well. Possible chemical combinations: A—Chlorine. B—Coagulant; aluminum sulfate (pH 5.5–8.0), ferrous sulfate (pH 8.5–11.0), ferric chloride (pH 5.0–11.0), sodium aluminate and or other (activated silica) (polyelectrolytes). C—Alkalinity adjustment; lime, soda ash, or polyphosphate. D—Activated carbon. E—Dechlorination; sulfur dioxide, sodium sulfite, sodium thiosulfate. F—Fluoridation treatment. X—Chlorine dioxide.

Note: The chlorinator should be selected to pre-chlorinate surface water at 20 mg/l and post-chlorinate at 3 mg/l. Provide for a dose of 3 mg/l plus chlorine demand for ground water.

WASTEWATER

Definitions

Excreta is the waste matter eliminated from the body, including feces, urine, and sweat. *Domestic sewage* is the used water from a home or community and includes toilet, bath, laundry, lavatory, and kitchen-sink wastes. Sewage from a community may include some industrial wastes, ground water and surface water, hence the more inclusive term *wastewater* is coming into general usage. Normal sewage from a private sewage disposal system contains about 99.8% water and 0.2% total mineral and organic solids. Domestic sewage contains less than 0.1% total solids. The strength of wastewater is commonly expressed in terms of 5-day biochemical oxygen demand (BOD), suspended solids and chemical oxygen demand (COD).

The biochemical oxygen demand of sewage, sewage effluents, polluted waters, industrial wastes or other wastewaters is the oxygen in parts per million (ppm) or milligrams per liter (mg/l) required during stabilization of the decomposable organic matter by aerobic bacterial action. Complete stabilization requires more than 100 days at 20°C. Incubation for 5 or 20 days is not unusual but as used in this chapter BOD refers to the 5-day test unless otherwise specified.

Suspended solids are those which are visible and in suspension in water. They are the solids which are retained on the asbestos mat in a Gooch crucible.

The *chemical oxygen demand* (COD) is also used, particularly in relation to certain industrial wastes. The COD is the amount of oxygen expressed in parts per million (ppm) or milligrams per liter (mg/l) consumed under specific conditions in the oxidation of organic and oxidizable inorganic material. The test is relatively rapid. It does not oxidize some biodegradable organic pollutants (pyridine, benzene,

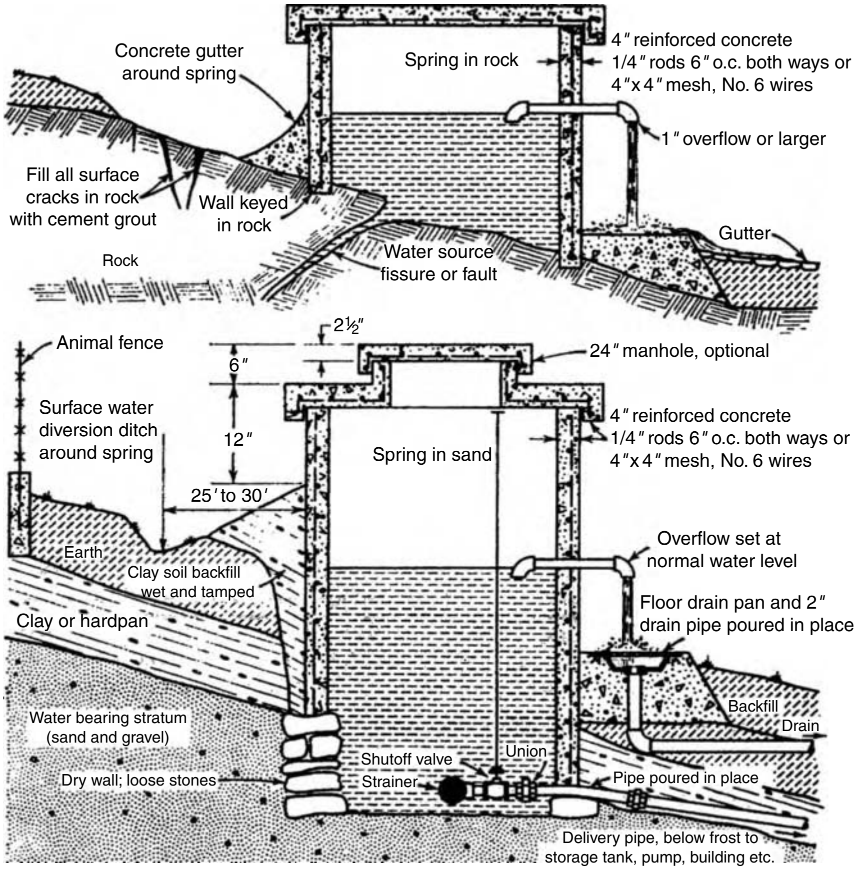


FIGURE 5 Properly constructed springs. Reprinted, with permission, from *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., copyright, 1958.

toluene) but does oxidize some inorganic compounds which are not measured, that is, affected by the BOD analysis.

When storm water and domestic sewage enter a sewer it is called a *combined sewer.* If domestic sewage and storm water are collected separately, in a *sanitary sewer* and in a *storm sewer,* then the result is a *separate sewer system.* A *sewage works* or *sewage system* is a combination of sewers and appurtenances for the collection, pumping and transportation of sewage, sometimes, also called *sewerage,* plus facilities for treatment and disposal of sewage, known as the *sewage treatment plant.*

The *privy,* or one of its modifications, is the common device used when excreta is disposed of without the aid of water. When excreta is disposed of with water, a *water-carriage* sewage disposal system is used; generally all other domestic liquid wastes are included.

Stream Pollution and Recovery

Although the 5-day BOD is the best single measure of wastewater or polluted water organic loading, aquatic organisms when measured with the biochemical oxygen demand, the chemical oxygen demand where indicated, the dissolved oxygen, and the sanitary survey taken all together are the best indicators of water pollution. Other chemical and bacterial analyses provide additional information.

In a freshly polluted water, during the first stage (5 to 20 days at 20°C), mostly carbonaceous matter is oxidized. This is demonstrated by an immediate increase in the stream BOD and oxygen utilization in the area of pollution discharge, followed by the second or nitrification stage where a lesser but uniform rate of oxygen utilization takes place for an extended period of time. This is accompanied by a related characteristic change in the stream biota. The amount of dissolved oxygen in a receiving water is the single most important factor determining the waste assimilation capacity of a body of water.

Stream pollution (organic) is apparent along its length by a zone of degradation just below the source, a zone of active decomposition and, if additional pollution is not added, a zone of recovery. In the zone of degradation the oxygen in the water is decreased, suspended solids may be increased and the stream bottom accumulates sludge. The fish life changes from game and food species to coarse. Worms, snails, and other biota associated with pollution increase. In the zone of active decomposition, the dissolved oxygen is further reduced and may approach zero. The water becomes turbid and gives off foul odors. Fish disappear, anaerobes predominate in the bottom mud and sludge worms become very numerous. In the zone of recovery the process is gradually reversed and the stream returns to normal. The zones mentioned may not be discernible or experienced where sewage has been given adequate treatment before discharge.

Small Waterborne Sewage Disposal Systems

The provision of running water in a dwelling or structure immediately introduces the requirement for sanitary removal of the used water. Where public sewage works are available, connection to the sewer will solve a major sanitation problem. Where public or central sewage works are not provided or anticipated, such as in predominately rural areas, then consideration must be given to the proposed method of collection, removal, treatment, and disposal of sewage on each lot as shown in Figure 9. With a suitable soil, the disposal of sewage can be simple, economical, and inoffensive; but careful maintenance is essential for continued satisfactory operation. Where rock or ground water is close to the surface or the soil is a tight clay, it would be well to investigate some other property.

The common system for sewage treatment and disposal at a private home in a rural area consists of a proper septic tank for the settling and treatment of the sewage, and a subsurface leaching system for the disposal of the septic tank overflow, provided the soil is satisfactory. The soil percolation test and soil characteristics are used as means for determining soil permeability or the capacity of a soil to absorb settled sewage. This and the quantity of sewage from a dwelling are the bases upon

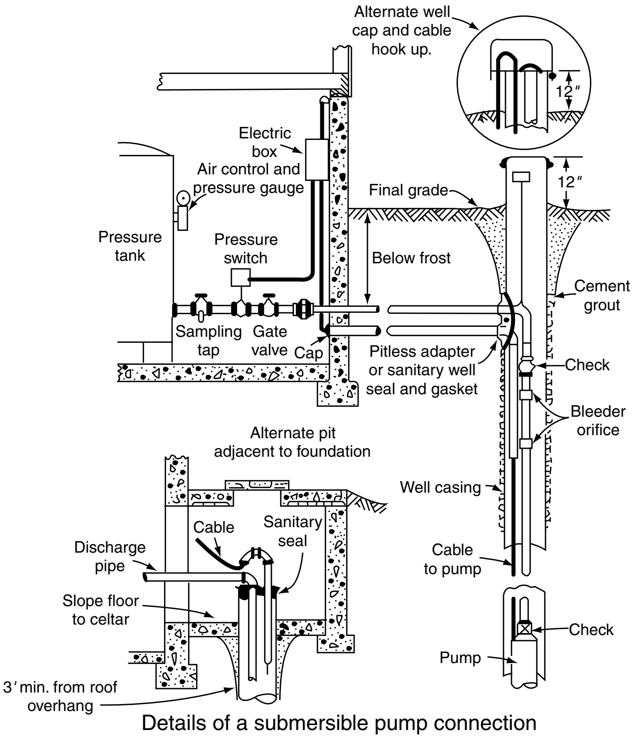


FIGURE 6 Sanitary well caps and seals. Reprinted with permission, from *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., copyright, 1958.

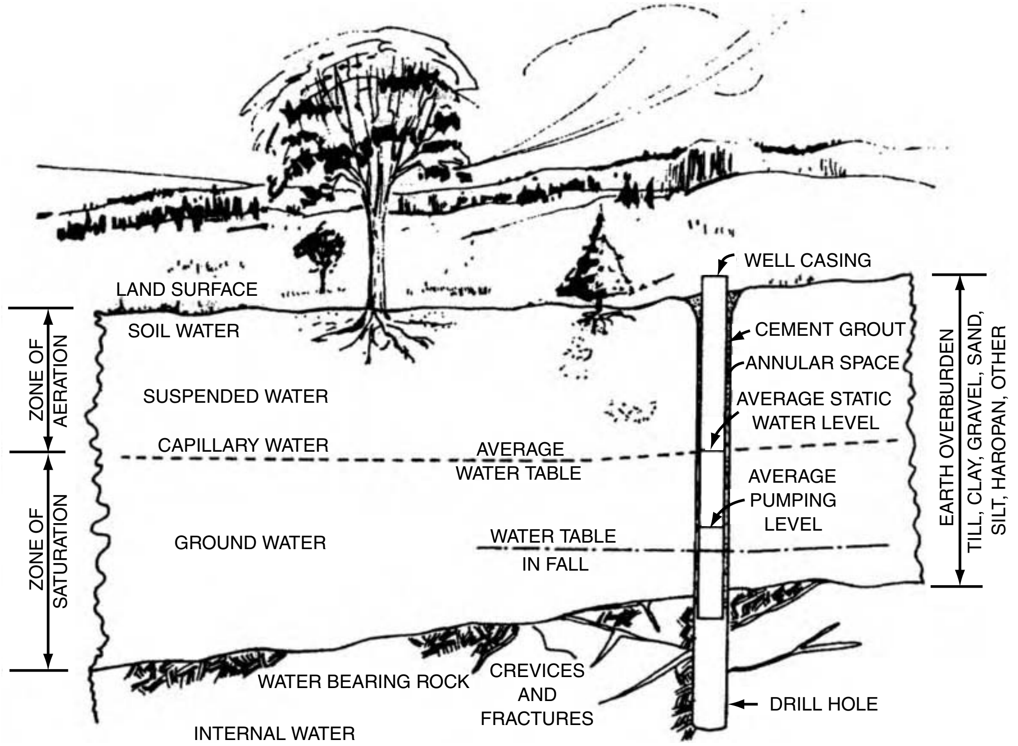


FIGURE 7 A geologic section showing ground water terms. From *Rural Water Supply,* New York State Department of Health, Albany, NY, 1966.

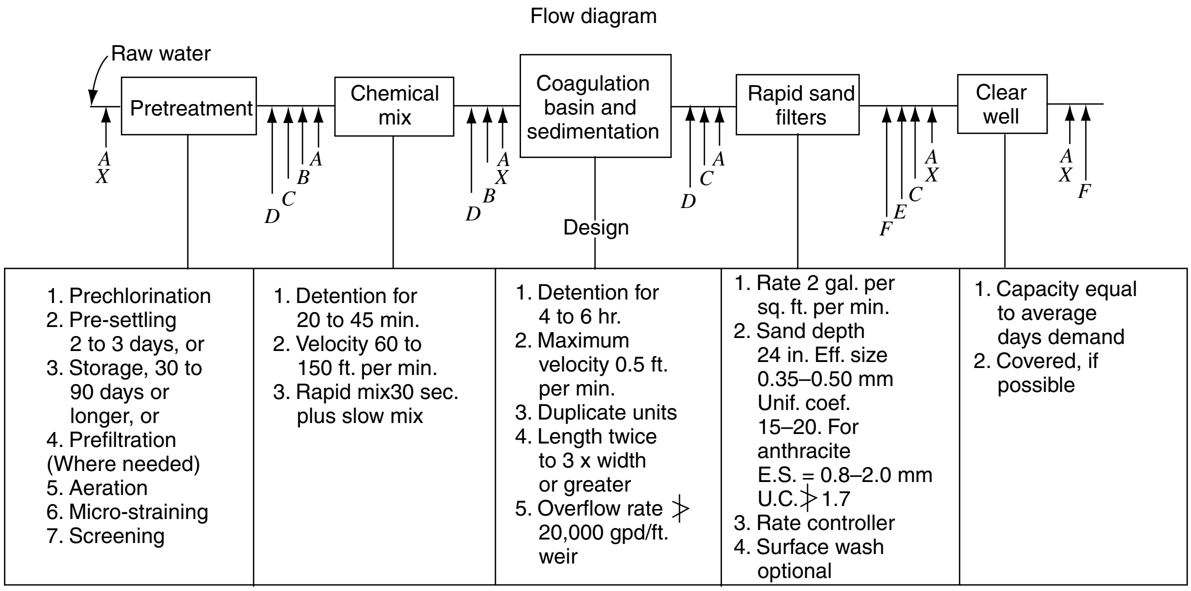
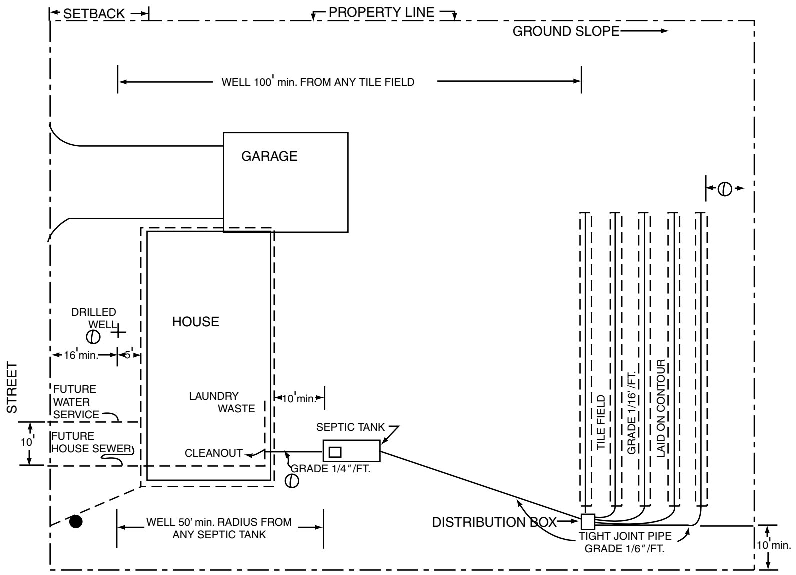


FIGURE 8 Rapid sand filter plant flow diagram. From *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., New York, 1958.



NOTES: 1. Watertight Footing Drain Within 25' Of Well
2. Tile Field To Be 50' Or More From Any Lake, Swamp, Ditch Or Watercourse & 10' Or More From Any Water Line Under Pressure
3. Coef from Pipe, Load Coated Joints Within 50' of Any Well
4. Discharge Footing, Roof And Celter Drainage Away Sewage system & Well

FIGURE 9 Private water supply and sewage disposal layout.

which a subsurface sewage disposal system is designed. Sand filters, evapotranspiration beds, aeration systems, stabilization ponds or lagoons, recirculating toilets and various types of privies are used under certain conditions.

Large Sewage Works

The design of sewers and treatment plants should take into consideration the broad principles of comprehensive community planning discussed at the beginning of this chapter. Regional and area-wide sewerage planning (preliminary) should take place within the context of the general overall comprehensive community plan. It recognizes the extent of present and future service areas, the established water quality and effluent standards, and the alternative solutions with their first costs and total annual costs. This information is needed to assist local officials in making a decision to proceed with the design and construction of a specific sewerage system and treatment plant.

These are essential first steps to ensure that the proposed construction will meet the community, state and national goals and objectives.

Water Quality Standards The purpose of stream and other surface water quality standards is to protect, improve where needed, and maintain the highest practicable water quality for certain specified best uses. Such uses are usually grouped under "public water supply," "bathing and recreational," "fishing, aquatic life and wildlife," and "natural drainage, agricultural and industrial water supply." In general, sludge deposits, solid refuse, floating solids, oils, grease, and scum are not allowed under any uses. Waters are also required to be free from organic and inorganic chemical or radiologic constituents in concentrations or combinations which would be harmful to human, animal, or aquatic life for the specified water use. For sources of drinking water supplies, the sewage and industrial waste treatment before discharge to a stream is such that with conventional (coagulation, sedimentation, rapid sand filtration) water treatment the Public Health Service Drinking Water Standards are not exceeded.

For sources of drinking water with chlorination treatment alone, the maximum permissible average most probable number (MPN) of coliform bacteria per month should not exceed 50 per 100 ml with not more than 20% of the samples exceeding 240. With coagulation, rapid sand filtration, and chlorination the average MPN should not exceed 5000 with not more than 20% of the samples exceeding 5000. Fecal coliforms should not exceed 20% of the total coliform counts. For sources of water for bathing and other contact recreation uses the average MPN should not exceed 2400 (or 1000) per 100 ml with not more than 20% (or 10%) of the samples exceeding 5000 (or 2000). The fecal coliform count (log mean) should not exceed 200 with not more than 10% of the samples collected during any 30-day period exceeding 400. In all cases the coliform counts should be interpreted in the light of a sanitary survey and related factors.

Water quality standards also specify a minimum dissolved oxygen content of 5.0 mg/l for trout waters, 5.0 or 4.0 mg/l for non-trout waters and a minimum 2.0 or 3.0 mg/l for any other waters. Similar standards as above are used for marine waters. If waters are used for shellfish harvesting for direct human consumption, the coliform bacteria shall not exceed a median MPN of 70 per 100 ml and not more than 10% of the samples shall ordinarily exceed 230.

Effluent Standards It should be obvious that the realistic compliance with water classification standards requires control over all sewage and industrial wastewater discharges to the classified waters. In Britain, the Ministry of Housing and local government reaffirmed in 1966 the Royal Commission "General Standard" as a "norm" for sewage effluents, i.e., 5-day BOD 20 mg/l and suspended solids 30 mg/l with a dilution factor of 9 to 150 volumes in the receiving watercourse having not more than 4.0 mg/l BOD. A higher effluent standard of 10 mg/l BOD and suspended solids may be required if indicated.[21] In any case, sewage and industrial waste effluents should not contain any matter likely to render the receiving stream poisonous or injurious to fish, or if used for drinking water supply injurious to man.

TABLE 6

Sewage-treatment plant unit combinations and efficiencies

Treatment plant	Total % reduction—approximate: Suspended solids	Total % reduction—approximate: Biochemical oxygen demand
Sedimentation plus sand filter	90–98	85–95
Sedimentation plus standard trickling filter, 600 lb BOD per acre-foot maximum loading	75–90	80–95
Sedimentation plus single stage high rate trickling filter	50–80	35–65[a]
Sedimentation plus two stage high rate trickling filter	70–90	80–95[a]
Activated sludge	85–95	85–95
Chemical treatment	65–90	45–80
Pre-aeration (1 hr) plus sedimentation	60–80	40–60
Plain sedimentation	40–70	25–40
Fine screening	2–20	5–10
Stabilization, pond aerobic		70–90
Anaerobic laboon	60–70	40–70

[a] No recirculation. Efficiencies can be increased within limits by controlling organic loading, efficiencies of settling tanks, volume of recirculation, and the number of stages: however, effluent will be less nutrified than from standard rate filter, but will usually contain dissolved oxygen. Filter flies and odors are reduced. Study first cost plus operation and maintenance. From *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., New York, 1958.

Wastewater Treatment The need for wastewater treatment can be better justified for aesthetic, economic, recreation, and fish reasons rather than for reasons of reduction of health hazards. Treatment of sewage and industrial wastes does not usually remove the need for nor reduction in required treatment for drinking water. On the contrary, greater emphasis needs to be placed on the improvement and maintenance of drinking water quality. The increasing urbanization and industrialization has and will continue to result in the discharge of viruses, organic, and inorganic chemicals, which are not readily removed by conventional water treatment, and whose long-term health effects are not yet known. Many people may be misled into thinking that money and effort spent in stream pollution abatement justifies relaxation of effort, money to be spent and support for drinking water treatment. Nothing could be further from the truth!

Wastewater treatment usually involves screening and sedimentation for primary treatment. When followed by oxidation and disinfection it is given secondary treatment. If secondary treatment plant effluent is passed through a sand filter, oxidation pond, or some combination of chemical treatment and filters plus chlorination, in which additional organic and inorganic materials are removed, together with bacteria, protozoa and viruses, then it would be receiving tertiary treatment. Some possible combinations of sewage

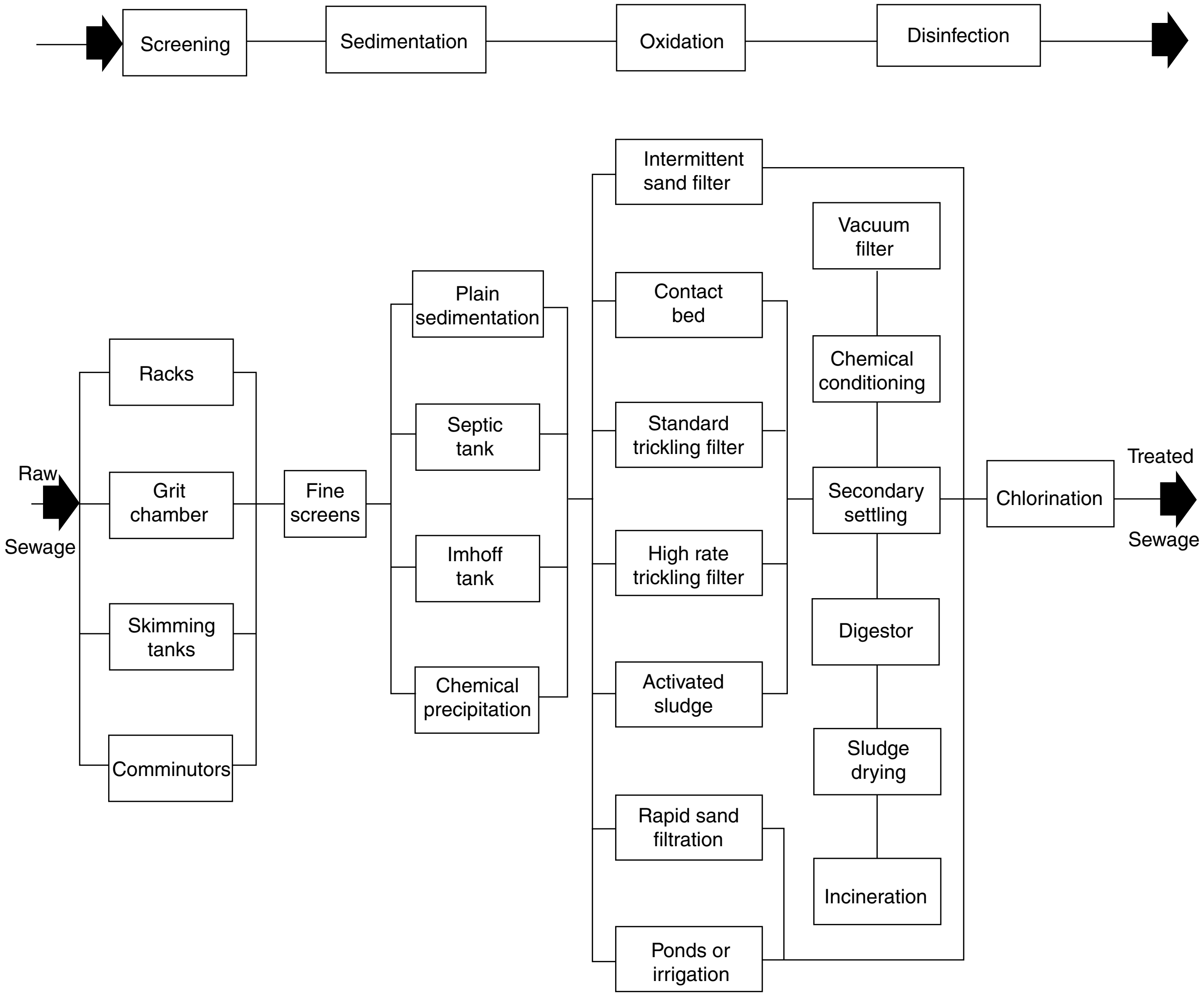


FIGURE 10 Sewage-treatment processes. From *Environmental Sanitation,* Joseph A. Salvato, Jr., John Wiley & Sons, Inc., New York, 1958.

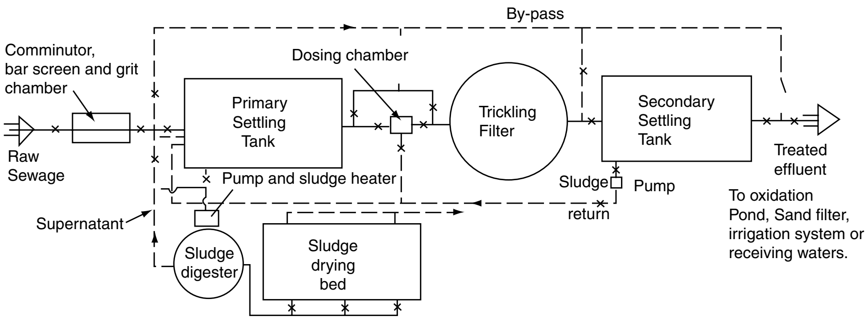


FIGURE 11 A secondary sewage treatment plant.

treatment unit processes are shown in Figures 10 and 11; their efficiencies are shown in Table 6.

There are two major factors which can determine the success or failure of a sewer and treatment plant construction program once completed. First is the positive exclusion of surface water, ground water, roof leaders, and rooting drains by ordinance and inspection. Many a sewage treatment plant constructed has been so hydraulically overloaded as to render it almost useless as a stream pollution abatement facility. Next is proper operation of the treatment plant by competent adequately paid personnel. Here the consulting engineers and the regulatory agencies share a joint responsibility to ensure that public officials are properly informed and that the huge investments in water pollution control accomplish the intended objectives.

SOLID WASTES

General

Aesthetic, land use, health, water pollution, air pollution, and economic considerations make proper solid waste storage, collection, and disposal municipal, corporate, and individual functions which must be taken seriously. Indiscriminate dumping of solid waste and failure of the collection system in a populated community for two or three weeks would soon cause many problems. Odors, flies, rats, roaches, crickets, wandering dogs and cats, and fires would dispel any remaining doubts of the importance of proper solid waste storage, collection, and disposal.

The complexities of proper solid waste management are not readily appreciated. There are numerous sources and types of solid wastes ranging from the home to the farm and from garbage to radioactive wastes, junked cars, and industrial wastes. Handling involves storage, collection transfer, and transport. Processing includes incineration, densification, composting, separation, treatment, and energy conversion. Disposal shows the environmental inter-relation of air, land, and water, and the place of salvage and recyling. All these steps introduce constraints—social, political, economic, technological, ecological, legal, information, and communications which must be considered in the analysis of the problem and in coming up with acceptable solutions.

All these factors show the need for comprehensive planning. It is necessary to come up with alternative solutions together with the environmental and social impact of each. In this way one is more likely to come up with a practical and acceptable plan that will have a good chance of being implemented.

Composition, Weight, and Volume

The solid waste of a municipality may consist of garbage, rubbish, ashes, bulky wastes, street refuse, dead animals, abandoned vehicles, construction and demolition wastes, industrial solid wastes, special hazardous wastes, animal and agricultural wastes, sewage treatment plant residues, and septic tank cleanings. Garbage is the putrescible organic material discarded or remaining as a result of the storage, processing, preparation, and consumption of food. Rubbish includes paper, wood, glass, metal, leaves, plastics, tin cans, stones, and dirt. Bulky wastes include stoves, refrigerators, auto parts, furniture, large creates, trees, and tree branches.

Various estimates have been made of the quantity of solid waste generated and collected per person per day. The amount of municipal solid waste *generated* in 1968 was estimated to be 7 pounds per capita per day and expected to increase at about 4% per year. The amount collected was found to average 5.32 pounds. The figures given in Tables 7a and 7b are averages, subject to adjustment dependent upon many local factors, including time of the year, habits and economic status of the people, whether urban or rural, education, and location. The estimates should not be used for design purpose. Each community should be studied by itself and actual weighings made to obtain representative information.

TABLE 7a
Average solid waste collected, 1968 (pounds per person per day)

Solid waste type	Urban	Rural	National
Household	1.26	0.72	1.14
Commercial	0.46	0.11	0.38
Combined	2.63	2.60	2.63
Industrial	0.65	0.37	0.59
Demolition, construction	0.23	0.02	0.18
Street and alley	0.11	0.03	0.09
Miscellaneous	0.38	0.08	0.31
Totals	5.72	3.93	5.32

Anton J. Muhich, "Sample Representativeness and Community Data," *Proceedings, Institute for Solid Wastes,* APWA, Chicago, IL., 1968.

TABLE 7b
Approximate composition of domestic solid wastes[a]

Component	% by weight
Food waste	15
Paper products	50
Plastics, rubber, leather	3
Rags	2
Metals	8
Glass and ceramics	8
Wood	2
Garden wastes	5
Rock, dirt, miscellaneous	7

[a] Moisture content approximately 30%.

Collection

Collection cost has been estimated to represent about 80% of the total cost of collection and disposal by sanitary landfill and 60% when incineration is used. In 1968 collection cost alone was found to average 10 to 13 dollars per ton.

The frequency of collection will depend upon the quantity of solid waste, time of the year, socio-economic status of area served, and municipal or contractor responsibility. In business districts refuse, including garbage from hotel and restaurants, should be collected daily except Sundays. In residential areas, twice a week refuse collection during warm months of the year and once a week at other times should be the maximum permissible interval. Slum and ghetto areas usually require at least twice a week collection. The receptacle should be either emptied directly into the garbage truck or carted away and replaced with a clean container. Refuse transferred from can to can will invariably cause spilling, with resultant pollution of the ground and attraction of rats and flies. If other than curb pick-up is provided, the cost of collection will be high. Some property owners are willing to pay for this extra service.

Transfer Station

The urban areas around cities have been spreading leaving fewer nearby acceptable solid waste disposal sites. This has generally made necessary the construction of incinerators in cities or on the outskirts, or the transportation of wastes longer distances to new sites. However, as the distance from the centers of solid waste generation increases, the cost of direct haul to a site increases. A distance is reached (in terms of time) when it becomes less expensive to construct an incinerator, or a transfer station near the center of solid waste generation where wastes from collection vehicles can be transferred to large tractor-trailers for haul to more distant disposal sites.

Rail haul and barging to sea also involve the use of transfer stations. They may include one or a combination of grinding, compaction to various densities, and baling.

Treatment and Disposal of Solid Wastes

Solid waste treatment and disposal methods include the open dump, hog feeding, incineration, grinding, and discharge to a sewer, milling, compaction, sanitary landfill, dumping, and burial at sea, reduction, composting, pyrolization, wet oxidation, and anaerobic digestion. The common acceptable methods are incineration, sanitary landfill and in some part of the world composting.

Open Dump The open dump is all too common and needs no explanation. It is never satisfactory as usually maintained. Refuse is generally spread over a large area providing a source of food and harborage for rats, flies, and other vermin. It is unsightly, an odor and smoke nuisance, a fire hazard, and often a cause of water pollution. It should be eliminated or its operation changed to a sanitary landfill.

Hog Feeding Where garbage is fed to hogs, careful suspervision is necessary. The spread of trichinosis in man, hog cholera, the virus of foot-and-mouth disease, and vesicular exanthema in swine is encouraged when uncooked garbage is fed to hogs. In some instances, tuberculosis, swine erysipelas, and stomatitis may also be spread by raw garbage. The boiling for thirty minutes of all garbage fed to hogs will prevent transmission of trichinosis and economic loss of the swine industry due to hog illness and death.

Grinding The grinding of garbage is fast becoming a common method of garbage disposal. It is highly recommended from a convenience and public health standpoint, but the disposal of other refuse remains to be handled. The putrescible matter is promptly removed thereby eliminating this as a source of odors and food for rats, flies, and other vermin. In one system, the home grinder is connected to the kitchen sink drain. Garbage is shredded into small particles while being mixed with water and is discharged to the house sewer. In another system, garbage is collected as before but dumped into large, centrally located garbage grinding stations that discharge garbage to the municipal sewerage system. In small communities, the garbage grinding station may be located at the sewage-treatment plant. The strength of the sewage is increased and additional sludge digestion and drying facilities will be required when a large amount of garbage is handled.

Disposal at Sea Where dumping at sea is practiced, all garbage and other refuse is dumped into large garbage scows or barges. The scows are towed by tugs and the garbage is taken out to sea and dumped a sufficient distance out to prevent the refuse being carried back to shore and causing a nuisance. Bad weather conditions hamper this operation and unless this method is kept under very careful surveillance, abuses and failures will result. Because of the cost of maintaining a small navy and difficulties in satisfactorily carrying out this operation, coastal cities have reverted to sanitary landfill and incineration. In recent years, consideration has been given to the compaction of refuse to a density greater than 66.5 pounds per cubic foot prior to transport and then disposal by burial in the ocean at depths greater than 100 feet, based upon oceanographic conditions, to ensure there will be no mixing with surface water. More research is needed to determine stability of the compacted refuse and effect on marine life. See *High density compaction.*

Garbage Reduction In the reduction method of garbage disposal, the garbage is cooked under pressure. Fats melt out and are separated from the remaining material. The fat is used in the manufacture of soaps or glycerines and the residue is dried, ground, and sold for fertilizer or cattle feed. Odor complaints are associated with this process and, where a solvent such as naphtha is used to increase the extraction of fat, a greater fire hazard exists. The use of synthetic detergents and chemical fertilizers and high operating costs have led to the abandonment of this process.

Composting Composting is the controlled decay of organic matter in a warm, moist environment by the action of bacteria, fungi, molds, and other organisms. This may be an aerobic and/or anaerobic operation. Moisture is maintained at 40–65%, 50–60% is best. Composition of the refuse, disposal of refuse not composted, demand for compost and salvaged material, odor production and control, public acceptance and total cost are factors to be carefully weighed. Compost is a good soil conditioner but a poor fertilizer. The process is very attractive, but because of the factors mentioned has not met with success in the United States.

The composting operation involves a combination of steps. These may include: (a) weighing, (b) separation of non-compostables and salvage by hand and by a magnetic separator, (c) size reduction to 2 inches or less by means of a shredder, grinder, chipper, rasp mill, hammermill, (d) ballistic and magnetic separation, (e) biological digestion by any one of a number of composting methods, (f) screening and possible standardization of fertilizer value, and (g) disposal by bagging for sale, or to a sanitary landfill.

Pyrolization Pyrolization as applied to solid wastes (metal and glass removed) is an experimental thermochemical process for conversion of complex organic solids, in the absence of added oxygen, to water, combustible gases, tarry liquids, and a stable residue. Intermediate products may be collected or may be used to contribute heat to support the process. The end products would be carbon, water and carbon dioxide if carried to completion. If the raw material contains sulfur and nitrogen, these oxides will be formed with resultant air pollution unless provision is made for their removal. Temperatures of 900–1700°F have been used. In a variation, some oxygen and a temperature up to 2100°F is used. It is a process of destructive distillation similar to that used for making charcoal and for the recovery of organic by-products such as turpentine, acetic acid, and methanol from wood.[22] The Lantz converter and the Urban Research and Development Corporation unit are variations of the pyrolytic process. The Lantz system uses a temperature of 1200–1400°F with ground refuse.

High temperature incineration High temperature incineration is carried out at 3400°F (Melt-zit), 2500–2600°F (FLK Slagging Incinerator, Germany), 2600–3000°F (Torrax system), 3000°F (American Design and Development Corporation). These units are in the developmental and pilot stage. Combustibles are destroyed and non-combustibles are reduced to slag or sand-like grit. A 97% volume reduction is reported.

Wet oxidation Wet oxidation and anaerobic digestion of refuse are in the experimental stage. In wet oxidation, the refuse can be processed under high pressures and temperatures, or the refuse can be ground and aerated while in suspension in a liquid medium. In anaerobic digestion, decomposable material is separated from refuse, ground and then digested at controlled temperature in the absence of air. The resultant gases are mostly methane and carbon dioxide and the residue can serve as a soil conditioner.

Size reduction Shredders, hoggers, and chippers are devices which reduce refuse, bulky items, and other solid wastes to a manageable size for disposal by landfill or for processing through an incinerator.

High density compaction High density compaction of solid wastes is accomplished by compression to a density of more than 66.5 pounds per cubic foot. The resulting bales may be enclosed in chicken wire, hot asphalt, vinyl plastic, plain or reinforced cement, or welded sheet steel depending on the method of disposal or intended use of the bales or blocks. Rogus reports on the Tezuka-Kosan process in Tokyo that liquid release during compression ranged from 2–5% by weight and that it will require treatment. The bales had a density of 70–109 pounds per cubic foot; they sank in sea water; had good structural cohesiveness; resisted corrosion; were reasonably free of odors, and insect, rodent hazard; and showed no evidence of aerobic or anaerobic decomposition.[23] This system has not yet been adopted in the United States but offers attractive possibilities for solid waste reuse, economic hauling, and more acceptable disposal.

Other methods of achieving high density are subjecting refuse to pressure in a hydraulic pressure chamber and forcing baled refuse through a tube extending 30–50 feet below the water surface. Plastic toys and containers do not remain compressed or submerged. The effect of time on decomposition with possible gas formation, disintegration and floating of the refuse is not known. The long-term effect of the submerged material on the marine ecology is also an unknown.[24]

Incineration

Incineration is a controlled combustion process for burning solid, liquid, or gaseous combustible waste to gases and a residue containing little or no combustible material *when properly carried out.* It is a volume reduction process suitable for about 70% of the municipal solid wastes.

Incineration, if properly designed and controlled, is satisfactory for burning combustible refuse provided air pollution standards can be met. Continuous operation six days a week and a controlled temperature are needed for efficiency, prevention of excessive air pollution and odor control.

An operating design temperature range of 1500–1800°F is generally recommended. Excessively high temperatures and extreme variations cause cracking and spalling with rapid deterioration of fire tile and brick linings (refractories). Batch feed or one-shift operation promote spalling and loosening of tile linings. Other types of lining and design may permit higher operating temperatures. Actually, the temperature in the furnace will range from 2100–2500°F. When the gases leave the combustion chamber the temperature should be between 1400 and 1800°F and the gas entering the stack 1000°F or less. The temperature will have to be lowered to 450–500°F before the gas is filtered or to 600°F or less if electrical precipitators are used.

Sanitary Landfill

A sanitary landfill is an engineering method of disposing of solid waste on land by spreading the waste in thin layers, compacting it to the smallest practical volume and covering it with a layer of earth at the end of each day or more frequently in a manner which prevents environmental pollution.[25] The cost of sanitary landfill versus incineration for various populations is shown in Figure 12.

Social and political factors An important aspect of refuse disposal site selection, in addition to engineering factors, is the evaluation of public reaction and education of the public so that understanding and acceptance is developed. A program of

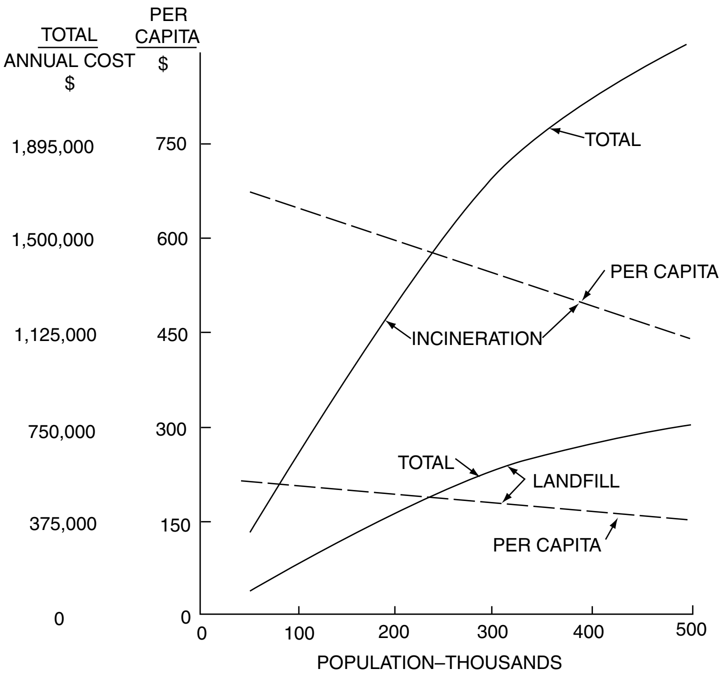


FIGURE 12 General cost comparison of sanitary landfill and incineration. From *Municipal Refuse Collection and Disposal.* Office for Local Government, Albany, NY, 1964. Adjusted to 1970 costs.

public information is also needed. Equally important are the climate for political cooperation, cost comparison of alternative solutions, available revenue, aesthetic expectations of the people, organized community support, and similar factors.

Experience shows that where open dumps have been operated, there will be opposition to almost any site proposed for sanitary landfill or an incinerator for that matter. However, local officials will have to study all of the facts and make a decision to fulfill their responsibility and exercise their authority for the public good in spite of any expressed opposition. Usually the critical factor is convincing the public that a nuisance-free operation can and will in fact be conducted. A private contract may make possible regional disposal, whereas an intergovernmental operation may be politically impractical.

Planning Local officials can make their task easier by planning ahead together on a county or regional basis for 20–40 years in the future, and by acquisition of adequate sites at least 5 years prior to anticipated needs and use. The availability of federal and state funds for planning for collection, treatment, and disposal of refuse on an area-wide basis such as a county should be explored. The planning will require an engineering analysis of alternative sites including population projections, volume, and characteristics of all types of solid wastes to be handled, cost of land and site preparation, expected life of the site, haul distances from the sources of refuse to the site, cost of operation, and possible value of the finished sites. Consideration would be given to the climate of the region, prevailing winds, zoning ordinances, geology, and topography. Location and drainage to prevent surface water or ground water pollution, access roads to major highways, and availability of suitable cover material are other considerations.

Special attention must be given to the diversion of surface water, to the highest ground water level so as to keep the bottom of the fill well above it, and to the soils available. The type of soil used for cover and its placement should minimize infiltration. The intervening soil between the bottom of the landfill and maximum ground water level should prevent or control leachate travel, or if this cannot be reasonably assured, an artificial membrane or clay soil barrier with an underdrain collection system may be needed.

Once a decision is made it should be made common knowledge and plans developed to show how it is proposed to reclaim or improve and reuse the site upon completion. This should include talks, slides, news releases, question and answer presentations, and inspection of good operations. Artist's renderings are very helpful in explaining construction methods and final use of the land. Figures 13 and 14 show the area and trench sanitary landfill methods.

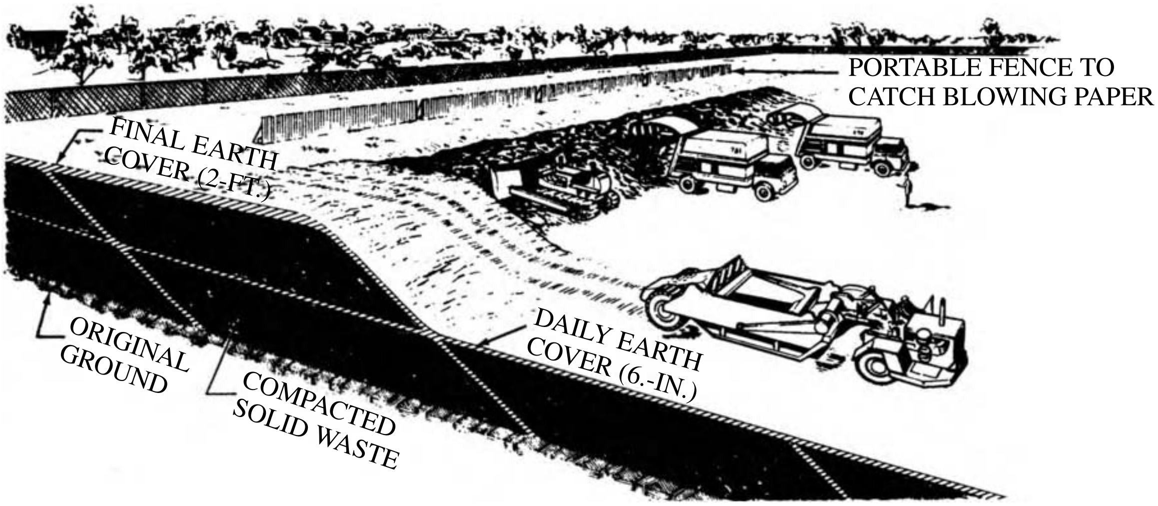


FIGURE 13 Area method. The bulldozer spreads and compacts solid wastes. The scraper (foreground) is used to haul the cover material at the end of the day's operations. Note the portable fence that catches any blowing debris. This is used with any landfill method.

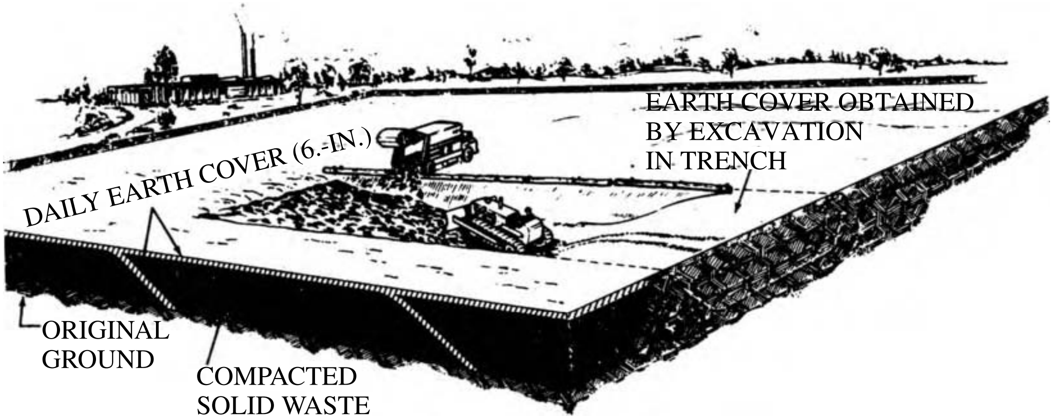


FIGURE 14 Trench method. The waste collection truck deposits its load into the trench where the bulldozer spreads and compacts it. At the end of the day the dragline excavates soil from the future trench; this soil is used as the daily cover material. Trenches can also be excavated with a front-end loader, bulldozer, or scraper. From *Sanitary Landfill Factors,* Thomas J. Sorg and H. Lanier Hickman, U.S. Department of Health, Education and Welfare, Public Health Service Publication N. 1792, Superintendent of Documents, Washington, DC, 1970.

RECREATION AREAS AND TEMPORARY RESIDENCES

Bathing Beaches and Swimming Pools

The sanitation of bathing places is dictated by health and aesthetic reasons. Few people would knowingly swim or water ski in polluted water; and insanitary surroundings are not conductive to the enjoyment of "a day at the beach." People are demanding more and cleaner beaches and pools, and a camp, hotel, club, or resort without a pool or beach is not nearly as popular as one that has this facility.

Health Hazards

From our knowledge of disease transmission, it is known that certain illnesses can be transmitted by improperly operated or located swimming pools and beaches through contact and ingestion of polluted water. Among these are typhoid fever, dysentery, and other gastrointestinal illnesses; conjunctivitis, trachoma, leptospirosis, ringworm infections, schistosomiasis, or swimmer's itch; upper respiratory tract diseases such as sinus infection, septic sore throat, and middle ear infection. The repeated flushing of the mucous coatings of the eyes, ears, and throat, and the excessive use of alum or lack of pH control, expose the unprotected surfaces to possible inflammation, irritation, and infection. Contraction of the skin on immersion in water may make possible the direct entrance of contaminated water into the nose and eyes.

Stevenson[26] reported that "an appreciably higher overall illness incidence may be expected in the swimming group over that in the nonswimming group regardless of the bathing water quality." It was further stated that "eye, ear, nose, and throat ailments may be expected to represent more than half of the over-all illness incidence, gastrointestinal disturbances up to one-fifth, and skins irritations and other illnesses the balance." Although based on limited data, swimming in a lake water with an average coliform content of 2300 per 100 ml caused "a significant increase in illness incidence . . ." and swimming in a river "water having a median coliform density of 2700 per 100 ml appears to have caused a significant increase in such (gastro-intestinal) illness." The study also showed the greatest amount of swimming was done by persons 5 to 19 years of age.

Diesch and McCulloch[27] summarized incidences of leptospirosis in persons swimming in waters contaminated by discharges of domestic and wild animals including cattle, swine, foxes, racoons, muskrats, and mice. Pathogenic leptospires were isolated from natural waters, confirming the inadvisability of swimming in streams and farm ponds receiving drainage from cattle or swine pastures.

Joyce and Weiser report that enteroviruses which are found in human or animal excreta, if introduced into a farm pond by drainage or direct flow, can constitute a serious public health hazard if used for recreation, drinking or other domestic purposes.[28]

Many other studies[29–30] have been made to relate bathing water bacterial quality to disease transmission with inconclusive or negative results.

British investigators[29] have drawn the following general conclusions:

1) That bathing in sewage-polluted sea water carries only a negligible risk to health, even on beaches that are aesthetically very unsatisfactory.
2) That the minimal risk attending such bathing is probably associated with chance contact with intact aggregates of faecal material that happen to have come from infected persons.
3) That the isolation of pathogenic organisms from sewage-contaminated sea water is more important as evidence of an existing hazard in the populations from which the sewage is derived than as evidence of a further risk of infection in bathers.
4) That, since a serious risk of contracting disease through bathing in sewage-polluted sea water is probably not incurred unless the water is so fouled as to be aesthetically revolting, public health requirements would seem to be reasonably met by a general policy of improving grossly insanitary bathing waters and of preventing so far as possible the pollution of bathing beaches with undisintegrated fecal matter during the bathing season.

The findings of the Public Health Activities Committee of the ASCE Sanitary Engineering Division[31] are summarized in the following abstract:

"Coliform standards are a major public health factor in judging the sanitary quality of recreational waters. There is little, if any, conclusive proof that disease hazards are directly associated with large numbers of coliform organisms. Comprehensive research is recommended to provide data for establishing sanitary standards for recreational waters on a more rational or sound public health basis. British investigations show that even finding typhoid organisms and other pathogens in recreational waters is not indicative of a health hazard to bathers but is only indicative of the presence of these diseases in the population producing the sewage. The Committee recommends that beaches not be closed and other decisive action not be taken because current microbiological standards are not met except when evidence of fresh sewage or epidemiological data would support such action."

In view of the available information, emphasis should be placed on elimination of sources of pollution (sewage, storm water, land drainage), effective disinfection of treated wastewaters, and on the proper interpretation of bacterial examinations of samples collected from representative locations.

In swimming pools there is a possibility of direct transmission of infection from one bather to another if the water does not have an active disinfectant such as free available chlorine. Proper operation and treatment therefore are of prime importance.

FOOD PROTECTION

Site Location and Planning

It is extremely important that a preliminary investigation be made of a site proposed for a new plant to determine the suitability of a site for a given purpose. This should include restrictions related to zoning, sewage, and wastewater disposal, solid wastes disposal, air pollution control, and noise.

Time and money spent in study, before a property is purchased, is a good investment and sound planning. For example, a factory that requires millions of gallons of cooling or processing water would not be located too distant from a lake or clear stream, unless it was demonstrated that an unlimited supply of satisfactory well water or public water was available at a reasonable cost. An industry having as an integral part of its process a liquid, solid, or gaseous waste would not locate where adequate dilution or disposal area was not available, unless an economical waste treatment process could be devised.

Such factors as topography, drainage, highways, railroads, watercourses, exposure, swamps, prevailing winds, dust, odors, insect and rodent prevalence, type of rock and soil, availability and adequacy of public utilities, the need for a separate power plant, and water-, sewage-, or waste-treatment works and special air pollution control equipment should all be considered and evaluated before selecting a site for a particular use.

Engineering and Architectural Considerations

The design of structures and facilities are properly engineering and architectural functions that should be delegated to individuals or firms who have become expert in such matters. The design and construction material used is dependent upon the type of structure, its geographical location and operation. Some materials can be used to greater advantage because of location, availability of raw materials, labor costs, type of skilled labor available, local building codes, estimate, and other factors.

In many instances plan approval and permits are required from federal, state or local regulatory agencies. Check with the agency having jurisdiction before any decisions are made or any plans are prepared.

Basic Sanitation Requirements for Food Processing and Food Service Establishments

Regulation The regulation and supervision of food establishments such as restaurants, caterers, commissaries, pasteurizing plants, frozen dessert plants, frozen prepared food plants, vending machine centers, slaughter houses, poultry processing, bakeries, shellfish shucking and packing plants, and similar places is in the public interest. This responsibility is usually vested in the state and local health and agriculture departments, and also in the Food and Drug Administration, the U.S. Department of Agriculture and in the U.S. Public Health Service when interstate traffic of the food is involved. The industry affected as a rule also recognizes its fundamental responsibility.

Industry and regulatory control emphasis needs to be placed at the foci of food collection, processing and distribution and on self-inspection. See page 355. Proper design, equipment and layout, and construction must build out environmental sanitation problems and simplify sanitary operations. Quality control procedures must also be applied to convenience type food such as prepared frozen dinners. The food catering industries which now serve banquets, hospitals and nursing homes, airlines, and other mass feeding facilities are particularly vulnerable. Food preparation is leaving the home and small restaurant and is becoming a centralized production process leaving only final conditioning or heating at the point of service. Such operations make possible more centralized control and certain efficiencies, but also increase the potential of foodborne outbreaks affecting large numbers of people.

General requirements It is to be noted that certain basic sanitation requirements are common to all places where food is processed. McGlasson[32] has proposed a set of standards under the following headings:

1) *Location, construction, facilities, and maintenance* (a) Grounds and premises, (b) Construction and maintenance, (c) Lighting, (d) Ventilation, (e) Dressing rooms and lockers.
2) *Sanitary facilities and controls* (a) Water supply, (b) Sewage disposal, (c) Plumbing, (d) Rest-room facilities, (e) Handwashing facilities, (f) Food wastes and rubbish disposal, (g) Vermin control.
3) *Food-product equipment and utensils* (a) Sanitary design, construction, and installation of equipment and utensils, (b) Cleanliness of equipment and utensils.
4) *Food, food products, and ingredients* (a) Source of supply, (b) Protection of food, food products, and ingredients.
5) *Personal* (a) Health and disease control, (b) Cleanliness.

A similar set of basic standards were published in the Federal Register in April 1969.[33]

Specific requirements The general sanitation requirements should be supplemented by specific regulations applicable to a particular establishment or operation. Excellent codes, compliance guides, and inspection report forms are available from regulatory and training organizations.

Food preparation Temperature control and clean practices should be the rule in kitchens and food-processing plants if contamination is to be kept to a minimum. All food contact surfaces and equipment used in food preparation must be kept clean and in good repair.

Frozen meat, poultry, and other bulk frozen foods should be thawed slowly under controlled refrigeration (36–38°F is ideal) and *not left to stand at room temperature overnight to thaw.* Frozen vegetables and chops need not be thawed but can be cooked directly. Prepared foods, especially protein types, should be served immediately, kept temporarily at a temperature of 45°F or less, or on a warming table maintained at a temperature above 140°F until served. If not to be served, the food should be refrigerated in shallow pans to a depth of 2 to 3 inches within 30 minutes at a temperature of 45°F. Bulk foods such as roasts should be refrigerated within 30 minutes of preparation and cooled so that the internal temperature is 40°F or less within two hours, unless immediately served. Note that souring of hot foods is prevented by prompt refrigeration.

Food handlers are expected to have hygienic habits. A conveniently located wash basin in the kitchen or work room, and in the toilet room, supplied with warm running water, soap, and individual paper towels are essential.

Microbiological standards Natural microbial variations in different foods, and the statistical aspects of sampling present considerable difficulties in the establishment of firm standards. In addition, because of normal errors inherent in laboratory techniques, it is practical to allow some leeway in the standards proposed and hence refer to them as guidelines. One might use the geometric mean of say 10 samples or allow one substandard sample out of four. Parameters used include total aerobic count, toxigenic molds, number of coliforms, number of *E. coli,* coagulase-positive staphylococci, salmonellae, shigella, *Clostridium perfringens, Clostridium botulinum,* and betahemolytic streptococci as indicated.

Milk Control

The milk industry and the regulatory agencies have a joint responsibility in ensuring that all milk and milk products consistently meet the standards established for protection of the public health. Inspection duplication should be avoided; instead there should be a deliberate synergism of effort. With proper planning and cooperation the industry, local, state, and federal systems can actually strengthen the protection afforded the consumer. The role of industry and official agencies to accomplish the objectives stated is described below.

Certified industry inspection Industry quality control inspectors are qualified by the official agency (usually health or agriculture) based upon education, experience, and examination to make dairy farm inspection pursuant to the milk code or ordinance. Certificates issued are issued for a stated period of time, one to three years, may be revoked for cause, are renewed based on a satisfactory work record, and may require participation in an annual refresher course. Copies of all inspections, field tests, veterinary examinations, and laboratory reports are promptly forwarded to the official agency or to an agreed upon place and kept on field at least one year.

Cooperative State—Public Health Service program for certification of Interstate Milk Shippers The federal-state program is commonly referred to as the IMS Program. A State milk sanitation rating officer certified by the U.S. Public Health Service makes a rating of a milk supply. The name of the supply and rating is published quarterly by the Public Health Service. If the milk and milk products are produced and pasteurized under regulations which are substantially equivalent to the PHS Milk Ordinance, and is given an acceptable milk sanitation compliance and enforcement rating, the milk and milk products may be shipped to another area of jurisdiction which is participating in the IMS Program.

The procedures for rating a milk supply are carefully designed with detailed instructions to be followed by the industry and the rating officer. Independent evaluations are made by Public Health Service rating officers to confirm ratings given, or changes since the last rating, and ensure reproducibility of results.

Official local program supervision and inspection The official agency makes regular review of the industry inspection files mentioned above, takes whatever action is indicated, and at least annually makes joint inspections with the industry inspector of a randomly selected significant number of dairy farms, including receiving stations. The quality of work done is reviewed, the need for special training is determined, and recommendation concerning certificate renewal is made to the permit issuing official. A similar review is also made of the sample collection, transportation and the procedures, equipment, and personnel in the laboratory making the routine milk and water examinations.

In addition, the local regulatory agency collects official samples as required by the state milk code, advises the industry having jurisdiction of the results and corrective action required, participates in training sessions, and serves as the state agent in securing compliance with the state milk code. The authorized local, city, or county agency usually has responsibility for the routine inspection of processing plants, sample collection, and overall program supervision for compliance with the code. The agency sanitarian serves as a consultant to the industry in the resolution of the more difficult technical, operational, and laboratory problems. This whole procedure makes possible better use of the qualified industry inspector, and the professionally trained sanitarian, with better direct supervision over dairy farms and pasteurization plants, and more effective surveillance of milk quality. In some states the local activities are carried out in whole or in part by the state regulatory agency.

Official state surveillance and program evaluation The state department of health, and in some instances the state department of agriculture, share responsibility for milk sanitation, wholesomeness and adulteration. The responsibility is usually given in state law, sanitary code or milk code, and in rules and regulations promulgated pursuant to authority in the law. Most states have adopted the *Grade "A" Pasteurized Milk Ordinance,* 1965 Recommendations of the United States Public Health Service, or a code which is substantially the same. This makes possible a reasonable basis for uniformity in regulations and interpretation both interstate and intrastate. However short-term and alleged economic factors frequently limit reciprocity and interstate movement of milk. Both industry and regulatory agencies need to cooperate in the elimination of milk codes as trade barriers. Health agencies usually have no objection to reciprocity where milk quality compliance and enforcements is certified under the Public Health Service IMS Program. However not all states or farms participate in the program.

The IMS Program also gives the state regulatory agency a valuable tool to objectively evaluate the effectiveness of the local routine inspection and enforcement activities. The types of additional training needed by the qualified industry inspector, the assistance the dairy farmer should have, and the supervisory training needed by the regulatory agency sanitarian become apparent. Changes in technology and in practices are noted and need for clarification of regulations, laws, and policies are make known to the state agency.

A common function of a state regulatory agency is periodic in-depth evaluation of local milk programs. This includes the effectiveness with which the local unit is carrying out its delegated responsibilities as described above, the quality of work done, the staff competency and adequacy, the reliability of the official laboratory work, the record keeping, the equipment and facilities available, the number of inspections and reinspections made of dairy farms and processing plants and their adequacy of excessive frequency. The state agency usually has responsibility for approving all equipment used in milk production and service from the cow to the consumer. The standards recommended by national organizations are generally used as a basis for the acceptance of equipment.

Pasteurization Pasteurization is the process of heating every particle of milk or milk product to at least 145°F and holding at or above such temperature continuously for at least 30 minutes, or to at least 161°F and holding at such temperature continuously for at least 15 seconds, in approved and properly operated equipment. A temperature of at least 150°F for 30 continuous minutes (155°F for 30 minutes in practice) or at least 166°F for 15 seconds (175°F for 25 seconds in practice) is needed for milk products which have a higher milkfat content than milk or contain added sweeteners. Higher temperatures and shorter times are also being used. The heat treatment should be followed by Prompt cooling to 45°F.

The effectiveness of pasteurization in the prevention of illnesses that may be transmitted through milk-borne disease organisms has been demonstrated beyond any doubt. The continued sale and consumption of raw milk must therefore be attributed to ignorance of these facts. Pasteurization does not eliminate pesticide residues, anthrax spores, or toxins given off by certain staphylococci; but the production of toxins is nil when milk is properly refrigerated. The equipment used to pasteurize milk is of three general types. These are the holder, high-temperature short-time, and in-the-bottle pasteurizers. High-temperature pasteurizers include the vacuum type (temperature 162–200°F) and the ultra-high temperature type (190–270°F).

Quality control Quality control involves herd health, milk handling, transportation, processing, and distribution. Field and laboratory testing coupled with inspection, supervision, education surveillance, enforcement, and evaluation are the major methods used.

The tests used to control the quality of milk are explained in detail in *Standard Methods for the Examination of Dairy Products* Twelfth Edition. The major tests are mentioned here. Raw milk quality is determined by temperature, sediment, odor and flavor, appearance, antibiotics test, direct microscopic counts including clumps of bacteria, leucocytes, and streptococci; standard plate counts, abnormal milk tests, and thermoduric determination. Tests for brucellosis and animal health are also made. Pasteurized milk quality is indicated by the standard plate count, direct microscopic count, phosphatase test, coliform test, taste, and odor tests. Other common tests are for butterfat, total solids, and specific gravity.

REFERENCES

1. WHO Expert Committee on Environmental Sanitation, Wld. Hlth. Org. techn. Rep. Ser., 1950, No. 10, p. 5.
2. Stead, Frank, M., Levels in environmental health, *AJPH,* **50**, No. 3, March 1960, p. 312.
3. Wld. Hlth. Org. techn. Rep. Ser., 1954, No. 77, p. 9.
4. Environmental factors in health planning, *AJPH,* **58**, No. 2, February, 1968, pp. 358–361.
5. National environmental health programmes: Their planning, organization, and administration, Wld. Hlth. Org. techn. Rep. Ser., 1970, No. 439. pp. 10–11.
6. *Local and Regional Planning and Development,* Governor's Task Force Report, West Virginia, Regional Review, National Service to Regional Councils, Washington, D.C., January 1970, p. 2.
7. Weibel, S.R., F.R. Dixon, R.B. Weidner, and L.J. McCabe, Waterborne-disease outbreaks, 1946–60, *Jour. AWWA,* Aug. 1964, pp. 947–58.
8. Craun, Gunther F. and McCabe, Leland, J., Review of the causes of waterborne-disease outbreaks, *Jour. AWWA,* Jan. 1973, pp. 74–84.
9. Dennis, Joseph M., 1955–56 Infectious hepatitis epidemic in Delhi, India, *Jour. AWWA,* Oct. 1955, pp. 1288–98.
10. Wolman, Abel, Water supply and environmental health, *Jour. AWWA,* Dec. 1970, pp. 746–749.
11. Woodward, William E. *et al.,* Foodborne disease surveillance in the United States, 1966 and 1967, *AJPH,* Jan. 1970, pp. 130–37.
12. Lofroth, Goran, Methylmercury, Redaktionstjansten, Natural Science Research Council, Stockholm, Sweden, March 20, 1969.
13. A special report to the Secretary's Pesticide Advisory Committee, *Hazards in Mercury,* Department of Health, Education, and Welfare, and the Environment Protection Agency, Washington, D.C., November, 1970.
14. Cooke, N.E. and A. Beitel, Some aspects of other sources of mercury to the environment, Symposium Mercury in Man's Environment, Royal Society of Canada, February 15, 1971.
15. *Statistical summary of Municipal Water Facilities in the United States,* Public Health Service Publ. No. 1039, Washington, D.C., January 1, 1963.
16. McCabe, Leland J., James, M. Symons, Roger D. Lee and Gordon, G. Robeck, Survey of community water, supply, systems, *Jour. AWWA,* Nov. 1970, pp. 670–687.
17. *Hydrology Handbook* Manual 28, A. Soc. of Civil Engrs., N.Y., 1949.
18. *Public Health Service Drinking Water Standards 1962* PHS Pub. 956, U.S. Department of Health, Education and Welfare, Washington, D.C.
19. *Glossary Water and Wastewater Control Engineering,* APHA, ASCE, AWWA, WPCF, 1969.
20. *Manual of Recommended Water-Sanitation Practice,* Recommended by the United States Public Health Service, 1946, Public Health Bulletin No. 296.
21. Saunders, W.R., Water and wastes, the public health viewpoint, *Royal Society of Health Journal,* Sept/Oct. 1968, pp. 247–250.
22. Hoffman, Donald A. and Richard, A. Fitz, Batch retort pyrolysis of solid municipal wastes, *Environmental Science and Technology,* **2**, November 1969, pp. 1023–1026.
23. Rogus, Casimir A., High compression baling of solid wastes, *Public Works,* June 1969, pp. 85–89.
24. Golueke, Clarence G. and P.H. McGauhey, Future alternatives to incineration and their air pollution potential, *Proceedings: The Third National Conference on Air Pollution,* U.S. Department of Health, Education and Welfare, Washington, D.C., December, 12–14, 1966.
25. Brunner, D.R., D.J. Keller, C.W. Reid, Jr. and J. Wheeler, *Sanitary Landfill Guidelines-1970,* Public Health Service, Dept. of Health, Education and Welfare, Washington, D.C., 1970.
26. Stevenson, Albert H., Studies of bathing water quality and health, presented before the second session of the Engineering Section of the American Public Health Association at the 80th Annual Meeting in Cleveland, Ohio, Oct. 23, 1952.
27. Diesch, Stanley L. and William, F. McCulloch, Isolation of pathogenic leptospires from waters used for recreation, *Public Health Reports,* **81**, No. 4, April 1966, pp. 299–304.
28. Joyce, Gayle and H.H. Weiser, Survival of enteroviruses and bacteriophage, in farm pond waters, *Jour. AWWA,* Vol. 59, No. 4, April 1967, pp. 491–501.
29. Sewage contamination of coastal bathing waters in England and Wales, Committee on Bathing Beach Contamination of the Public Health Laboratory Service, *The Journal of Hygiene,* **57**, No. 4, December 1959, pp. 435–472.
30. Henderson, John M., Enteric disease criteria for recreational waters, *Journal of the Sanitary Engineering Division,* ASCE, SA6, Proc. Paper 6320, December 1968, pp. 1253–1276.
31. Coliform standards for recreational waters, Committee Report, *Journal of the Sanitary Engineering Division,* ASCE, SA4, Proc. Paper 3617, August 1963, pp. 57–94.
32. McGlasson, E.D., Proposed Sanitation Standards for Food-Processing Establishments, Association of Food and Drug Officials of the United States, Quarterly Bulletin, April 1967.
33. Title 21—Food and Drugs, Chapter 1, Food and Drug Administration, Department of Health, Education and Welfare, Part 128, Human foods; Current food manufacturing, processing, packing, or holding, Federal Register, April 25, 1969.
34. Low-cost water supply and sanitation technology; Pollution and health problems, WHO, New Delhi, 1985.
35. Watterson, A., Pesticide users' health and safety handbook, Gower, Aldershot, 1988.
36. Von Sonntag, C., Disinfection with UV radiation, in Process Technologies for Water Treatment, S. Stück, Ed., Plenum, New York, 1988.
37. White, G.C., Handbook of chlorination, 2nd ed., Van Nostrand Reinhold, New York, 1989.

JOSEPH A. SALVATO, Jr.
Troy, New York

ENVIRONMENTAL LAW

1

Environmental Law can be defined as those statutes, administrative regulations, executive orders, court decisions and international agreements that pertain to protection of Nature and human beings from the consequences of unwise resource extraction, production or development and the procedural means by which these 'laws' can be invoked by citizens and government.

Constitutionally, the United States exhibits a federal system: Law making and enforcement power is divided among the Federal government, the fifty State governments, regional units and local governments according to formulas contained in Federal and State constitutions and statutes. Environmental Law is a complex and delicate set of relationship among various levels of government. For example the *Clean Water Act* on the federal level establishes the general framework of water pollution control laws, empowering the Federal Environment Protection Agency (EPA) to set federal minimum performance standards for discharges to waterways, distribute grant monies for publicly owned treatment works and perform backup permitting, surveillance, and enforcement functions. But, in most cases it is the States that grant discharge permits (which can be stricter than the federal minimum standards), determine the recipients of federal grant funds, monitor discharger compliance and enforce against violators. Water quality management is frequently a function of substate regional planning agencies. Moreover, in some areas interstate regional organizations, such as the Delaware River Basin Commission, have regulatory authority over dischargers. Finally, most municipalities possess storm-water and other "public nuisance" ordinances.

Because of the multi-level nature of Environmental Law and its administration, there are numerous state, regional and local variations on the generally uniform federal environmental legal regimes. Consequently, this article will focus on Federal Environmental Law. Most federal environmental protection statutes are lengthy and complicated documents. When augmented by federal executive orders, administrative regulations and judicial interpretations, each federal statutory system would merit an article unto itself. Rather than that, the function of this article is to identify and briefly describe the major federal environmental protection statutes and to provide the reader with references so that he or she can seek out relevant regulations, orders, and decisions if more information is desired in a particular area. (For each statute mentioned the *United States Code* citation is given.)

2

The term "Environmental Law" did not come into general usage until the enactment of *The National Environmental Policy Act* (42 USC 4321 *et seq.*) in early 1970. However, long before NEPA Congress was occupied with conserving the Nation's natural resources from reckless despoilation. For example, Yellowstone National Park was created in 1872 and the US Forest Service, in the United States Department of Agriculture, was entrusted with the care of forested public domain lands in 1897. In addition, after World War II Congress began to legislate in the area of pollution control, culminating in the major federal pollution control statutes of the 1970's. The contemporary body of Federal Environmental Law consists of laws in the conservation and pollution control areas, along with key planning and coordination statutes, such as NEPA. But even more than most fields of law, Environmental Law is constantly changing and evolving as the Nation responds to shifting priorities and pressures on resources. Thus, current information is particularly important when it comes to Environmental Law.

3

In this section major federal environmental protection statues will be discussed under the following headings: Right-to-Know Statutes, Right-to-Sue Statutes, Product-Licensing Statutes; Standard-Setting Statutes; Cleanup-Response Statutes; and Conservation Statutes.

RIGHT-TO-KNOW STATUTES

NEPA

NEPA requires each federal agency that proposes a major federal action which may have a significant effect on the human environment to prepare and circulate—in draft and final form—an Environmental Impact Statement (EIS) discussing the environmental impact of the proposed action, reasonable alternatives, irretrievable commitments of resources, and the agency's balance among environmental protection and economic development, national security and other factors. Federal actions covered by NEPA are not only federal projects (roads, dams, etc.), but also federal licenses, permits, leases, contracts, and research projects that may have significant effects on the environment. NEPA does not mandate particular results: it does not require a federal agency to adopt the least environmentally damaging alternative. Courts frequently refer to NEPA as an "environmental full disclosure law," obliging an agency to take a "hard look" at the environmental consequences of its action. Once the EIS is complete (i.e., the agency makes full disclosure), the political process, not the courts, determines whether the agency will complete the proposed action. In most instances NEPA does guarantee that other federal agencies having expertise in the field, State officials and the general public will have an opportunity to formally comment on a proposed federal action before it is undertaken. When federal agencies have taken their responsibilities seriously, they have integrated environmental planning into the early stages of project planning, giving good faith consideration to alternatives and mitigation measures. However, all too often the EIS becomes a massive, unreadably technical rationalization of a project chosen for traditional reasons—economics, politics or bureaucratic inertia.

Energy Planning and Community Right-to-Know Act of 1986

Title III of the 1986 amendments to the Comprehensive Environmental Response, Compensation, and Liability Act (42 USC 9601 *et seq.*) established a regulatory program that will require disclosure of information to workers and the general public about the dangers of hazardous chemicals as well as development of emergency response plans for chemical emergencies.

Emergency response plans will be prepared by local emergency planning committees under the supervision of state emergency response commissions. Local committees are appointed by state commissions; and each committee must include representation from a wide range of community groups, including owners and operators of facilities in the planning area. Plans for responding to chemical emergencies were to have been completed by 1988. Facilities that release extremely hazardous chemicals in amounts over threshold limits must immediately notify the community emergency coordinator and the state commission.

Facilities that are required by the Federal Occupational Safety and Health Act to have "material safety data sheets" (MSDSs) for hazardous chemicals must submit a copy to the local committee, the local fire department, and the state commission. All MSDSs must be made available to the public. In addition, these facilities must prepare and submit detailed "emergency and hazardous chemical inventory forms" which describe amounts of hazardous chemicals present at the facility and their approximate locations.

Section 313 of CERCLA requires certain facilities to submit to EPA and the host state annual "Toxic Chemical Release Forms" reporting on the total quantity of listed chemicals used on-site and released to the environment—either accidentally or intentionally—through discharges to air, water, and land.

RIGHT-TO-SUE STATUTES

Citizen Suits

Almost every federal environmental protection statute contains a provision granting citizens, no matter where they are located, standing to bring suit in federal district court against violators of the statute or against EPA for failure to perform a mandatory duty. No specific injury need be alleged by a plaintiff in a statutory citizen suit, but plaintiff must allege that no government agency is effectively enforcing against the defendant. In a citizen suit, a judge may fine the defendant and award reasonable counsel and expert witness fees to a citizen plaintiff.

Judicial Review

Under the Federal Administrative Procedure Act (5 USC 701 *et seq.*), any "aggrieved person" may sue to obtain judicial review of a final decision made by a federal administrative agency. Potential damage to a person's interest in a clean and natural environment will confer standing on a plaintiff under this statute. Agency decisions are accorded a strong presumption of correctness by our legal system. Thus, in order to overturn an agency decision, a plaintiff must prove that it conflicts with the agency's statutory authority, is "arbitrary and capricious", or violates constitutional safeguards.

PRODUCT LICENSING STATUTES

Toxic Substances Control Act (TSCA)

Under TSCA (15 USC 2601 *eq seq.*), EPA is empowered to require the testing of new and existing chemicals that are potentially toxic, and to prohibit or place conditions on the manufacture, distribution, and usage of a chemical if it poses an unreasonable risk to human health or the environment.

The heart of TSCA is its requirement of a "premanufacture notice" (PMN) to EPA by any person who proposes to import, manufacture, or process a new chemical or to introduce a significant new use of an existing chemical. Once a PMN is received, EPA must act within one hundred eighty days to regulate or prohibit the manufacture of the chemical. If EPA does not act within one hundred eighty days, the manufacturing process may begin. However, if on the basis of the PMN EPA finds that production, use, or disposal of the chemical "may present an unreasonable risk" to human health or the environment, EPA may promulgate "test rules" requiring further testing of health effects, environmental effects, and chemical fate. If these tests confirm the existence of an unreasonable risk, EPA may act to limit or prohibit production, use, or disposal. "Reasonableness" in TSCA consists of a "risk-benefit" analysis in which the health and environmental risk is weighed against the social and economic benefits of the chemical.

As for existing chemicals, TSCA established an "Interagency Testing Committee" (ITC) that screens chemicals and recommends to EPA existing chemicals that are so potentially dangerous as to deserve further study. The ITC has developed several lists of potentially dangerous chemicals, including a "priority list" and a "suspicious list". Chemicals on these lists are subject to extensive monitoring and reporting requirements with regard to production, release, and exposure data. When a report by a manufacturer, processor, or distributor indicates that the substance or mixture might present an unreasonable risk to human health or the environment, EPA may require further tests or act to limit manufacture, use, or disposal of the chemical.

Finally, TSCA establishes a framework for regulating certain extraordinarily dangerous chemicals, for example asbestos, TRIS and PCBs. TSCA is one of the federal statues that could be relied upon to regulate products of genetic engineering.

Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA)

Under FIFRA (7 USC 136 *et seq.*), all new pesticides must be registered with EPA and all pesticides registered in the past must be reviewed with regard to potential adverse impacts on public health and the environment. A registration must be approved or renewed unless the pesticide will have "unreasonable adverse effects on the environment." EPA is authorized to restrict the use of certain pesticides to particular uses or to certified applicators, order specific packaging or labeling requirements, cancel a registration, or suspend the production and distribution of a pesticide.

If information disclosed in a pesticide application indicates toxicity exceeding EPA "risk criteria," a presumption against registration arises and the burden of proof shifts to the applicant. States are also authorized to regulate pesticides and pesticide applicators.

Food, Drug, and Cosmetic Act (21 USC 301 et seq.)

No "food additive", "color additive", drug or cosmetic may be sold unless it has been approved by the Food and Drug Administration (FDA). No FDA approval may be issued if the substance is found to cause cancer in humans or laboratory animals. EPA also sets tolerance levels for pesticide residues on raw agricultural products.

Atomic Energy Act (42 USC 2011 et seq.)

A Nuclear Regulatory Commission license must be obtained for the production or distribution of nuclear materials. Separate licenses must be obtained in order to construct or operate a nuclear power plant. The Food and Drug Administration's Bureau of Radiological Health regulates the uses of nuclear materials in the medical arts, and also regulates radioactivity levels in milk. The Environmental Protection Agency has responsibility for advising on environmental standards for radioactive materials, and for setting such standards for drinking water (under the Safe Drinking Water Act) and for ambient air and air emission sources (under and Clear Air Act).

STANDARD-SETTING STATUTES

Clean Air Act (42 USC 7401 et seq.)

This Act establishes federal nationwide ambient air quality standards for conventional air pollutants and provides for their attainment by reducing pollution from stationary and mobile sources. Stationary sources must meet emissions limitations set in State Implementation Plans (SIPs) administered by the States. Mobile sources pollution control is achieved by federal emissions limitations on new motor vehicles and state inspection and maintenance programs plus warranty provisions for vehicles in actual use. The Act also regulates new sources of air pollution, emissions of hazardous air pollutants, emissions limitations and transportation control plans in "Nonattainment Areas" (areas where the federal standards are being exceeded) and allowable pollution in "Prevention of Significant Deterioration" (PSD) areas where standards are presently being met.

Clean Water Act (33 USC 1251 et seq.)

Enacted in 1972 and amended several times since then, this statute sets out a goal of "fishable/swimmable" waters wherever attainable, and attempts to achieve it through (1) a construction loan program authorizing federal grants to State revolving loan funds for the construction and rehabilitation of publicly owned treatment works; (2) permit system for point sources of pollution incorporating progressively stricter technology-based effluent limitations set by industrial subcategory; (3) a requirement that publicly owned treatment works achieve the equivalent of secondary treatment; (4) areawide water quality management for the diminution of nonpoint sources of pollution and (5) setting of Total Maximum Daily Loads (TMDLs) where water quality standards are not achieved.

The Act makes it illegal to discharge a pollutant without a permit. Most States have assumed administration of the point source permit program from EPA. Also included in the *Clean Water Act* are provisions dealing with wetlands protection, sludge disposal, estuarine protection and ocean discharge. (Permits for ocean dumping are provided for in the *MarineProtection Research and Sanctuaries Act,* 33 USC 1401 *et seq.*) Concentrated Animal Feeding Operations (CAFOs) are regulated as point sources under the Act.

Safe Drinking Water Act (42 USC 300f et seq.)

Whereas the Clean Water Act regulates ambient water quality, the Safe Drinking Water Act controls the quality of tap water. The Act specifies procedures for setting maximum contaminant levels for drinking water provided by public water systems. It also governs underground injection of wastewater and limits federal development programs in areas designated as "sole source aquifers." Special programs are included for the protection of aquifer recharge areas and areas contiguous to wellheads.

Resource Conservation and Recovery Act (RCRA)

RCRA (40 USC 6901 *et seq.*) is the statute that governs the management of hazardous waste in the United States. RCRA's approach to hazardous waste management consists of 4 major elements:

- federal identification of hazardous wastes:
- a manifest system of tracing hazardous wastes from generator, to transporter, to treatment, storage, or disposal facility;
- federal minimum standards for hazardous waste treatment, storage, and disposal, enforced through a permit system; and
- state implementation of hazardous waste management programs at least equivalent to the federal program.

First enacted in 1976, RCRA was significantly amended in 1984. Under the 1984 amendments the land disposal of hazardous waste, including deep well injection, must be banned unless EPA determines that a particular method of land disposal will be consistent with protecting human health and the environment. A method of land disposal cannot be acceptable unless a petitioner demonstrates that there will be no migration from the land disposal unit for as long as the waste remains hazardous. If a disposability determination is not made within sixty-six months, land disposal of the hazardous waste is automatically banned.

The 1984 RCRA amendments also prohibited the landfilling of noncontainerized liquids, tightened performance standards for existing landfills, and established a regulatory program for underground storage tanks.

Occupational Safety and Health Act (29 USC 651 et seq.)

This statute imposes a duty on employers to furnish each employee with a place of employment that is free of recognized hazards that are causing or likely to cause illness, injury or death. The Department of Labor can promulgate safety and health standards, conduct inspections, issue citations, impose penalties, and require employers to maintain records. Health standards may be expressed as "permissible exposure limits" (PELs) which must be imposed where there is a significant risk to worker health and safety. PELs must be technologically and economically feasible.

Consumer Product Safety Act (15 USC 2051 et seq.)

Congress, through this Act, created the Consumer Product Safety Commission and authorized it to impose performance standards or packaging and labeling requirements on consumer products in order to prevent an "unreasonable risk of injury." This entails a benefit-cost analysis before the Commission can impose a restriction on a consumer product. In an extreme case, a hazardous product may be banned from commerce.

CLEANUP-RESPONSE STATUTES

Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)

CERCLA (42 USC 1901 *et seq.*), also known as "Superfund," established an 8.5 billion dollar fund to finance cleanup activities at currently operational and abandoned facilities. The Act also contains spill notification provisions covering listed hazardous substances.

When a release has occurred, EPA can use fund monies to clean up the site and then proceed against responsible parties for reimbursement, or move against responsible parties in the first instance. CERCLA imposes strict "joint and several" (recovery may be obtained from one or all) liability against (1) current owners or operators of facilities, (2) owners or operators at the time the hazardous substances were released, and (3) generators and transporters of the hazardous substances that were ultimately released by the facility. Only innocent purchasers who have made reasonable investigations are insulated from cleanup liability. CERCLA amendments enacted in 1986 have significant bearing on how quickly, and to what extent, superfund sites will be cleaned up. Remedial investigation/feasibility studies for facilities on the National Priority List must be commenced at the following rate: 275 with in three years; and additional 175 within four years; and an additional 200 within five years; for a total of 650 by 1991.

The amendments require EPA to select, to the maximum extent practicable, remedial actions that utilize permanent solutions and alternative treatment technologies or resource recovery technologies. A preference is established for remedial actions that utilize treatment to permanently and significantly reduce the volume, toxicity, or mobility of hazardous substances. Offsite transport and disposal without treatment is the least preferred option where practicable treatment technologies are available. If the selected remedy does not achieve the preference for treatment, EPA must publish explanation. For on-site remedial actions, the amendments require attainment of "legally applicable or relevant and appropriate Federal and state standards, requirements, criteria, or limitations" (ARARs), unless such requirements are waived. Maximum contaminant level goals under the Safe Drinking Water Act and water quality standards under the Clean Water Act must be met where relevant and appropriate.

Congress' failure to reauthorize CERCLA has caused a depletion of Superfund so that State cleanup funds are now of primary importance for site remediation. However, since 1990 CERCLA has been amended to provide various incentives for remediation of urban ("Brownfields") contaminated sites.

Oil Pollution Act of 1990 (OPA '90)

Galvanized into action by the Exxon Valdez oil spill in 1989, Congress enacted the Oil Pollution Act of 1990 (OPA '90), which integrated and strengthened prior federal law covering liability for and cleanup of oil spills. Modelled on CERCLA, OPA '90 establishes a billion dollar Oil Spill Liability Trust Fund financed by the imposition of a five cents per barrel tax on oil delivered at the refinery. Owners and operators of vessels and onshore and offshore facilities are strictly liable for cleanup and natural resource damages, subject to the defenses of (1) act of God, (2) act of war, or (3) negligence of a third party not associated with the owner or operator. Facilities must develop spill prevention, control, and countermeasure (SPCC) plans, have them approved by EPA or the Coast Guard, and implement them, or face heavy civil and criminal penalties. Single hull tankers are to be phased out, and licensing and supervision of officers and seamen are strengthened. The OPA '90 explicitly does not preempt state oil spill laws.

CONSERVATION STATUTES

The *Surface Mining Control and Reclamation Act* (30 USC 1201 *et seq.*) provides extensive federal regulation of private coal surface mining on private lands and also regulates excavation of coal by strip-mining on public lands. The basic regulatory mechanism of this act is a permit system, to be implemented by the States according to federal standards, with provision for federal enforcement if a state program does not meet federal standards. There are also flat prohibitions on surface mining in environmentally sensitive areas. The Act is administered by the Office of Surface Mining in the Department of Interior.

The *Fish and Wildlife Coordination Act* (16 USC 661 *et seq.*), originally passed in 1934, requires federal agencies proposing water resource development projects, either directly or by permit, to consult with the United States Fish and Wildlife Service, United States Department of the Interior, with a view towards conserving fish and wildlife resources. The reports of the Secretary of the Interior under the Act must include proposed measures for mitigating or compensating damage to wildlife resources resulting from the project. The proposing agency is only required to fully consider the Secretary's report.

The *Endangered Species Act* (16 USC IS31 *et seq.*) requires federal agencies to ensure that any action authorized, funded, or carried out by them are not likely to jeopardize the continued existence of any listed endangered or threatened species, or result in the destruction or adverse modification of its critical habitat. The key to the Act's implementation is mandatory consultation by the proposing agency with the Fish and Wildlife Service. When FWS makes a formal finding that the proposed action will violate the Act, and after a good faith effort to find alternatives has failed, an Endangered Species Committee, composed of Federal officials and a representative of the affected States, must balance preservation against development and decide the issue: but the Act weighs the balance in favor of preservation.

The *National Historic Preservation Act* (16 USC 470 *et seq.*) established the National Register of Historic Places. It requires federal agencies to consult with the Advisory Council on Historic Preservation whenever proposed federal projects might have adverse impacts on historic or archaeological sites listed or eligible for listing on the National Register. By regulation, the proposing federal agency must consult a State Historic Preservation Officer when determining how its activities will affect historic or archeological sites. The procedures also require that the SHPO, along with the Advisory Council and the proposing agency, reach written agreement in certain cases on how to mitigate any adverse effects expected from a federal project.

The *Wild and Scenic Rivers Act* (16 USC 1271 *et seq.*) was originally enacted in 1968 and has been subsequently amended several times. The statute is jointly administered by the Secretary of the Interior, through the National Park Service and, when national forests are involved, by the Secretary of Agriculture through the National Forest Service. The Act established "a national wild and scenic rivers system" and defines criteria for eligibility under each of the following three classifications: wild, scenic, and recreational rivers. The Act provides that "no department or agency of the United States shall assist by loan, grant, license or otherwise in the construction of any water resources project that requires a direct adverse effect on the values for which such river was established, as determined by the Secretary charged with its administration." Adjacent federal lands must be managed consistent with the river's conservation values.

The *Coastal Zone Management Act* (16 USC 1451 *et seq.*) provides for federal financial assistance to coastal state governments for the development of Coastal Zone Management Plans. Coastal Zone Management Plans have as their primary function land-use management for the Coastal Zone to assure the orderly and environmentally sound development of these ecologically sensitive areas. Under the Act, federal or federally assisted or licensed projects are required to be approved by the coastal state as consistent with an approved Coastal Zone Management Plan. At the federal level the Coastal Zone Management Program is administered by the Office of Coastal Zone Management of the National Oceanic and Atmospheric Administration of the Department of Commerce.

Sixty-one percent of all federal lands, some 470 million acres, is administered by the Department of Interior's Bureau of Land Management. The statute under which BLM administers the national resource lands is the *Federal Land Policy and Management Act* (43 USC 1701 *et seq.*). Among the Act's declared policies is that public lands should be retained in federal ownership unless disposal will serve the national interest. In pursuit of this policy, all public lands and resources are to be inventories periodically: all previous land classifications are to be reviewed; and all land use decisions are to be the result of a land-use planning process. When resources or land are to be disposed of or utilized, effort must be made to assure a fair market value return to the United States. Guidelines for land-use planning and management must assure attainment of multiple-use and sustained yield. Like most federal conservation statutes, the Act incorporates the *Multiple-Use and Sustained Yield Act* (16 USC 528 et seq.). Most important for our purposes, planning and management activities are to be undertaken so as to protect the environmental, ecological, scientific, scenic, historical, air, atmospheric and water resource values of the public domain. Areas of critical environmental concern are to be identified and protected through regulation and planning as soon as possible. Land-use plans must be consistent with all state and federal pollution control statutes.

The second largest administrator of public lands is the US Forest Service, which controls the use of over 187 million acres dispersed throughout the United States. Consolidation and clarification of the FS mission were accomplished by enactment of the *Forest and Rangeland Renewable Resources Planning Act* and its subsequent amendment by the *National Forest Management Act* (16 USC 16 *et seq.*). These Acts, like the *Federal Land Policy and Management Act,* established a comprehensive planning structure for the management of the national forest system in general as well as particular forests. Unit plans must integrate both economic and environmental considerations. The Act declares that each unit plan should protect the environment, provide for restocking, and perpetuate multiple-use and sustained yield. Clearcutting is prohibited unless it is determined to be the best means of meeting the objectives of the unit plan. If clearcutting is chosen, efforts must be made to mitigate its effects. Other standards in the Act prevent premature harvesting of trees unless such harvesting is compatible with other multiple uses of the forest. While developing unit plans, lands unsuitable for timber production are to be identified and timber production from such lands is to be prohibited for a period of ten years.

National Forest Lands and other federal lands can be withdrawn from logging if Congress designates them as Wilderness Areas under the *Wilderness Act* (16 USC 1131 *et seq.*). These lands must be administered to preserve wilderness character and must be devoted to scenic, historic, recreational, and similar uses.

As for mining on federal lands, the antiquated *General Mining Act of 1872* (30 USC 21) governing the location and patenting of claims to hardrock minerals, has not provided an adequate basis for mitigating the environmental abuses of hardrock mining. However, where energy resources are concerned, environmental protection is one factor to be considered in awarding mineral leases under the following statutes: *Mineral Leasing Act* (30 USC 181 *et seq.*); *The Outer Continental Shelf Lands Act* (43 USC 1331 *et seq.*); *Multiple Mineral Development Act* (30 USC 521 *et seq.*); *The Geothermal Steam Act* (30 USC 1001 *et seq.*); and *The Federal Coal Leasing Amendments* (30 USC 201 *et seq.*).

A number of federal transportation statutes deal with conservation of natural resources. Under both the *Federal-Aid Highway Act* (23 USC 101 *et seq.*) and the *Urban Mass Transit Act* (49 USC 1601 *et seq.*) the Secretary of Transportation cannot approve a transportation project which encroaches upon park land, wildlife refuges or historic sites unless there is “no feasible and prudent alternative” and the “program includes all possible planning to minimize harm to such areas.” Highway planning must include environmental factors and be consistent with State transportation plans prepared under the *Clean Air Act.*

The conservation of Alaska’s vast natural resources has been a high priority of conservationists for many years. In 1978 President Carter, under the authority of the *Antiquities Act* (16 USC 431 *et seq.*), designated 17 National Monuments encompassing 56 million acres of Alaskan land. The *Antiquities Act* empowers the Secretary of the Interior to designate certain federally owned lands as National Monuments, including “historic landmarks, historic and prehistoric structures and other objects of historic or scientific interest.” Once a designations has been made, the Secretary can prohibit any activity that would adversely affect the site. An additional 50 million acres of land in Alaska were withdrawn for wildlife refuges under the *Federal Land Policy and Management Act.* Finally, in late 1980, Congress passed the *Alaska National Interest Lands Conservation Act,* ratifying most of the previous executive withdrawals.

America’s National parks have been referred to as its “Crown Jewels.” The National Park Service was created in 1916 (16 USC 1 *et seq.*) to “promote and regulate the use of National Parks, monuments and reservations and to conserve the scenery and natural and historic objects and the wildlife therein and by such means as will leave them unimpaired for the enjoyment of future generations.” Parks are created by individual Congressional enactments, so that one must look to the individual organic acts in order to understand the legal restrictions applicable to each park. However, mining in National Parks was severely restricted by the *National Park Service Mining Activity Act* (16 USC 1901 *et seq.*). NPS also administers the Wild and Scenic Rivers Systems and the National Trails System (16 USC 1241 *et seq.*). Financing for purchases of National Park Lands has frequently been provided by the *Land and Water Conservation Fund Act* (16 USC 460 *et seq.*).

National Wildlife Refuges are administered under the *National Wildlife Refuges System Administration Act* (16 USC 668dd *et seq.*). Hunting, fishing, and other public recreation may be permitted in National Wildlife Refuges if they are compatible with the conservation purpose of the refuge.

Finally, there is extensive federal legislation relating to the protection of particular wildlife species, including habitat acquisition programs. Some of these statutes are the *Marine Mammal Protection Act* (16 USC 1361 *et seq.*); the *Wild Free-Roaming Horses and Burros Act* (16 USC 1331 *et seq.*); the *Bald Eagle Protection Act* (16 USC 668 *et seq.*); the *Migratory Bird Hunting Stamp Act* (16 USC 718 *et seq.*); and the *Water Bank Act* (16 USC 1301 *et seq.*).

4

In recent years, international environmental law has burgeoned to the degree that it is now one of the most active areas in the field. The international community has recognized that environmental threats are frequently transnational, multinational, regional or even global in scope. As a result, international agreements (treaties, conventions and protocols) have entered into effect governing activities involving diverse environmental issues, such as biodiversity (e.g., The Convention on Biological Diversity), climate change, ozone depletion (e.g., the "Montreal Protocol"), the protection of Antarctica, transboundary disposal of hazardous wastes, vessel-based water pollution and ocean dumping, transboundary air and water pollution, conservation of whales and other marine resources, trade in endangered species and nuclear weapons testing.

International agreements, unlike most national environmental laws, are essentially voluntary in nature: that is, they cannot be enforced by an international legislature, court system or police force. Such agreements can only be enforced by the nations that consent to be bound by a particular international compact. However, political methods, for example, international exposure of violating nations and potential trade boycotts, are often successful in achieving compliance by nations that either do not adopt international agreements or else adopt them and systematically violate their provisions.

REFERENCES

1. Plater, G. Z. B., Abrams, R., Goldfarb, W., Graham, R. L., Heingerling, L. and Wisch, D. L., *Environmental Law and Policy: Nature, Law and Society.* 3rd Edition. New York, NY. Aspen Publishers. 2004.
2. Government Institutes, Inc., *Environmental Law Handbook,* 17th Edition, Rockville, MD. 2003.
3. Environmental Law Institute, *Environmental Law Reporter.*
4. Bureau of National Affairs, *Environment Reporter.*
5. Guruswamy, L. D. and Hendricks, B. R., *International Environmental Law.* St. Paul, MN. West Publishing Co. 1997.

WILLIAM GOLDFARB
Rutgers University

EPIDEMIOLOGY

ORIGINS AND DEFINITIONS

Epidemiology (Waterhouse, 1998) is a science that basically borrows from the other sciences to form its own area of expertise. The actual word *epidemiology* can be broken down into three parts: first *epi,* which means "upon"; then *demo,* which is population; and finally *ology,* which refers to studying. So we can in a simple form say epidemiology is the study of events that occur upon or on populations or groups. Overall, epidemiology is not interested in the individual, but rather the population; however, these data are often used to relate and infer risks to an individual. The field of epidemiology interacts with other science areas and rarely functions on its own. For example, in the study of occupational diseases, there may be an interaction of occupational exposure and health effects in determining the risk of a specific disease (Stern, 2003). Biostatistics, the study of statistical relationships for biological systems, is an area often in close association with epidemiologists. It could even be argued that epidemiologists cannot easily function without using basic biostatistics. Thus, epidemiologists are routinely trained in the basics of biostatistics as well. In addition, it is not uncommon for some epidemiologists to have been originally trained or cotrained in other disciplines (e.g., environmental health).

The field of epidemiology can be broken down into different subject areas. In the simplest form it can be grouped as acute (e.g., accidents), chronic (e.g., type II diabetes), and infectious (e.g., malaria). However, it can also be grouped by subject name, such as occupational epidemiology, environmental epidemiology, cardiovascular epidemiology, and so forth. The other way of classifying epidemiology is by disease name, such as malaria epidemiology, epidemiology of heavy metals, and so forth. Thus, like most scientific fields of study, this area can be categorized in many different ways depending on one's prospective. In this chapter, we are concerned with the area of epidemiology that is most closely associated with environmental science and engineering. Traditionally environmental and occupational epidemiology were related to those in environmental science and engineering, but as the world changes and the concept of global epidemiology emerges, most if not all subfields or subjects of epidemiology are becoming interspersed among previously distinct and separate scientific and other fields of study (e.g., sociology). However, due to the necessity of brevity in this chapter, the focus will stay on the traditional subject areas of environmental and occupational epidemiology.

One of the biggest problems with environmental epidemiology is that studies rarely find a strong association for cause and effect. This is commonly thought to be a result of confounders and problems in conducting studies of this nature. These problems include the lack of a clear study population, low-level exposures, inaccurate exposure doses, and related confounding factors. Some of these concerns can be overcome in occupational studies where the population is better defined and exposures have been better documented, although the same issues can also occur in this area of epidemiology as well. However, these problems should not discourage us from conducting or evaluating epidemiological investigations. Readers should be aware of general texts on this subject, and a few are mentioned here as potential references (Lilienfeld and Stolley, 1994; Timmreck, 1998; Friis and Sellers, 1998), although this list is not complete.

Epidemiology begins with the application of numbers to a disease, set of cases, or event (like accidents), primarily in the sense of counting rather than measurement. Some can even say that counting is at the heart of epidemiology, because it provides us with how many of the cases or events exist or occurred (Lange *et al.,* 2003a). Disease, which is used here to include all events or occurrences that may be identified in an epidemiological study, are identified as either *incidence* or *prevalence.* These two terms are rates of occurrence or existence for the disease. The term *disease,* in this chapter, will also mean and include any event or case that is measured, such as cancer, injury, disorder, or a similar occurrence. Incidence is the number of cases that arose during a specific time period, usually a year; prevalence is the number of cases that exist at some point in time or within a time period of interest, again, usually a year. In most cases, prevalence will be a larger numerical value than incidence. This is true when people with the disease survive for a long period of time, which would be a time period longer than the time period established for the incidence rate. However, if the disease event is very short or can occur multiple times over a short period of time, incidence and prevalence can be similar. If the same disease event can occur more than once in the same person, it is possible that the incidence can be greater than prevalence. An example of this would be influenza (the flu, which is a viral disease) in a small population, say 15 people in an isolated location (e.g., a research station in the Arctic). If prevalence is counted as anyone having the disease during the time period and incidence of the occurrence of the disease, if all 15 had the flu and someone contracted the disease twice, incidence would be 16 times in a population of 15, with prevalence being 15 out of 15. As noted, very seldom will the incidence be larger than prevalence; this would only occur in rare or unusual events and would likely involve small populations. It is important to understand the difference between each of these terms, in that they represent different "values" for a disease or event in the population being studied. Table 1 provides an example of incidence and prevalence for data collected from a computer database of different diseases (Centers for Disease Control and Prevention [CDC], 2004).

In Table 1, the incidence and prevalence (I/P) are the same, since both involve the occurrence of death. For Parkinson's disease there was an increase in I/P for both the United States and Pennsylvania, while for cancer there was a decrease (1970–2000) for the United States and a steady state for Pennsylvania. Adjustment involves standardizing the population for such variables as age, race, and sex. These variables can also be considered confounders.

Epidemiology was recognized at first implicitly by a general appreciation of probabilities, rather than explicitly by recording each incident. This is noted by some of the first attempts to conduct epidemiological investigations where the number of events was noted but no rate of the event was determined. Just knowing the number of cases alone, without a rate of occurrence, does not allow comparison with other events. However, lack of a rate does not necessary minimize an epidemiological study, although in the modern day, rates are often essential. But, in parallel cases with the base population among whom the cases have occurred, in order to obtain in ratio from the rate incidence or occurrence of the disease, suitably refined, according to the circumstances of the situation and in ways we shall discuss later, such a rate can be used as a measure for purposes of comparison in the same place between different time periods, or between different places at the same time, or in a variety of other ways. Rates are represented in units of a population, like per 100,000 people. By an appropriate extension we can measure the impact of disease, whether in general or of a particular type, on the population. But we also find that the characteristics of the population itself can alter the manifestation of the disease, so that the science of epidemiology can be symmetrically defined as measuring the impact of disease on a population, or of a population on a disease—perhaps better expressed by saying that the concern of epidemiology is with the measurement of the interaction of disease and population. Thus, at the heart of epidemiology is counting (Lange *et al.*, 2003a), which is then concerted to a rate as expressed as either incidence of prevalence.

TABLE 1

All races and all gender death rates for Parkinson's disease and cancer of bronchus and lung unspecified for 1979–1998 using 1970 and 2000 standardized populations

Standard Population	Region	Crude*	Age-adjusted*
Parkinson's Disease			
2000	US	4.8	4.9
1970	US	2.9	2.2
2000	Pennsylvania	3.5	3.3
1970	Pennsylvania	3.5	2.2
Cancer of Bronchus and Lung Unspecified			
Standard Population	Region	Crude*	Age-adjusted*
2000	US	52.5	55.2
1970	US	46.0	60.7
2000	Pennsylvania	54.4	60.7
1970	Pennsylvania	46.0	60.7

Source: From CDC (2004), CDC Wonder (database on disease occurrence).

* Rates are per 100,000.

The issues of rates can be illustrated through two historical studies. The first did not employ rates in determining a cause of scurvy, while the other employed rates to locate the source of the infectious agent in causing cholera. These studies illustrate how rates can be used in evaluating disease, although the importance of basic observation cannot be forgotten or lost in a study.

In the study by James Lind on scurvy (Timmreck, 1998), in 1753, he noted that some sailors developed this disease while others did not. Lind examined the diet of those with and without the disease as part of the investigation into the cause of scurvy. Although he did identify a crude rate in a population of sailors initially studied (80 out of 350 had the disease), this rate or its comparison was not employed in his study design. To evaluate the differences in reported diets, he provided oranges and lemons to two sailors and followed their progress. After a few days he noted that their scurvy subsided and concluded that these dietary supplements were most effective at treating and preventing the disease. In modern epidemiology we would most likely look at the rates of disease occurrence and cure rather than using observational numbers, as had been used by Dr. Lind. However, Dr. Lind did make observations of cause and effect and time and place, as well as sources of causation in the disease process (Timmreck, 1998). It is worth noting that today the size of this study would likely be considered too small for publication in a scientific journal. However, this demonstrates the importance of observation even for small study populations.

What most consider the first true epidemiology study that employed rates was conducted by John Snow in the 1850s and concerned an outbreak of cholera. Dr. Snow actually conducted two studies on the epidemiology of cholera: the first was a descriptive study in the SoHo district of London (this is in the Broad Street area), and the second was a classical investigation in determining rates of disease.

In the first study he observed that two different populations were affected by cholera, one with a low number of deaths and the other with a high number. By mapping locations of deaths, commonly used today in geographic and ecological epidemiology studies, he concluded that there were different sources of exposure (Paneth, 2004). The population with a low number of deaths was obtaining water from a brewery source that had its own well, which as we now know was not contaminated, and those in the second population, having a high number of deaths, were obtaining it from the Broad Street pump. From these data, he plotted the occurrences and extent of the outbreak, which we now look at as the duration of the epidemic. Near the end of the epidemic, Snow had the Broad Street pump handle removed for to prevent the reoccurrence of the disease. From his investigation, a foundation of causative agents (which was not known at the time), population characteristics, environment, and time were connected in evaluating the disease process with an applicability of prevention.

During an epidemic in 1853, Snow examined the sources of water. At the time, there were three water companies serving the area, Southwark, Vanxhall, and Lambeth. Southwark and Vanxhall collected water from a polluted section of the Thames river, while Lambeth collected water upstream of the pollution. By using deaths published by the registrer general in London, Snow was able to deduce that those obtaining their water from Southwark and Vanxhall had a much higher death rate that those getting water from Lambeth. Snow obtained the addresses of those that died, and by knowing the water source and the population in the area, he was able to calculate death rates for the various water sources. He determined that those having Southwark and Vanxhall water experienced a death rate of 315 per 10,000 and those with Lambeth had a rate of 37 per 10,000 (Lilienfeld and Stolley, 1994). This provided evidence that obtaining water from the polluted area of the river resulted in a high rate of death from cholera and that cholera was a waterborne disease. Snow's discovery, through epidemiology, occurred approximately 40 years before Robert Koch, in 1884, identified *Vibrio cholera* as the causative agent of cholera. This certainly established a relationship of disease with the environment, but also showed the importance of representing epidemiological data in the form of a rate. Today, rates are commonly reported as a number per 100,000 or million. However, any rate expression is acceptable. Even when cause of a disease is not known, as shown by Snow, a great deal can be learned about the agent through epidemiology.

Today, the pump handle from the Broad Street well is in possession of the John Snow Society. One survey reported that Snow was the most influential person in medicine, with Hippocrates being second (Royal Institute of Public Health, 2004). Certainly this report does suggest that there may be bias in the survey, with a larger number of votes coming from the John Snow Society, but it illustrates the importance of his contribution and the influence that epidemiology has had on medicine. It should be mentioned that Snow was one of the originators of the field of anesthesiology as well. Thus, his contribution is not limited to pure epidemiology.

From these examples, it becomes clear that the methods of epidemiology are in essence those of statistics and probability. It is also clear that much of medicine is based on observation within the field of epidemiology; diagnosis depends upon a recognizable cluster of signs and symptoms characteristic of a disease, but this is only so because of their statistical similarity extended over many cases. And in a like manner, the appropriateness and efficiency of treatment methods summarize the result of practice and observation. Many of the developments of modern medicine, both in methods of diagnosis and treatment, depend upon epidemiological procedures for their assessment and evaluation, such as in clinical trials (see below) and many experimental studies, as was illustrated by Dr. Snow's study of water sources.

Generally, epidemiological studies can be divided into four groups: ecological, cross-sectional, case-control, and cohort. Ecological and cross-sectional studies are hypothesis-generating investigations, while case-control and cohort studies can establish a causal effect. Case-control and cohort studies can provide odds ratios (ORs) and relative risks (RRs). In most cases, the OR and RR will be equal to each other, and represent the risk associated with exposure and occurrence of disease.

MORTALITY AND THE FIRST LIFE TABLES

It is in the description and measurements of mortality that we first meet quantitative epidemiology. The *London Weekly Bills of Mortality* begun early in the sixteenth century continued irregularly during that century and were resumed in 1603, largely to give information about the plague. John Graunt published an analysis and comparison of them in the middle of the seventeenth century (*Natural and Political Observations upon the Bills of Mortality*), and later Sir William Petty published *Five Essays in Political Arithmetic*, a book that was devoted rather less to numerical data that was Graunt's. Graunt had examined deaths by causes and age, which led to the interest at this time in the construction of life tables. A life table aims to show the impact of mortality by age through a lifetime. Starting with an arbitrary number of people (e.g., 1,000—known as the "radix") who are regarded as having been born at the same time, the life table thus opens with 1,000 persons at exactly age zero. A year later this number will be diminished by the number of infant deaths that have occurred among them, leaving as survivors to their first birthday a number usually designated $\int_1$. Similarly, the deaths occurring in the second year of life reduces the number still further, to $\int_2$. By the same process the diminution of numbers still alive continues until the age at which none survive. The first actual life table was constructed in 1693 by Edmund Halley, the mathematician (best known perhaps for the comet named after him), and it was based on 5 years' experience of deaths in the German city of Breslau. Since it recorded deaths by age, without reference to birth, the radix was obtained from a summation that the population was in dynamic equilibrium. Although there were other life tables constructed around this time, when life-insurance companies began to be founded, it was not possible to construct an accurate life table without using rates of mortality rather than numbers of deaths. Rates required denominators to be both appropriate and accurate, and the obvious source was a census.

CENSUSES

Apart from censuses of Roman and biblical times, the first modern census was taken in Sweden in 1751. The first in the United States was in 1790, and the first in England was in 1801. Censuses traditionally were taken for two main purposes, military and fiscal. Their epidemiological value in supplying denominators for the construction of rates of mortality was very much an incidental usage. Just as the concern about the plague gave a new impetus to the regular production of the London *Bills of Mortality,* so the anxiety about attacks by cholera was an important factor in setting up national registration of deaths in England and Wales in 1837. But from that time onward, mortality rates were published annually in England and Wales, and their implications, medical, social, geographical, and occupational, were very effectively analyzed and discussed by William Farr, the first medical statistician appointed to advise the register general, which collected information on Mortality.

CAUSES OF DEATH AND THE ICD

With the advent of routine death registrations and censuses throughout Europe and North America, the publication of mortality rates in successively increasing detail stimulated comparison, and demanded at the same time an agreed basis for terminology. This led to the setting up in the middle of the nineteenth century of international Statistical Congresses to produce a classification of causes of death. Gradually these lists of causes became generally adopted by individual countries, and in order to keep up with medical advances, the list was required to be revised every 10 years. From a list of causes of death it was extended to include diseases and injuries not necessarily resulting in death, so that it could be used for incidence by hospitals as a diagnostic index. The ninth revision of the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD) came into force in 1979 and has recently been replaced by the ICD-10, on January 1, 1999. The ICD was originally formalized in 1893 as the Bertillon Classification of International Causes of Death. The ICD-10 is copyrighted by the World Health Organization (WHO). The WHO publishes the classification and makes it available to countries of the world. In the United States, the U.S. government developed a clinical modification for purposes of recording data from death certificates.

The degree of detail it is now possible to convey through the use of the latest ICD code is very great, but of course it is entirely dependent upon the subtlety of the information available to the coder. However, the hierarchical design of the code does permit expression of a rather less specific diagnosis when the data are inadequate or vague. One of the biggest problems with this type of system is that the data are extracted from death certificates, which may not accurately reflect the true cause of death.

The WHO collects mortality data from its member states and publishes mortality rates by cause, sex, and age group, in the *World Heath Statistics Annual.* Individual countries also publish their own mortality data, often including more detailed subdivisions, for instance of geographical areas. The same offices in nearly all countries are responsible for collecting and publishing statistics of births and marriages, and probably also for the censuses, which recur at intervals of 5, 7, or 10 years, according to the practice of the country.

THE SEER PROGRAM

Another evaluator of specific mortality is the Surveillance, Epidemiology, and End Results (SEER) Program of the U.S. National Cancer Institute (NCI). This report provides information on cancer incidence and survival using various geographic locations of the United States. The concept of these areas is to represent occurrence for the overall population. SEER registries now include in its collection about 26 percent of the U.S. population. Information collected by the SEER registries includes patient demographics, primary tumor site, morphology, stage at diagnosis, first course of treatment, and follow-up status. Currently this is the only source of population-based data on cancer that includes its stage and diagnosis and survival rates for the stages of cancer. This is also a Web-based source and is provided by the National Center for Health Statistics. Analyses of SEER data are commonly published in the literature, including for determining trends of disease (Price and Ware, 2004).

OTHER DATA SYSTEMS

There are other Internet-based data systems that provide information on rates on deaths in the United States. This includes the CDC Wonder system (CDC, 2004). This system provides both crude and age-adjusted death rates as categorized by the ICD-9 and ICD-10 (specific causes or diseases). Thus, by using this system, rates can be determined by county and state and for the United States as a whole for any year or group of years. Such systems allow evaluation of varying rates over time and determination of trends. These data can also be used in ecological epidemiological studies to evaluate trends.

COMPARISON OF MORTALITY RATES AND STANDARDIZATION

When comparing the experience of two different countries with respect to mortality from a specific disease, the rates for each age group can be contrasted perhaps most easily in graphical form. But comparing their crude rates of mortality from the disease, in an endeavor to simplify the comparison, is only legitimate in the unlikely event of their age structures being identical. Thus, in most studies there is an age adjustment (Baris *et al*., 1996). This adjustment is based on a large population, which is usually based on the national or state population. Use of crude rates alone, without age adjustment, may lead to inaccurate interpretation of the rate of disease and does not allow these rates to be compared to other studies (Lange, 1991).

The overall mortality rates increase sharply with age after puberty (Figure 1): the increase is in fact close to exponential in its shape, as is clear from its linear form when plotted on a logarithmic vertical scale (Figure 2). Consequently, if one of the two populations to be compared has a greater proportion of the elderly than the other, its crude rate will exceed the other, even if their age-specific rates are identical throughout the age range. The crude rate is the ratio of the total deaths to the total population (this may be for both sexes together or separately by sex), and more deaths will result from the larger population of the elderly groups. However, it is possible to obtain a legitimate comparison using a single figure for each population by the simple method of applying the separate age-specific rates observed in the first population to the numbers of the population in the corresponding age groups of the second. In this way we find the numbers of deaths that would have occurred in the second population if it had experienced the mortality rates by age of the first. These "expected" deaths can be totaled and expressed similarly to a crude rate by dividing by the total of the second population. This comparison is legitimate because the population base is now identical in its age structure and cannot distort the results. The process has been called by some "standardization," and the rates of the first population are described as having been standardized to the second. Clearly it would be equally possible to reverse the procedure by standardizing the second to the first population. A different pair of rates would of course be obtained, but it would in general be found that their ratio was similar to the ratio of the first pair.

An example of the differences of crude and age-adjusted rates can be observed by using the CDC Wonder system (CDC, 2004). Table 1 (see previous discussion) shows the crude and age-adjusted death rates for Parkinson's disease (ICD code 332) and cancer of bronchus and lung unspecified (ICD code 162.9). These rates are standardized for 2000 and 1997 for the United States and Pennsylvania. As can be seen from the table, there is a difference in rates between crude and age-adjusted as well as for different standardized populations for the United States and Pennsylvania. This also illustrates that there are different rates for disease in specific populations, like Pennsylvania versus the United States. Such rates can be used to evaluate trends for disease by time and geography. When evaluating and reading epidemiological studies, it is important to note that the title of tables and figures should first be carefully read so as to understand the information presented.

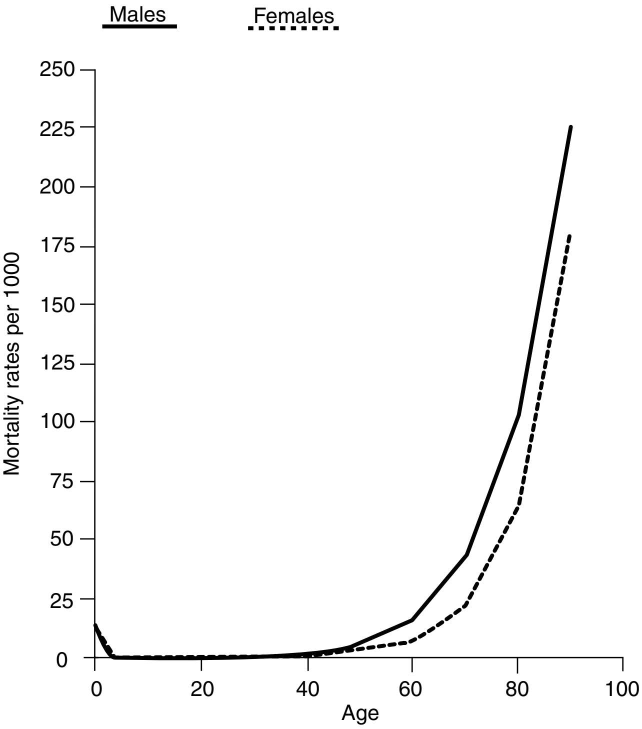


FIGURE 1 Mortality rates by age and sex (arithmetical vertical scale).

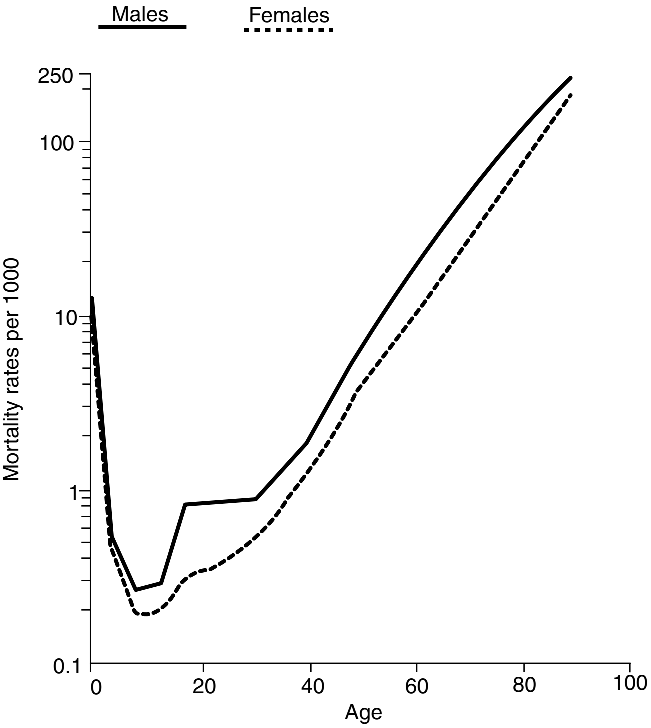


FIGURE 2 Mortality rates by age and sex (logarithmic vertical scale).

WORLD STANDARDIZED RATES

Another method of standardization, essentially similar to that described above, makes use of standard population, defined in terms of the numbers in each age group. The rates of each population are applied to this standard population to obtain a set of expected mortality deaths and thus a rate standardized to the standard population. It is becoming increasingly common today to use a constructed "world standard population" for this purpose, so that rates so obtained are described as "world standardized rates" (WSRs). This concept was created originally by the late Professor Mitsui Sigi, a Japanese epidemiologist, when attempting to compare cancer mortality rates between different countries throughout the world. The age structure of a developing country (often typified as Africa) has a triangular form when depicted as a pyramid, at least before the onset of AIDS (see Figure 3), with a small proportion of the elderly, but its proportion increasing regularly toward the lowest age groups. A typical pyramid for a developed country (typified as European) is that in Figure 4, which shows a rather more stable pattern until the ultimate triangle at the upper end.

These forms of standardization have been disrupted by HIV, which is the causative agent of AIDS. In Botswana for the year 2020 it has been predicted that there will be a larger population around the age group 60–70s than for 40–50s as a result of AIDS (Figure 5). The dramatic effect on the population structure of this virus will change how age adjustment must be performed for many of the affected countries. Thus, in the future, age adjustment will not be as straightforward as described in many standard epidemiology textbooks.

INDIRECT STANDARDIZATION

When the objective is to compare the mortality rates of various subpopulations, such as geographical, occupational, or other subdivisions of a single country, a different method is commonly used. What has already been described is known as the "direct method" of standardization, using a standard population to which the rates for various countries are applied. The "indirect method" of standardization makes use of a standardized set of mortality rates by age group, and these rates are applied, age by age, to each of the subpopulations, providing thereby a total of expected deaths; the actual total of deaths observed in each subpopulation is then divided by the expected total to provide what is known as the "standardized mortality ratio" (SMR). The standard set of mortality rates used is that of the overall population's experience, and almost invariably that population is the sum of all the subpopulations. Clearly if some SMRs are greater than 100 (it is conventional to multiply the SMR by 100, which has the convenience of making apparent the percentage difference from expectation), then some will be below, since the weighted mean of the SMRs must be 100.

For the purposes of comparisons of this type, the indirect method has a number of advantages over the direct method. Several of the subpopulations may be quite small in size, especially in some age groups where the numbers observed may be very small, so that age-specific mortality rates can fluctuate widely. The mortality rates of the parent population, on the other hand, are inherently more stable than those of any fractional subpopulation. The structure by age of each subpopulation will in general be easily obtainable, often from the census, with reasonable accuracy, and so will the total number of deaths. The ratio of observed to expected deaths—the SMR—is then easily interpreted as a percentage above or below expectation. An assessment of the statistical significance of its difference from 100 can be obtained by assuming a distribution similar to the Poisson, so that the standard error would be $100E$, where E is the expected number of deaths: deviations from 100 of more than twice this quantity would be regarded as statistically significant at

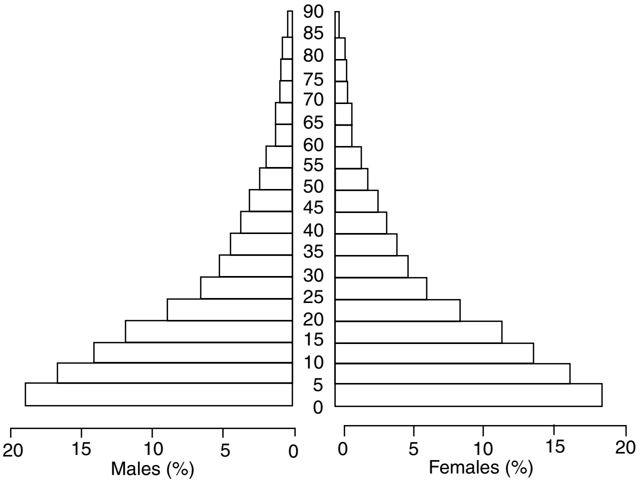


FIGURE 3 Population pyramid: a developing country.

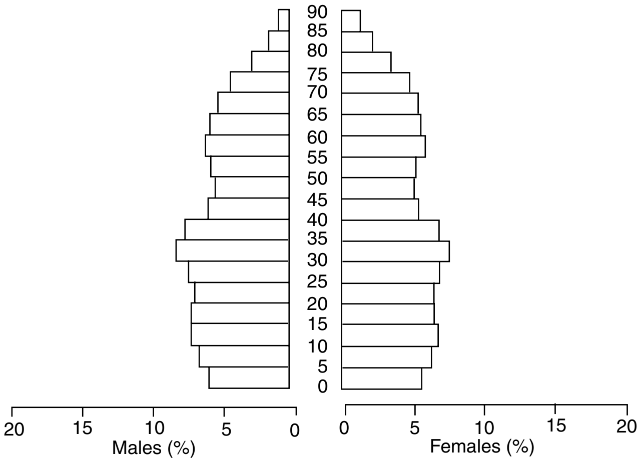


FIGURE 4 Population pyramid: a developed country.

the 5% level. In many studies a confidence interval (CI) at 95% is presented. Even if an SMR is above or below 100, a CI that has an overlap with 100 is often considered to be in the range of nonsignificant. In most cases, statistical significance exists when the summary value and its CI do not overlap 100.

OCCUPATIONAL MORTALITY COMPARISONS

It will be obvious that precisely the same methods can be applied to mortality rates from any single disease—or group of diseases, such as cancer—as to total mortality from all causes. By appropriate choice of cause groups it is possible to examine the pattern of mortality in a particular industry or occupation—for example, to highlight any excesses or deficits, when compared to the overall experience of the total population. But such a comparison often needs to be made with caution and circumspection; the total population includes the handicapped, the chronically sick, and the unemployable, none of whom will be found in the industrial population. This leads to the healthy-workers effect (HWE) whereby the overall mortality experience of the industry is often better than that in the total population, partly for the reasons just given and partly because there may well have been a medical examination to select only healthy new recruits to the industry. Another effect, known as the survivor-population effect (SPE) or survivor effect, arises because those workers in an industry who find the work too strenuous or beyond their capacity will leave to find more suitable work; those who remain in the industry—the survivors—will again be a group selected to be of better health, stronger, and more competent at the work. A thorough ongoing epidemiological review of the industry or of a sufficiently large factory within it will generally allow these effects to be separately measured and assessed, together with the specific hazards, if any, that may be characteristic of the industry.

Many occupational epidemiology studies (McMichael, 1976) now carefully evaluate the influence of the HWE and SPE. Both the HWE and SPE are considered a form of bias. In many ways both the HWE and SPE are similar or the same occurrence. However, it can be inferred that the SPE involves, at least initially, those that are best able to tolerate the work conditions or are best able to cope with exposure to occupational stress, most notably at the beginning of an occupational activity. The SPE will likely include the HWE for those that remain at an occupation for a longer period of time and would include an adaptive response as would be related to injuries. Many of the factors associated with these effects are commonly called confounders. Some of these would include personal confounders like smoking. Not all events are equally affected by the HWE. For example, the HWE has been suggested to have a weak-to-nonextant influence on cancer mortality, while having a stronger impact on mortality from cardiovascular disease (McMichael, 1976). However, by employing appropriate methodology, confounders and the HWE can be controlled for (Mastrangelo *et al.*, 2004). It should be noted that the most important confounders in epidemiology are age, sex, social and economic status, and smoking, although many others may be important as well depending on the study. The importance of a confounder is best illustrated by cigarette consumption (smoking) and lung cancer (Lee *et al.*, 2001).

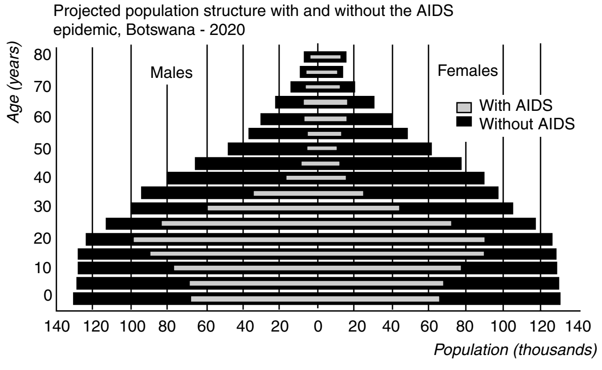


FIGURE 5 Botswana is predicted to have more adults in their 60s and 70s in 20 years' time than adults in their 40s and 50s.

LIFE TABLES

We have already referred to some of the early essays on the production of a life table, and to the difficulties of having to use various records, because the appropriate mortality rates were not yet available. When death registration was reasonably complete and census sufficiently accurate, it was possible to construct a much better life table. William Farr, for his first life table, used the census of 1841 and the deaths of the same year. In his second table he broadened his basis, using both the 1841 and 1851 censuses, and the deaths of a period of 7 years (1838–1844). Modern practice usually combines the deaths of 3 years, to reduce the effects of minor epidemic or climatic variations, and uses the census of the middle year for the denominators. Mortality rates by sex and single years of age then enable the construction of a full life table, advancing in single years from 0 to about 110 years of age. The successive l_x figures denote the numbers of living to the exact age x from the radix at l_0 of 100,000. The larger radix is justified by the greater degree of accuracy now available. Essentially the mode of calculation is the same:

$$l_{x+1} = l_x - d_x$$

where d_x = number of deaths between ages x and the day before attaining age $x + 1$, and

$$d_x = l_x \cdot q_x$$

where q_x = mortality rate at exact age x.

Single-year mortality rates are generally obtained as the ratio of the number in a calendar year of deaths whose age was given as x to the mid-year population aged x: for each of these quantities the age given as x would range from exact age x (the xth birthday) to the day before the $(x + 1)$th birthday, and would thus average $x + 1/2$. This mortality rate is designated m_x, such that $m_x = d_x/p_x$, where p_x = midyear population aged x.

If we go back 6 months to the beginning of the calendar year, the average age of those encumbered in the middle of the year at $x + 1/2$ would become x, but they should also be augmented by half of the deaths (also of average age), on the plausible assumption that they were divided approximately equally between the two halves of the year. This is of course because none would have died by the beginning of the year, and furthermore their average age would then be x rather than $x + 1/2$. Now we can obtain the mortality rate at exact age x since

$$q_x = d_x/(p_x + 1/2d_x)$$

Dividing through by p_x, this becomes

$$q_x = m_x/(1 + 1/2m_x)$$

thus relating the two mortality rates.

SURVIVAL RATES ADJUSTED FOR AGE

Strictly speaking, the life table is a fiction, in the sense that it represents an instantaneous picture or snapshot of the numbers of living at each single year of age, on the assumption that the mortality rates at the time of its construction remain unchanged at each period of life. Mortality rates have generally tended to fall, though they are rather more stable, on a worldwide basis, than they have been earlier in the century. However, there are modern-day exceptions, as is seen in the old Soviet Union countries where life expectancy is declining (Men *et al.*, 2003). Even though life expectancy was lower than that for Western Europe, a dramatic decline has been observed after the fall of the Soviet Union around 1991. This decline in life expectancy, an increase in premature deaths, has been attributed to social factors and alcohol use, resulting in increased incidence of ischemic heart disease, infectious diseases (e.g., tuberculosis), and accidental deaths (Men *et al.*, 2003). Changes in mortality in the old Soviet Union show the dynamics of epidemiology. However, for the world overall, especially Westernized nations, this means that as time goes on the life table is more pessimistic in its predictions than is the reality of life experience. Nevertheless the life table can be put to a number of uses within the field of epidemiology, quite apart from its commercial use in the calculation of life-insurance premiums for annuities. One of these uses is in the computation of age-adjusted survival rates. Frequently in comparing the experience of different centuries, whether geographically separated or over periods of time, with respect to survival from cancer, a 5-year period is taken as a convenient measure. Cancer patients are not of course immune to other causes of death, and naturally their risk of them will increase progressively with age. In consequence, a comparison using 5-year survival rates of two groups of cancer patients, one of which included a greater proportion of elderly patients than the other, would be biased in favor of the younger group. By using the life table it is possible to obtain 5-year survival rates for each group separately, taking full account of their makeup by sex and age, but considering only their exposure to the general experience of all causes of death. The ratio of the observed (crude) 5-year survival rate of the cancer patients to their life-table 5-year survival rate is known as the "age-adjusted" or "relative" survival rate. Changes in survival, by age adjustment, resulting from a dramatic health effect, as seen in Africa from AIDS (Figure 5), can greatly impact the regional or national survival table. When this procedure is done for each group, they are properly comparable since allowance has been made for the bias due to age structure. Clearly the same mode of adjustment should be used for periods other than 5 years, in order to obtain survival rates free of bias of specific age structures. If the adjusted rate becomes 100% it implies that there is no excess risk of death over the "natural" risk for age; a rate above 100% seldom occurs, but may imply a slightly lower risk than that natural for age.

OTHER USES OF THE LIFE TABLE

The ratio of l_{70} to l_{50} from the life table for females will give the likelihood that a women of 50 will live to be 70. If a man marries a woman of 20, the likelihood that they will both survive to celebrate their golden wedding (50 years) can be obtained by multiplying the ratio l_{75}/l_{25} (from the male life table) by l_{70}/l_{20} (from the female life table). These are not precise probabilities, and furthermore they include a number of implicit assumptions, some of which have already been discussed. Similar computations are in fact used, however, sometimes in legal cases to assess damages or compensation, where their degree of precision has a better quantitative basis than any other.

INFANT MORTALITY RATES

In the construction of life tables, as has been noted, it is necessary to use a mortality rate centered on an exact age rather than the conventional rate, centered half a year later. Only one of the mortality rates in common use is defined in the life-table way, and that is the infant mortality rate (IMR), which measures the number of children born alive who do not survive to their first birthday. The numerator is thus the number of deaths under the age of 1 year, and the denominator is the total number of live births; usually both refer to the same calendar year, although some of its deaths will have been born in the previous year, and likewise some deaths in the following year will have been among its births. The rate is expressed as the number of infant deaths per thousand live births, and it has changed from an average of 150 in much of the last century (but attaining much higher figures in some years) down to below 10 in many countries today. It has been very dependent on general social conditions: low wages, poor housing, and bad nutrition, all having shown close correlation with high IMRs. When infections were rife, and brought into the home by older children, the rate was higher. But with the improvement of infection prevention and treatment, much related to sanitation, vaccination, and antibiotics, infant mortality has occurred close to the time of birth. For this reason, the national neonatal mortality rate (NMR) has been used, a neonate being defined as up to the age of 28 days. The same denominator is used as for the IMR, and the difference between them is known as the postneonatal mortality rate. Defined in this way, as it is, it contravenes the proper definition of a rate, which should refer to the ratio of the number to whom some event has happened (e.g., death) to all those who were at risk for that event. The denominator of the postneonatal mortality rate is the number of live births, just as it is for the IMR and the NMR. But all those who succumbed as neonates are no longer at risk in the postneonatal period, and thus should be excluded from the denominator. The difference, however, is usually small, and it is more convenient to use two rates, which add to the overall IMR.

Further reductions in the deaths at this period of life have focused attention nearer to the time of birth. Deaths in the first week of life (up to the age of 7 days) have been recorded for many years now, as well as separately for each of those 7 days, and even for the first half hour of life. Clearly many of the causes of those very early deaths will have originated in the antenatal and intrauterine period. They will share causes with those born dead (stillbirths), and indeed they are combined together in the prenatal mortality rate. This includes both stillbirths taken together. The stillbirth rate (SBR) alone must of course use the same denominator, since all births were at risk of death in the process of birth, to which the stillbirths fall victim. All of these rates have been devised to highlight specific areas of importance, especially in pediatrics. Closely related is the measurement of the material morbidity rate (MMR). Here the numerator is the deaths of women from maternal or puerperal causes, and the denominator, interestingly, is the total number of births, live and still. A moment's reflection will show that it is the occasion of birth (whether live or still) that puts a woman at risk of this cause of death, and that if she has twins—or higher orders of multiple births—she is at risk at the birth of each, so that the correct denominator must include all births.

FERTILITY RATES

The information collected on the birth certificate usually permits the tabulation of fertility rates by age and number of previous children. Age-specific fertility rates are defined as the number of live births (in a calendar year) to a thousand women of a given age. If they are expressed for single years of age, and they are separated into male and female births, then we add together all the rates for female births to give what is known as the gross reproduction rates (GRRs). If this quantity is close to unity, then it implies that the number of girl children is the same as the number of women of reproductive age, and the population should thus remain stable in number. But no allowance has been made for the number of women who die before the end of their reproductive life, and thus will fail to contribute fully to the next generation. When this allowance is made (using the female mortality rates for the appropriate ages) we obtain the net reproductive rates (NNRs). Note, however, that there remains an assumption that may not be fulfilled—that the age-specific rates remain unchanged throughout the reproductive age range (usually taken as 15 to 45), that is, for a period of 30 calendar years. Indices such as the NRR were devised as attempts to predict or forecast the likely future trends of populations. The crude birth rates (CBRs), defined as the ratio of the number of births to the total of the population, is like the crude death rate in being very sensitive to the age structure of the population. Nonetheless, their difference is called the rates of natural increase (RNI) and provides the simplest measure of population change:

$$\text{CBR} - \text{CDR} = \text{RNI}$$

The measure excludes the net effect of migration in changing the population numbers: in some countries it is very rigidly controlled, and in others it may be estimated by a sampling process at airports, seaports, and frontier towns.

POPULATION TRENDS

Previously it has been noted that both the GRR and NRR make the assumption of projecting the rates observed in 1 calendar year to cover a 30-year period (15 to 45). It would of course be possible to follow a group of women, all of the same age, from when they were 15 up to the age of 45 in the latest year for which figures are available. Such a group would be called a "cohort"—the term used in epidemiology for a group defined in a special way. To cover this cohort would necessitate obtaining fertility rates for up to 30 years back, and in any case that cohort would of course have completed its reproductive life. The highest fertility rates are commonly found at younger ages: it is possible to show graphically a set of "cohort fertility rates" by age labeled by their year of birth (often a central year of birth, since the cohort may be more usefully defined as a quinquennial group). If they are expressed in cumulative form (i.e., added together) and refer only to female birth, it will become clear how nearly they approach unity, from below or above, if the population is increasing. No adjustment for female mortality in the period is required, since the rates are, for each year (or quinquennium), calculated for those women of that cohort alive at that time. The method therefore represents the most useful prediction of future population trends, which can be projected further forward by assumptions that can be made explicit in their graphical depiction.

COHORT ANALYSIS OF MORTALITY

A similar breakdown of age-specific mortality rates can be made, in order to reveal different patterns of relationship to the passage of time. Figure 1, for instance, shows mortality rates by sex and age in a single calendar year—the age in which death took place. Mortality rates are given for 5-year age groups, which is the usual practice, so that if a similar curve were to be drawn on the same graph for the calendar year 5 years earlier, you could join together the point representing, say, the age group 60–64 on the original curve to the point for 55–59 5 years earlier. This line would then represent a short segment of the cohort age-specific mortality curve born in the period 60–64 years before the date of the first curve. By repeating the process, it is clearly possible to extend the cohort curves spaced 5 years apart in their birth years. Figure 6 shows how the cohort mortality makes clear the rising impact of cigarette smoking in the causation of lung cancer, since successive later-born cohorts show increases in the rates, until those of 1916 and 1926, which begin to show diminishing rates. The cohort method is thus of particular relevance where there have been secular changes similar to that of cigarette smoking.

MEASUREMENT OF SICKNESS (MORBIDITY)

If, instead of death, you look for ways of measuring sickness in the population, once again you are confronted by several major differences in both interpretation and presentation. In the first place, illness has a duration in a sense that is absent from death. Secondly, the same illness can repeat in the same individual, either in a chronic form or by recurrence after complete remission or cure. And thirdly, there are grades of illness or of its severity, which at one extremity may make its recognition by sign or symptom almost impossible without the occurrence of the individual. The tolerance of pain or disability, or their threshold, differ widely between people, and therefore complicate its measurement. In the case of absence from work, where a certificate specifying a cause may (or may not) be required, various measures have been used. A single period of absence is known as a "spell," and thus the number of spells per employee in a year, for instance, can be quoted, as well as the mean length of spell, again per employee, or perhaps more usefully, by diagnosis. Inception rate, being the proportion of new absences in a given period (1 year, or perhaps less) is another measure, which again would be broken down into diagnostic groups. Prevalence is yet another measure, intended to quantify the proportion of work by sickness (perhaps by separate diagnostic groups) at a particular time. This may be, for instance, on one particular day, when it is known as "point prevalence," or in a certain length of time (e.g., 1 month), which is known as "period prevalence." Most prevalence rates are given for a year, and the definition often referred to is the number of cases that exist within that time frame. On the other hand, incidence is the number of cases that arose in the time period of interest, again usually a year. When sickness-absence certificates are collected for the purpose of paying sickness benefits, they have been analyzed to present rates and measures such as those discussed here, often against a time base, which can show the effect of epidemics or extremes of weather—or may indicate the occurrence of popular sports events! But such tabulations are either prepared for restricted circulation only, or if published are accompanied by a number of caveats concerning their too-literal interpretation.

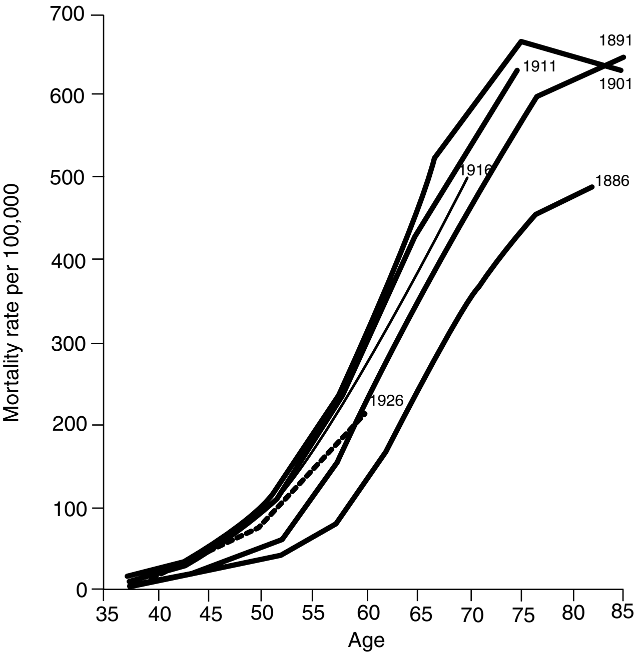


FIGURE 6 Lung-cancer incidence in birth cohorts.

Incidence and prevalence rates are related to each other, and it is not unusual to have both reported in a single study (Mayeux et al., 1995). An example of prevalence and incidence for Parkinson's disease for the total population and different ethnic groups is shown in Tables 2 and 3. For prevalence, the study identified 228 cases of the diseases (Parkinson's) for the time period 1988–1989, with the final date of inclusion being December 31, 1989. Not included in the table is the mean age of cases (prevalence) (73.7 years, standard deviation 9.8) for patients having ages 40 to 96 years. Mayeux also reported that the mean age of occurrence (symptoms) was 65.7 (standard deviation 11.3), with differing ages for men (64.6, standard deviation 12.7) and women (67.4, standard deviation 10.6), with these differences having a p value of 0.06, or 6%. It should be noted that if a statistical significance of 5% is used for establishing a difference, the age difference in years between men and women when symptoms of Parkinson's disease were first observed (occurrence or onset of diseases), thus, is not different. However, this raises an important issue that using a cutoff value, say 5%, does not provide a definitive determination for evaluating data, in this case the importance of disease occurrence between the two sexes. Most when examining these data would suggest that even though the difference is not significant, an important difference between men and women appears to exist for the onset of disease, with men's being much earlier. This would indicate, for example, that screening for this disease be initiated at an earlier time period for men.

Table 2, by Mayeux *et al*. (1995), also indicates a difference of disease onset by age. There is a dramatic prevalence rate (total) for populations under 45 (1.3 per 100,000) as compared to those over (e.g., 99.3 per 100,000 for the age group 45–64 years). The prevalence also overall increases with age. There are also differences in prevalence among ethnic groups and sex within these groups. This demonstrates the importance of studies examining race and sex as important factors in disease (Ness *et al*., 2004; Lange *et al*., 2003b). This study (Mayeux *et al*., 1995) examined the age-adjusted prevalence rate for combined men and women by ethnic group (race), and a significant difference (p value < 0.01) was observed among blacks (57 per 100,000), whites (116 per 100,000), and Hispanics (130 per 100,000). This indicates that there is not only a difference in the onset of Parkinson's

TABLE 2
Prevalence of idiopathic Parkinson's disease in a New York neighborhood based on a community diseases registry, 1988–1989

Ethnic group and sex	Age group (years)					Total	
	<45	45–64	65–74	75–84	⩾85	Crude	Age-adjusted
Black men							
No.	0	1	7	6	1	15	—
Population	19,395	4,265	1,216	530	150	25,556	—
Prevalence rate	0	23.4	575.7	666.7	58.7	92.0	92.0
						(29.0–88.4)[‡]	(54.7–129.0)
White men							
No.	0	1	7	8	3	19	—
Population	20,285	6,020	2,296	1,305	443	30,349	—
Prevalence rate	0	16.6	304.9	613.0	667.0	62.6	54.7
						(34.5–90.7)	(28.4–81.0)
White women							
No.	1	14	12	34	12	73	—
Population	26,447	7,036	2,446	1,710	636	38,275	—
Prevalence rate	3.8	199.0	490.6	1,074.7	1,886.8	167.3	86.0
						(147.0–234.0)	(131.0–114.0)
Hispanic men							
No.	0	1	7	6	1	15	—
Population	19,395	4,265	1,216	530	150	25,556	—
Prevalence rate	0	23.4	575.7	666.7	58.7	92.0	92.0
						(29.0–88.4)	(54.7–129.0)
Hispanic women							
No.	0	1	7	6	1	15	—
Population	19,395	4,265	1,216	530	150	25,556	—
Prevalence rate	0	23.4	575.7	666.7	58.7	92.0	92.0
						(29.0–88.4)	(54.7–129.0)
Total							
No.	0	1	7	6	1	15	—
Population	19,395	4,265	1,216	530	150	25,556	—
Prevalence rate	0	23.4	575.7	666.7	58.7	92.0	92.0
						(29.0–88.4)	(54.7–129.0)

Source: From Mayeux et al. (1995), The frequency of idiopathic Parkinson's disease by age, ethnic group and sex in northern Manhattan, 1988–1993, American Journal of Epidemiology, 142:820–27; with permission from Oxford Press.

disease by age and sex, but by ethnic group as well. These differences in Table 2 illustrate the importance of controlling for various confounders such as sex, age, and race in epidemiological studies.

When examining the incidence rate of Parkinson’s disease (Table 3), this study reports 83 new cases during the 3-year period. For the incident cases, the mean age was 76.3 years, with a standard deviation of 9.5 and a time period of symptoms (duration of 1.4 years). There was reported no difference in the mean age (75.2 years) for men and women for the onset of symptoms. From Table 3, the annual incidence of Parkinson’s disease in New York City is 13 per 100,000. Unlike that reported for prevalence, there were no incidences of Parkinson’s diseases below the age of 45 years during this study period. However, as noted for prevalence, the incidence rate of disease increases with age. There is also a difference in the rates for men and women and among ethnic groups evaluated.

The data in Tables 2 and 3 can be used to evaluate the numbers of cases and the occurrence of disease in a

TABLE 3

Annual incidence of idiopathic Parkinson’s disease over a 3-year period in a community diseases registry, New York City, 1988–1989

Ethnic group And sex	Age group (years)					Total	
	45–64	65–74	75–84	⩾85		Crude	Age-adjusted
Black men							
No.	1	7	6	1	15	—	—
Population	4,265	1,216	530	150	25,556	—	—
Prevalence rate	—	23.4	575.7	666.7	58.7	92.0	92.0
						(29.0–88.4)[‡]	(54.7–129.0)
White men							
No.	1	7	8	3	19	—	—
Population	6,020	2,296	1,305	443	30,349	—	—
Prevalence rate	16.6	304.9	613.0	667.0	62.6	54.7	—
						(34.5–90.7)	(28.4–81.0)
White women							
No.	14	12	34	12	73	—	—
Population	7,036	2,446	1,710	636	38,275	—	—
Prevalence rate	199.0	490.6	1,074.7	1,886.8	167.3	86.0	—
						(147.0–243.0)	(131.0–114.0)
Hispanic men							
No.	1	7	6	1	15	—	—
Population	4,265	1,216	530	150	25,556	—	—
Prevalence rate	23.4	575.7	666.7	58.7	92.0	92.0	—
						(29.0–88.4)	(54.7–129.0)
Hispanic women							
No.	1	7	6	1	15	—	—
Population	4,265	1,216	530	150	25,556	—	—
Prevalence rate	23.4	575.7	666.7	58.7	92.0	92.0	—
						(29.0–88.4)	(54.7–129.0)
Total							
No.	1	7	6	1	15	—	—
Population	4,265	1,216	530	150	25,556	—	—
Prevalence rate	23.4	575.7	666.7	58.7	92.0	92.0	—
						(29.0–88.4)	(54.7–129.0)

Source: From Mayeux et al. (1995), The frequency of idiopathic Parkinson’s disease by age, ethnic group and sex in northern Manhattan, 1988–1993, American Journal of Epidemiology, 142:820–27; with permission from Oxford Press.

population. This provides information on the distribution of the disease and possibly allocation of resources in its prevention and treatment. Although the use of prevalence and incidence was demonstrated for a chronic disease, Parkinson's disease, it can also be used for occupational and environmental diseases and events. Application of preralence incidence are illustrated in Table 2 and 3.

SICKNESS SURVEYS

In some cases an estimate of the amount of sickness in the population has been made by market-research techniques, whereby people in the street are interrogated about their health in the last week (or month, but the shorter period is preferred for purposes of better accuracy). This method was used by the British government during World War II and was described as the "Survey of Sickness." There are several obvious omissions that are likely to distort the findings, such as the chronically ill who cannot get out into the street, or those who return straight to work after short illness and are not available to be questioned in the street. Nevertheless it is a cheap method that may have sufficient consistency to validate time trends of indices based on it. For more severe illnesses, statistics of admissions to hospitals, diagnoses, treatment methods, length of stays, etc. may serve to supplement the measurement sickness in the population. Not all hospitals may be able to provide useful figures, however, nor may their catchment areas be sufficiently clearly defined. However, with the advent of computers and development of databases for diseases, these issues of disease are better defined, especially in Westernized countries. Because of the wide range of sickness itself, the means of dealing with it, and variation in reporting, some form of sampling from the various sources is likely to yield the most useful results. However, even with attempts at standardization, ICD codes, there is often a wide variation in the incidence and prevalence of disease, even in the same community. One solution to this problem, in providing an accurate estimate of disease, is to employ the capture-recapture method (CRM). Since the problem in determining the occurrence of sickness and disease is at the heart of counting, the CRM has been suggested as the most accurate method for counting (Lange *et al.*, 2003c). However, since this method was originally derived for counting wildlife, it is best known as an ecological- and population-biology method, and has not been widely adopted by epidemiologists. Its recent use in counting hazardous-waste sites (Lange *et al.*, 2003a) demonstrates this method's versatility for counting, including in the area of epidemiology.

DISEASE REGISTERS

For a number of diseases, attempts have been made to make and maintain lists or registers of those affected. Infectious diseases are among the most obvious to fall into this category, since isolation of those affected was for so long the only effective deterrent to their spread. The advent of specific treatment methods, or more usefully of vaccines and immunization, has greatly reduced their impact, except perhaps for rarities such as Lhasa fever, or, in another category, AIDS. The pulmonary-tuberculosis register was an important one until treatment became readily available. Notification of the local health authority of any of the range of diseases has often been a requirement of general practitioners, in order to obtain early warning of an impending epidemic, and to monitor the occurrence of those diseases. Some heart conditions have formed the subjects of registers, though limited usually by time or space. Furthermore, a number of diseases, usually rare and often genetic, have registers or societies of affected patients, which can be useful not only for the discussion and possible alleviation of common problems, but also for the purposes of research. Most recently, disease registries, or what could be included as registries, have emerged for specific cohorts such as migrant agricultural workers (Zahm and Blair, 2001) as well as for specific occurrences of disease (Lange *et al.*, 2003d).

CANCER REGISTERS

Cancer registers, or more commonly registries, form a very distinct and important group of disease lists. They fall into three categories: (1) special registries concerned only with certain sites of the disease (e.g., bone cancer or gastrointestinal cancers); (2) hospital-based registries, which record all those cases seen at a particular hospital or groups of hospitals; and (3) population-based registries, which endeavor to collect records of every case of cancer within a specific population. The last group is the most important as a source of morbidity data about cancer. Cancers form a special group of diseases of exceptional importance, at least historically, which attract a great deal of interest and research, and for which epidemiological methods are of outstanding relevance. The date of diagnosis of cancer can be used as the basis of morbidity rates analogously to the mortality rates, so that from a population-based registry, rates of morbidity by sex, age, and site can be constructed, using the sex and age structure of the population for appropriate denominators. Cancer registries exist now in many parts of the world, though more in developed than in the underdeveloped countries, and their incidence rates by sex, age, and site have been collectively published in the successive volumes of *Cancer Incidence in Five Continents*, beginning in 1966 and subsequently at intervals of approximately 5 years. They have formed a valuable source of comparative data about the different patterns of cancer found geographically, and in combination with other data sources can lead to the generalization hypotheses of etiology. To their use in a variety of other ways, such as in the evaluation of occupational and environmental carcinogenic hazards, in the conduct of comparative clinical trails (especially in those patients not included), and among the sequelae of certain types of chronic disease, we shall refer later.

The pattern of cancer displayed in relation to sex, age, and site, whether in terms of mortality or morbidity, can provide important information on carcinogenic hazards in the area to which it refers. The different proportions by site throughout the world, where data are available, affect climatic, geographical, and lifestyle variations but are not always simple to analyze. Furthermore, some sources of data may consist of numerator information only (e.g., deaths or diagnosed cases of disease) without the corresponding population figures by sex and age, which enable the construction of rates of mortality and morbidity. Rather than have to ignore such partial information, it is possible to present it in the form of proportionate mortality or morbidity rates. In its simplest form this method expresses the omitting age, since it may not be available; then it may facilitate comparison with another source of data that may have affinities, perhaps in the likely age structure or in climate, to one under study.

PROPORTIONATE MORTALITY RATE

A situation somewhat similar to what has just been described can occur when a factory may be able to provide details (of cause, sex, and age) of the deaths of former employees over a period of time, but without adequate additional information to permit further analysis (see below). The pattern of their mortality, by cause and sex, can then be compared with the general patterns of the area where the factory is situated. Usually the deaths will be categorized into broad groups (e.g., cardiovascular, neoplasms, respiratory) (Lange *et al.*, 2003b), but if there is a reason to examine certain sites of cancer individually, they can be included. Each observed death is allotted an age group in the general population; the proportions of deaths (as fractions of 1) occurring in each of the chosen cause categories are entered and summed after all the observed deaths have been included. The accumulated fractions in each group will then constitute the expected number of deaths to compare with the number observed. Accumulated by age into numbers of expected and observed deaths by sex and cause, the comparison can be evaluated statistically for its significance (see below). If the observed deaths are spread over a number of years, the expected deaths should be strictly obtained from the corresponding calendar years, though they can usually be taken in quinquennial groups without serious loss of accuracy. It will be evident that an assumption implicit in the method is that the factory population has been sufficiently similar to the general population to justify the comparison. Thus, proportionate mortality rates (PMRs) are commonly used for occupational cohorts.

ANALYTICAL EPIDEMIOLOGY

It is conventional to divide epidemiology into two distinct branches: descriptive and analytical. Up to this point we have been concerned mainly with the description of the health status of a population by means of rates of mortality and morbidity, by sex, age, and cause; for geographical and other subgroups; and in calendar time. Some of the methods of comparison we have discussed, necessitating standardization of the calculation of expected figures, have touched on the analytic division, though the definitions are not always clear-cut. Other methods that will be discussed are experimental in their design, such as clinical trials where the treatments of a disease by two different regimes forms the basis of a comparison of their relative efficiency, the numbers of patients being decided by consideration of the statistical power required or attainable.

SCATTERGRAMS

A method that has been frequently used in searching for factors related, possibly in an etiological way, to the incidence of disease is to display in graphical form a correlation diagram—a scatter diagram or "scattergram" where the incidence (or mortality rate) of the disease is measured on one scale and the factor of interest on the other. Figure 7 gives an example of the method, whereby each point represents a country whose (standardized) rate of colon cancer is set against the per-capita consumption of fat in that country. It is clear that there is a relationship between these two measures, such that as one increases so does the other. This apparent movement together may suggest a possible causative relationship, such that the higher the average consumption of fats, the greater the risk of colon cancer. But note first that it is the average per-capita factor, which is obtained from the total fat consumed in a country divided by population. Clearly, individuals in the population will vary in their mean levels of consumption; some average more while others less than the average. If there is a causative relationship, we would expect

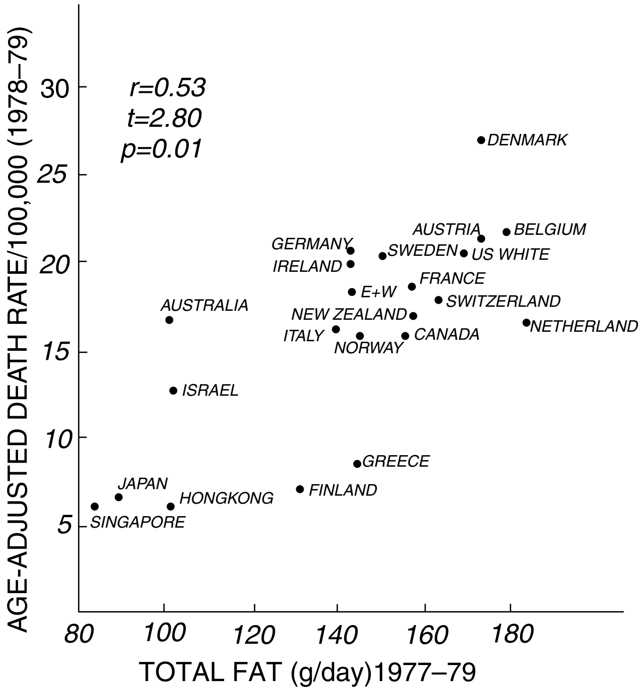


FIGURE 7 Cancer of the colon and fat consumption (scattergram).

to find higher rates of colon cancer among those consuming more than the average. But this information we do not posses, though it is precisely what is required to demonstrate the relationship conclusively.

LATENCY PERIOD

Another consideration, which is of particular relevance in the field of cancer and other areas of epidemiology, is that there is a latency period, often of at least 20 years, between exposure to a substance (like a carcinogen) and the development of a clinically observable disease (again, like cancer). Thus the per-capita fat consumption should refer to figures of 20 or more years ago. If the cancer incidence is not showing rapid secular change, this consideration may not be important. It has also been shown that there is a similar relationship between the consumption of protein and colon cancer. It may well be that both fat consumption and a diet high in protein are among the causative factors of colon cancer, but it may also be that some other factor or group of factors, correlated with these two in particular, is more directly relevant. Lifestyle factors or social status may each be included, and both subsume at the same time a wide variety of other measures, some of which may be more legitimately described as directly etiological. In sum, therefore, a correlation diagram can lead to the generation of hypotheses of causation but cannot of itself prove the relationship. The gross correlation between national averages and disease incidence needs to be investigated on individual cases of the disease, for each of which measures of consumption of putatively carcinogenic items of diet can be obtained, preferably over a large period of time in retrospect. Data of this kind, if sufficient in quantity and reliability in substance, can form the basis of an informative etiological study, probably using multivariate analytical techniques.

REGRESSION AND CORRELATION ANALYSIS

The relationship between the incidence of lung cancer and the number of cigarettes smoked is now well known, and has been verified many times in a variety of situations. For most of these studies it is possible to obtain a graph of the mortality (or morbidity) rate against the number of cigarettes smoked per day, yielding a straight-line relation of the form $y = a + bx$ between them, where y is the morbidity rate and x the cigarettes smoked per day, a and b being appropriate constants. The values of a and b can of course be readily obtained from the data, being the parameter of the simple linear regression between the two quantities. Many books on statistics prescribe the technique of fitting regression lines.

INTERACTION

A relationship of a similar kind has been shown between the incidence of esophageal cancer and the number of cigarettes smoked. The same disease is also related in the same way to quantity of alcohol consumed. We could combine these two findings to obtain a regression equation using two quantities as influencing a third (the morbidity rate), of the form

$$z = a + bx + cy$$

where z is the morbidity rate,

x is the number of cigarettes smoked,
y is the quantity of alcohol consumed,
and a, b, and c are appropriate constants.

The form of this equation assumes independence between the actions of each quantity on the morbidity rate. It has been found in actual study that the combined effect of both quantities in the same individuals results in an enhanced rate of morbidity, above the additive effect of the two separately. This enhancement, amounting to a multiplicative rather than an additive effect, is known as "synergism" (Figure 8). A similar example of a synergistic effect is found in the occupational field by the combination of the effects of exposure to asbestos and cigarette smoking on the development of lung cancer. Figure 9 shows the separate effects of each agent in terms of a rate set at 1 for those exposed to neither and also the rate for both together, which corresponds to multiplying rather than adding the separate rates. The establishment of a genuinely synergistic effect requires both extensive and reliable data.

ANALYSIS OF OCCUPATIONAL DISEASE

In studies of occupational disease, the basic question to be answered is whether there is an excess of cases of the disease

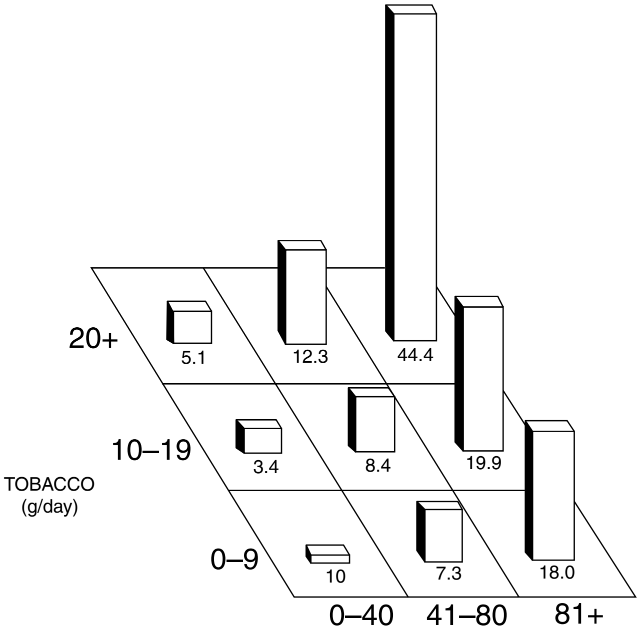


FIGURE 8 Cancer of the esophagus in relation to alcohol and smoking.

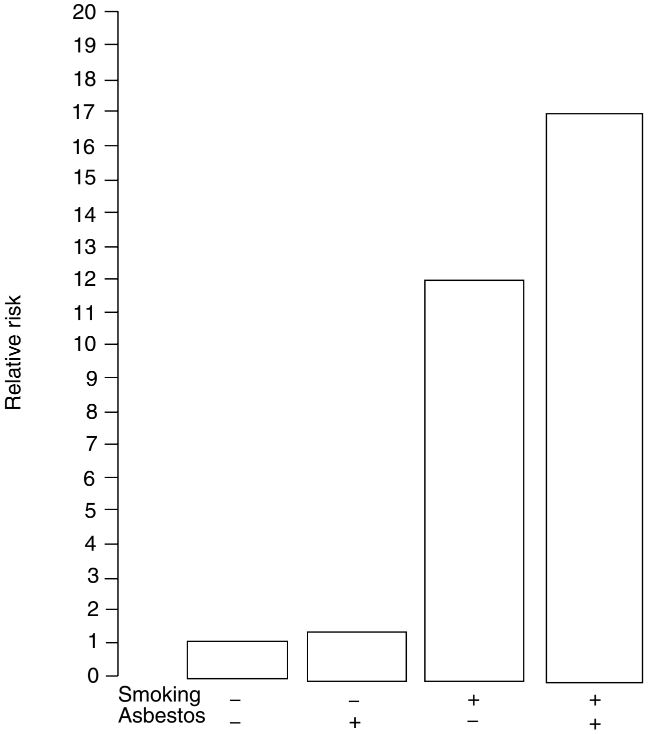


FIGURE 9 Cancer of the lung in relation to asbestos and smoking.

among the workers exposed to the putative causative factor, whether that is a substance or a process that is used in certain parts of a factory. In its most elementary form, the results of an investigation can be put in the form of a 2 × 2 table, as in Figure 10.

The first two cells of this table, horizontally, include the numbers of those who were exposed (+) to the presumed hazard, the first cell containing the number of those who developed the disease in question (a), the second those who did not (b). In the lower line are those who were not exposed (−), the first cell again including all those for this group who developed the disease (c) and the second those who did not (d). Clearly if the ratio of the left to the right is the same in both rows, there is no evidence of an effect. This is an example of the 2 × 2 or fourfold table to which the X^2 test can easily be applied, with one degree of freedom, to assess whether any difference in the proportion, horizontally or vertically, attains statistical significance, at whatever level may be chosen. The conventional levels of significance are 0.05 (5%), 0.01 (1%), and 0.001 (0.1%), each referring to the probabilities that the observed result could have occurred by chance alone (the levels are sometimes quoted in the form of percentages, multiplying their probabilities by 100). For each cell of the table of Figure 10 an "expected" figure can be calculated from the marginal and grand totals, by dividing, for instance, each row total in the proportions of the column totals: thus the expectation for the top left cell is the product of the first row's total divided by the grand total. The difference between the observed number in each cell is squared and divided by the expectation for that cell, and the sum of these four quantities constitute X^2. Tables provided in almost all books on statistics will enable the level of significance to be obtained for the value of X^2 and for one degree of freedom. Two caveats should be noted: first, that no expectation should be less than 5—if it is, a larger size of sample is required—and second, that when numbers are small (yet satisfy the proceeding conditions), Yates's correction should be made, which reduces the absolute size of the difference between observed and expected by the quantity 1/2. It will have been noted that this difference is the same magnitude in each of the four cells, though it changes sign, but that is irrelevant since it is squared. It is the absolute magnitude of this common difference that should be reduced by 1/2.

		Disease	
		+	−
Exposure	+	a	b
	−	c	d

FIGURE 10

ODDS RATIO AND RELATIVE RISK

In the circumstances set out above and in Figure 10, the ratio $c/(c + d)$ is the risk of disease in the unexposed group, which we can call P_0, and P_1/P_0 is known as the "relative risk," RR or r. In many cases the disease in question will be rare, even among the exposed, so that a and c will be small relative to b and d. If $Q_0 + 1 - P_0$ and $Q_1 + 1 - P_1$ express the risk of not contracting the disease, then they will both be close to 1, since P_0 and P_1 are supposed to be small. The quantity $(P_1/Q_1)/(P_0/Q_0)$ is known as the "odds ratio," since it is the ratio of the odds of occurrence of the disease in the exposed to the unexposed groups. Since we are presuming the Q's to be close to 1, the odds ratio can be put as P_1/P_0, which is the same as the relative risk, r. We shall see later that this fact permits the estimation of the ratio of the incidence of disease in the exposed and unexposed groups from a case-control type of investigation, though their absolute incidences are not obtainable.

ETIOLOGICAL STUDIES

A situation that is formally very similar to what we have just been considering arises if we suspect a certain factor may be one that is involved in the etiology of the disease. We shall again be comparing persons with and without the disease and those affected in the form of a 2 × 2 table like Figure 10, where we replace "Factor" for "Exposure." If the factor is indeed an etiological one, it will be found more frequently

		Lung Cancer		
		+	−	
Smoking	+	688	650	1388
	−	21	59	80
		709	709	1418

FIGURE 11

in association with the presence of the disease, and less so with its absence.

A case-control investigation led naturally to this end point, though it may take a variety of different forms. Take one of the earliest studies of smoking and lung cancer, Doll and Hill (1950), which was followed by a cohort of English physicians using a questionnaire (Doll and Hill, 1954) and then continued as a prospective study up to the present day (Doll *et al.*, 2004). For this study 709 patients with lung cancer, in 20 hospitals, were matched with the same number of patients from the same hospital, but not having cancer or a respiratory disease. The matching was for the same hospital, of the same sex, and within the same 5-year age group. All the patients were interviewed according to a standard questionnaire. The simplest form of the results is shown in Figure 11.

The expectations in the two lower cells of the square are each 40, and in the upper cells each 699. The difference therefore in each cell is 19, with the result that the X^2 value, whether or not Yates's correction is used (it is not necessary here), is very large, and the probability that the association between smoking and lung cancer might be a chance one is extremely unlikely. In the study itself the result of smoking (in numbers of cigarettes smoked per day) was quantified, and the results given separately by sex. For the purpose of this illustration the study is summarized in Figure 11, but clearly the additional evidence afforded by the quantification data, which for each sex showed a steadily increasing risk of lung cancer at each successive level of smoking, reinforces the basic etiological relationship of smoking to lung cancer.

CASE-CONTROL STUDIES

In any case-control study ("case-referent study" is a synonymous term) the choice of appropriate controls is of special importance. In the study discussed above, the controls were matched for sex and age group—the two most commonly used characteristics for matching—and also for hospital, lest there should be some factor associated with that. There were two exclusions: cancer and a respiratory condition, which could confuse the contrast between cases and controls. When there are few cases available, and in some other circumstances, it may be advisable to use more than one control per case. Beyond four controls per case little further advantage can be gained, but two, three, or four controls for every case may be useful, though expensive. In general the more closely the controls resemble the cases in terms of characteristics, the more efficient the contrast, except that one or more of those characteristics may be of genuine etiological significance, but because it is possessed by both case and control, it is impossible to distinguish.

CLINICAL TRIALS

The underlying philosophy is that of the experimentalists of the scientific renaissance, who began in physics or in chemistry to look at the effects of a single factor alone, varying its contribution to the ultimate effect while endeavoring to keep other factors constant. The method could then be repeated for other factors, and thus the independent effects assessed, as well as those where separation proved impossible because of close correlations. The aim of the clinical therapeutic trial, for instance, is to obtain two groups of patients so similar in all known relevant respects that any difference in their responses can be reasonably attributed to their different treatments. Not only sex and age but the type and severity of the disease and its history, together possibly with socioeconomic or lifestyle factors, if relevant, need to be taken into account in ensuring the parallelism of the two treatment groups. It is important that the full treatment regimen in both groups (experiment and control) be decided in advance and adhered to precisely. There must be of course a provision for emergencies, and therefore escape clauses or alternative regimens should form part of the design of the trial. A pilot trial, perhaps amounting to around 5% of the full trial, can greatly help to reveal aspects previously overlooked, and if the modifications it suggests are not too great, it may be possible to include it as the start of the main trial. For the reason that those directly concerned with the conduct of the trial, or with its assessment, may form premature opinions about its outcome and hence introduce a bias if they know the actual treatment that patients receive, it is customary to run many clinical trials "blind"—that is, in such a way that the clinicians are unaware of the treatment given. If the active treatment consists of tablets, the control could be a placebo presented in the same form; if it is an injection, the control can receive an injection of normal saline; etc. The trial may also be "double-blind," when neither clinician nor patient knows the identify of the "apparent" treatments. Of course, a singly blind trial may imply that the patient is unaware of his treatment but the clinician does know. There are also occasions when the two treatments have differences that cannot be distinguished, such as surgery for one and radiotherapy for another.

STRATIFICATION

We have stressed already the importance of the close similarity—almost identity—of the two groups of patients, referring to obvious characteristics such as sex and age. Other relevant features should also, if feasible, be similarly balanced between the groups, and each of them may be described as a stratum, so that the trial is described as stratified. An example may be the stage of disease (if distinct stages are discernible), while another might be for an ethnic group. Finally, when the numbers of strata have been decided, the patients within each ultimate subgroup should be randomly allocated to the treatment groups. The more strata that are used, the more patients will be required to enable a proper balance between treatments.

FOLLOW-UP

When a clinical trial of treatments for a cancer is conducted within the territory of a cancer registry, and, as almost invariably happens, not all eligible patients have been included in the trial, the data from the registry concerning those patients excluded (for whatever reason) can help to indicate whether the trial represents a reasonably unbiased selection of the totality of patients with the specific tumor in the period covered. If the registry also usually undertakes follow-up of its cases, it can also provide this service for the clinical trial, although the time intervals may be different, and there may also be certain monitoring requirements for trial patients to be added.

ETHICAL CONSIDERATIONS

Some clinical trials have been set up for a fixed period of time, or with a certain number of patients, sufficient to evaluate a difference of a previously determined size between the treated and control groups. For these, therefore, there is an endpoint at which the outcome is assessed. Often, the desired magnitude of the difference was not attained, and so the trial was in that respect inclusive. Sometimes the reverse was the case, and those on the new treatment fared so much better that it could be considered unethical to proceed with the trial, but instead, the control group should be given the new treatment also. The ethical question raised is a difficult one, since it has sometimes happened that the favorable difference has not continued into a longer follow-up period. There are many other aspects to questions of ethics in the field of clinical trials, which are not appropriately considered here other than to note that they exist.

SURVIVAL: LIFE-TABLE FORM

Often the results of a clinical trial will be displayed in what is described as life-table form. This can be constructed at specific intervals of time (e.g., 3 months, 6 months, then annually) or can be related directly to the times of events, such as the death of a patient or his withdrawal from the trial. This latter event could be because that period represents his whole experience of the trial from entry to the current date: this situation is sometimes described as a "censored" survival time, since it will be true that his survival is, at its least, that time, and is likely to be greater if follow-up can continue. If withdrawal is made on grounds related to the conditions on the trial itself, such as a side effect of treatment, it will clearly have an effect on the ultimate interpretation of the trial as a whole. These patients may be removed from the trial, but their number and the cause of removal must be reported.

When the life table is constructed in relation to the occurrence of events, the patients in a specific group (e.g., treatment, or control) are arranged in order of their survival time (including censored times). If there are n patients in one such group, they will of course start at time zero ($t_0 = 0$), as 100% alive. When the first event occurs, at time t_1, suppose it is that one patient has died. Then the probability of dying at that time can be regarded as $1/n$ ($= q_1$), and the probability (p_1) of surviving as its complement ($p_1 = 1 - q_1$). Suppose the next event is the death of two patients at time t_2; the number of patients at risk at the time of the event is $(n - 1)$, and so q_2. The survival probability to time t_2 is therefore the product of p_1 and $p_2 = s_2$. Now, at time t_3, one patient arrives at his censored survival period. The probability of dying then is $q_3 = 0$ (since he is still alive) and thus $p_3 = 1$, and $s_3 = s_2$, the same proportion. At the time q_4 is $5(n - 4)$, since four had been removed from the numbers continuing up to the maximum follow-up time. The resultant curve is in the form of a series of steps, at varying intervals (the t's), and represents the survival curve for that group.

RETROSPECTIVE AND PROSPECTIVE METHODS

The case-control study discussed earlier can be regarded as a clinical trial, but in the field of etiology rather than in therapy. There is the same call for close similarity between case and control as between the parallel cases of the therapeutic trial. The etiological study, however, is essentially retrospective, beginning with cases of a disease, and appropriate controls, and comparing their past experiences in an attempt to discover relevant differences that could be etiological factors. The therapeutic trial, on the contrary, is prospective, in the sense that, having set up the two-treatment group, the outcome of the trial is awaited and assessed. Another type of etiological study is known as a "cohort" study because its basis is a group or cohort of people defined in a certain way and exposed to some putative hazard (an etiological factor of a resultant disease or disability). In the realm of occupational health, for instance, we may need to investigate whether the work of a particular factor or substance thereof might include some carcinogenic aspects.

We have already referred to the phenomenon of latency as one of the difficulties encountered when investigating carcinogenic hazards. This is the term used for describing the long period elapsing between the first exposure to the hazard and the development of a recognizable tumor. In other words, this is the time from the first exposure to the time of diagnosis of the disease and can be applied to any disease. The duration of the latency period varies from a few years (short period of time), possibly even less, up to decades, like 20, 30, and over 40 years. Some of the shortest latency periods are found with the leukemias and lymphomas, and the longer with solid tumors. Often this long period of latency has led to the failure to recognize an occupational cancer because its manifestation has been relatively late in a man's working life and has been regarded therefore as a spontaneous incidence with age; this has also been true of postretirement cases.

COHORT ETIOLOGICAL STUDIES

In its basic design, the cohort etiological study is prospective, since it notes the essential details of all those exposed to the suspected hazard and awaits the possible development of cancers. It is possible to estimate the number of cancers of various sites that would be expected to occur in a group (cohort) constituted in the same way by sex and age and observed for the same period of time, from rates of incidence obtained from a cancer registry in the same area. This would of course compare morbidity, though it would be equally possible to use mortality: in general, rather more information becomes available from the morbidity study, largely because of differential survival rates. The same design of study (i.e., a cohort etiological study) can be conducted in a retrospective way by the use of records, if well maintained by the factory. If, for instance, a full list is available of those employed, say, 30 years ago, when a new process or chemical (substance) was first used, and all the personnel changes since then to present time are also available, then the situation is almost precisely equivalent to that postulated above. We set ourselves 30 years back in time, collect all the original personnel data, and continue to do so until we are back to the present day. For this reason the method is often called the "historic cohort method." The only other information required is the present health status of each worker employed over that period, whether still employed, retired, or dead. From this information, which is seldom entirely complete, a comparison is made with the expectations of deaths and cases of disease, obtained by applying morbidity and mortality rates to the exposure data.

CALCULATION OF EXPECTED DEATHS

Since both mortality and morbidity rates change rapidly with age, and also to a lesser extent with calendar years, the calculation of expectations requires a systematization of the exposure data by sex, race, age, and calendar years. Usually it is sufficient to use 5-year age groups and also groups of 5 calendar years, classifying the data into "person-years" within the groups. Thus, if in the age group 55–59 in the early 1970s, the incidence rate of stomach cancer in white males was 75 per 100,000 (note that this is a rate), and there were 1,000 person-years (that is, men who were within that age group and in the early 1970s, for up to a maximum of 5 years each), the total contribution to the expectation of stomach cancer from this group would be 0.75. The overall expectation for white-male stomach-cancer cases could be a result from the summation of similar figures to cover the full range of ages and calendar years required for the known workforce. In all these computations it is usual to include only their observed lifetimes, so that a man who dies during the period of observation contributes to person-years only until the time of his death. An alternative method ignores the actual durations of life, substituting the life-table expectations from their sex and age at the time of entry, including only of course up to the endpoint of the study. If one of the effects of exposure at the factory were to cause a shortening of the normal lifespan, this method would serve to reveal it. In general, however, it is best to use either method with circumspection, since other features may also contribute to distortion of the expectations, such as the HWE or SPE, both described earlier in this chapter. The appropriate rates, whether of mortality or morbidity, need to also be carefully chosen. There may be no choice, if only the rates for the country as a whole are available: this is often true for mortality rates, though there may be some regional variants. Cancer registry data may be more regionalized and therefore more appropriate to the location of the factory or study population.

COMPARISON: THE SMR

The comparison of the observed numbers of deaths (or cases of disease) with the number expected is usually expressed in the form of an SMR, since the aim of the method of calculation of expectations is to obtain figures that have allowed for all the pertinent factors that distinguish those at risk, such as sex, age, race, calendar-year period, and region, amounting to a similar process to that of standardization described earlier. In general the comparisons will be made separately for different hazards, and also for different disease groups. Frequently they are evaluated for their statistical significance by use of the Poisson test, where the difference between the observed and expected numbers is set against the square root of the expected to provide a ratio that, if the expected is sufficiently large (12 or more), can be regarded approximately as a *t*-test, but otherwise needs specific calculation or recourse to tables of the function. In interpreting the contrast between observed and expected, it is important to keep in mind the source of the expectations, and its relevance to the comparison: if it based on mortality rates for a whole country, it will include those physically handicapped or otherwise unable to work, which favors the factory population and leads to the HWE, and also will include other groups of the population (e.g., social classes, other industries and occupations) that may serve to distort the comparison. SMR values are commonly used to evaluate risks associated with occupational and industrial situations.

THE PROPORTIONAL-HAZARDS MODEL

A different approach to the evaluation of specific hazards has been devised that makes use in effect of a series of internal comparisons. It is known as the method of regression models in life tables (RMLT), the name given to it by the originator, D.R. Cox, or as the proportional-hazards model. Essentially it is a form of multiple regression analysis, into which it is incorporated as many as possible of the variants that distinguish the workers in the factory: sex, age, race, duration of employment, type of job, area of work, exposure levels, etc. The comparisons are then made within the factory itself, so that the workers in an area of potential hazard are evaluated against the experience of the rest of the factory, taking into account all those factors included in the regression and likely to influence the behavior of each individual. Because of its independence of any external standard as a basis for the control, this method has a clear advantage over other kinds of comparisons. Its disadvantages stem from mainly information required, and an adequate varied range of areas of work. Given the requisite data, it provides probably the most powerful form of epidemiological analysis in this field.

SOME OTHER APPLICATIONS OF EPIDEMIOLOGY

The methods of epidemiology are increasingly being used in the investigation of a wide variety of health deficiencies or of areas of group ill health. The study of the relationship between Legionnaires' disease and defective maintenance of air-conditioning plants owes much to such methods. Another example, not fully worked out, is that of the sick-building syndrome, where a number of symptoms of malaise, together with respiratory and eye afflictions, have been related to ventilation problems and lighting conditions in some kinds of modern buildings. Incidents of food poisoning have been among the "classical" applications of methods perhaps more akin to those of crime detection than a strictly epidemiological type. More recently studies of outbreaks of food-borne disease—some due to inadequate cooking, some to the increased demand for precooked convenience foods, some to more intensive animal husbandry and processing of the meat—have made extensive use of methods of epidemiological analysis. Other obvious areas of application include inquiries into sudden outbreaks of other forms of disease, especially when localized in time or space, and the study of road accidents, whether in general or of specific "hot spots."

CONFIDENTIALITY

The usefulness of epidemiology, whether in its descriptive or analytical manifestation, depends essentially on the extent and consistency of the data on which it is based. A current trend in a number of countries has been that of "data protection," ostensibly to protect the privacy of the individual, and resulting in the "anonymization" of many data items. For a great many, if not all, items of information concerning an individual, the name is the most useful identification to link them together. Consequently the movement toward data protection has been against the interests of epidemiology, and has hampered much of its development. In many of the countries of Europe a residue of the Code of Napoleon has made the link between the cause of death and the name of the deceased an item of strict confidentiality, so that the death certificate is in two parts: a statement of the fact of death of a named individual, for legal purposes, and a separate statement of the cause(s) of death of an anonymous person, of stated sex, age, and race, for purposes of preparation of the country's mortality statistics. Consequently, epidemiological study has been considerably hindered by this dissociation. It is ironic, therefore, that whereas in those countries there have been attempts to liberalize this restriction in the interest of useful epidemiological studies, the reverse has been the trend in some countries where access to such data has previously been unrestricted. It should be possible to devise a compromise that permits suitably motivated and qualified investigators to obtain free access to all data necessary to their study without jeopardizing the general right of the individual to the confidentiality of the data concerning himself.

REFERENCES

Baris D., Armstrong B.G., Deadman J., Theriault G. (1996). A mortality study of electrical utility workers in Quebec. *Occupational and Environmental Medicine.* 53:25–31.

Centers for Disease Control and Prevention. (2004). CDC Wonder, Compressed Mortality data request source, 1979–1998, ICD-9, accessed March 10, 2004, http://wonder.cdc.gov/.

Doll R, Hill AB. (1950). Smoking and carcinoma of the lung. *British Medical Journal.* ii:739–48.

Doll R., Hill A.B. (1954). The mortality of doctors in relation to their smoking habits. iii:1451–55.

Doll R., Peto R., Boreham J., Sutherland I. (2004). Mortality in relation to smoking: 50 years' observations on male British doctors. *British Medical Journal.* 328:1519–28.

Friis R.H., Sellers T.A. (1998). *Epidemiology for Public Health Practice.* Aspen Publication, Gaithersburg, MD (second edition).

Lange J.H. (1991). Reanalysis of epidemiological data for selenium anticancer activity. *Toxicology and Industrial Health.* 7:319–25.

Lange J.H., LaPorte R.E., Talbott E.O., Chang Y.F. (2003a). Capture-recapture method: the gold standard for incidence and prevalence. (Letter.) *New Zealand Medical Journal.* 20:116.

Lange J.H., Mastrangelo G., Thomulka K.W. (2003b). Fireside Hypothesis—a comment on confounders. (Letter.) *Medical Hypotheses.* 60:776–77.

Lange J.H., Chang Y.F., LaPorte R.E., Mastrangelo G. (2003c). Hazardous waste site frequency: use of capture recapture. *Toxicology and Industrial Health.* 19:109–14.

Lange J.H., Mastrangelo G., Fedeli U., Fadda E., Rylander R., Lee E. (2003d). Endotoxin exposure and lung cancer mortality by type of farming: is there a hidden dose-response relationship? *Annuals of Agriculture and Environmental Medicine*. 10:229–32.

Lee P.N., Forney B., Fry J.S. (2001). Revisiting the association between environmental tobacco smoke exposure and lung cancer risks. III. Adjustment for the biasing effect of misclassification of smoking habits. *Indoor and Built Environment.* 10:384–98.

Lilienfeld D.E., Stolley P.D. (1994). *Foundations of Epidemiology.* Oxford Press, Oxford, UK.

Mastrangelo G., Marzia V., Milan G., Fadda E., Fedeli U., Lange J.H. (2004). An exposure-dependent reduction of lung cancer in dairy farmers. *Indoor and Built Environment.* 13:35–52.

Mayeux R., Marder K., Cote L.J., Denaro J., Hemenegildo N., Mejia H., Tang M-X, Lantigua R., Wilder D., Gurland B., Hauser A. (1995). The frequency of idiopathic Parkinson's disease by age, ethnic group and sex in northern Manhattan, 1988–1993. *American Journal of Epidemiology.* 142:820–27.

McMichael A.J. (1976). Standardized mortality ratios and the health worker effect: scratching beneath the surface. *Journal of Occupational Medicine.* 18:165–68.

Men T., Brennan P., Boffetta P., Zaridze D. (2003). Russian mortality trends for 1991–2001: analysis by cause and region. *British Medical Journal.* 327:964–66.

Ness R.B., Haggerty C.L., Harger G., Ferrell R. (2004). Differential distribution of allelic variants in cytokine genes among African Americans and White Americans. *American Journal of Epidemiology.* 160:1033–38.

Paneth N. (2004). Assessing the contributions of John Snow to epidemiology 150 years after removal of the Broad Street pump handle. *Epidemiology.* 15:514–16.

Price B., Ware A. (2004). Mesothelioma trends in the United States: an update based on surveillance, epidemiology, and the end results program data for 1973 through 2003. *American Journal of Epidemiology.* 159:107–12.

Stern F.B. (2003). Mortality among chrome leather tannery workers: an update. *American Journal of Industrial Medicine.* 44:197–206.

Royal Institute of Public Health. (2004). John Snow Society, accessed March 10, 2004, http://www.riph.org/johnsnow_news.html.

Timmreck T.C. (1998). *An Introduction to Epidemiology.* Jones and Bartlett Publishers, Boston, MA (second edition).

Waterhouse J.A.H. (1998). Epidemiology. In Pfafflin JR and Ziegler EN, eds., *Encyclopedia of Environmental Science and Engineering,* Gordon and Breach Science Publishers, Amsterdam (fourth edition), 381–96.

Zahm S.H., Blair A. (2001). Assessing the feasibility of epidemiological research on migrant and seasonal farm workers: an overview. *American Journal of Industrial Medicine.* 40:487–89.

J.H. LANGE
Envirosafe Training and Consultants

EUTROPHICATION

INTRODUCTION

For a considerable time, scientists have been aware of the natural aging of lakes, a process so slow that it was considered immeasurable within the lifetime of human beings. In recent years, however, that portion of the nutrient enrichment or eutrophication of these and other natural bodies of water contributed by man-made sources have become a matter of concern. Many bodies of water of late have exhibited biological nuisances such as dense algal and aquatic weed growths whereas in the past they supported only incidental populations of these plants.

Excessive nutrients are most often blamed in the scientific literature for the creation of the plant nuisances. Among the nutrients, dominant roles have been assigned by most researchers to nitrogen and phosphorus. These elements can be found in natural waters, in soils, in plants and animals, and in precipitation. Man-made sources for these nutrients are in domestic wastes and often in industrial wastes.

This chapter concerns itself with the nature of algae, the environmental factors affecting their growth, the nature of the entrophication problem (sources, relative quantities of nutrients contributed by these sources, threshold limits for the growth of aquatic plants), and various techniques for the removal of those nutrients usually associated with the eutrophication problem.

THE PHYSICAL NATURE OF ALGAE

Most bodies of water which can be considered eutrophic exhibit various predominant forms of algae at different times of the year. Algae that are important to investigators concerned with the eutrophication problem may be classified into four groups which exclude all but a few miscellaneous forms. The four groups are:

1) Blue-green algae (Myxophyceae)
2) Green algae (Chlorophyceae)
3) Diatoms (Bacillariophyceae)
4) Pigmented flagellates (Chrysophyceae, Euglenophyceae)

The basis for this classification is the color of the organism. Blue-green and green algae are self descriptive, whereas diatoms are brown or greenish-brown. Pigmented flagellates can be brown or green. They possess whip-like appendages called flagella, which permit them to move about in the water. It is not inferred by the above list that all algae are restricted to these colors. Rhodophyceae, for example, which are primarily marine algae, are brilliant red.

Aquatic biologists and phytologists do not agree on the number of divisions that should be established to identify algae. Some authorities use as many as nine divisions while others use seven, five and four. Nevertheless, the four divisions as suggested by Palmer will be used as they are adequate for the ensuing discussions.

BLUE-GREEN ALGAE

Blue-green algae as a group are most abundant in the early fall at a temperature range of 70 to 80°F. Data obtained from water sources in the southwestern and southcentral United States indicate that for this section of the country maximum growth occurs at the end of February and throughout much of April, May and June. When blue-green algae becomes predominant, it frequently indicates that the water has been enriched with organic matter, or that previously there had been a superabundance of diatoms. Blue-green algae are quite buoyant due to the oil globules and gas bubbles which they may contain. For this and other reasons they live near the surface of the water often producing offensive mats or blankets. Since these algae are never flagellated, they are not considered swimmers although a few, such as oscillatoria and spirulina, are able to creep or crawl by body movements. Some of the common blue-green algae are anabaena, aphanizomenon, rivularia, gomphosphaeria and desmonema.

GREEN ALGAE

Green algae are most abundant in mid-summer at a temperature range of 60 to 80°F. For water bodies in the southwestern and southcentral United States, maximum growth occurs during the first half of September with little variation throughout the remainder of the year. Like the blue-green algae, green algae usually contain oil globules and gas bubbles which contribute to the reasons why they are found near the surface of the water. Green algae are distinguished by their green color which comes from the presence of chlorophyll in their cells. Many of the green algae are flagellates and due to their swimming ability they are frequently found in rapidly moving streams. Some of the common green algae are chlorella, spirogyra, chlosterium, hydrodictyon, nitella, staurastrum and tribonema.

DIATOMS

Diatoms are usually most prevalent during the cooler months but thrive over the wide temperature range of from 35 to 75°F. For water bodies in the southwestern and southcentral United States, diatoms thrive best in May, September and October with the maximum growth observed in mid-October. It is generally recognized that many diatoms will continue to flourish during the winter months, often under the ice. The reason for the increase in growth twice a year is due to the spring and fall overturn, in which food in the form of carbon, nitrates, ammonia, silica and mineral matter, is brought to the surface where there is more oxygen and a greater intensity of light.

Diatoms live most abundantly near the surface, but unlike the buoyant green and blue-green algae, they may be found at almost any depth and even in the bottom mud. Diatoms may grow as a brownish coating on the stems and leaves of aquatic plants, and in some cases they grow along with or in direct association with other algae. In rapidly moving streams they may coat the bottom rocks and debris with a slimy brownish matrix which is extremely slippery. Lastly, diatoms are always single-celled and nonflagellated.

PIGMENTED FLAGELLATES

No classification of algae has caused more disagreement than that of the pigmented flagellates. The difficulty arises from the fact that hey possess the protozoan characteristics of being able to swim by means of flagella, and the algae characteristic of utilizing green chlorophyll in association with photosynthesis. Thus they could be listed either as swimming or flagellated algae in the plant kingdom, or as pigmented or photosynthetic protozoa in the animal kingdom. One of many attempts to resolve this problem has been the proposal to lump together all one-celled algae and all protozoa under the name "Protista." This method, however, has not met with general acceptance. For the sanitary engineer the motility of the organism is of lesser importance than its ability to produce oxygen. The pigmentation characteristic associated with green chlorophyll and oxygen production is sufficient criteria for separating these organisms into a class by themselves. Thus a distinction is made between pigmented flagellates (algae) and nonpigmented flagellates (protozoa).

Pigmented flagellates are more abundant in the spring than at any time of the year although there is generally considerable variation among the individual species. Apparently flagellates are dependent on more than temperature. They are found at all depths, but usually are more prevalent below the surface of the water than at the surface.

For present purposes pigmented flagellates can be divided into two groups: euglenophyceae which are grass-green in color and chrysophyceae which are golden-brown.

Euglenophyceae are usually found in small pools rich in organic matter, whereas chrysophyceae are usually found in waters that are reasonably pure.

Some of the more common pigmented flagellates are euglena, ceratium, mallomonas, chlamydomonas, cryptomonas, glenodinium, peridinium, synura and volvox.

MOTILITY

Of additional value in the classification of algae are their means of motility. Three categories have been established, namely:

Nekton—algae that move by means of flagella.
Plankton—algae that have no means of motility.
Benthic algae—algae that attach themselves to a fixed object.

NEKTON

Nekton are the most active algae and are often referred to as "swimmers." Due to their activity they use more energy and in turn release more oxygen during the daylight hours. Their cells are supplied with one, two, or more flagella which extend outward from the front, side or back of the cell. These flagella enable the organisms to move about freely in the aquatic environment and to seek food which, in the case of turbulent water, is constantly changing in location.

In general nekton have the most complex structure of the three categories and come nearest to being simple animals. Nekton are the predominant algae found in swiftly moving rivers and streams.

According to Lackey, results of tests performed on waters of the Ohio River show that certain nekton are the only algae that provide reliable clear-cut responses to the presence of pollution and thus are true indicator organisms. Five flagellates have been singled out on the genus level as being common and easily recognized. They are (1) cryptomonas, (2) mallomonas, (3) synura, (4) uroglenopsis, and (5) dinobryon. Dinobryon is perhaps the most easily recognized due to its unique shape which resembles a shaft of wheat. Samples taken from several rivers indicate that these algae react adversely to the presence of sewage and are found in abundance only in clean water. Unfortunately not all experts agree on what constitutes clean water and what algae serve as indicator organisms. Patrick states that the "healthy" portion of a stream contains primarily diatoms and green algae. Rafter states that the absence of large amounts of blue-green algae is an indicator of clean water. Palmer lists 46 species which have been selected as being representative of "clean-water algae," and these consist of diatoms, flagellates, green algae, blue-green algae and red algae. In addition Palmer lists

47 species of algae condensed from a list of 500 prepared from reports of more than 50 workers, as being representative of "polluted-water algae." These consist of blue-green algae, green algae, diatoms and flagellates.

PLANKTON

Plankton are free-floating algae which are most commonly found in lakes and ponds, although they are by no means limited to these waters. Most species are unicellular; however, they tend to become colonial when their numbers increase, as in the formation of a heavy concentrated growth known as a "bloom." An arbitrary definition of a bloom is that concentration of plankton that equals or exceeds 500 individual organisms per ml. of raw water. Blooms usually show a predominance of blue-green algae although algae from other classes can also form blooms.

An algae bloom often becomes sufficiently dense as to be readily visible on or near the water surface, and its presence usually indicates that a rich supply of nutrients is available. Other environmental factors may stimulate the formation of blooms, and a bloom of the same organism in two bodies of water may or may not result from identical favorable environmental conditions. These growths are extremely undesirable in bodies of water, in general, and in potential water supply sources in particular for the following reasons:

1) They are very unsightly.
2) They interfere with recreational pursuits.
3) When the water becomes, turbulent, fragments of the mat become detached and may enter a water treatment system clogging screens and filters.
4) When the algae die (as a result of seasonal changes or the use of algicides), decomposition occurs, resulting in foul tastes and odors.
5) They may act as a barrier to the penetration of oxygen into the water which may result in fish kills.
6) They may reduce the dissolved oxygen in the water through decay or respiration within the bloom.
7) Some blooms release toxic substances that are capable of killing fish and wild life.
8) They may cause discoloration of the water.
9) They attract waterfowl which contribute to the pollution of the water.

Some of the common blue-green algae that form blooms are anabaena, aphanizomenon, oscillatoria, chlorella and hydrodictyon. Synedra and cyclotella are common diatoms that form blooms and synura, euglena and chlamydomonas are common flagellates that form blooms. Filamentous green plankton, such as spirogyra, cladophora and zygnema form a dense floating mat or "blanket" on the surface when the density of the bloom becomes sufficient to reduce the intensity of solar light below the surface. Like blooms, these blankets are undesirable, and for the same reasons cited earlier. However, in addition, blankets also serve as a breeding place for gnats and midge flies, and after storms they may wash up on the shores where they become offensive. In many cases hydrogen sulfide and other gases which are able to spread disagreeable odors considerable distances through the air are liberated. In large amounts, hydrogen sulfide has been known to seriously discolor the paint on lakeside dwellings.

BENTHIC ALGAE

Benthic algae are those algae which grow in close association with a supply of food. That is, they seek out an aquatic environment where nutrients are adequate, then attach themselves to a convenient stationary object such as a submerged twig or rock. They may be found in quiet ponds and lakes or in fast-moving rivers and streams. In some cases they break away from their attachments and form unsightly surface mats, or they may re-attach themselves somewhere else. Chlamydomonas is such an organism, where in one growth phase it may be found attached to a fixed object, and in another phase it may be dispersed throughout the water.

Benthic algae include diatoms, blue-green algae, green algae and a few species of red fresh-water algae. None of the pigmented flagellates are benthic. Most attached algae grow as a cluster of branched or unbranched filaments or tubes and are fastened at one end to some object by means of an anchoring device. Others take the shape of a green felt-like mat (gomphonema), a thin green film or layer (phytoconis), or a soft fragile tube (tetraspora). Some of the most common benthic algae are cladophora, chara, nitella, ulothrix, cymbella, vaucheria and gomphonema.

ENVIRONMENTAL FACTORS AFFECTING GROWTH OF ALGAE

The effects of certain environmental factors on the growth of the aforementioned forms of algae have been fairly well defined. The most important parameters to be considered in the growth pattern are light intensity, temperature, pH and nutritional requirements.

LIGHT INTENSITY

Light is essential to all organisms which carry on photosynthesis; however, requirements or tolerance levels differ greatly with the organism. For example, terrestrial species of vaucheria grow equally well in fully-illuminated soil and densely-shaded soil, while a number of blue-green algae grow only in shaded habitats. In addition, some algae are unable to endure in the absence of sunlight caused by several consecutive cloudy days, whereas certain submerged algae are unable to withstand exposure to full sunlight. Thus, an algae kill may be noted during a drought where shallow water prevents depth-dwelling algae from escaping the intensity of the sunlight penetration. Often muddy rivers are virtually algae-free due to the lack of penetration of the sunlight, the Missouri and the Mississippi being two such rivers.

The distribution of algae at the various depths in a body of water is directly correlated with the intensity of illumination at the respective depths. This distribution would be difficult to express in general terms when dealing with algae on their species level. In addition, the depths at which these species would be found would change with such variables as growth phase of the organism, temperature and the absorptive and reflective characteristics of the water. It can be stated, however, that certain fresh-water red algae and blue-green algae are found only at considerable depths and that some diatoms exist in the bottom mud. In the most general terms it can also be stated that algae are found at all levels, but most commonly near the surface.

The vertical distribution may also be related to the division of light rays into various spectral colors. This division varies with the concentration of dissolved color material, plankton and particulate matter, with the seasons, and with the depth. In colored water the violet-blue end of the spectrum is absorbed more readily. As depth increases light rays divide differently with greater absorption occurring at the red end of the spectrum.

The depth to which light penetrates has a direct influence on photosynthetic activity. The seasonal variation in this light and the resulting availability of certain dominant wave lengths may be the reason for fluctuations in the composition of the algal population from spring to fall. Much more work is needed in this area.

TEMPERATURE

In general, temperature is not the key factor in determining the nature of the algal flora. Most species are able to grow and reproduce if other environmental conditions are favorable. According to Patrick, however, the above statement is not true in the case of diatoms, where temperature changes are more important than any other environmental factor in influencing their rate of growth. Additional work in this area by Cairns indicates that certain diatoms grow best only at a specific temperature, and that at some temperatures they will not grow at all.

Most algae are not affected by minor changes in pH brought about by the seasonal variations, growths of carbon-dioxide producing organisms, etc. Large changes such as would be caused by the introduction of industrial wastes or acid mine waters, will greatly affect algae, usually causing a decrease in population.

The majority of algae thrive when the pH is near 7.0. Some blue-green algae prefer high pHs. Anacystis and coccochloric are found at about pH 10.0 with little or no growth below pH 8.0. Other algae such as eugleny mutabilis, cryptomonas erosa and ulothrix zonata prefer low pHs.

NUTRITIONAL REQUIREMENTS AND TOXIC ELEMENTS FOR ALGAE

Calcium

Calcium is not an essential element for most algae, although some cannot develop without it.

Calcium and Magnesium

As bicarbonates they are a supplemental supply of carbon dioxide for photosynthesis. This accounts for the greater abundance of algae in hard-water lakes than in soft-water lakes.

Iron

Most algae grow best when the ferric oxide content of the water is between 0.2 to 2.0 mg per liter. Above 5 mg per liter there is a toxic effect unless it is overcome by the buffering action of organic compounds or calcium salts. Certain diatoms (eunotia and pinnularia) are found in iron-rich water. Effluent from steel mills may be toxic to most algae if the resulting iron concentration exceeds the toxic limitation.

Copper

Copper is extremely toxic to algae in the range of 0.1 to 3.0 ppm as copper sulfate; the sulfate form being used as an algicide. Some algae are able to tolerate large amounts of copper ion and are considered copper-sulfate resistant. Protococcus, for example, is not destroyed by 10 ppm of copper sulfate.

Phenol

At a concentration of up to 1.9 mg per liter, phenol apparently has no toxic effect on diatoms.

Nitrates, Phosphates and Ammonia

These are essential food elements necessary for growth. Nitrogen may be obtained from nitrates, nitrites or simple ammonia compounds. The primary source of these nutrients is from sewage treatment plant effluents, although nitrogen may be derived from the atmosphere, land runoff, etc. (See section on EUTROPHICATION.) In general as little as 0.3 to 0.015 ppm of nitrates and phosphates will produce blooms of certain species of algae, other conditions being favorable.

Oil

Streams polluted with oil are usually low in algae. One variety of diatoms may be dominant in such waters.

Salinity

Increases in salinity up to about one percent do not affect the algae population. Significant increases, such as caused by salt-brine wastes, may destroy most of the algae present, however. Certain fresh-water algae may become adapted to water with slowly increasing salinity.

Hydrogen Sulfide

At a concentration of 3.9 ppm, hydrogen sulfide is toxic to most diatoms. Some resistant species are achnanthese affinis, cymbella ventricosa, hantzschia amphioxys and nitzschia palea.

Silica

Silica is necessary for the growth of diatoms whose cell wall is composed of silica. Presently no limits have been determined (to the author's knowledge).

Vitamins

Several vitamins in small quantities are a requisite to growth in certain species of algae. Chief among these vitamins are vitamin B-12, thiamine and biotin. These vitamins are supplied by bottom deposits, soil runoff and by the metabolites produced by other organisms.

Micronutrients

Substances such as manganese, zinc, molybdenum, vanadium, boron, chlorine, cobalt, etc. are generally present in water in the small concentrations sufficient for plant growth.

Carbon Dioxide

Carbon dioxide is necessary for respiration. If it is deficient, algae may remove carbon dioxide from the atmosphere.

Chlorine

Chlorine is toxic to most algae and is used as an algicide in the range of 0.3 to 3.0 ppm. It is used as an algicide in the treatment plant and distribution system. Some algae, cosmarium for example, are resistant to chlorine. Protococcus, which is resistant to copper sulfate, is killed by 1 ppm of chlorine. Therefore, algae resistant to the copper ion may not be resistant to the chlorine ion and vice versa.

Calcium Hydroxide (lime)

An excess of lime in the water, as may be introduced during pH adjustment for coagulation, results in the death of certain algae. Five ppm of lime with an exposure of 48 hours has been lethal to melosira, nitzschia and certain protozoa and crustacea.

THE EUTROPHICATION PROBLEM

Of the factors previously discussed which promote the growth of algae, that factor which man has altered is the nutrient concentration in may of the natural waterways.

In simplest terms, eutrophication is the enrichment of waters by nutrients from natural or man-made sources. Of the many nutrients which are added to the waters by man-made sources, nitrogen and phosphorous are most often cited by researchers as being the key nutrients responsible for the promotion of algae growth. In nearly all cases when the nitrogen and phosphorus level of a body of water increases, there will be a corresponding increase in the growth of algae and aquatic plants. Such growth greatly speeds up the aging process whereby organic matter invades and gradually displaces the water until eventually a swamp or marsh is formed. Unfortunately, the process of eutrophication is often difficult to reverse in bodies of water such as large lakes where the flushing or replacement time for the waters can be in the order of years.

The following sections provide the relative magnitude of natural and man-made sources of nutrient material associated with plant growth.

SOURCES OF NUTRIENTS

While it is recognized that certain algae require a number of chemical elements for growth, it is also known that algae can absorb essential as well as superfluous or even toxic elements. Although every essential element must be present in algae, this does not mean that every element is essential. On the other hand, the absence of certain nutrient elements will prevent growth. Nutrients may be classified as (1) "absolute nutrients," which are those which cannot be replaced by other nutrients, (2) "normal nutrients," which are the nutrients contained in the cell during active growth, and (3) "optimum maximum growth." It may also be well to assign a broad meaning to the word "nutrient" and define it as anything that can be used as a source of energy for the promotion of growth or for the repair of tissue.

In evaluating the effects of nutrients on algae, care must be exercised to consider the interaction between nutrients and other physical, chemical or biological conditions. Rapid growth of algae may be stimulated more by factors of sunlight, temperature, pH, etc., than by an abundance of nutrient material. Tests performed with nitzschia chlosterium, in order to study the interaction of environmental factors showed that two identical cultures of the organism, when supplied with a reduced nutrient level, had a lower optimum light intensity and optimum temperature for maximum growth. Thus light intensity and temperature data should accompany data on nutrient concentration and growth rate.

Of all the possible nutrients, only nitrogen and phosphorus have been studied in depth both in the field and in the laboratory. This is because of the relative difficulties associated with the study, analysis and measurement of trace elements, compounded by the minute impurities present in the regents and distilled water. In addition, nitrates and phosphates have a long history of use in agricultural fertilizers where determination of their properties have been essential to their economical use.

The following are the most common sources of nitrogen and phosphorus in bodies of water:

1) Rainfall—Based on experimental data, it has been found that rainwater contains between 0.16 and 1.06 ppm of nitrate nitrogen and between 0.04 and 1.70 ppm of ammonia nitrogen. Computations based on the nitrogen content of rainwater show that for Lake Mendota, Wisconsin, approximately 90,000 pounds of nitrogen are available each year as a result of rainfall. Thus it can be seen that rainfall plays a significant role in building up the nitrogen content of a lake or reservoir especially if the surface area is large.

An examination of the phosphorus content of rainwater of different countries shows that a number of concentrations may exist ranging from 0.10 ppm to as little as an unmeasurable trace, the latter reported in the Lake Superior region of the United States. In view of the wide variation in the determinations, little can be stated at present regarding the degree of phosphorus build-up in impoundments resulting from rainwater.

2) Groundwater—Studies conducted on sub-surface inflows to Green Lake, Washington, show that this water contains approximately 0.3 ppm of phosphorus. Other reports, however, claim that the amount of phosphorus in groundwater is negligible.

Investigations into the nitrate content of groundwater produced variable results; however, it can be stated that 1.0 ppm is a reasonable figure. The results of the above studies on both nitrogen and phosphorus can be summarized by stating that groundwater should not be discounted as a possible source of nutrients and that quantitative values should be obtained for the specific locality in question.

3) Urban Runoff—Urban runoff contains storm water drainage, overflow from private disposal systems, organic and inorganic debris from paved and grassed areas, fertilizers from lawns, leaves, etc. In view of the variable concentration of the above material, precise figures cannot be obtained on the phosphorus or nitrate content that would be meaningful for all areas.

Studies conducted in 1959 and 1960 by Sylvester on storm water from Seattle street gutters shows the following nutrients:

Organic nitrogen—up to 9.0 ppm

Nitrate nitrogen—up to 2.8 ppm

Phosphorus—up to 0.78 ppm soluble and up to 1.4 ppm total.

4) Rural Runoff—Rural runoff for the purposes of definition may be considered as runoff from sparsely-populated, wooded areas with little or no land devoted to agriculture. Investigations by Sylvester showed that the phosphorus content of drainage from three such areas in the state of Washington contained 0.74, 0.77 and 0.32 lb./acre/year, or a total concentration of 0.069 ppm. The corresponding nitrate nitrogen concentration and organic nitrogen concentration amounted to 0.130 and 0.074 ppm, respectively.

5) Agricultural Runoff—Agricultural runoff is one of the largest sources of enrichment material and may be derived from two sources—wastes from farm animals and the use of nitrogen and phosphorus-containing fertilizers.

Farm-animal wastes add both large quantities and high concentrations of nutrients to adjacent streams and rivers. The large concentrations are due primarily to the practice of herding animals in relatively confined areas. A comparison on the nutrient value of animal wastes and human wastes has been made in a study by the President's Science Advisory Committee. According to the findings, a cow generates the waste equivalent of 16.4 humans, a hog produces as much as 1.9 humans and a chicken produces as much as 0.14 humans.

The use of chemical fertilizers in the United States has grown almost 250% in the decade from 1953 to 1963. In 1964 the use of phosphorus-containing fertilizers and the use of nitrogen-containing fertilizers reached approximately 1.5 and 4.4 million tons, respectively, per year. Most all of this fertilizer is distributed to soil already high in natural-occurring nitrogen. When nitrogen fertilizer and natural soil nitrogen combine, there is a great increase in crop production, but also a greater opportunity for loss of this nitrogen in runoff. This loss will increase if the fertilizer is not properly applied, if it is not completely utilized by the crops, if the crops have a short growing season (the land being non-productive for a time), if the land is irrigated, and if the land is sloped.

The addition of nitrogen-bearing fertilizers also increase the quantity of mineral elements in the soil runoff which are necessary for the growth of aquatic plants and algae. When applied, the nitrogen in the fertilizer is converted into nitric acid which combined with the minerals in the soil, such as calcium and potassium, rendering them soluble and subject to leaching.

6) Industrial Wastes—The nutrient content of industrial waste effluents is variable and depends entirely upon the nature, location and size of the industry. In some cases the effluents are totally free of nitrogen and phosphorus.

The meat packing industry is one of the chief producers of nitrogen-bearing wastes. The greatest producer of phosphate-bearing wastes is most likely the phosphate-manufacturing industry itself. Most phosphate production in the United States is concentrated in Florida and as a result many severe localized problems of eutrophication have resulted in that state.

Fuel processing industries and petroleum refineries discharge vast quantities of nitrogen into the atmosphere both in the gaseous state and solid state as particulate matter. This nitrogen is then washed from the atmosphere by the rain and carried back to earth. In 1964, the 500 billion tons of coal used in the United States released about 7.5 million tons of nitrogen into the atmosphere, most of which has returned to be combined with the soil. This greatly exceeds the use of nitrogen in the form of fertilizers which, as previously stated, amounted to 4.4 million tons for that year. Thus, through the atmosphere we are bringing more nitrogen into the soil than we are taking out, and much of this excess ultimately gets washed out into our waterways.

7) Municipal Water Treatment—The water treatment plants themselves are to a degree responsible for adding to the eutrophication problem as approximately 33% of the municipal water in the United States is treated with compounds containing phosphorus or nitrogen. Some of the commonly used nutrient-bearing chemicals or compounds are ammonia (in the use of chloramines) organic polyelectrolytes, inorganic coagulant aids, sodium hexametaphosphate, sodium tripolyphosphate, and sodium pyrophosphate.
8) Waterfowl—It has been estimated that wild ducks contribute 12.8 pounds of total nitrogen/acre/year and 5.6 pounds of total phosphorus/acre/year to reservoirs or lakes. A number of studies have been conducted on waterfowl, but it may be concluded that, although there may be some bearing on localized eutrophication, in general the overall effect is negligible.
9) Domestic Sewage Effluent—Undoubtedly the greatest contributor toward the eutrophication of rivers and lakes is the discharge from sewage treatment plants. Conventionally treated domestic sewage usually contains from 15 to 35 ppm total nitrogen and from 6 to 12 ppm total phosphorus. In addition there are a large number of minerals present in sewage which serve as micro-nutrients for algae and aquatic plants.

Phosphorus in domestic sewage may be derived from human wastes, waste food (primarily from household garbage-disposal units), and synthetic detergents. Human wastes have been reported in domestic sewage at the rate of 1.4 pounds of phosphorus/capita/year. The largest source of phosphorus, however, is from synthetic detergents which amounts to approximately 2.1 pounds/capita/year. Sawyer indicates that detergent-based phosphorus represents between 50 and 75% of the total phosphorus in domestic sewage. It should be noted that both the use of household garbage-disposal units and detergents is fairly recent, and accordingly they may be considered as contributing strongly to the development of the recently magnified eutrophication problem.

Not all the phosphorus entering a sewage treatment plant will leave the plant since chemical removal does occur during the treatment process. Calcium and metallic salts in large concentrations form insoluble phosphates which are readily removed. Very often phosphate-precipitating agents are present in waters containing industrial wastes, and when these agents are received at the plant, removals in the neighborhood of 60% may be realized.

Nitrogen in domestic sewage is derived from human wastes and from waste food primarily from household garbage-disposal units. Human wastes, the major source of nitrogen, contributes an average of about 11 pounds of nitrogen/capita/year. Some reduction in the nitrogen also takes place during the treatment of the sewage. Many plants treat the sludge anerobically which permits significant release of the nitrogen. In general the removal amounts to between 20 and 50%. The higher percentage of removal occurs when fresh wastes are given complete treatment with no return of sludge nutrients to the effluent.

EUTROPHICATION STUDIES

In recent years a considerable number of studies have been made on eutrophication and related factors. Most of the studies can be grouped into the following categories:

1) nutrient content of runoff, rainwater, sewage effluent, bottom mud, etc.
2) nutrient analysis and physical distribution of nutrients in bodies of water before and/or after enrichment.
3) methemoglobinemia (illness in infants due to drinking high nitrate-content water)
4) toxicological and other effects on fish of high nitrate/high phosphate-content water
5) the chemical composition of plants in both eutrophied and non-eutrophied waters
6) the nutrient values of various fertilizers, manures and other fertilizing elements
7) the nutrient value of various soils
8) the effects of eutrophication on aquatic plants, animals and fish
9) studies on specific algae under either controlled laboratory conditions or in a particular body of water, using artificial or natural environmental conditions
10) methods for the removal or reduction of nitrogen and phosphorus
11) nutrient thresholds for growth of algae and aquatic weeds
12) the effects of eutrophication on the oxygen balance.

Of the above list, only studies conducted in the areas of (11) and (12) will be presented below. Work done in regard to (1) has already been presented. The removal or reduction of nitrogen and phosphorus (10) will be discussed separately as part of the subject matter in “CONTROL METHODS.”

NUTRITIONAL THRESHOLDS FOR THE GROWTH OF ALGAE

Studies conducted by Chu indicate that for growth on artificial media most planktonic algae flourish if the total nitrogen content ranges from 1.0 to 7.0 ppm and the total phosphorus content ranges from 0.1 to 2.0 ppm. If the nitrogen is reduced below 0.2 ppm and the phosphorus below 0.05 ppm, the growth of algae appears to be inhibited. The same inhibiting effect is created when the nitrogen or phosphorus content is raised above 20.0 ppm. The lower limit of the optimum range of nitrogen varies with the organism and with the type of nitrogen. For ammonia nitrogen the optimum range varies from 0.3 to 5.3 ppm and for nitrate nitrogen the optimum range falls between 0.3 and 0.9 ppm. Below these values the growth rate decreases as the concentration of nitrogen decreases.

Apparently the use of the various forms of nitrogen by algae is not constant throughout the year. Tests conducted at Sanctuary Lake in Pennsylvania (1965) indicate that the order of preference for the three forms of nitrogen—ammonia-nitrogen, nitrate-nitrogen, and nitrite-nitrogen—are defined by three seasonal periods, which are:

Spring	(1) Ammonia nitrogen
	(2) Nitrate nitrogen
	(3) Nitrite nitrogen
Midsummer	(1) Ammonia nitrogen
	(2) Nitrite nitrogen
	(3) Nitrate nitrogen
Fall	(1) Ammonia nitrogen
	(2) Nitrate nitrogen
	(3) Nitrite nitrogen

The amount of nitrogen in the aquatic environment is important to algae because it determines the amount of chlorophyll that may be formed. Too much nitrogen, however, inhibits the formation of chlorophyll and limits growth.

Laboratory studies on algae conducted by Gerloff indicate that of all the nutrients required by algae, only nitrogen, phosphorus and iron may be considered as limiting elements, and of these three, nitrogen exerts the maximum limiting influence. Approximately 5 mg of nitrogen and 0.08 mg of phosphorus were necessary for each 100 mg of algae produced. The corresponding nitrogen/phosphorus ratio is 60 to 1.

Hutchinson cites phosphorus as being the more important element since it is more likely to be deficient. When phosphorus enters a body of water, only about 10% is in the soluble form readily available for algal consumption. During midsummer total phosphate may increase greatly during the formation of algal blooms, while soluble phosphate is undetectable due to rapid absorption by the growing algae. Very often during warm weather these blooms are stimulated by the decomposition and release of soluble phosphates from the bottom sediments, deposited by the expired blooms of previous seasons. Thus when phosphates are added to a lake, only a portion of the phosphates are used in producing blooms. The blooms thrive and consume phosphates for only a short time, and a significant amount finds its way to the bottom sediments where it will be unavailable to further growth of aquatic vegetation.

Prescott examined a number of algae and concluded that most blue-green algae are highly proteinaceous. Aphanizomenon flos-aquae, for example, was shown to contain 62.8% protein. Green algae were found to be less proteinaceous. Spirogyra and cladophora, for example, contain 23.8 and 23.6% respectively. Thus it can be concluded that the nitrogen requirement (for the elaboration of proteins) depends on the class of algae, and that blue-green algae would require more nitrogen than green algae.

Provasoli examined 154 algal species to determine the requirements for organic micronutrients. He found that although 56 species required no vitamins, 90 species were unable to live without vitamins such as B_{12}, thiamin and biotin, either alone or in various combinations. He concluded that these vitamins are derived from soil runoff, bottom muds, fungi and bacterial production (B_{12}), and from a natural residual in the water.

Ketchum and Pirson conducted a series of examinations on the inorganic micronutrient requirements of algae and concluded that a number of elements are necessary for growth. No numerical values were assigned to the requirement levels. Those elements shown to be essential were C, H, O, P, H, S, Mg, Ca, Co, Fe, K and Mo. Those elements which may be essential (subject to further study) were Cu, An, B, Si, Va, Na, Sr, and Rb.

In summation, absolute values and nutrient thresholds cannot be set at this time because too little is known regarding the requirements of individual species. It might be stated in general terms, however, that nitrogen and phosphorus are two essential nutrient elements related to the production of blooms, and that if they are present in the neighborhood of 0.2 ppm and 0.05 ppm, respectively, algal growths will increase significantly.

NUTRITIONAL THRESHOLDS FOR THE GROWTH OF AQUATIC PLANTS

Studies conducted by Harper and Daniel indicate that submerged aquative plants contain 12% dry matter of which 1.8% are nitrogen compounds and 0.18% are phosphorus compounds. Hoagland indicates that when the nitrate content of water is high, nitrates may be stored in aquatic plants to be reduced to the usable ammonia nitrogen form as required. Subsequent investigations show that ammonia nitrogen can be substituted for nitrate nitrogen and used directly. Light apparently is not a necessary factor in the reduction of the nitrogen.

Muller conducted a number of experiments on both algae and submerged aquatic plants, and concludes that excessive growths of plants and algae can be avoided in enriched waters if the concentration of nitrate nitrogen is kept below 0.3 ppm, and if the concentration of total nitrogen remains below 0.6 ppm.

OXYGEN BALANCE

Recently, attention has been given to the effect of the intense growths of algae on the oxygen balance of natural waterways. It has been established that the dissolved oxygen concentrations may exhibit wide variation throughout the course of the day. This variation is attributed to the ability of algae to produce oxygen during the daylight hours, whereas they require oxygen for their metabolic processes during the hours of darkness.

In addition, since algae are organic in nature, they exert a biochemical oxygen demand (BOD) on the stream oxygen resources as does other materials which are organic.

Extensive tests were run on the Fox River in Wisconsin by Wisniewski in 1955 and 1956 to examine the influence of algae on the purification capacity on rivers. In the most general terms, the studies indicate that algae increase the B.O.D. by adding organic matter capable of aerobic bacterial decomposition and by the respiration of the live cells which utilize oxygen during the absence of light. In the presence of light, algae produce oxygen and as a result may cause a "negative" B.O.D. for a production of oxygen in excess of that required for the normal B.O.D. requirements or aerobic bacterial stabilization. In addition to the above, the following specific conclusions were drawn from the tests:

1) The oxidation rate resulting from the respiration of live algae was much lower than that obtained by the biological oxidation of the dead algae.
2) The ultimate B.O.D. of live algae was practically the same as for dead algae.
3) A linear relationship was found to exist between the five-day B.O.D. of suspended matter and volatile suspended solids concentration.
4) The B.O.D. increases with increases in suspended solids, the latter consisting largely of algae.

Additional work was done in this area and reported in 1965 by O'Connell and Thomas. They note that the oxygen produced by photosynthetic plants is affected greatly by changes in the availability of light due to cloud cover, turbidity in the water, etc., and therefore it may be too variable to be used as a reliable factor in evaluating the oxygen resources of a river. Another variable may be the loss of oxygen to the atmosphere during the daylight hours, caused by excess oxygen production and localized supersaturation.

An important consideration is the type of photosynthetic plants which are prevalent in a river. According to the above authors, if benthic algae and/or rooted aquatic plants are predominant (in lieu of phytoplankton), there will be little beneficial effect on the oxygen balance. In addition nighttime absorption of oxygen through respiration can seriously reduce daily minimum concentrations of dissolved oxygen. Determination of the effects of the benthic algae oscillatoria along a fivemile stretch of the Truckee River in Nevada indicated that on the average of the organism produced 72.5 pounds/acre/day of oxygen through photosynthesis. Oxygen uptake for these same organisms amounted to an average of 65.4 pounds/acre/day.

An examination of the oxygen profiles indicated that the oxygen variation throughout the day ranged from 2 (at night) to 13 (during daylight) parts per million.

It is dissolved oxygen variations such as the above which has been responsible for the disappearance of high quality game fish in many of our natural waterways.

CONTROL METHODS TO PREVENT EUTROPHICATION

There are a number of methods which attempt to limit the amounts of nutrients in bodies of water once the point of eutrophy has been reached. Some of these include dredging and removing bottom sediments with an inert liner, harvesting the algae, fish, aquatic weeds, etc., and diluting the standing water with a water of lower nutrient concentration. Although these methods may have their proper application, if eutrophication is to be decelerated, nutrient removal must start before wastes are permitted to enter the receiving waters.

Regarding the specific nutrients necessary to be removed, most researchers have placed the blame of eutrophication in waters to the inorganic forms of phosphorus and nitrogen. A smaller number of researchers are claiming that the algae–bacteria symbiosis relationship might be responsible for the rapid growth of blooms and that the amount of algae present in natural waters is in direct balance with the amount of carbon dioxide and/or bicarbonate ions in the waters. They further argue that an external supply of the above elements is necessary for the growth of algae populations. Since neither theory has been proved conclusively to date, the control methods given will be for the removal of nitrogen and phosphorus since it is these nutrients which most researchers lay to the blame of eutrophication and which have been therefore subsequently studied in detail.

NITROGEN REMOVAL

Land Application

It has been found that nitrogen-bearing waters, when percolated through soil are subjected to physical adsorption and biological action which removes the nitrogen in the ammonium form. It appears, however, that the nitrate form of nitrogen remains unaffected. At present this process is only at the theoretical stage, and to the author's knowledge no full-scale application has been attempted. Considerable land area would be involved which may prove a deterrent.

Anaerobic Denitrification

In this process, the nitrate present in sewage is reduced by denitrifying bacteria to nitrogen and nitrous oxide gases which are allowed to escape into the atmosphere. In order to satisfy the growth and energy requirements of the bacteria, methanol in excess of 25 to 35% must be added as a source of carbon.

The removal efficiency ranges from 60 to 95%. The major advantage to anaerobic denitrification is that there are no waste products requiring disposal. This process is still primarily in the experimental stage at this date.

Ammonia Stripping

Ammonia stripping is an aeration process modified by first raising the pH of the wastewater above 10.0. At this pH the ammonia nitrogen present is readily liberated as a gas and is absorbed into the atmosphere. Aeration is usually accomplished in a packed tray tower through which air is blown.

This process is suited to raw sewage where most of the nitrogen is either in the ammonia form or may be readily converted to that form. In secondary treatment processes the conversion of ammonia nitrogen to nitrate nitrogen can be retarded by maintaining a high organic loading rate on the secondary process.

Efficiency of nitrogen removal by ammonia stripping is excellent with 80 to 98% reported. There is also the advantage that there are no waste materials which must be disposed of.

PHOSPHORUS REMOVAL

Chemical Precipitation

Precipitation of phosphorus in wastewater may be accomplished by the addition of such coagulants as lime, alum, ferric salts and polyelectrolytes either in the primary or secondary state of treatment, or as a separate operation in tertiary treatment. In general, large doses in the order of 200 to 400 ppm of coagulant are required. However, subsequent coagulation and sedimentation may reduce total phosphates to as low as 0.5 ppm, as in the case of lime. Doses of alum of about 100 to 200 ppm have reportedly reduced orthophosphates to less than 1.0 ppm.

Phosphorus removal by chemical coagulation generally is efficient with removals in the order of 90 to 95% reported. Additional benefits are gained in the process by a reduction in B.O.D. to a value of less than 1.0 ppm. Both installation and chemical costs are high, however, and the sludges produced are both voluminous and difficult to dewater.

Sorption

Sorption is the process of passing wastewater downward through a column of activated alumina whereby the common form of phosphate are removed by ionic attraction. Regeneration of the media is accomplished by backwashing with sodium hydroxide followed by acidification with nitric acid.

Contrary to alum treatment, this process has the advantage in that sulfate ions are removed and thus the sulfate concentration is not increased. Since no salts are added, the pH and the calcium ion concentration remain unchanged. The process is efficient with more than 99% removal reported. The process should be limited to wastewater with a moderate amount of solids so as not to clog the media.

REMOVAL OF NITROGEN AND PHOSPHORUS

Biological (secondary) Treatment

In the secondary method of sewage treatment, bacteria utilize soluble organic materials and transform them into more stable and products. In the process nitrogen and phosphorus are removed from the wastes, utilized to build new cellular materials, and the excess is stored within the cell for future use. For each pound of new cellular material produced, assuming the material to be in the form of $C_5H_7NO_2$, about 0.13 pounds of nitrogen and about 0.026 pounds of phosphorus would be removed from the sewage. In the actual operation of this process not all of this nitrogen is removed unless additional energy material in the form of carbohydrates is added. Although it may be possible to eliminate all the nitrogen, a considerable amount of soluble phosphorus may remain, possibly because of the high ratio of phosphorus to nitrogen in sewage, attributable to synthetic detergents. Much of this phosphorus can be removed by absorption on activated sludge floc when it is later separated and removed.

This process offers a 30 to 50% removal of nitrogen and about a 20 to 40% removal of phosphorus without the special addition of carbohydrates.

Reverse Osmosis

The process of reverse osmosis consists of passing wastewater, under pressures as high as 750 psi, through a cellulose acetate membrane. The result is the separation of water and all ions dissolved therein. In actual practice the process has been plagued with difficulties primarily due to membrane fouling or premature failure of the membrane. In addition some nitrate and phosphate ions escape through the membrane.

Removal efficiency ranges from between 65 to 95% (for nitrogen).

Electrodialysis

Like reverse osmosis, electrodialysis is a non-selective demineralization process which removes all ions which would include the nitrate and phosphate ions. Essentially an electric current is used in conjunction with a membrane inserted in the line of current flow to separate the cations and anions.

The problems that have developed in the operation of this process include membrane clogging and precipitation of low-solubility salts of the membrane. Acidification of the water and removal of some of the solids prior to treatment has been effective in minimizing these problems, although it adds to the cost.

Removal efficiency ranges from between 30 to 50% (for nitrogen).

Ion Exchange

In the ion exchange process wastewater is passed through a media bed which removes both anionic phosphorus and anionic nitrogen ions and replaces them with another ion from the media. Regeneration of ion exchangers is commonly accomplished with inexpensive sodium chloride, and frequently the salt is salvaged by recycling the backwash water.

Difficulties in the process may be caused by fouling of the exchange resin due to organic material and reduction in the exchange capacity due to sulfates and other ions. The former may be reduced by removing the organic matter from the resin with sodium hydroxide, hydrochloric acid, methanol and bentonite.

The efficiency and cost of nitrogen and phosphorus removal by ion exchange depends largely on the degree of pretreatment and/or the quality of the water to be treated. Removal of nitrogen ranges from between 80 to 92%.

A number of ions exchange resins are available for nitrogen removal alone. These include zeolites, strong base anion resins (Amberlite IRA-410) and nuclear sulfonic cation resins (Nalcite HCR and Amberlite IR-120).

Algae Harvesting

A two-phase process which involves (1) growing algae in special shallow wastewater ponds where they feed on the absorb nutrients, and (2) removing the algae which then contain the nutrients within their systems. Algal predominance will depend upon the type of nutrients available and the concentrations. Frequently the flagellates euglena and chlorella will predominate where the nutrient concentration is high, and filamentous green algae such as spirogyra, vaucheria and ulothrix will predominate where the nutrient concentration is more moderate. The most desirable algae would be those that are large, those that would grow rapidly and those that would require vast quantities of food for energy, such as the swimming algae. In addition to removing nutrients, algae produce oxygen which reduces the B.O.D., and certain flagellates which ingest inorganic solids, are able to stabilize some of the organic material.

One of the major difficulties experienced with this process involves the harvesting procedure. A number of methods have been tried which include screening, settling, centrifuging and chemical screening. All have been found to present some form of difficulty, although it appears from the standpoint of performance and economy, the screening method may be the least unsatisfactory. Another problem is that complete nitrogen removal is seldom achieved unless a carbon source, such as carbon dioxide or methanol, is supplied. Still another problem is the need for disposing of a huge, sloppy mass of slimy and odoriferous dead algae.

Pilot studies show that a high-rate continuous-flow process is feasible when light is not limiting, and that orthophosphate concentrations can be reduced 90% to less than 1.0 ppm within 6 to 12 hours. Nitrogen removal is variable with estimates ranging from 40 to 90% efficiency depending upon the feed rate, pond design, and climatic conditions. The major drawbacks to this process aside from those mentioned, are the large land requirements needed and the necessity to rely on climatic conditions. In the latter case, artificial illumination may prove of value depending on power rates in the area.

CONCLUSIONS AND OBSERVATIONS

It might be concluded after examining a vast array of material that the bio-physical and bio-chemical factors which affect algae are extremely complex, and it is difficult to predict, with exacting certainty, future events relating to algal growth. The complexity is greatly magnified by the interaction which apparently exists between the numerous factors themselves. When these factors are examined and evaluated, the conclusions reached by one observer are not always in complete agreement, or may disagree entirely, with the conclusions reached by another observer. Commonly, specific organisms may produce different reactions because of their phase of life, seasonal changes, or because of other complex and little understood metabolic functions.

In few areas is there less accord than in the literature written about algae. (See references—The Physical Nature of Algae, #1 to 33, at the end of this article.) Authorities vary sharply in their opinions as to classification, physical descriptions, toxicity thresholds, etc. When this disagreement is added to the previously expressed uncertainties, it can be seen that our current knowledge is subject to various interpretations. Thus it is evident that there is a need for more work; work to develop new and useful information which will receive universal acceptance and which will clarify and expand our present knowledge. Conferences dealing with unique and local problems involving algae and eutrophication are ongoing. (See References—Eutrophication, #53, 55, 56 and 57, at the end of this article.)

ACKNOWLEDGMENTS

My sincere appreciation is extended to Mr. Robert G. Wieland who helped in the preparation of the manuscript and to the Research Foundation of the Newark College of Engineering for their aid in the typing of the manuscript.

REFERENCES

THE PHYSICAL NATURE OF ALGAE

1. Palmer, Mervin C., *Algae in Water Supplies,* Robert, A., Taft Sanitary Engineering Center, Cincinnati, Ohio, Public Health Service Pub. No. 657, 1959, p. 8.
2. Babbitt, H.E. and J.J. Doland, *Water Supply Engineering,* McGraw-Hill, N.Y., 1955, p. 457.
3. Kudo, R.R., *Protozoology,* Chas. C. Thomas Publisher, 3rd Ed., 1950.
4. Palmer, Mervin C. and W.M. Ingram, Suggested classification of algae and protozoa in sanitary science, *Sewage and Industrial Wastes,* **27,** 1955, pp. 1183–1188.
5. Silvey, J.K.G. and A.W. Roach, Studies on microbiotic cycles in surface waters, *Jour. of the Amer. Water Works Assn.,* Jan., 1964, p. 61.
6. Lackey, James B., Two groups of flagellated algae serving as indicators of clean water, *Jour, of the Amer. Water Works Assn.,* June, 1941, p. 1000.
7. Patrick, Ruth. A proposed biological measure of stream conditions, Proc. of the 5th Industrial Waste Conference, Purdue, Univ. Engineering Bull. No. 34, 1950, pp. 379–399.
8. Rafter, G.W., *The Microscopial Examination of Potable Water,* Van Nostrand Co., N.Y., 1960.
9. See Reference 1. above, pp. 41 and 42.
10. See Reference 1. above, p. 38.
11. Lackey, J.B., Plankton as related to nuisance conditions in surface waters, Amer. Assn, for the Advancement of Science, 1949, pp, 56–63.
12. Krauss, Robert W., Fundamental characteristics of algal physiology, Seminar on Algae and Metropolitan Wastes, Robert, A., Taft Inst., Cincinnati, Ohio, Apr., 27–29, 1960, p. 41.

13. Ingram, W.M. and G.W. Prescott, Toxic fresh-water algae, *Amer. Midland Naturalist,* **52,** 1954, pp. 75–87.
14. Derby, R.L. and D.W. Graham, Control of aquatic growths in reservoirs by copper sulfate and secondary effects of such treatment, *Proc. Amer. Soc. of Civil Engineers,* Vol. 79, 1953.
15. Borchardt, J.A., Eutrophication causes and effects, *Jour. of the Amer. water Works Assn.,* June 1969, pp. 272–275.
16. Thomas, S. and J.W. Burger, Chlamydomonas in storage reservoirs, *Jour. of the Amer. Water Works Assn.,* July, 1933.
17. Smith, Gilbert M., *The Fresh-Water Algae of the United States,* McGraw-Hill, N.Y., 1933, p. 13.
18. Mackenthum, K.M. and W.M. Ingram, *Limnological Aspects of Recreational Lakes,* U.S. Public Health Service Pub. No. 1167, 1964, pp. 18–19.
19. Lauff, George H., The role of limnological factors in the availability of algae nutrients, Contrib. No. 22, Univ. of Georgia Marine Inst., Sapelo Is., Ga.
20. Patrick, Ruth, Factors effecting the distribution of diatoms, *Botanical Review* No. 14, 1948, pp. 473–524.
21. Cairns, J., Jr., Effects of increased temperatures on aquatic organisms, *Industrial Wastes,* **1,** 1956, pp. 150–152.
22. Gerloff, G.C. *et al.,* The mineral nutrition of coccochlores peniocystis, *Amer. Jour. of Botany,* **37,** 1950, pp. 835–840.
23. Lackey, J.B., Aquatic life in waters polluted by acid mine waste, U.S. Public Health Service Report No. 54, 1939, pp. 740–746.
24. Claassen, P.W., Biological studies of polluted areas in the Genesee river system, N.Y. State Dept. of Conservation, Supplement to the 16th Annual Report for 1926, 1927.
25. *Water Quality and Treatment,* Amer. Water Works Assn., N.Y., 1951, p. 119.
26. See Reference 1, above, p. 57.
27. Lackey, J.B., Algal density as related to nutritional thresholds, Seminar on Algae and Metropolitan Wastes, Robert, A., Taft Inst., Cincinnati Ohio, Apr., 27–29, 1960, p. 59.
28. See Reference 17, above, p. 15 (see reference to Moore, 1917, and Moore and Carter, 1923).
29. See Reference 17, above, p. 57.
30. Provasoli, Luigi, Micronutrients and heterotrophy as possible factors in bloom production in natural waters, Seminar on Algae and Metropolitan Wastes, Robert, A., Taft Inst., Cincinnati, Ohio, Apr., 27–29, 1960, p. 48.
31. See Reference 25, above, pp. 112–113.
32. See Reference 1, above, p. 65.
33. See Reference 25, above, p. 111.

EUTROPHICATION

1. Fruh, Gus E., Biological responses to nutrients—eutrophication: Problems in fresh water, Water Resources Symposium No. 1, Univ. of Texas, Austin, 1968.
2. Ketchum, B.H., Mineral nutrition of phytoplankton, *Annual Review of Plant Physiology,* No. 5, 1954, p. 55.
3. Maddux, W.S. and R.F. Jones, Some interactions of temperature, light, intensity and nutrient concentration during the continuous culture of Nitzschia Closterium and Tetraselmis sp., *Jour. of Limnology and Oceanography,* Vol. 9, 1964.
4. Carrol, D., Rainwater as a chemical agent of geological progress—A review, Geological Survey Water—Supply 1535-G, 1962.
5. Fruh, Gus E. and G.F. Lee, The effect of eutrophication upon the water resources management of the Yahara river basin, report of the Water Chemistry Program, Univ. of Wisconsin, Madison, Wisc.
6. Putnam, H.D. and T.A. Olson, an investigation of nutrients in western lake Superior, School of Public Health Report, Univ. of Minnesota, 1960.
7. Sylvester, R.O. and G.C. Anderson, A lake's response to its environment, American Society of Civil Engineers, Sanitary Engineering Div., **90,** 1964.
8. Sawyer, C.N., Causes, effects and control of aquatic growths, *Jour. of the Water Pollution Control Fed.,* **34,** 1962, 279.
9. Reference 1, above, p, 51.
10. Sylvester, R.O., An engineering and ecological study for the rehabilitation of Green Lake, Univ. of Washington, Seattle, Wash., 1960.
11. Restoring the quality of the environment, report of the Environmental Pollution Panel, President's Science Advisory Committee, The White House, Wash., D.C., 1965.
12. Sauchelli, V., *Chemistry and Technology of Fertilizers,* ACS Monograph Series No. 148, Reinhold Pub. Co., N.Y., 1960.
13. Sources of nitrogen and phosphorus in water supplies, Task Group Report, *Jour. of the Amer. Water Works Assn.,* Mar. 1967, p. 347.
14. Andrews, W.R., *The Response of Crops and Soils to Fertilizers and Manures,* State College, Mississippi, 1947.
15. Reference 13, above, p. 349.
16. Reference 13, above, p. 350.
17. Paloumpis, A.A. and W.C. Starrett, An ecological study of benthic organisms in three Illinois river flood plain lakes, *Amer. Midland Naturalist,* Vol. 64, 1906.
18. Hume, N.B. and C.E. Gunnerson, characteristics and effects of hyperion effluent, *Jour. of the Water Pollution Control Fed.,* **34,** 1962, p. 15.
19. Hawk, P.B. *et al.,* Practical Physiological Chemistry, 13th Ed., The Blakiston Co., Phila., Pa., 1955.
20. Sawyer, C.N., Problems of phosphorus in water supplies, *Jour. of the Amer. Water Works Assn.,* Nov., 1965, p. 1431.
21. Hurwitz, E. *et al.,* Phosphates—Their fate in a sewage treatment plant waterway system, *Water and Sewage Works,* **112,** 1965, p. 84.
22. Standard practice in separate sludge digestion—committee report, *Proc. of the Amer. Soc. of Civil Engineers,* **63,** 1937, p. 39.
23. Reference 13, above, p. 348.
24. Mackenthun, Kenneth M., *Nitrogen and Phosphorus in Water, An Annotated Selected Bibliography of Their Biological Effects,* U.S. Public Health Service, 1965.
25. Chu, S.P., The influence of the mineral composition of the medium on the growth of planktonic algae—Part I, *Jour. of Ecology,* **30,** No. 2, 1942, pp. 284–325.
26. Chu, S.P., The influence of the mineral composition of the medium on the growth of planktonic algae—Part II, *Jour. of Ecology,* **31,** No. 2, 1943, pp. 109–148.
27. Dugdale, V.A. and R.C. Dugdale, Nitrogen metabolism in lakes III; Tracer studies of the assimilation of inorganic nitrogen sources, *Limnology and Oceanography,* **10,** No. 1, pp. 53–57.
28. Gerloff, G.C. and F. Skoog, Nitrogen as a limiting factor for the growth of microcystis aereginosa in southern Wisconsin Lakes, *Jour. of Ecology,* **38,** No. 4, 1957, pp. 556–561.
29. Hutchinson, G.I., *A Treatise on Limnology,* John Wiley, N.Y., 1957.
30. Prescott, G.W., Biological disturbances resulting from algal populations in standing waters, the ecology of algae, Spec. Pub. No. 2, Pymatuning Laboratory of Field Biology, Univ. of Pittsburgh, 1959, pp. 22–37.
31. Provasoli, Luigi, Micronutrients and heterotrophy as possible factors in bloom production in natural waters, algae and metropolitan wastes, Trans. of the 1960 Seminar, Taft Sanitary Engineering Center, Cincinnati, Ohio.
32. Reference 2, above, pp. 55–74.
33. Harper, H.J. and H.R. Danial, Chemical composition of certain aquatic plants, *Botanical Gazette,* **96,** 1939, p. 186.
34. Hoagland, D.R., Lectures on the inorganic nutrition of plants, Chronica Botanica Co., Waltham, Mass., 1944.
35. Muller, W., Nitrogen content and pollution of streams, Gesundheitsing., *Water Pollution Abstracts,* **28,** No. 2, Abs. No. 454, **74,** p. 256.
36. Wisniewski, Theodore F., Algae and their effects on D.O. and B.O.D., Parts I–III, water and sewage works, June–Aug., 1958.
37. O'Connell, R.L. and N.A. Thomas, Effects of benthic algae on stream dissolved oxygen, *Jour. of the Sanitary Engineering Div.—Amer. Society of Civil Engineers,* June, 1965, pp. 1–16.
38. Lange, Willy, Effect of carbohydrates on the symbiotic growth of planktonic blue-green algae with bacteria, *Nature,* Sept., 1967.
39. Kuentael, L.E., Bacteria, carbon dioxide and algal blooms, *Jour. of the Water Pollution Control Fed.,* October, 1969.
40. King, D.L., The role of carbon in eutrophication, *Jour. of the Water Pollution Control Federation,* Dec., 1970.
41. Herbert, C.P. and G.S. Schroepfer, The travel of nitrogen in soils, *Jour. of the Water Pollution Control Fed.,* 39, No. 1, Jan., 1967.

42. Eliassen, Rolf *et al.,* Removal of nitrogen and phosphorus, 23rd Purdue Industrial Waste Conference, Purdue Univ., Lafayette, Ind., May, 8, 1968.
43. Missingham, G.A., Occurrence of phosphorus in surface waters and some related problems, *Jour. of the Amer. Water Works Assn.,* Feb., 1967, p. 207.
44. Culp, R.L., Wastewater reclamation by tertiary treatment, *Jour. of the Water Pollution Control Fed.,* **35,** No. 6, June, 1963.
45. Reference 20. above, p. 1438.
46. Yee, W.C., Selective removal of mixed phosphates by activated alumina, *Jour. of the Amer. Water Works Assn.,* Feb., 1966, p. 246.
47. Bennett, G.E. *et al.,* Water reclamation study program, Water Quality Control Lab., Stanford Univ., Stanford, Calif., Aug., 1968.
48. Rolich, G.A., Chemical methods for the removal of nitrogen and phosphorus from sewage treatment plant effluents, algae and metropolitan wastes, Trans. of the 1960 Seminar, Taft Sanitary Engineering Center, Cincinnati, Ohio.
49. McKinney, Ross E., *Microbiology for Sanitary Engineers,* McGraw-Hill, N.Y., 1962, p. 243.
50. Bogan, R.H., The use of algae in removing nutrients from domestic sewage, algae and metropolitan wastes, Trans. of the 1960 Seminar, Taft Sanitary Engineering Center, Cincinnati, Ohio.
51. Jones, R.A. and G.F. Lee, 1982. Recent advances in assessing the impact of phosphorus loads on eutrophication-related water quality. *Water Res.,* **16:** 503–515.
52. Rast, W., R.A. Jones and G.F. Lee. 1983. Predictive capability of US OECD phosphorus loading/eutrophication response models. *Journ. Water Pollut. Control Fed.* **55,** 990–1003.
53. White, E., 1983. Lake eutrophication in New Zealand—A comparison with other countries of the Organization for Economic Cooperation and Development, *New Zealand Journal of Marine and Freshwater Research,* **17,** 437–444.
54. Birge, W.J., J.A. Black and A.G. Westerman, "Short-Term Fish and Amphibian Embryo-Larval Tests for Determining the Effects of Toxicant Stress on Early Life Stages and Estimating Chronic Values for Single Compounds and Complex Effluents," *Environmental Toxicology and Chemistry,* Vol. 4, 1985, pp. 807–821.
55. Brown, O.B., R.H. Evans, J.W. Brown, H.R. Gordon, R.C. Smith and K.S. Baker, "Phytoplankton Blooming Off the US East Coast: A Satellite Description," *Science,* **229:** 163–167 (1985).
56. Suszkowski, D.J. and J.M. Mansky, "The Disposal of Sediments Dredged from New York Harbors," Management of Bottom Sediments Containing Toxic Substances, 10th U.S./Japan Experts Meeting, 1985.
57. New Jersey Department of Environmental Protection (NJ DEP). "Green Tide: Occurrence, Identification and Impacts," report of the New Jersey Department of Environmental Protection (1987).

ROBERT DRESNACK
New Jersey Institute of Technology

F

FLUIDIZED BED COMBUSTION

INTRODUCTION

The technology for reacting suspended coal particles with a gas flowing through them dates back to the 1920s when the Winkler gas generator was developed in Germany. The petroleum industry was responsible for the commercial expansion of fluidization techniques in the U.S. (1940s), particularly in the use of solids which catalytically crack vaporized heavy oils to produce gasoline and other petroleum fuels. The application of fluidized bed combustion (FBC) technology (to various solid fuels) is widespread in the U.S. and in other countries for all types of industrial processes. More than 350 atmospheric fluidized bed units are operating in North America, Europe and Asia. FBC is part of the answer to the question—how do we control our major emissions from coal sources? Briefly an FBC boiler is a finely divided bed of solid fuel particles in admixture with limestone particles which are suspended or conveyed by primary combustion air moving in the vertical upward direction. The limestone reacts with sulfur dioxide to remove it from the flue gas. The low uniform temperature (ca 1550°F) has a beneficial effect on nitrogen oxide suppression. The emission from coal combustion schemes of nitrogen oxides (NO_x) and sulfur dioxide (SO_2), together with carbon oxides (CO and CO_2), particulate matter and solid wastes must always be compared when evaluating various alternative schemes. The potential consequences of gaseous emissions, include the greenhouse effect and acid rain, which have received much publicity in recent years. The practical FBC limit of SO_2 removal is currently about 95%. Nitrogen oxide formation is lower than with conventional pulverized coal (PC) boiler NO_x control.

TYPES OF FLUIDIZED BED COMBUSTORS (FBCs)

FBCs are generally referred to as either circulating (CFB) or bubbling beds. However, the bubbling type may be classified according to whether reaction takes place at atmospheric (AFB) or under pressurized conditions (PFB).

A. Circulating Fluidized Bed Combustors

In the basic CFB combustor, coal or some other type of fossil fuel, e.g., natural gas or petroleum, is injected into the combustor together with a calcium based material such as limestone or dolomite to be used as a sorbent for SO_2. The bed material is entrained by fluidizing air usually in the velocity range of 12–30 ft/sec. The entrained material is forced into a refractory-lined cyclone located between the combustor and the convective pass. The separated larger particles are reintroduced at the bottom of the combustion chamber or, as in some designs, to an external heat exchanger. The mean bed particle size is usually between 50 and 300 microns. Combustion temperature will vary but generally is kept between 1550°F and 1650°F.[1] In this temperature range SO_2 sorption is optimized and the formation of nitrogen oxides is minimized.

The heavier solids fall to the bottom of the cyclone and are recirculated at a ratio of between 15:1 and 100:1 (solids to feed). The carbon content of the bed is usually about 3–4%. Calcium sulfate, ash, and calcined limestone make up the bulk of the recirculated material. The flue gas exits the top of the cyclone, travels through the convective pass and typically goes into an economizer (heat exchanger—superheated steam produced) and into a tubular air preheater. From there the gas may enter an electrostatic precipitator or a bag house dust collector (for removal of fine particulate matter from the gas). An induced draft fan is finally employed to force the gas up a stack and into the atmosphere.

Combustion air is provided at two levels of the combustor. Primary air enters through the bottom of the combustor and is evenly distributed by a gas distributor plate. Secondary air enters through a number of ports in the sidewalls of the combustor. Hence, there are two staged areas of combustion within the combustor. In the lower combustor, combustion takes place under reducing conditions. In the upper combustor nitrogen oxides are further reduced as is particulate matter. The admission of secondary air is also beneficial in controlling the temperature of the combustor as well as in maintaining the transport (entrainment) of the bed material throughout the length of the combustor. The density of the bed naturally varies with the combustor height, with density increasing towards the bottom.

Steam may be produced at several locations. Water-walls fixed to the upper portion of the combustor extract heat generated by the combustor. The convective pass also emits heat associated with the hot flue gas and solids which pass through it. External heat exchangers are also employed for the steam production. These heat exchangers (EHE's) are unfired, dense fluidized beds, which extract heat from the solids which fall to the bottom of the cyclone(s). More than one cyclone may be employed. The heat exchange is accomplished before the material is returned to the combustor. The external heat exchanger is a device which can, thus, be used as an effective additional method for controlling combustor temperature. The heat transfer coefficients to the water-walls usually lie between 20 and 50 Btu/hr. ft.^{2}°F.

B. Bubbling Fluidized Bed Combustors

Bubbling fluidized bed combustors are characterized by distinct dense beds. The bed material may be recirculated as in the case of CFB's, but at substantially lower recycle ratios (between 2:1 and 10:1). Particle velocities are usually between 2–15 ft/sec and a small amount of bed material is separated out (elutriated) as compared with CFB's. The mean bed particle size generally lies between 1000 and 1200 microns.

As with CFB's, the fuel used is usually coal or some other type of fossil fuel. Limestone or some other sorbent material is also used to decrease SO_2 emissions. The feed material may be fed either over the bed or under the bed. The manner of the feeding is an important design criterion in that it effects boiler control, emissions control (especially for SO_2) and combustion efficiency. Many bubbling bed designs incorporate overbed feeding in which the feed is "thrown" into the combustor by pressurized air. This overbed method can often be a disadvantage because throwing distance is limited. Hence, a long, narrow boiler is often required.

The underbed method of feeding is often associated with plugging and erosion problems. However, these problems can be avoided with proper design considerations. The Tennessee Valley Authority (TVA) has designed a 160 MW bubbling bed unit at its Shawnee station in Kentucky. The facility was constructed at a cost of $232,000,000 (1989). EPRI believes that most retrofits would fall into the $500–1000/kW range (1989 dollars) and that the levelized generation cost would be 5–10% less than a conventional unit with downstream flue gas treatment.[1a] The coal used for this unit is crushed to less than ¼ inch and dried with flue gas to less than 6% moisture. The fuel then passes through a fluidized bottle splitter with a central inlet and fuel lines arranged concentrically around the inlet. The feed material is forced into the combustor from the bottles, which are pressurized, by blowers. Each bottle acts as an individual burner and can be used to control load in the same way as cutting a burner in and out.[2] When overbed feeding is used, the fine material in the fuel has a tendency to elutriate too swiftly. If the fuel is fed underbed, the fines will have a longer residence time. Excess CO generation can result with the excessive burning of fines. This in turn can lead to overheating which could cause superheater controls to trip-off. Ash-slagging is another potential problem associated with overheating. Sometimes it may be necessary to recycle the fly-ash in order that carbon is more thoroughly burned and sorbent more completely utilized.

In-bed combustor tubes are generally used to extract heat (create steam). The heat transfer coefficient range is higher than that of CFB's, i.e., 40–70 Btu/hr.ft^2°F. Erosion of the tubes is a problem which is ever present in the bubbling bed combustor. The problem worsens as bed particle velocity increases. Horizontally arranged tubes are more susceptible to erosion than are vertical tubes. Various methods of erosion protection include metal spray coatings, studding of the tube surfaces with small metal balls, and wear fins. Occasionally recycled cold flue gas is used in lieu of tubes.

Waterwalls located in the upper portion of the combustor are also used (as with CFB's) to extract heat. The lower portion is refractory lined. Combustor free-board is usually between 15 and 30 ft. The typical convective pass, cyclone, air heater, particle separator scenario closely resembles that of the CFB.

C. Pressurized Fluidized Bed Combustors (PFBCs)

The pressurized fluidized bed combustor is essentially analogous to the bubbling bed combustor with one exception—the process is pressurized (10 to 16 atmospheres) thereby allowing the flue gas to drive a gas turbine/electric generator. This gas turbine along with a stream-driven turbine creates a very efficient "combined cycle" arrangement.

PFBCs may also be "turbocharged," i.e., before the flue gas enters the gas turbine, heat is extracted via a heat exchanger. Steam created by the energy transfer is used to drive the compressor which pressurizes the system. There is no energy excess to drive an electric generator in this case.

Deeper beds (typically 4 m.) may be used in PFBCs because they are pressurized. The residence time of a particle in the bed is longer than that of a particle in the shallower bed of a bubbling bed combustor. The fluidizing velocity (typically 1 m/s) is also lowered because of pressurization. As mentioned before, lower velocities minimize the amount of in-bed tube erosion.

Two other benefits of pressurization are a reduced bed cross-sectional area and reduced boiler height.

Since combustor efficiency and sorbent utilization are excellent, recycle is rarely needed. However, when very unreactive fuels are burned, recycling of fines may be necessary.

Since PFBCs are pressurized, certain design characteristics must be taken into consideration, especially in regard to the gas turbine. This turbine supplies the combustion and fluidizing air for the bed. Unlike conventional AFBC's the turbine inlet air is dependent upon certain temperature and pressure conditions since this inlet air is actually the exhaust gas from the combustor. To compensate for variations in load and subsequent changes in the exhaust gas conditions the gas turbine must be flexible. An effective turbine should be able to accept low gas temperatures, be minimally affected by unremoved fines in the gas, compensate for low load conditions, and allow the gas velocity through the hot gas clean up (HGCU) system and excess air to remain near constant over much of the load range. Most FBC systems incorporate a free-wheeling low pressure and constant velocity high pressure shaft design to accomplish the aforementioned requirements. The HGCU system generally consists of one or several cyclones. Sometimes a back-end filter at conventional pressuers and temperatures is used in addition.

The gas turbine accounts for approximately 20% of a FBC's total power output while the steam turbine creates the remainder.[5] The steam turbine is powered from steam created via combustor tubes and is totally independent of the exhaust gas and gas turbine. Steam turbine performance is therefore only affected by fuel/feed conditions. Two types of fuel feeding are generally used for FBC's—dry and wet. For fuels with high heating values the fuel is mixed with water to create a paste (20–25% water). With this method there is naturally no need for coal drying, and evaporated water creates additional mass flow through the gas turbine. Dry fuel feeding is more beneficial with low heating value fuels.

FEDERAL AIR EMISSIONS STANDARDS

The standards of performance for fossil-fuel-fired steam generators (constructed after August 17, 1971) were last revised by the federal government as of July 1, 1988.

Regulated facilities include fossil-fuel-fired steam generating units of more than 73 megawatts (heat input rate 250,000,000 Btu/hr.) and fossil-fuel and wood-residue-fired steam generating units capable of firing fossil fuel at a heat input rate of more than 73 megawatts. Existing fossil-fuel-fired units which have been modified to accommodate the use of combustible materials other than fossil fuels are regulated in a different manner.

Within 60 days after the maximum production rate is attained by a regulated facility, the facility must conduct performance tests and provide the E.P.A. with the results of the tests. The tests must also take place before 180 days after the initial start-up of a facility. Each test is specific and used for the determination of such things as nitrogen oxide emission. These test methods and procedures may be found in 40 C.F.R. (Code of Federation Regulations) Part 60.46.[6] After a performance text is completed, a facility must not discharge pollutants into the atmosphere at levels greater than those established and listed in the federal regulations.

Gases may not contain more than 43 nanograms of particulate matter per joule heat input (0.10 lb. per million Btu) where particulate matter is defined as a finely divided solid or liquid material, other than uncombined water as measured by the reference methods specified in 40 C.F.R. Part 60.46. These gases must also not exhibit greater than 20% opacity except for one six-minute period per hour of not more than 27% opacity. Opacity is defined as "the degree to which emissions reduce the transmission of light and obscure the view of an object in the background."

Less stringent standards have been developed for the three following facilities.[6]

The Southwestern Public Service Company's Harrington Station No. 1 in Amarillo, Texas must meet an opacity of not greater than 35%, except that a maximum of 42% opacity is permitted for not more than six minutes in any hour. The Interstate Power Company's Lansing Station Unit No. 4 in Lansing, Iowa must meet an opacity of not greater than 32%, except than a maximum of 39% opacity is permitted for not more than six minutes in any hour. The Omaha Public Power District's Power Station in Nebraska City, Nebraska must meet an opacity of not greater than 30%, except that a maximum of 37% opacity is permitted for not more than six minutes in any hour.

Gases may not contain more than 30 nanograms per joule heat input (0.80 lb. per million Btu) of sulfur dioxide (SO_2) derived from liquid fossil fuel or liquid fossil fuel and wood residue. 520 ng/joule heat input (1.2 lb. per million Btu) is the maximum allowable SO_2 discharge from gases derived from solid fossil fuel or solid fossil fuel and wood residue.

When different fossil fuels are burned simultaneously in any combination, the SO_2 emission standard is calculated by the following formula:

$$\mathrm{PS}_{\mathrm{so}_2} = (y(340) + z(520)) / (y + z)$$

where $\mathrm{PS}_{\mathrm{so}_2}$ is the prorated standard in ng/joule heat input derived from all fossil fuels or fossil fuels and wood residue fired, y is the percentage of total heat input derived from liquid fossil fuel, and z is the percentage of total heat input derived from solid fossil fuel.

The SO_2 emission standard for Units 1 and 2 at the Central Illinois Public Service Company's Newton Power Station must comply with the 520 ng/joule requirement if the units individually comply with the 520 ng/joule requirement or if the combined emission rate from both unites does not exceed 470 ng/joule (1.1 lb/million Btu) combined heat input to both units.

It is interesting to note that the federal SO_2 emission limit for West German coal fired boilers is 2.5 lB./Mbtu (avg.) for boilers of between 18 and 110 MW and 0.51b./MBtu (avg.) for boilers of over 110 MW.[7]

Gases may not contain more than 86 ng/joule heat input (0.20 lb/million Btu) of nitrogen dioxide (NO_2) derived from gaseous fossil fuel. 129 ng/joule heat input (0.30 lb/million Btu) is the maximum allowable NO_2 discharge from gases derived from liquid fossil fuel, liquid fossil fuel and wood residue, or gaseous fossil fuel and wood residue. 300 ng/joule (0.70 lb/million Btu) is the maximum allowable NO_2 from solid fossil fuel or solid fossil fuel and wood residue (except lignite or a solid fossil fuel containing 25%, by weight, or more of coal refuse). 260 ng/joule (0.60 lb/million Btu) is the maximum allowable NO_2 from lignite or lignite and wood residue with the exception that 340 ng/

joule is the limit for lignite which is mined in North Dakota, South Dakota, or Montana and which is burned in a cyclone fired unit.

When different fossil fuels are burned simultaneously in any combination, the nitrogen oxide emission standard is calculated by the following formula:

$$PS_{NO_x} = \frac{w(260)+x(86)+y(130)+z(300)}{w+x+y+z}.$$

Where PS_{NO_x} is the prorated standard in ng/joule heat input for nitrogen oxides (except nitrous oxide) derived from all fossil fuels or fossil fuels and wood residue fired, w is the percentage of total heat input derived from lignite, x is the percentage of total heat input derived from gaseous fossil fuel, y is the percentage of total heat input derived from liquid fossil fuel and z is the percentage of total heat input derived from solid fossil fuel (except lignite).

There is no standard for nitrogen oxides when burning gaseous, liquid, or solid fossil fuel or wood residue in combination with a fossil fuel that contains 25%, by weight, coal refuse. Coal refuse is defined as "the waste products of coal mining, cleaning and coal preparation operations (e.g., culm, gob, etc.) containing coal, matrix material, clay, and other organic and inorganic material."[6]

The NO_x emission standards for West Germany and Japan are even more stringent than those of the U.S.[7] For new and existing West Germany boilers of over 110 MW, the limit is 0.16 lb./MBtu (6% O_2). For Japanese boilers built after 1987, the limit is 0.33 lb./MBtu.

PROMINENT FBC INSTALLATIONS IN THE U.S.

Recently, in order to reduce SO_2 emissions, Northern States Power Company (NSP) converted its Black Dog pulverized coal-fired boiler to that of a bubbling bed combustor. This unit is the largest of its kind in the world; its capacity is 130 megawatts.

NSP received a new Emissions Permit from the Minnesota Pollution Control Agency (MPCA) for the upgraded unit. The emissions standards set forth in this permit are less stringent than those of the federal standards for particulate matter and SO_2. In the event that utilities should become regulated, the operating parameters of the system or the system itself would have to be modified.[10]

The most recent literature available to the author (April 1988) stated that limestone was being added to the bed in order to lower SO_2 emissions sufficiently to help NSPS standard. The control of particulate matter was difficult at the onset. However, this problem was resolved by changing the bed material to an inert fired-clay material. NO_x emissions requirements have easily been met.

The Tennessee Valley Authority (TVA) has built a 160 MW bubbling bed combustor for the utility's Shawnee steam plant in Paducah, Kentucky. It has been operating sporadically since autumn of 1988.

A pilot plant (20 MW) was completed in 1982 and had brought forth some very promising results. With a Ca : S ratio of 2 to 2.5 (typical range) and a recycle ratio of 2 to 2.5 the SO_2 retention was approximately 90%.[11] This result has been matched by the scaled-up plant. The pilot plant has both an underbed and overbed feed system. Overbed feed does not produce as great a combustion efficiency as that achieved by the underbed method. This would be expected due to the lack of control over fines in the feed. NO_x emissions were less than 0.25 lb/million Btu.[11] The NSPS for NO_x is 0.7 lb/million Btu for solid fuel.

The original underbed feed system was determined to be inadequate because of plugging and erosion problems. The system was redesigned and proved to be successful. The feed system is one of pressurized bottles mentioned earlier in this report under "Bubbling Bed Combustors."

As stated in the "Introduction," fluidized bed combustion can be used for many different types of industrial processes. An example of this is the installation of the direct alkali recovery system at Associated Pulp and Paper Mills' Burnie, Tasmania mill.

In this process, sodium carbonate (residual) found in soda-quinone black liquor (a waste product) reacts with ferric oxide to produce sodium ferrite in the combustor (bubbling bed). The sodium ferrite is then contacted with water to yield sodium hydroxide (desired) and ferric oxide. The ferric oxide is returned to the combustor to be reused. It is interesting to note that most of the steam produced in this process is created from the extraction of heat from the exhaust gas and not from bed tubes. The exhaust gas is cleaned via a fabric filter and the dust collected is palletized. The pellets are later used in the process. The fluidizing air is heated from the heat extracted from the hot sodium ferrite after it has been removed from the combustor.

Since there is no sulfur involved in this process the exhaust gas is easily cooled, thereby allowing greater production of high-pressure steam.[13]

The title for the world's largest CFB probably belongs to the nuclear generating station owned by Colorado-Ute Electric Association. The original 25-year-old plant was replaced because it was uneconomical to operate. The capacity of the new plant is 110 MW.

In May of 1988 on EPRI (Electric Power Research Institute) assessment began and is scheduled to continue until May of 1990. As of April 1988, the unit was reported to be easy to operate, responsive to load variations, and easily restarted following a trip. However in 1989 operational difficulties were reported. SO_2 emissions standards were expected to be easily met and NO_x emissions were well under the limit. Final determination of the optimum Ca : S ratio still needed to be determined. Particulate matter emissions are expected to be less than 0.03 lbs/million. Btu because of the addition a new baghouse to the existing three baghouses.

Some valuable information has been learned from the unit thus far, e.g., control of coal feed size has been important in maintaining the bed quality and agglomerations can be avoided if the feed is started in short bursts prior to being continuous; this is to allow the temperature rise to be more uniform.[14]

One of the larger *commercial* units in the U.S. is located in Colton, California and was installed for Cal-Mat Co. The 25 MW CFB was constructed because electric utility rates were rising and the availability of power was uncertain. The company manufactures cement—a process requiring much electricity. Since the company had easy access to coal and limestone as well as a large quantity of heat from its kilns, CFB technology became an effective solution to their energy needs. Bottom ash waste and flyash could also be used in the cement-making process.

As could be expected, the air pollution controls instituted by the state of California are very strict. However, a permit was granted to CalMat in a relatively short period of time because of the fine performance demonstrated by this unit. The exhaust gases were found to contain SO_2 at 30 lb/hr., NO_x at 57 lb/hr. and CO at 24 lb/hr.[15] There were initial problems with equipment and systems, however, these were eventually eliminated. Bed retention and temperature control problems have also been resolved through modifications of the air flows and nozzles.

PFBC's have been installed in Sweden, the U.S. and Spain.

Two PFBC modules of 200 MW each have been installed in Vartan, Stockholm. The first unit is due for start-up in late 1989. The Swedish emission standards are very strict and include special restrictions on noise and dust since the units are located very close to a residential area.

A 200 MW combined cycle PFBC will be installed by American Electric Power (AEP) at its Tidd Power Plant at Brilliant, Ohio. Test results from joint studies proved PFBC technology to have environmental benefits surpassing those of traditional boilers with flue gas desulfurization systems (FGD), selective catalytic reduction, etc.

A 200 MW PFBC will be installed by Empresa Nacional de Electricidad S.A. (ENDESA) at its Escatron Power Plant as a retrofit for an existing unit. 90% sulfur removal and a NO_x decrease of 30% are expected. Many different technologies were considered but PFBC was chosen because of the high sulfur/ash/moisture black lignite coal that they burn.[16]

NO_X/SO_2 FORMATION AND CONTROL

Fossil fuels naturally contain sulfur in varying percentages. As fuel is burned the sulfur combines with oxygen to form SO_x, and primarily, SO_2. When emitted into the atmosphere this SO_2 can combined with water vapor to form sulfuric acid (and sulfurous acid to a lesser extent). This is a part of the basic mechanism by which acid rain is created.

In order to control sulfur dioxide emissions, the oldest and still most common method used is to react the gas with limestone or a similar calcium based material. Crushed limestone ($CaCO_3$) can be fed continuously to a conventional coal boiler or fluidized bed where it calcines to lime (CaO) and then reacts with SO_2 in the presence of oxygen to form calcium sulphate ($CaSO_4$). This material precipitates to the bottom of the combustor and is removed.

Coal particle size has a definite impact on desulfurization. Bed composition also has an effect on sulfur removal. A typical bed might be composed of coarse partially sulfated limestone and ash (produced by combustion). The particle size of the coal and limestone would probably be equal however combustor operation conditions such as fluidizing velocity will dictate the particle size.

An alternate scenario might be to pulverize the limestone and introduce it to a bed composed of ash or some other type of refractory material. Fines naturally have shorter residence times than do coarse materials and, hence, would probably have to be recycled to increase efficiency.

A series of experiments were carried out by Argonne National Laboratory[17] using three different types of limestones to test their effects on sulfur capture during combustion. The average particle size of the limestone was 500–600 micrometers. The Ca : S ratio was 2.3–2.6 and the combustion temperature was 1600°F. SO_2 removal was 74 to 86%. The test proved that the amount of SO_2 removal was relatively independent of the type of limestone used. The test also proved that particle size did not have much of an effect on SO_2 removal. The explanation offered for this observation was that although larger particles are less reactive than smaller particles, the increased residence time in the combustor of larger particles compensates for the lower reactivity.

Dolomite was also evaluated for SO_2 capture. In two experiments, Tymochtee dolomite was added to a bed composed of alumina at Ca : S ratios of 1.5 and 1.6. The average particle size was 650 micrometers. The SO_2 removals were 78% and 87% respectively. MgO is contained within the dolomite matrix and is believed to keep the particles more porous such that sulfation is greater, especially in larger particles.

Combustion temperature had a marked effect on SO_2 removal in these experiments. Dolomite No. 1337 was most effective in reducing SO_2 at 1480°F. Limestone No. 1359 was most effective in the range of 1500–1550°F. Both sorbents achieved approximately 91% SO_2 removal. The average particle size was approximately 500 micrometers. Pulverized limestone No. 1359 with an average particle size of 25 micrometers was most effective in the range of 1550–1600°F. The extent of calcination is more dependent upon bed temperature for finely pulverized limestone. The greater the calcination, the greater the reactivity with SO_2.

An unusual finding occurred in that Tymochtee dolomite was observed to be most effective in SO_2 removal at 1800°F. For all of the other sorbents the SO_2 removal was very poor at this temperature. Explanations for this phenomenon have been proposed. One explanation suggests that above a certain temperature the sorbent's pores close thereby ending sulfonation. Depending upon the sorbent's structure and composition, this temperature would be different for each sorbent. Another explanation involves the effect of fluidized bed gas circulation on bed chemistry. An emulsion phase and a gas bubble phase exist in any gas-solid fluidized bed. Excess gas which is not needed for fluidization circulates back and forth between the two phases. This gas does not react with the sorbent until it reaches the emulsion phase. All of the oxygen in the lower portion of the emulsion phase reacts with the fuel to from CO. As the bubbles rise through the bed, air exchanges between the bubbles and the emulsion phase. The upper portion of the emulsion phase contains excess oxygen. The following reaction was thus proposed as one which takes place in the lower portion of the combustor:

$$CaSO_4 + 4CO \rightleftharpoons CaS + 4CO_2. \qquad (1)$$

One or more of the following reactions were proposed to occur in the upper portion of the combustor:

$$CaS + 1\frac{1}{2}O_2 \rightleftharpoons CaO + SO_2 \qquad (2)$$

$$CaS + 3CaSO_4 \rightleftharpoons 4CaO + 4SO_2 \qquad (3)$$

$$CaS + 2O_2 \rightleftharpoons CaSO_4. \qquad (4)$$

Reactions 2 and 3 would limit the unit's ability to remove sulfur because of the regeneration of SO_2. This regeneration of SO_2 is so dependent upon temperature that it could very possibly be an explanation as to why SO_2 removal generally suffers at high temperatures.

Nitrogen also occurs naturally in fossil fuels. This nitrogen reacts with oxygen during combustion and later forms acid rain in very much the same manner as with sulfur. Oxides of nitrogen (NO_x) are also responsible for the creation of "smog." As nitrogen dioxide (NO_2) absorbs light of certain wavelengths it dissociates photochemically to form nitric oxide (NO) and atomic oxygen. This atomic oxygen is very reactive and readily combines with O_2 to from ozone (O_3). Ozone in turn oxidizes hydrocarbons in the air to form aldehydes. Ozone and the aldehydes are components of smog. NO_2 is the reddish-brown gas which can often be seen on the horizons of cities such as Los Angeles.

The principal oxide of nitrogen formed during combustion is nitric oxide. Nitrogen in the fuel combines with oxygen in the fluidizing air as follows:

$$\frac{1}{2}N_2 + \frac{1}{2}O_2 \rightleftharpoons NO.$$

The kinetics of NO decomposition are slow enough so that equilibrium levels are not achieved. Various experiments conducted by Argonne National Laboratory as well as by other researchers have proven that most of the nitrogen forming NO_x is from the fuel and not from the air. This has been easily demonstrated by substituting an inert gas (such as argon) for nitrogen in the fluidizing air and then comparing the results to those of combustion with standard fluidizing air.

As previously mentioned in this report two-stage combustion is an effective method of decreasing NO_x emissions. As with SO_2 reduction bed composition has an important effect on NO_x. It has been determined through experimentation and experience that limestone also decreases NO_x emissions. Skopp and Hammons[18] observed that when using a limestone bed two factors were changing with time which could have been responsible for decreasing NO emissions: the $CaSO_4$ concentration in the bed was increasing and so was the SO_2 concentration. The increase in $CaSO_4$ suggested that it could be a selective catalyst for reduction of NO. The increasing SO_2 concentration suggested that there might be a reaction occurring between it and the NO which was lowering the NO. This was investigated by conducting experiments using synthetic $NO—SO_2—N_2$ gas mixtures. The results showed that no reaction in the gas phase occurred. There was also no reaction between the NO and SO_2 over $CaSO_4$. However, there was a reaction occurring over a bed of 20% sulfated lime. This reaction was found to have a negative temperature dependence. The following mechanism was proposed by Skopp and Hammons[18] as an explanation for their results:

$$CaO + SO_2 \rightleftharpoons CaSO_3$$

$$CaSO_3 + 2NO \rightleftharpoons CaSO_3(NO)_2$$

$$CaSO_3(NO)_2 \rightleftharpoons CaSO_4 + N_2O$$

$$N_2O \rightleftharpoons N_2 + \frac{1}{2}O_2.$$

Esso researchers investigated the possibility of NO being produced by CO catalyzed by $CaSO_4$. The rate of this reaction was found to increase with increasing temperature.

Argonne researchers[17] investigated the use of metal oxides, among them, aluminum oxide (Al_2O_3), zirconium oxide (ZrO_2) and cobalt oxide (Co_3O_4). At the time these experiments were conducted, the literature had indicated that these metal oxides were effective in reducing or catalytically decomposing NO. The results showed that the addition of Al_2O_3 and ZrO_2 did nothing to reduce NO formation during combustion in a fluidized bed. The addition of Co_3O_4 actually increased rather than decreased the formation of NO.

A study was conducted by McCandless and Hodgson[20] for the U.S.E.P.A. on the use of metal sulfides as a way to reduce NO emissions. The following is well known as the "Thiogen" process and has been used in the recovery of sulfur from SO_2

$$CaS + 2SO_2 \rightleftharpoons CaSO_4 + S_2$$

$$4CaS + 6SO_2 \rightleftharpoons 4CaSO_3 + 3S_2.$$

Based on this process it was determined that the following reaction might also be possible

$$CaS + 4NO \rightleftharpoons CaSO_4 + 2N_2.$$

Preliminary studies indicated that the reaction did proceed and could be an effective method for NO_x control. Nineteen metal sulfides were used. All but one reduced NO to N_2 at temperatures between 194°F and 1202°F. However, a weight loss did occur indicating that an undesirable side reaction was taking place—probably the formation of SO_2. Some metal SO_4 was formed in most of the tests. However, the alkaline earth sulfides were determined to be the most stable.

It was also found that the temperature at which the reduction reaction occurs can be lowered if certain catalysts such as NaF and $FeCl_2$ are mixed with the sulfides. Reaction temperature was again reduced when the sulfide/catalyst combination was impregnated on alumina pellets. Tests were also conducted involving synthetic flue gas containing 1000 ppm NO, 1% O_2, 18% CO_2 and the remainder N_2. Using this gas in combination with the CaS showed that NO was significantly reduced above temperatures of 1112°F, by using the sulfide/catalyst combination. The results of the experiments showed that between 0.372 and 0.134 grams of NO were reduced per gram of metal sulfide. Between 0.76 and 0.91 grams was achieved when using the impregnated alumina pellets. The authors recommended that more research be done to evaluate the economical implications of using these materials.

Several other interesting facts known about NO_x control and found in the literature are that increasing fluidizing velocity decreases NO_x, NO_x is not significantly affected by excess air, and NO_x production increases at lower temperature, especially below approximately 1500°F.

For conventional coal-fired boilers the most common approach to control NO_x and SO_x simultaneously is the combination of selective catalytic reduction (SCR) and wet-limestone or spray dryer flue gas desulfurization (FGD). The SCR process converts NO_x to N_2 and H_2O by using ammonia as a reducing agent in the presence of a catalyst. The catalytic reactor is located upstream from the air heater and speeds up the reaction between the NO_x and the ammonia, which is injected into the flue gas in vapor form immediately prior to entering the reactor. The reduction reactions are as follows:[21]

$$4NH_3 + 4NO + O_2 \rightleftharpoons 4N_2 + 6H_2O$$
$$4NH_3 + 2NO_2 + O_2 \rightleftharpoons 3N_2 + 6H_2O.$$

It can be seen that the amount of NO_2 removed primarily depends on the amount of NH_3 used. Although SCR technology has proved to be an effective means to reduce NO_x with removal results as high as 90% in some European facilities, the U.S. does not consider the technology economically feasible. In addition to the high cost there are the undesirable effects of unreacted NH_3, by-product SO_3 and increased CO production to consider. There are also catalyst deactivation problems caused by contamination by trace metals in the fly ash and by sulfur poisoning. The Japanese have improved on the design of catalysts and their arrangement within the reactor. However, these modifications are still too new to evaluate their merit.[22] U.S. industry also feels that more data has to be generated for the medium to high sulfur coals most commonly used in this country. Since characteristics such as high sulfur, low flyash alkalinity and high iron content are common in U.S. coal, and these qualities do influence SO_3 production, SCR would not appear to be one of the likely options for U.S. industry at least in the near future.

Exxon has developed a process called "Thermal deNO_x" which makes use of ammonia injection into the flue gas at temperatures of between 1600 and 2200°F. This process is claimed to remove NO_x by up to 90%. years, CFB's have become the dominant FBC choice in industry. The most common problems that have been associated with bubbling beds include erosion of the inbed tubes. This can be reduced through the use of studding, fins, etc. as previously mentioned in this report. However, CFB's are also prone to erosion, i.e., the waterwalls, as well as the refractory lining. Agglomeration is another common problem associated with bubbling beds. Sand can fuse in localized hot spots to form clinkers or "sand babies" especially when the fuel has a high concentration of alkali compounds.[9] In severe cases, agglomerations can cause the bed to defluidize, block air ports, and make bed material removal more difficult. Sulfur removal is more difficult with bubbling beds. In general, large quantities of double-screen stoker coal must be used to attain the high sulfur removal rate displayed by CFB's. Most overbed feed bubbling beds in existence must use coals which contain less than 10% fines. This can often be quite costly. As previously mentioned, underbed feed also has problems associated with it. Since low fluidizing velocities are required with underbed feed, the bed plan area must be larger and, subsequently, contain a higher density of feed ports. This serves to complicate the already unreliable feed system. In order to utilize the sorbent better, the recycle ratio has to be increased. However, above a certain recycle ratio, and in-bed tubes might have to be removed in order to maintain combustor temperature, compromising the CFB design.

NO_x control is better with CFB's than with bubbling beds. This is because of the aforementioned stage combustion which is physically unachievable in bubbling beds due to the large bed plan area and low fluidizing velocity. On average, 0.1 lb/million Btu less NO_x is produced by CFB's than by bubbling beds.

As of the present, there are no federal regulations governing CO emissions. However, some states have promulgated regulations. As would be expected with overbed feed bubbling bed combustors, the CO emissions are high. While emissions of over 40 ppmv are common with bubbling beds, CFB's are usually under 100 ppmv.[3] This is due to better circulation and recycle. There is not much data on CO emissions for underbed feed bubbling beds. However, it evidently reduces CO more than does overbed. Unfortunately, with CFB's there is a trade-off between SO_2/NO_x and CO. Staged combustion will increase CO emissions as the primary to secondary air ratio becomes smaller. SCR/SNR specific to CO also may cause an increase in NO_x.

TABLE 1

Major chemical components of composite residues: pilot and full-scale CFBC units

Component	Concentration (wt percent)				
	Test no.1[a]	Text no. 2[a]	Test no. 3[a]	Baghouse[b]	Full-scale residues
$CaSO_4$	26.1	31.6	21.7	27.7	26.1
CaS	0.8	0.6	0.7	5.2	0.45
Free CaO	24.8	28.3	24.7	15.7	23.5
$CaCO_3$	3.4	3.2	5.2	6.8	4.6
Fe_2O_3	10.7	9.5	9.1	10.6	15.9
Other mainly SiO_2 and C	27.3	20.8	33.8	19.5	24.2
LOI[c] (corrected)	11.4	5.4	5.1	8.0	4.3
Sum	104.5	99.4	100.3	—	99.0

[a] Composite pilot-scale residues.

[b] Calculated from TGA and other analyses.

[c] LOI indicated loss on ignition.

TABLE 2

Geotechnical properties of the pilot-scale rig composite and baghouse samples and full-scale unit residues

	Pilot-scale composite residues	Pilot-scale baghouse residue	Full-scale residue
Specific gravity	2.83–3.07	2.58	2.95
Mean size, D_{50} mm	0.2	0.04	0.04
Optimum water content percent[a]	14.5–17.5	32	26.5–30.5
	Unconfined compressive strength, kPa[b]		
Curing period, days			
0	230–360	—	—
3	—	—	2470
7	150–290	—	4120
8	—	309	—
10	260–425	385	—
12	—	461	—
28	—	—	4660
Freeze/thaw cycles			
4	—	201	—
6	540–1020	Sample destroyed	—
7	—	—	3200
15	—	—	880

[a] As determined by standard Proctor test.

[b] Samples were cured at 100 percent relative humidity at 23 ± 2°C for periods shown.

Other problems associated with bubbling beds are scale-up and turn-down. Scale-up is limited because of the feed distribution problem and turn-down is usually more frequent because of the erosion problem. Incidentally, it should have been mentioned before, that, in general, FBC's take longer to start up and turn down than conventional boilers because of the large amounts of bed material which must be heated or cooled.

The major advantage which PFBC's hold over AFBC's is that NO_x, SO_2, and CO are weakly linked. Thus, 60 mg/MHJ NO_x can be attained while at the same time only 50 mg/MJ SO_2 (or less) and 10 mg/MJ (or less) CO are produced. Another advantage of PFBC's is that the waste contains negligible lime, sulfides, and sulfites. The decreased lime concentration makes the waste less reactive and probably renders it nonhazardous. The decreased sulfites makes the material a good candidate for use in cement kilns and concrete. The maximum sulfur content per ASTM standards is 1.2% by weight or 3.1% by weight as sulfites. The material has been found useful for building roads, manufacturing gravel and formed bricks or tiles, and for roofing and flooring material.[3]

Other advantages of PFBC's include compactness because of smaller bed requirements, plant cycle efficiencies of 40–42% and subsequent reduced fuel costs, and unit modularity for ease in increasing future capacity.[8]

Disadvantages include in-bed tube erosion and potential damage to the gas turbine if the hot gas clean-up is ineffective. It should be noted that the technology is too new to accurately assess its advantages versus its disadvantages.

CHARACTERIZATION OF SOLID WASTES FROM FLUIDIZED BEDS

The characterization and use of fluidized-bed-combustion coal/limestone ash is discussed in the articles of Behr-Andres and Hutzler[23] and Anthony *et al.*[24] The former dealt with the use of the mixture in concrete and asphalt. The latter presented chemical and physical properties for the waste (see Tables 1 and 2 above).

Hot-gas cleanup (HGCU) technologies have emerged as key components of advanced power generation technologies such as pressurized fluidized-bed combustion (PFBC), and integrated gasification combined cycle (IGCC). The main difference between HGCUs and conventional particulate removal technologies (ESP and baghouses) is that HGCUs operate at higher temperatures (500 to 1,000°C) and pressures (1 to 2 MPa), which eliminates the need for cooling of the gas. See Website (2005): http://www.worldbank.org/html/fpd/em/power/EA/mitigatn/aqpchgas.stm

REFERENCES

1. Robert H. Melvin and Reid E. Bicknell, "Startup and Preliminary Operation of the Largest Circulating Fluid-Bed Combustion Boiler in a Utility Environment—NUCLA CFB Demonstration Project," Paper Presented at the 50th American Power Conference, Chicago, Illinois, April 18–20, 1988 p. 2.
2. Jason Makansi and Robert Schweiger, "Fluidized-bed boilers," *Power,* May 1987, p. 9.
3. *Efficiency and Emissions Improvements by Means of PFBC Retrofits* (Finspong, Sweden: Asea PFBC Component Test Facility, S-61220, 1988), p. 2.
4. Taylor Moore, "Fluidized bed at TVA," EPRI Journal, March 1989, p. 27.
5. *Asea Babcock PFBC Update,* 1, No. 3 (Fall 1988), n. pag.
6. *Code of Federal Regulations,* Vol. 40, Part 60, Revised as of July 1988.
7. Charles Sedman and William Ellison, "German FGD/DeNO$_x$ Experience," Presented at the Third Annual Pittsburgh Coal Conference, Pittrsburgh, Pennsylvania, September 1986.
8. *Asea Babcock PFBC Update,* 1, No. 2 (Summer 1988), n. pag.
9. Jason Makansi, "Users pause, designers wrestle with fluid-bed boiler scaleup," *Power,* July 1988, p. 2.
10. David Osthus, John Larva, and Don Rens, "Update of the Black Dog Atmospheric Fluidized-Bed Combustion Project," Paper Presented at the 50th American Power Conference, Chicago, Illinois, April 18–20, 1988, p. 1.
11. Bob Schweiger, ed., "Fluidized-bed boilers achieve commercial status worldwide," *Power,* Feb. 1985, p. 9.
12. R.A. Cochran and D.L. Martin, "Comparison and Assessment of Current Major Power Generation Alternatives," Presented at the Power-Gen Exhibition and Conference for Fossil and Solid Fuel Power Generation in Orlando, Florida, Boston, Massachusetts, Dec. 1988, p. 2.
13. Sheldon D. Strauss, "Fluidized bed keys direct alkali recovery," *Power,* Feb. 1985, p. 1.
14. Melvin and Bicknell, p. 6 (see 1).
15. Bob Schweiger, ed., "U.S.'s largest commercial CFB burns coal cleanly in California," *Power,* Oct. 1986, p. 2.
16. *Efficiency and Emissions Improvement by Means of PFBC Retrofits,* p. 10.
17. A.A. Jonke *et al.*, "Reduction of Atmospheric Pollution by the Application of Fluidized-Bed Combustion," Argonne National Laboratory, Publication No. ANL/ES-CEN-1002, 1970, n. pag.
18. A. Skopp and G. Hammons, "NO$_x$ Formation and Control in Fluidized-Bed Coal Combustion Processes," ASME Winter Annual Meeting, Nov./Dec., 1971.
19. Jason Makansi, "Meeting future NO$_x$ caps goes beyond furnace modifications," *Power,* September 1985, p. 1.
20. *Reduction of Nitric Oxide with Metal Sulfides,* Research Triangle Park: U.S. E.P.A., EPA-600/7078–213, Nov. 1978, pp. 1–5.
21. *Ibid,* p. 3.
22. Ed Cichanowicz, "Selective catalytic reduction controls NO$_x$ in Europe," *Power,* August 1988, p. 2.
23. Christina B. Behr-Andres and Neil J. Hutzler, "Characterization and use of fluidized-bed-combustion coal ash," Journal of Environmental Engineering, November/December 1994, p. 1488–506.
24. E.J. Anthony, G.G. Ross, E.E. Berry, R.T. Hemings. and R.K. Kissel, "Characterization of Solid Wastes from Circulating Fluidized Bed Combustion", Trans. of the ASME Vol. 117, March 1995, 18–23.

JAMES SANDERSON
Environmental Protection Agency
Washington, D.C.

FISH ECOLOGY: see POLLUTION EFFECTS ON FISH; THERMAL EFFECTS ON FISH ECOLOGY

FOSSIL FUEL CLEANING PROCESSES

The amount of pollutants, especially sulfur oxides and particulates emitted to the atmosphere may be reduced by treating fuels prior to combustion. This approach may be more energy efficient than treatment of flue gases as per *Vapor and Gaseous Pollutant Fundamentals.* More than thirty million tons of sulfur dioxide are discharged annually in the United States, 75% of which is the result of fuel burning.

FOSSIL FUEL PRODUCTION, RESERVES AND CONSUMPTION

The world's production of oil in 1980 was 66 million barrels per day with a projected value of 77 MBPD for the year 2000. The relatively small anticipated increase reflects increased conservation and alternate fuel source application. The overall oil output of the USSR was about 14 MBPD[1] as compared to about 12 MBPD for combined US and Canadian production (1980).

About 2500 trillion cubic feet of natural gas reserves are estimated to exist worldwide. The US reserves are 200 TCF with an annual consumption of about 20 TCF. Soviet bloc production was about 15 TCF in 1980. Most oil and natural gas reserves fall in a crescent shaped area extending from Northern Algeria Northward to West Siberia. Lynch[2] felt that the level of surplus capacity would remain stable for the early '90s with the then world stock level of about 100 gigaliters (1.3 giga barrels).

Coal is consumed at a rate of 600 million tons annually in the US utility industry. Only a small portion of Eastern US coals fall in the low (less than 1% sulfur) category—see Table 1. The US, USSR and China own about two thirds of the world's 780 billion tons of presently recoverable coal reserves. The US has about one quarter of the total. Coal accounts for 90% of the US's proven reserves.[3]

Consumption of fuel might be measured in "quads" or quadrillion Btu's. It has been estimated that US electric consumption was 13 quads and nonelectric industrial about 16 quads for the year 1980.[3] Total US fossil fuel consumption is about 76 quads, most in the non-industrial sector. Worldwide energy consumption is predicted to double over the next 25 years according to the World Energy Council.[3a] The predicted fossil fuel usage in terms of billions kwh electric generation in the year 2015 is for coal-2000, natural gas-1000, nuclear-400, and petroleum-less than 100. Renewables are estimated at 400 billion kwhs. Divide these numbers by 100 to estimate the number of quads; assuming a plant efficiency of current Rankine cycle plants (about 34%) or by 170 if a combined cycle (Brayton = Rankine) is assumed.

SULFUR REMOVAL

Typical legislative actions have been the setting of limits on the allowable sulfur content of the fossil fuel being burned or on the SO_2 emission rates of new sources. In California, regulations have limited the use of fuel oil to those of 0.5% or less sulfur. Since 1968, a limit of a 0.3% sulfur oil has been in effect in New York City. In 1980, Massachusetts set a 1% sulfur limit on the coal to be burned. This limit is being considered for other Atlantic seaboard states as coal conversion is increasingly encouraged. Chemical and physical desulfurization of fossil fuels can be used to produce levels of sulfur which comply with government standards. To reduce a 3% sulfur coal to a 1% sulfur coal may add about 10% to the cost of coal F.O.B., but may save on transportation and flue gas desulfurization costs.

The amount of sulfur dioxide emitted worldwide might double in the next decade due to increased energy demands (approximately 3.5% annually) and the use of more remote crudes having higher sulfur concentration.

The chemical and power industries must strike a delicate balance between the public's dual requirement of increased quantities and preparation of fossil fuel. More fuel must now be desulfurized more completely and/or more sulfur dioxide must be removed from stack gases. The techniques for cleaning fossil fuels used throughout the petroleum, natural gas and coal production industries are covered in this article. Treatment of stack gases to effect particulate and sulfur removal are discussed separately in other articles.

PROCESSES INVOLVING THE BASIC FUELS

The two most commonly combusted energy sources are coal and fuel oil having typical sulfur ranges of 1–4% and 3–4%, respectively; a 3% sulfur oil produces about the same SO_2 emission as a 2% sulfur coal when based on a comparable energy release. Fuel oil desulfurization is used by most major oil producers. Hydrogenation, solvent extraction, absorption and chemical reaction are used to varying extents at petroleum refiners. Finfer[4] claims a possible sulfur reduction from 2.5 to 0.5% by a hydrodesulfurization process. Coal contains sulfur which may be combined with either the organic or

TABLE 1
Ash content and ash fusion temperatures of some U.S. coals and lignite

Rank	Low Volatile Bituminous	High Volatile Bituminous			Subbituminous		Lignite
Seam	Pocahontals No. 3	No. 9	Pittsburgh	No. 6			
Location	West Virginia	Ohio	West Virginia	Illinois	Utah	Wyoming	Texas
Ash, dry basis, %	12.3	14.10	10.87	17.36	6.6	6.6	12.8
Sulfur, dry basis, %	0.7	3.30	3.53	4.17	0.5	1.0	1.1
Analysis of ash, % by wt	—	—	—	—	—	—	—
SiO_2	60.0	47.27	37.64	47.52	48.0	24.0	41.8
Al_2O_3	30.0	22.96	20.11	17.87	11.5	20.0	13.6
TiO_2	1.6	1.00	0.81	0.78	0.6	0.7	1.5
Fe_2O_3	4.0	22.81	29.28	20.13	7.0	11.0	6.6
CaO	0.6	1.30	4.25	5.75	25.0	26.0	17.6
MgO	0.6	0.85	1.25	1.02	4.0	4.0	2.5
Na_2O	0.5	0.28	0.80	0.36	1.2	0.2	0.6
K_2O	1.5	1.97	1.60	1.77	0.2	0.5	0.1
Total	98.8	98.44	95.74	95.20	97.5	86.4	84.3
Ash fusibility	—	—	—	—	—	—	—
Initial deformation temperature, F							
Reducing	200+	2030	2020	2000	2060	1990	1975
Oxidizing	2900+	2420	2265	2300	2120	2190	2070
Softening temperature, F							
Reducing		2450	2175	2160		2180	2130
Oxidizing		2605	2385	2430		2220	2190
Hemispherical temperature, F							
Reducing		2480	2225	2180	2140	2250	2150
Oxidizing		2620	2450	2450	2220	2340	2210
Fluid temperature, F							
Reducing	2620	2370	2320		2250	2290	2240
Oxidizing		2670	2540	2610	2460	2300	2290

inorganic (pyritic and sulfate) matter. The organics may be removed by various cleaning processes, but little reduction in organic sulfur has been found to occur by physical cleaning methods. Currently an extraction process, followed by hydrogenation, is being tried. Some coals have been reduced to S contents below 2%, and typical sulfur reduction estimates are in the range of 20–40% reduction.[5,6,7] Even if these reduced levels are achieved, a need for further removal of sulfur from the flue gases might exist. Cleaning, when combined with flue gas desulfurization as a method of SO_2 control, could eliminate the need for reheat and considerably reduce the sludge handling requirements of the plant.

Fuel Oil Desulfurization (General)

Before the ecological need for fuel oil desulfurization was recognized, oil stocks were desulfurized for a number of other reasons:

1) To avoid poisoning and deactivation of platinum catalysts used in most catalytic reforming processes.
2) To reduce sulfurous acid corrosion of home burner heating equipment.
3) To demetalize crude stocks (sulfur removal from crude is generally accompanied by a concomitant removal of such trace metals as sodium, vanadium and nickel).
4) To recover pure sulfur.
5) To reduce or eliminate final product odor.

By definition, hydrodesulfurization is the removal of sulfur by a catalytic reaction with hydrogen to form hydrogen sulfide. As carried out in the petroleum industry, the hydrodesulfurization process is not a specific chemical reaction. Various types of sulfur compounds (mercaptans, sulfides, polysulfides, thiophenes) with varying structures and molecular weights are treated. Obviously, they react at various rates.

In addition, during the course of desulfurization, non-sulfur containing molecules may be hydrogenated and in some cases cracked.

The flow design of hydrosulfurization process systems is relatively simple. Preheated oil and hydrogen under pressure are contacted with catalyst. The effluent from the reactor is passed to one or more separators to remove most of the effluent hydrogen and light hydrocarbon gases produced in the operation. These gases are generally recycles with or without prior removal of light hydrocarbons by absorption. The separator liquids may be stripped, rerun or otherwise treated to obtain hydrogen sulfide free products of the desired boiling range.

Except in the case of residuum processing, plant design options are few in number and relatively simple. For example, in the processing of distillates, correlation systems have been developed which relate degree of desulfurization to about three parameters which define the charge stock, reactor temperature, temperature, pressure, feed space velocity, hydrogen rate and a catalyst activity parameter.

When residuum stocks are considered, however, generalizations are not so easily made. The wide variance in residuum properties (i.e., atmospheric or vacuum type, viscosity, Conradson carbon content, metal content and the paraffinic or aromatic nature of residuum) makes each case a special one as far as process design. Catalyst poisoning due to metals deposition on the catalyst surface can reduce overall desulfurization yields. Catalyst must then be regenerated or replaced, thus adding to overall cost of the particular system employed.

An alternative to desulfurization exists, that being the use of natural low sulfur fuel oils. They may be used alone or in blends with higher sulfur content material. The major source of low sulfur fuel oil is North African crudes, principally from Libya and Nigeria, and some Far Eastern crude from Sumatra. Fuels made from these crudes will meet even very low sulfur regulations calling for 0.5% sulfur or less. However, the highly waxy nature of these paraffinic materials makes handling difficult and costly. Therefore, the blend becomes a more palatable course of action.

Blends of natural low-sulfur fuels oils with other high sulfur fuel oils will be adequate in some cases to meet more moderate sulfur regulations. The fuel oil fractions of North African crudes contain about 0.3% S. Thus significant amounts of higher sulfur fuel oils can be added to make blends calling for 1–2% sulfur. These blends have physical properties which obviate the need for specialized handling (a must for existing industrial installations).

Before delving into specific desulfurization technology and applications, pertinent terms will be defined. Figure 1

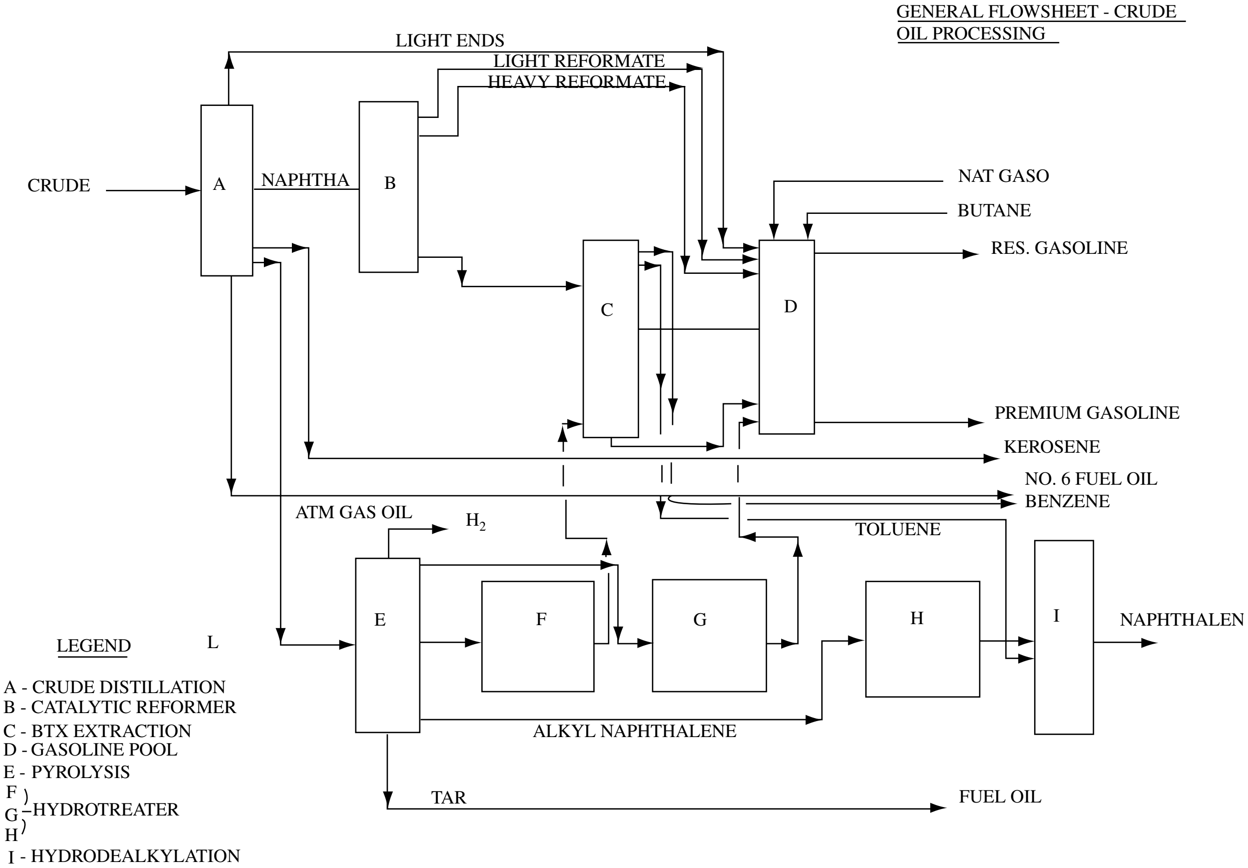


FIGURE 1

schematically represents a general flowsheet for crude oil processing. Crude oil, as received from the source is first atmospherically distilled. Light ends and mid-distillates from this operation are further processed to yield gasolines and kerosene. Atmospheric residuum can be directly used for No. 6 fuel oil, or further fractionated (*in vacuo*) to produce vacuum gas oil (vacuum distillate) and vacuum residuum. After atmospheric distillation, the average crude contains about 50% of atmospheric tower bottoms, which is nominally a 650°F$^+$ oil. The vacuum distillation yields roughly equal parts of vacuum gas oil and vacuum residuum. The bottoms from this unit is nominally a 975°F$^+$ oil, although the exact cut point will vary for each vacuum unit.

Desulfurization of vacuum residuum would be applicable where a refinery has use for the virgin vacuum gas oil other than fuel oil, and sulfur restrictions or increased prices make desulfurization of vacuum bottoms attractive. Another situation is where desulfurizing the vacuum gas oil and blending back with vacuum bottoms no longer produces a final fuel oil meeting the current sulfur specification.

Present in the residuum (vacuum) is a fraction known as asphaltenes. This portion is characterized by a molecular weight of several thousand. The majority of the organometallic compounds are concentrated in the asphaltene fraction. Although many of the metals in the periodic table are found in trace quantities, vanadium and nickel are usually present in by far the highest amounts. Residual oils from various crudes differ from each other considerably in regard to hydrodesulfurization. These differences reside to a great extent in the asphaltene fraction.

Light Oil Desulfurization

The G O-Fining Process The G O-Fining process is designed for relatively complete desulfurization of vacuum gas oils, thermal and catalytic cycle oils, and coker gas oil. It represents an extremely attractive alternative where a lesser degree of sulfur removal from the fuel oil pool and/or a very low sulfur blending stock is required. The feed to the G O-Finer System is atmospheric residuum. This stream is vacuum fractionated and the resulting vacuum gas oil (VGO) is desulfurized using a fixed bed reactor system. Resultant VGO is then reblended with vacuum bottoms to yield a desulfurized fuel oil or used directly for other applications. Figure 2 shows quantitative breakdown of various process streams for a 50,000 barrel per stream day (BPSD) operation utilizing a 3% sulfur Middle East atmospheric residuum feed. The process has the capability of producing 49,700 BPSD of 1.72% S fuel oil.

There are currently a number of G O-Fining units in commercial operation.

Investment and operating costs will vary depending on plant location and crude stock characteristics, but for many typical feedstocks (basis 50,000 BPSD) total investment is about 16.3 million dollars and operating costs average out at 60¢/barrel fuel oil (1989).

UOP's gas desulfurization process Another light oil desulfurization process is UOP's gas oil desulfurization scheme. Unlike the previously discussed G O-Fining process, UOP's scheme (already commercial) is designed for almost complete (∼90%) desulfurization of a 630 to 1050; F blend of light and vacuum gas oils (approximate sulfur content of feed—1.5%). Vacuum residuum is neither directly nor indirectly involved anywhere in the process.

In almost all other respects, however, UOP's process parallels G O-Fining. The current plant facility is of 30,000 BPSD capacity with above mentioned feed.

Comparison of UOP and G O-Finer costs show that both are of the same order of magnitude and differ markedly only in initial capital investment. This is in part attributable to the fact that a G O-Fining facility requires atmospheric residuum fractionation whereas UOP's does not.

Stocks of high-sulfur content are difficult to crack catalytically because all or most of the catalysts now in commercial use are poisoned by sulfur compounds. In recent years the trend has been toward processes that remove these sulfur compounds more or less completely. The high sulfur

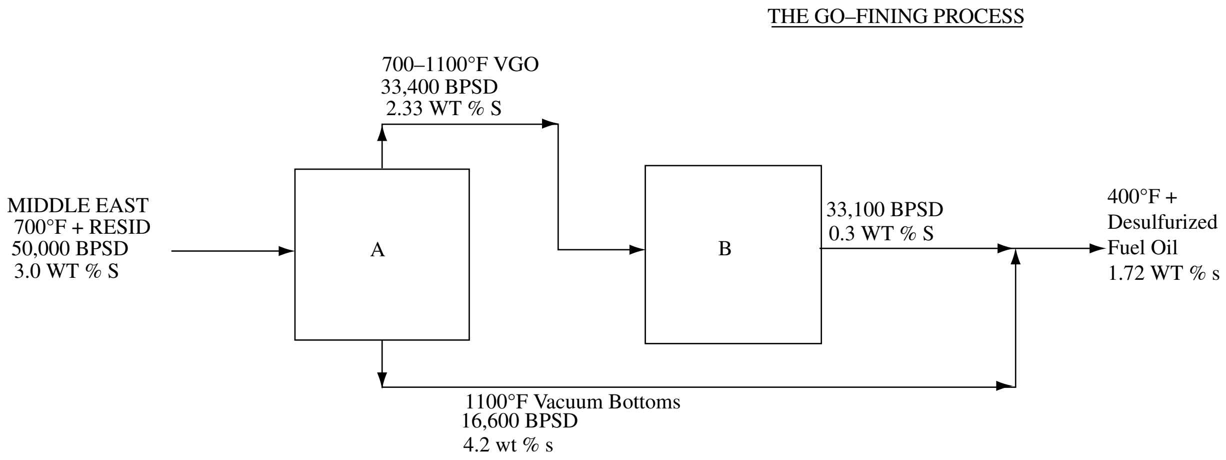


FIGURE 2

contents of petroleum stocks are mainly in the form of thiophenes and thiophanes and these can be removed only by catalytic decomposition in the presence of hydrogen. The Union Oil Company has developed a cobalt molybdate desulfurization catalyst capable of handling the full range of petroleum stocks encountered in refining operations. Even the more refractory sulfur compounds associated with these stocks are removed. This catalyst exhibited excellent abrasion resistance and heat stability, retaining its activity and strength after calcination in air at temperature as high as 1470°F.[8] Cobalt molybdate may be considered a chemical union of cobalt oxide and molybdic oxide, $CoO \cdot MoO_3$. The high activity of this compound is due to an actual chemical combination of these oxides with a resultant alteration of the spacing of the various atoms in the crystal lattice.[8] Catalyst life is two to five years. Catalyst poisons consisted of carbon, sulfur nitrogen and polymers. Regeneration is accomplished at 700 to 1200°F using air with steam or flue gases.

The fundamental reactions in desulfurization are as follows:

General Reaction

$$C_nH_mS_p + xH_2 \rightarrow C_nH_{m+2x-2p} + pH_2S$$

Desulfurization of ethyl mercaptan

$$C_2H_5SH + H_2 \rightarrow C_2H_6 + H_2S \ \Delta H = -19.56 \text{ kg cal/mole}$$

Desulfurization of diethyl sulfide

$$(C_2H_5)_2S + 2H_2 \rightarrow 2C_2H_6 + H_2S \ \Delta H = -36.54 \text{ kg cal/mole}$$

Desulfurization of thiophene

$$C_4H_4S + 4H_2 \rightarrow C_4H_{10} + H_2S \ \Delta H = -73.26 \text{ kg cal/mole}$$

Desulfurization of amylene

$$C_5H_{10} + H_2 \rightarrow C_5H_{12} \ \Delta H = -33.48 \text{ kg cal/mole.}$$

The change in heat content for all these reactions is negative, indicating that they take place with evolution of heat. The sulfur content in Middle East Gas Oil, a typical feed, is 1.25% by weight. The pilot plant data shows that the heat effect is not serious and whole process can be treated as isothermal.

The chemical reaction process on the catalyst is postulated to proceed on the surface of the catalyst by interaction of the sulfur-bearing molecules and hydrogen atoms formed through activated absorption of hydrogen molecules.[9] Oil molecules are more strongly absorbed than hydrogen molecules, and therefore may preferentially cover part of the surface, leaving less surface available for dissociation of hydrogen molecules. In the presence of diluent, namely, N_2, it can also compete for free sites on the surface, and accordingly may cause a reduction in the concentration of hydrogen on the surface, thus giving the lower rate constant when working with H_2–N_2 mixture.

Conversion of the sulfur compounds to hydrogen sulfide and saturated hydrocarbons occurs by cleavage of the sulfur to carbon bonds; essentially no C—C bonds are broken.

Residuum Desulfurization

The H-Oil-process (*Cities Service*) In order to meet the need for an efficient method of desulfurizing residual oils without the complexities encountered in the myriad of existing fixed bed catalytic systems, Cities Service developed what is known as the H-Oil system.

Although fixed bed catalytic reactors had been extensively used for desulfurizing distillate oils, desulfurization of residual oil in a fixed bed reactor presented several difficulties:

1) the high temperature rise through the bed tended to cause hot spots and coking,
2) the presence of solids in the feed and the formation of tar-like coke deposits on the catalyst tended to cause a gradual build-up of pressure drop over they catalyst bed and
3) because of the relatively rapid deactivation of the catalysts, system shut down for catalyst replacement occurred often, on the order of six times yearly.

To overcome these problems an ebbulated bed reactor was designed. Figure 3 is a simplified drawing of reactor workings.

The feed oil is mixed with the recycle and makeup hydrogen gas and enters the bottom of the reactor. It passes up through the distributor plate which distributes the oil and gas evenly across the reactor.

The reaction zone consists of a liquid phase with gas bubbling through and with the catalyst particles suspended in the liquid, and in random motion. It is a back-mixed, isothermal reactor, with a temperature gradient between any two points in the reactor no greater than 5°F.

Due to the catalyst suspension in liquid phase, catalyst particles do not tend to adhere to one another, causing blockage of flow. Any solids present in the feed pass directly through the reactor. Reactor pressure drop is constant.

One of the more important aspects of the ebbulated bed reactor system is that periodic shutdowns for catalyst replacement is not necessary. Daily catalyst replacement results in a steady state activity.

Table 2 shows examples of H-Oil desulfurization performance with atmospheric and vacuum residuals. In addition, investment and operating cost data are shown to illustrate the important effect of feed stock characteristics on overall economics.

Cases 1–3 describe processing of three atmospheric residual feeds. The Kuwait Residuum treated in case 1 is a high sulfur oil containing relatively low metals content (60 PPM). Therefore, the rate of catalyst deactivation is low and operating conditions are set to minimize hydrocracking and maximize desulfurization. In fact, only 2–3% naphtha and 9–10% middle distillate are produced. The actual chemical hydrogen consumption is fairly close to the estimated needed to remove the sulfur. For many atmospheric residuals which are not too high in metals, this case is typical to give maximum production of low sulfur fuel oil at minimum conversion and hydrogen consumption.

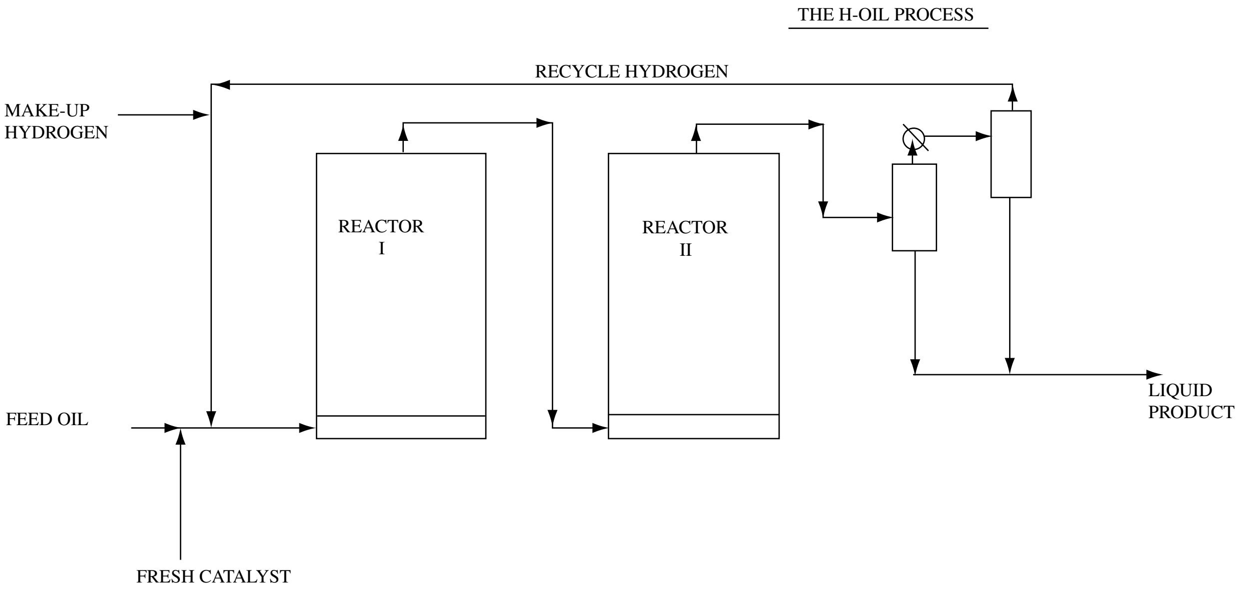


FIGURE 3

TABLE 2
H-OIL desulfurisation of atmospheric and vacuum residuals

Type-Feed (A-atmos) (V-vacuum)	Case 1A	Case 2A	Case 3A	Case 4V	Case 5V	Case 6V
Source	Kuwait	W. Texas	Venezuela	Kuwait	W. Texas	Venezuela
Feedstock data	—	—	—	—	—	—
Sulfur (Wt%)	3.8	2.5	2.2	5.0	2.2	2.9
Vanadium and nickel (PPM)	60	40	320	90	55	690
975ºF+	—	—	—	—	—	—
Vol%	45	45	52	80	70	75
Sulfur, Wt%	5.3	3.2	2.8	5.3	2.7	3.2
Yield, quality (400ºF+)	—	—	—	—	—	—
Vol%	99.3	96.0	94.2	94.7	92.9	92.9
% S	0.9	0.4	0.9	1.8	0.6	1.2
Chemical H_2	—	—	—	—	—	—
Consumption (SCI-/BBL)						
For S removal (est.)	290	210	140	340	170	200
Total	490	670	470	660	640	920
Economics (Relatives)	—	—	—	—	—	—
Capital inv. est.	6.7	7.8	6.9	7.9	8.3	8.9
OP cost, (20,000 BPSD UNIT)	33	40	39	44	46	65

In case 2, although metals content is also low (~40 PPM), hydrogen consumption is exceptionally high. This is due to the fact that conversion was not minimized and 7% naphtha and 13% middle-distillate was produced by hydrocracking.

Case 3 is characteristic of high metals content (~320 PPM) oils from that area. As noted previously, catalyst deactivation increases with metals content. Therefore, catalyst addition rates are higher, resulting in increased operating costs. To compensate for the reduced catalyst activity, higher operating temperatures and/or residence times are used.

Cases 4–6 summarize vacuum residua operations. Desulfurization rates for vacuum residua are lower than for atmospheric. The asphaltenes and metallic compounds reside in the vacuum residuum, consequently increasing catalyst deactivation rates and therefore catalyst costs per barrel. In all the cases depicted (4–6) hydrogen consumption, relative

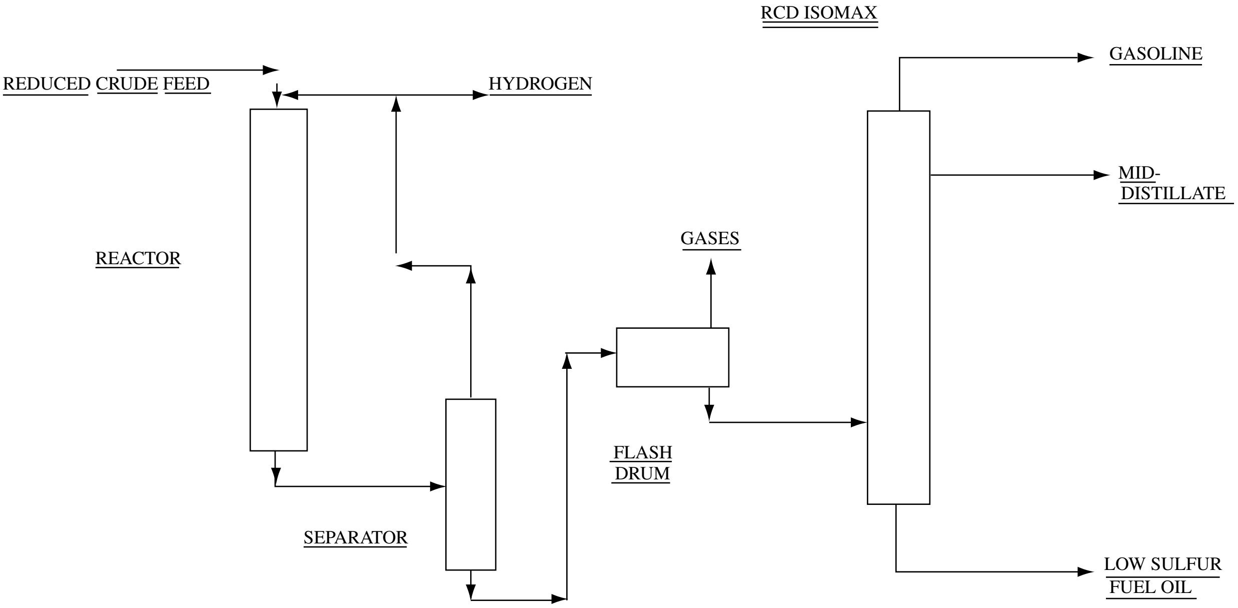


FIGURE 4

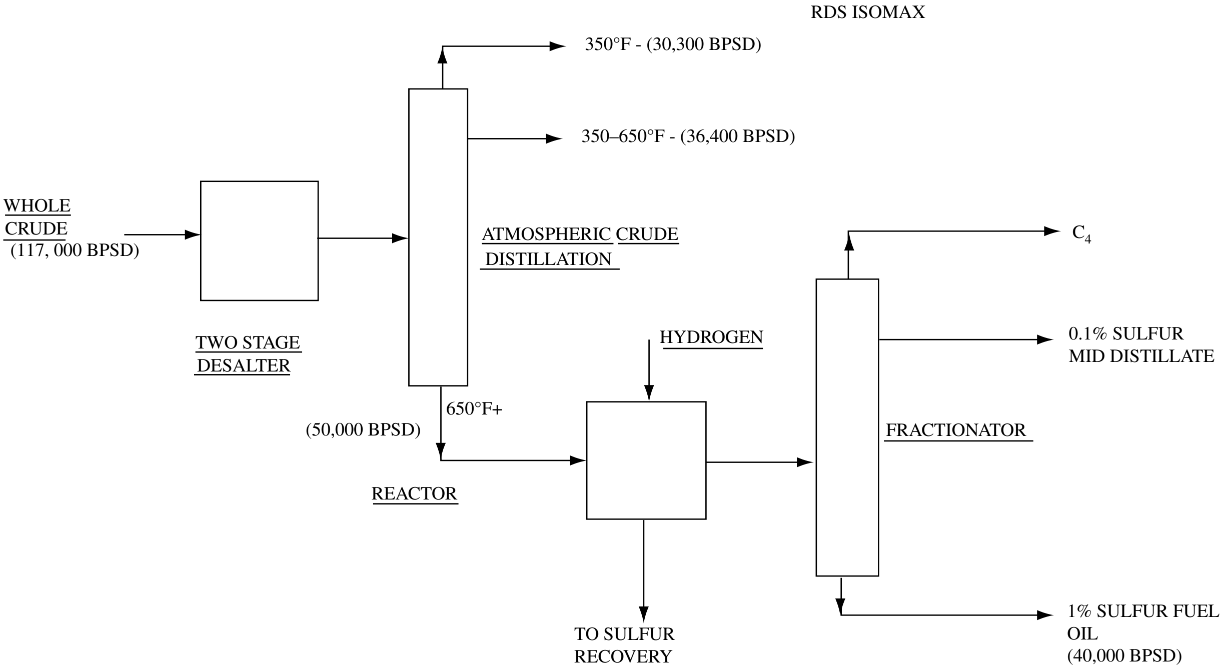


FIGURE 5

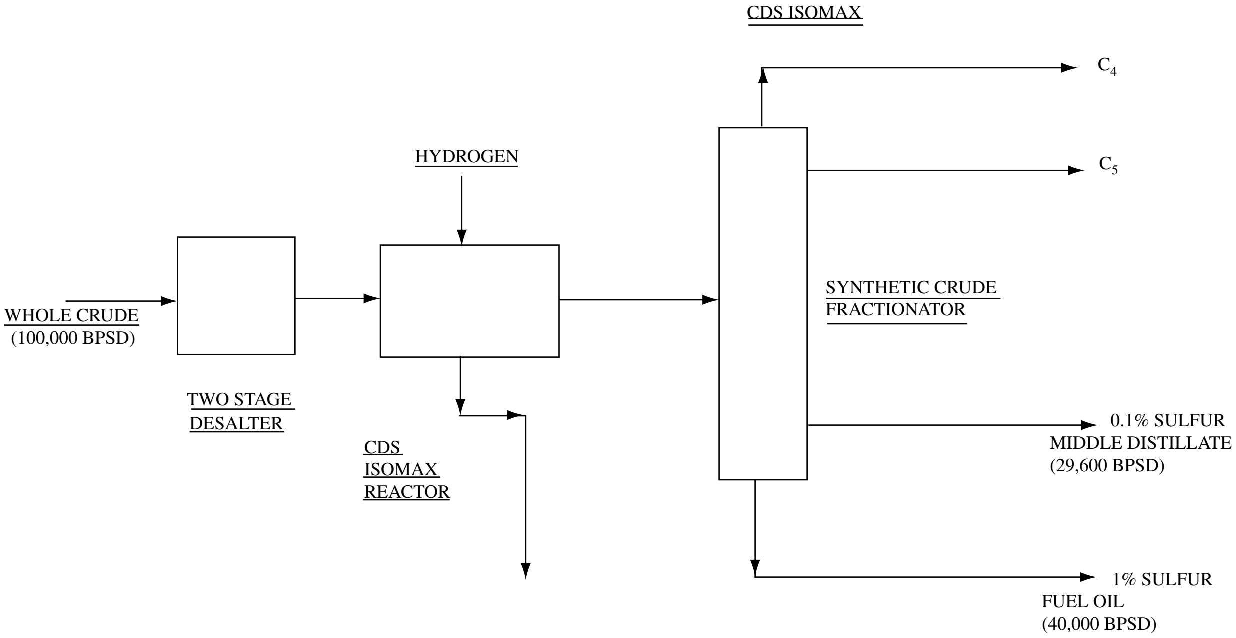


FIGURE 6

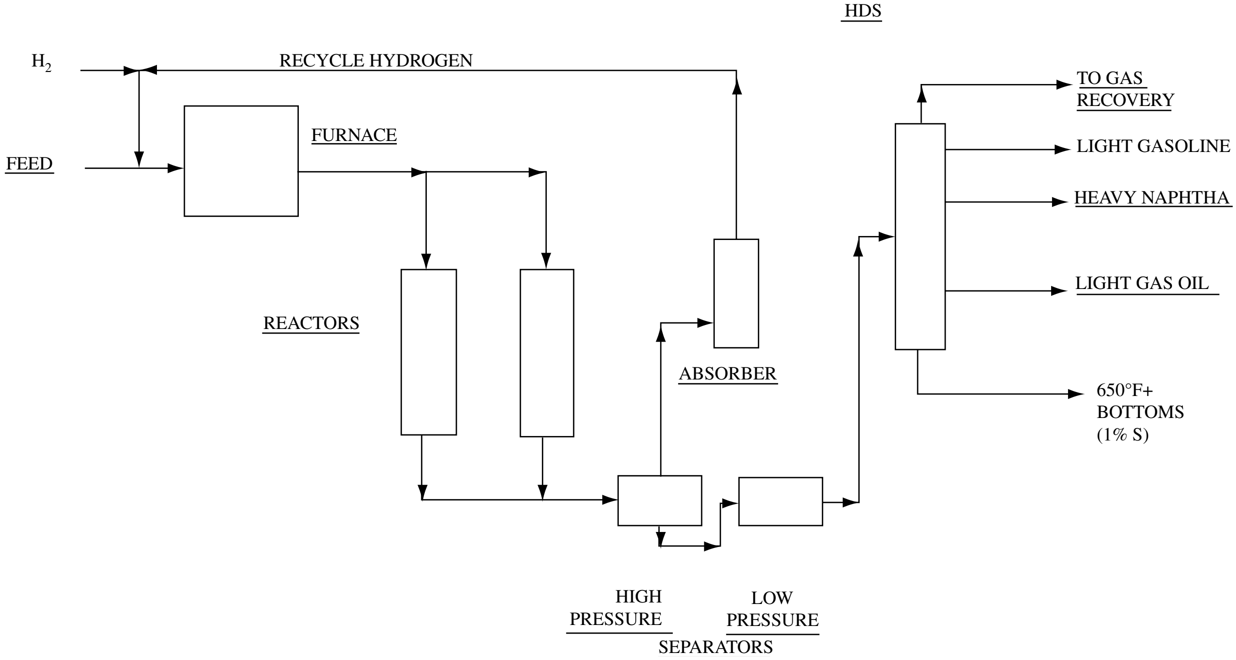


FIGURE 7

to that needed for desulfurization, is high indicating that high sulfur content of feed precludes setting of operating conditions to minimize conversion. In fact, naphtha production ranges from 7–15%, mid-distillates from 15–23%.

The Isomax processes A broad spectrum of fixed bed desulfurization and hydrocracking processes are now in operation throughout the world. They are characterized by their ability to effectively handle a wide range of crude feedstocks. In addition, some of the processes are capable of directly desulfurizing crude oil while others treat only residual stocks.

Rather than discuss each process individually, a comparative summary of the major ones is presented in Table 3.

There are many other processes which in one way or another effect a reduction in the amount of sulfur burned in our homes and businesses. All of them use some type of proprietary catalytic system, each with its own peculiar optimum operating ranges with regard to feed composition and/or reactor conditions.

The hydrodesulfurization process is still relatively expensive (in 1989 more than 75¢/BBL) by petroleum processing standards. The capital investment for large reactors which operate at high pressures and high temperatures, the consumption of hydrogen during the processing and the use of large volumes of catalyst with a relatively short life all contribute to the costs. In addition, processing costs also depend on the feedstock characteristics.

But when one considers the awesome annual alternative of 30 million tons of sulfur dioxide being pumped into the atmosphere, the cost seems trifling indeed.

Desulfurization of Natural Gas

Approximately 33% of the natural gas in the United States and over 90% of that processed in Canada is treated to remove normally occurring hydrogen sulfide. The recovered sulfur, which now accounts for about 25% of the free world's production is expected to increase in the future.

Current processes may be classified into four major categories:

1) Dry Bed—Catalytic Conversion,
2) Dry Bed—Absorption—Catalytic Conversion,
3) Liquid Media Absorption—Air Oxidation,
4) Liquid Media Absorption—Air Conversion.

Dry bed catalytic conversion (*the Modified Claus Process*) The Modified Claus Process is used to remove sulfur from acid gases which have been extracted from a main sour gas stream. The extraction is done with one of the conventional gas treating processes such as amine or hot potassium carbonate.

The process may be used to remove sulfur from acid gas streams containing from 15 to 100 mole % H_2S. The basic schemes use either the once through process or the split stream process. Figure 8 shows flow characteristics of the once-through scheme, which in general gives the highest overall recovery and permits maximum heat recovery at a high temperature level.

Split stream processes are generally employed where H_2S content of the acid gas is relatively low (20–25 mole %) or when it contains relatively large amounts of hydrocarbons (2–5%).

Pertinent design criteria for dry bed catalytic conversion plants include the following:

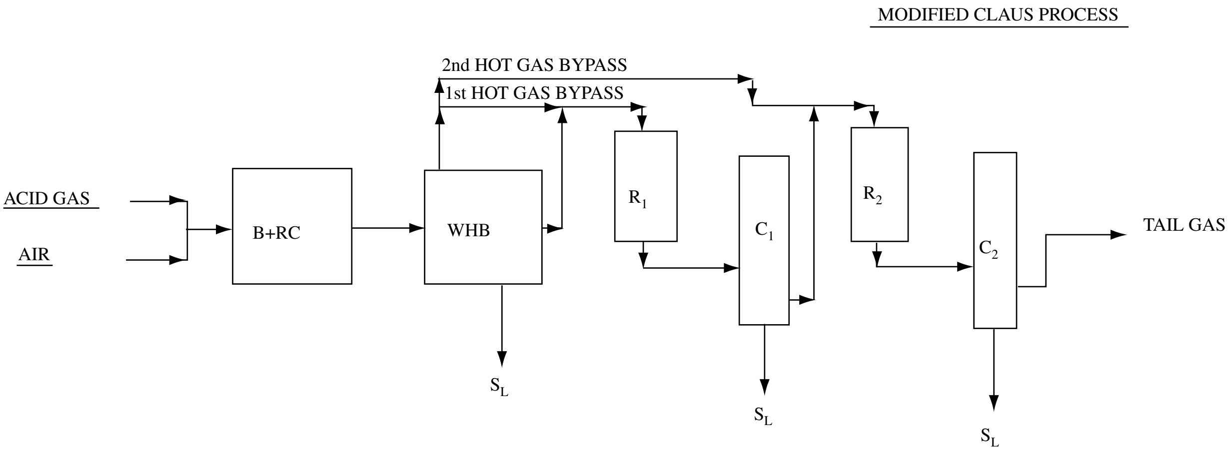
1) Composition of Acid Gas Feed,
2) Combustion of Acid Gas,
3) For a Once-Through Process, Retention Time, of Combustion Gases at Elevated Temperatures,
4) Catalytic Converter Feed Gas Temperature,
5) Optimum Reheat Schemes,
6) Space Velocity in the Converters,
7) Sulfur condensing Temperatures.

TABLE 3

Process	RCD Isomax	RDS Isomax	CDS ISomac	HDS
Licensers	UOP	Chevron	Chevron	Gulf R & D
General feed type	Atmospheric	Atmospheric	Whole crude	Residuum
Feed characteristics				
Name	Kuwait	Arabian light	Arabian light	—
Sulfur content	3.9	3.1	1.7	5.5
Process diagram	Figure 4	Figure 5	Figure 6	Figure 7
Fuel oil product				
Quantity (BPSD)	40,000	40,000	40,000	40,000
Sulfur content	1.0	1.0	1.0	2.2
Economies (Relative)				
Investment[a]	9.7	24.5	156.7	10.0
Operating costs[b]	51	40–60	40–60	—

[a] Includes only cost for Isomax reactor/distillation and auxiliary equipment.

[b] Includes utilities, labor, supervision, maintenance, taxes, insurance, catalyst, hydrogen, etc.

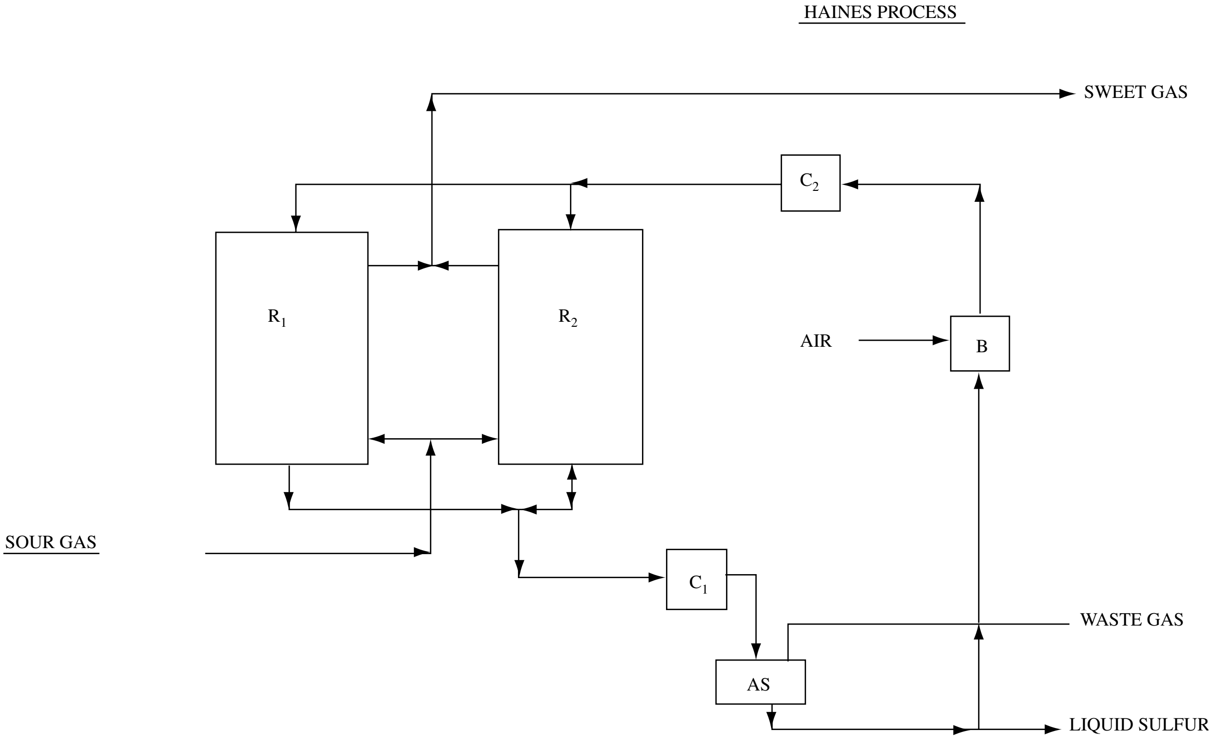


MODIFIED CLAUS PROCESS

LEGEND

B+RC - BURNER + REACTION CHAMBER
WHB - WASTE HEAT BOILER
R - CATALYTIC CONVERTOR
C - CONDENSER
S_L - LIQUID SULFUR

FIGURE 8



HAINES PROCESS

LEGEND

R - ZEOLITE BED ABSORBERS
C_1 - SULFUR CONDENSER
C_2 - ARIAL COOLER
AS - ACCUMULATOR/SEPARATOR
B - SULFUR BURNER

FIGURE 9

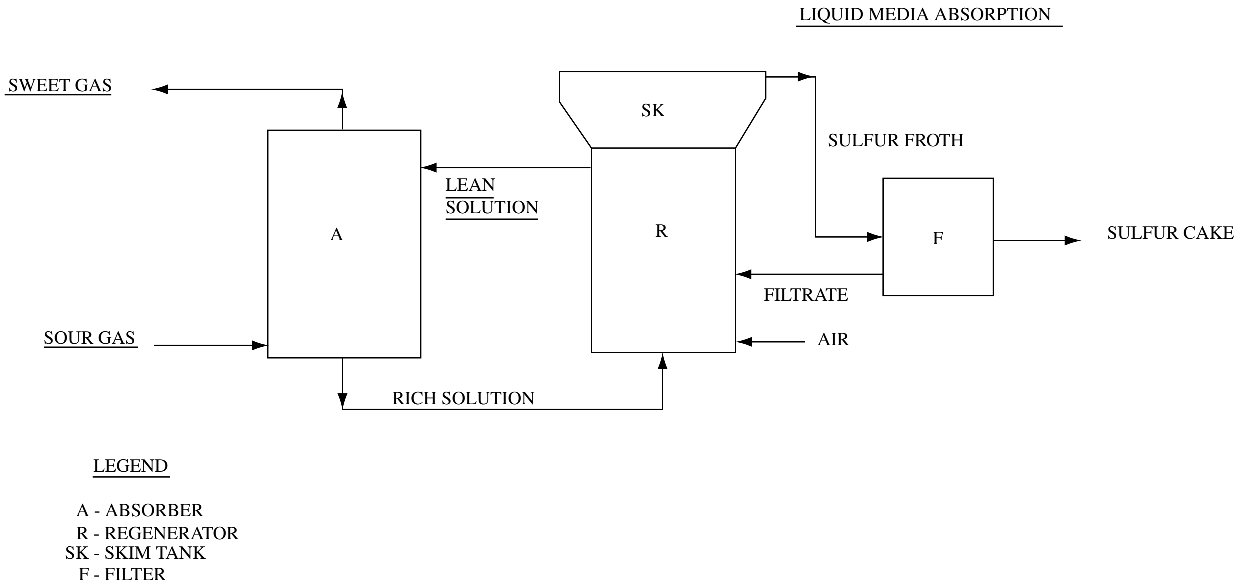


FIGURE 10

The sulfur from a claus plant may be produced in various forms, such as liquid, flaked or prilled. Form of choice is determined by transportation mode and end usage.

Dry bed absorption—catalytic conversion (*Haines Process*) In the Haines Process (see Figure 9), sulfur is removed from sour natural gas using zeolites or molecular seives.

Hydrogen sulfide is absorbed from the gas on a bed of zeolites until the bed becomes saturated with H_2S. The bed is then regenerated with hot SO_2 bearing gases, generated from burning a part of the liquid sulfur production.

The zeolite catalyzes the reaction of H_2S form sulfur vapor. The regeneration effluent gases are cooled and sulfur is condensed.

Liquid media absorption—air oxidation Most commonly in use in Europe for removal and recovery of sulfur from manufactured gases such as coal gas or coke oven gas, several processes in this category are available today.

Typically, the process scheme involves absorption of H_2S in a slightly alkaline solution containing oxygen carriers. Regeneration of the solution is by air oxidation. The H_2S is oxidized to elemental sulfur. Figure 10 shows the typical processing scheme.

The air also acts as a flotation agent for the sulfur which is collected at the regenerated solution surface as a froth. The sulfur sludge is either filtered or centrifuged to yield a sulfur cake which may be recovered as a wet paste or dry powder.

Liquid media absorption—direct conversion (*Townsend Process*) The Townsend Process is perhaps one of the newest sulfur removal processes on the scene today. Still in formative stages of development, its potential advantages point to it directly challenging the other conventional system containing an amine plant, a dehydration plant and a claus sulfur plant. The Townsend plant offers the added advantage of recovering a higher percentage of the sulfur.

The Townsend Process uses an aqueous solution of an organic solvent such as triethylene glycol to contact the sour gas and simultaneously sweeten the gas, dehydrate the gas and convert the gas to elemental sulfur.

ENVIRONMENTAL CONCERNS

Trowbridge[10] states that reductions in lead content of fuel oil are occurring world wide. Unleaded mo gas originally introduced in California in 1975 is now available throughout the U.S., where permissible maximum lead levels were established at 0.1 grams/gal after 1986. Trowbridge describes additional environmental concerns and fuel treatment procedures which are summarized below:

- ozone formation—reduction of gasoline vapor pressure
- carbon monoxide emissions—add oxygenates to blend
- benzene emissions—reduce fraction in reformate by extraction for example.

DESULFURIZATION OF COAL

Although coal buring accounts for only about 1/4 of the nation's energy, approximately 2/3 of all the sulfur dioxide emitted in the United States is traceable to its usage. And since coal supplies far outstrip gas and oil reserves, interest in coal desulfurization is great.

Presently, however, there are no processes sufficiently developed, either technically or commercially, which have any significant impact on the industry. The following is, therefore, a brief summary of the more pertinent processes currently under evaluation.

A typical high-sulfur, bituminous coal (4.9%) sulfur may be characterized by the composition shown in Table 4.

The solvent refine process Solvent refined coal (SRC) is one name among others given to a reconstituted coal which has been dissolved, filtered, and separated from its solvent. It is free of water, low in sulfur, very low in ash, and sufficiently low in melting point that it can be handled as a fluid.

In the process (see Figure 11), coal is ground to 80% through a 200 mesh screen and slurried in an initial solvent oil. The slurry is pumped to a pressure of 1,000 lb/sq.in. and passed upwards through a heater to bring the slurry to a temperature of 45°C.

The flow rate is between a half and one space velocity. As the material leaves the heater, more than 90% of the carbon in the coal is in solution. (Anthracite coal is an exception.)

A small amount of hydrogen is introduced into the slurry prior to preheating. The hydrogen prevents repolymerization of the dissolved coal. A portion of the organic sulfur in the coal unites with hydrogen to form hydrogen sulphide, which is later separated as a gas. The unutilized hydrogen is recycled back to the preheater. After the coal has been dissolved, the ash is filtered at a pressure of about 50 lb/sq.in. The pyritic sulfur leaves the process mixed with the other separated mineral matter.

The filtrate or coal solution, now greatly reduced in ash and sulfur is flash evaporated to obtain the necessary recycle solvent. It has been demonstrated that no make up solvent is required. In fact, an excess of solvent oil, released from the coal structure itself during solvation is recoverable, if desired. The remaining hot liquid residue is discharged and cooled to form a hard, brittle solid of solvent refined coal.

Table 5 shows a comparative analysis of a raw coal (% S—3.27%) and the solvent refined process, along with rough economies.

Obviously, solvent refined coal represents a marked improvement over the raw starting material. Sulfur content is reduced by about 30%, Ash content by 93% and heating value (in BTU/LB) is significantly increased. However, the cost of these improvements produces a fuel whose cost is higher than low sulfur oil.

Devolatilization I carbonization to a low sulfur char United Engineers and Company have investigated the use of a fluidized bed desulfurization process. Figure 12 shows general flow scheme for the label scale process.

TABLE 4

Coal Analysis (on dry basis)	
Volatile combustible matter	42%
Ash	10%
Fixed carbon	48%
Moisture	—
Sulfur breakdown (by % —forms available)	
Sulfate	0.33
Pyritic	2.53
Organic	2.04
	4.9%

TABLE 5

	Raw coal	Solvent refined coal
Ash	6.91%	0.41
Carbon	71.31	89.18
Sulfur	3.27	0.95
Volatile matter	44.00	51.00
Heat content (BTU/LB)	13978	15956
Cost ($/Ton)	40.00	120.00
Cost (c/MMBTU)	156.00	469.00

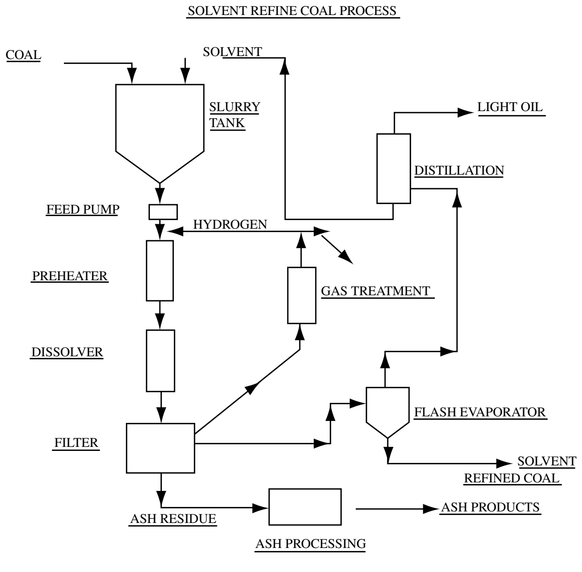


FIGURE 11

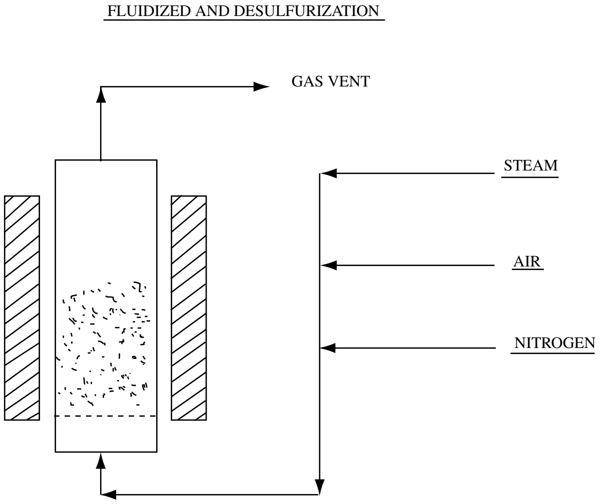


FIGURE 12

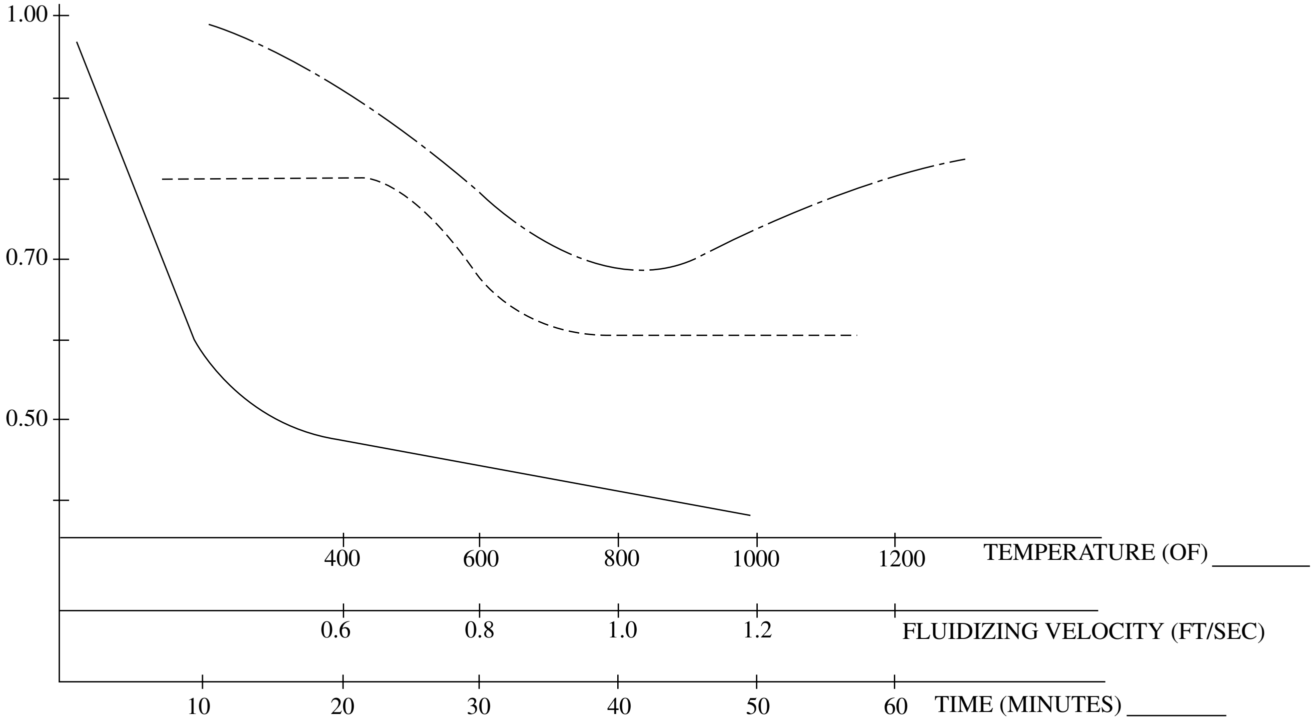


FIGURE 13

Ground coal (1/16″) is introduced into the heated, tubular reactor. Gas flow is then adjusted to maintain bed fluidity. Reactor temperatures approach 950°F.

The char is then removed through the bottom of the reactor.

The factors most directly affecting desulfurization are particle size, reactor residence time, reaction temperature and fluidizing velocity. Figure 13 summarizes relationships between extent of desulfurization (expressed as the ratio of sulfur in char to that in feed-ΔS) and above mentioned parameters.

Other processes, in various stages of development include causticized fluidized bed desulfurization, gasification to high and low BTU Gases and extr action of pyritic sulfur from raw coal. The Bureau of Mines, bituminous coal research and others have sought to remove sulfur (pyritic) by washing, using various techniques including centrifugation, flotation and magnetic separation methods. None of these has the potential to remove more than half the sulfur and each leads to significant product losses.

Bartok *et al.*[11] point out the advantages of systems which combine advanced processing techniques with advanced combustion recovery cycles. Gasification combined cycle, for example, couples the precombustion clean-up of coal (via gasification) with gas and steam turbines for power generation. As the efficiency of such plants improves their economics will improve relative to conventional coal fired plants while having a superior environmental impact (much lower nitrogen oxides, for example).

COAL ASH REMOVAL

Detroit Edison Company demonstrated the benefits of operating with low ash coals. In addition to the lower ash removal requirements, they lead to reduced transportation costs, higher heating values and improved boiler performance. One possible disadvantage of water cleaning method is additional drying which might be required prior to combustion.

Ash removal is considered primarily for coal production. Ash contents of oils are relatively small and of natural gas are negligible. The popular separation techniques usually depend on specific gravity differences (physical) or froth flotation (chemical) methods. The dispersing medium for specific gravity separation may be water, air or suspended matter in water. Coal specific gravities may range from 1.1 to 1.8, whereas impurities typically have specific gravities above 2.0 for carbonates and silicates and as high as about 5.0 for pyrites.

Bowling *et al.*[12] state that the levels of ash-forming mineral matter in most coals can be reduced by a combination of physical and chemical methods, to yield ultraclean coals with ash yields of 0.1–1%.

BLENDING WITH PETROLEUM COKE

To satisfy a worldwide power production growth rate of more than 2.5% a year, with even higher rates in the developing

Ultimate & Proximate Analyses and Ash Contend of Delayed PC, Typical Weight %, As Received

Species	Avg.	Range
Carbon	80	75–86
Hydrogen	3.3	3.0–3.6
Nitrogen	1.6	1.3–1.9
Sulfur	4.5	3.4–5.3
Ash	0.27	0.0–0.6
Oxygen	<0.1	0.0–0.1
Moisture	10.6	5.5–15.0
HHV, kBtu/ lb	13.5	12.6–14.5
Volatile Matter	10	8–6
Ash Analysis, ppm		
Vanadium	<2000	500–000
Nickel	336	250–50
Iron	84	50–50

nations in Asia and Latin America, nations are reviewing their fuel alternatives. One such fuel is petroleum coke (PC), being produced at a worldwide rate of over 50 tons per year in 2000. PC is a byproduct of oil refining that can be burned along with other fossil fuels. The cement industry is currently one of the widest consumers of blended PC. When co-firing liquid or gaseous fuel, PC will usually require additional control equipment for particulate and sulfur oxides reduction. In boilers cofiring coal particulate controls may be adequate since PC ash contents are low, however, the flue gas may require scrubbers to reduce sulfur oxides, since PC sulfur contents are high. See table for typical PC properties. The low ash content and low grindability reduce solids handling costs. The high vanadium content may, however, contribute to secondary plumes problems in the absence of sulfur removal equipment.

REFERENCES

1. Myerboff, A.A., *American Scientist* **69,** 624 (1981).
2. Lynch, M.C., *Chem. Eng. Prog.* **84,** 20–25 (Mar. 1988).
3. Balzhiser, R.E., *Chem. Eng.* 74–96 (Jan. 11, 1982).

3a. Hansen, Teresa, *Electric Power and Light,* November 1996, pp. 15–16.

4. Finfer, E.Z., *J. Air Poll. Control Asso.* **15,** 485–488 (1965).
5. Harris, R.D. and R.G. Moses, “Coal Desulfurization” at MECAR Symposium, Feb. 24, 1966.
6. Beckberger, L.H., E.H. Burk, J.H. Frankovich, J.O. Siemssen, and J.S. Yoo, EPRI Report #CS-3902, Feb. 1985, Elect. Pwr. Res. Inst., Palo Alto, CA 94304.
7. Parkinson, J.W. and E.R. Torak, EPRI Report #CS-3808, Feb. 1985, Elec. Pwr. Research Inst., Palo Alto, CA 94304.
8. Bartenope, D.P. and E.N. Ziegler, *Enc. of Env. Sci. and Eng.* **2,** 837 (1975), 1st Edition, Gordon and Breach Sci. Pub., New York (London).
9. Hoog, H.J. *Inst. Petrol.* **36,** 738 (1950).
10. Trowbridge, T.F., *Chem. Eng. Prog.* **84,** 26–33 (Mar. 1988).
11. Bartok, W., R.K. Lyon, A.D. McIntyre, L.A. Ruth and R.E. Sommerlad, *Chem. Eng. Prog.* **84,** 54–71 (Mar. 1988).
12. Bowling, K.M., H.W. Rottendofr and A.B. Waugh, *J. Inst. of Energy* **155,** 179–184 (1988).
13. Genereux, R.P. and Doucette, B., Power, July/August 1996.

EDWARD N. ZIEGLER
Polytechnic University

FRESH WATER: see EUTROPHICATION, WATER—FRESH

GASEOUS POLLUTANT CONTROL:
see VAPOR AND GASEOUS POLLUTANT FUNDAMENTALS

G

GEOGRAPHIC INFORMATION SYSTEMS

The term "Geographic Information System" (GIS) varies as a matter of perspective, and ranges in scope from specific computer software packages to software, hardware and data, to software, hardware, data, and support personnel. The most exhaustive definition is given by Dueker and Kjern as:

> Geographic Information System—A system of hardware, software, data, people, organizations, and institutional arrangements for collecting, storing, analyzing, and disseminating information about areas of the earth.

While this is an all-inclusive definition, the software packages at the heart of a GIS have their roots in the work of two researchers at Ohio State University, Marble and Tomlinson, in the mid to late 1960s. These men coined the phrase "Geographic Information System," and defined a GIS as having the following four components:

1. Data Input and the ability to process data.
2. Data Storage and Retrieval with the ability to edit.
3. Data Manipulation and Analysis.
4. Data Reporting Systems for the display of tabular and graphic information.

Data input, storage/retrieval, and reporting were (and still are) common to two other computer software packages—Computer Assisted Mapping (CAM) and Database Management Systems (DBMS) which separately managed graphical and tabular information. The development of GIS combined these packages to provide what many refer to as "intelligent maps," which are maps with extended information. Extended information can include information from areas such as census, tax assessment, natural resources availability/quality, which may be linked to a map, but are managed as part of a separate database. However, the true distinction of a GIS is Marble and Tomlinson's third component—the ability to analyze spatial information. The analysis capability enables a GIS to automatically evaluate information from several sources as a function of their spatial context. CAM systems may provide some of the same information as a GIS through a series of separate maps, but they require manual interpretation.

There are six basic analytical questions which many GIS software packages are able to address, partially or in full. These are:

1. Location—"What is it?"—What types of features exist at a certain place, such as "What is the population of a given census tract?"
2. Condition—"Where is it?"—Finding a site with certain characteristics, such as "Which agricultural fields are within 100 meters of a stream?"
3. Trend—"What has changed?"—Evaluation of spatial data as a function of time.
4. Routing—"Which is the best way?"—A variety of problems to determine paths through a network, such as finding the shortest path or optimum flow rates.
5. Pattern—"What is the pattern?"—A function which allows environmental and social planners to account for spatial distribution. Examples include the spread of diseases, population distribution versus urban development, or targeting specific consumer trends.
6. Modeling—"What if?"—Allow model development and evaluation, including "Monte Carlo" evaluation, where a variety of factors influence a situation, and their relationship is determined by varying one factor while holding the others constant.

A GIS is able to perform the above operations because they maintain the topology, or spatial associations, of the elements of their database. For example, an individual looking for 617 East Central Avenue on a map is able to see that this address is between 6th and 7th streets, on the south side of the street. Furthermore, if the individual wants to drive to this address and the map indicates that these streets are one-way in given directions, they are able to plan their route accordingly. A GIS may perform the same operation because it has the same information—that 6th and 7th intersect Central Avenue, the directed flow of traffic on these streets, and the distinction of areas on the left and right side of a line. This case is an example of what is called a "vector format" GIS, where space is represented by a two dimensional manifold of polygons (city blocks) which are bounded by a set of directed line segments (streets).

The other format by which a GIS may represent spatial data is referred to as "raster format." In this format, an area or volume is represented by a two or three dimensional matrix of uniform data elements, called "grid cells." Grid cells are the smallest areas that this type of GIS may resolve and are artificial constructs uniform in size and shape and are considered to have homogeneous data properties. An example would be a Soil map represented as a series of 30 meter by 30 meter plots of ground, where each plot is considered to be one particular type of soil, or a uniform blend of multiple soil types. An advantage of the raster format is the ability to show gradual changes and trends over a spatial area, because the matrix indices of the grid cell coordinates inherently imply the topology of the GIS data set. One of the disadvantages of the raster format is that data in the "real world" may not conveniently break down into uniform grid cells. Consequently, using this format involves an optimization problem to select a grid cell resolution that adequately represent the given data set, while minimizing the database size. Decreasing the size of gird cells in a raster exponentially increases the database size.

With the growing usage of Geographic Information Systems, there is a corresponding need for standardization. Standardization does not mean that all GIS databases contain the same format and quality of information; instead it provides a common terminology and format for maintaining appropriate metadata, or data about data. Currently in the United States, the official federal standard regarding GIS is the Spatial Data Transfer Standard, which provides a specific terminology for GIS through a set of definitions and format for maintaining an appropriate spatial data quality report for metadata. The importance of providing a standard, such as the SDTS, is that it provides the end user with and ability to assess whether or not the output from a GIS has the quality necessary to carry out further operations or analysis.

Dueker and Kjern's definition of a GIS is the most appropriate to consider, because the successful adoption of a GIS does not rely only on purchasing a software package. A GIS relies on people gathering the appropriate data, which is then input and manipulated in the GIS software environment, and which has knowledgeable individuals interpreting and analyzing results for their acceptability. Therefore, these elements must be considered as part of a Geographic Information System.

REFERENCES

Cooke, D.F., 1992, *Spatial Decision Support System: Not Just Another GIS, Geo Info Systems,* May 1992, pp. 46–49.

Cowen, D.J., 1988, *GIS versus CAD versus DBMS: What Are the Differences Photogrammetric Engineering and Remote Sensing,* Vol. 54, No. 11, pp. 1551–1555.

Dueker, K.J., 1987, *Geographic Information Systems and Computer Aided Mapping, APA Journal,* Vol. 53, No. 3, pp. 383–390.

Joffe, B.A. and B.S. Bergstrand, 1992, *Managing Engineering Drawings in a GIS Environment, Geo Info Systems,* April 1992, pp. 58–63.

Huxhold, W.E., 1991, *Information in the Organization, An Introduction to Urban Geographic Information Systems,* Oxford University Press, New York, pp. 3–24.

National Institute of Standards and Technology (NIST), 1992, *Federal Information Processing Standard Publication 173: Spatial Data Transfer Standard, Part 1,* U.S. Department of Commerce.

TODD HEPWORTH
University of Wisconsin

GREENHOUSE GASES EFFECTS

INTRODUCTION

The possibility that man-made emissions of carbon dioxide and other infra-red absorbing gases may enhance the natural greenhouse effect and lead to a warming of the atmosphere and attendant changes in other climate parameters such as precipitation, snow and ice cover, soil moisture and sea-level rise, constitutes perhaps the most complex and controversial of all environmental issues and one that is likely to remain high on both the scientific and political agenda for a decade or more. The issues have been obscured by a good deal of exaggeration and distortion by the media, and by some scientists, so that governments, the public, and scientists in other disciplines are confused and sceptical about the evidence for global warming and the credibility of the predictions for the future.

Until very recently the atmospheric concentrations of carbon dioxide had been increasing and accelerating since regular measurements began in 1958. Only during the last few years has there been a levelling off, probably because of the world-wide recession, the run-down of industry in the former Soviet bloc, and the substitution of gas for coal. This pause is likely to be only temporary, and if the concentrations resume their upward trend, they will eventually lead to significant climate changes. The important questions concern the likely magnitude and timing of these events. Are they likely to be so large and imminent as to warrant immediate remedial action, or are they likely to be sufficiently small and delayed that we can live with them or adapt to them?

Careful reconstruction of historical records of near-surface air temperatures and sea-surface temperatures has revealed that globally-averaged annual mean temperatures have risen about 0.6°C, since 1860 (see Figure 1). There is general consensus among climatolologists that this can now confidently be ascribed to enhanced greenhouse warming rather than to natural fluctuations. The last decade has been the warmest of this century and 9 of the 10 warmest years have occurred since 1990. Moreover, as described later, any temperature rise due to accumulated concentrations of greenhouse gases may have been masked by a concomitant increase in concentrations of aerosols and by delay in the oceans.

ROLE OF CARBON DIOXIDE IN CLIMATE

Carbon dioxide, together with water vapor, are the two main greenhouse gases which regulate the temperature of the Earth of its atmosphere. In the absence of these gases, the average surface temperature would be −19°C instead of the present value of +15°C, and the Earth would be a frozen, lifeless planet. The greenhouse gases act by absorbing much of the infrared radiation emitted by the Earth that would otherwise escape to outer space, and re-radiate it back to the Earth to keep it warm. This total net absorption over the whole globe is about 75 PW an average of 150 Wm^{-2}, roughly one-third by carbon dioxide and two-thirds by water vapour.

There is now concern that atmospheric and surface temperatures will rise further, owing to the steadily increasing concentration of carbon dioxide resulting largely from the burning of fossil fuels. The concentration is now 367 ppmv, 31% higher than the 280 ppm which prevailed before the industrial revolution and, until very recently was increasing at 0.5% p.a. If this were to continue, it would double its pre-industrial value by 2085 AD and double its present value by 2135. However, if the world's population continues to increase at the present rate, the concentration of carbon dioxide may well reach double the present value in the second half of this century.

Future concentrations of atmospheric carbon dioxide will be determined not only by future rates of emissions, which can only be guessed at, but also by how the added CO_2 is partitioned between the atmosphere, ocean and biosphere. During the decade 1940–9, the rate of emission from the burning of fossil fuels and wood is estimated at 63 ± 0.5 GtC/yr (gigatonnes of carbon per year). The atmosphere retained 3.2 GtC (about half of that emitted), leaving 3.4 GtC/yr to be taken up by the oceans and terrestrial biosphere. Models of the ocean carbon balance suggest that it can take up only 1.7 ± 0.5 GtC/yr so that there is an apparent imbalance of 1.4 ± 0.7 GtC/yr. Some scientists believe that this difference can be accounted for by additional uptake by newly growing forests and by the soil, but this is doubtful, and the gap is a measure of the uncertainty in current understanding of the complete carbon cycle. Reliable quantitative estimates of the combined effects of the physical, chemical and biological processes involved, and hence of the magnitude and timing of enhanced greenhouse warming await further research.

Nevertheless, very large and complex computer models of the climate system have been developed to simulate the present climate and to predict the likely effects of, say, doubling to atmospheric concentration of CO_2, or of increasing it at an arbitrary rate. This approach bypasses the uncertainties

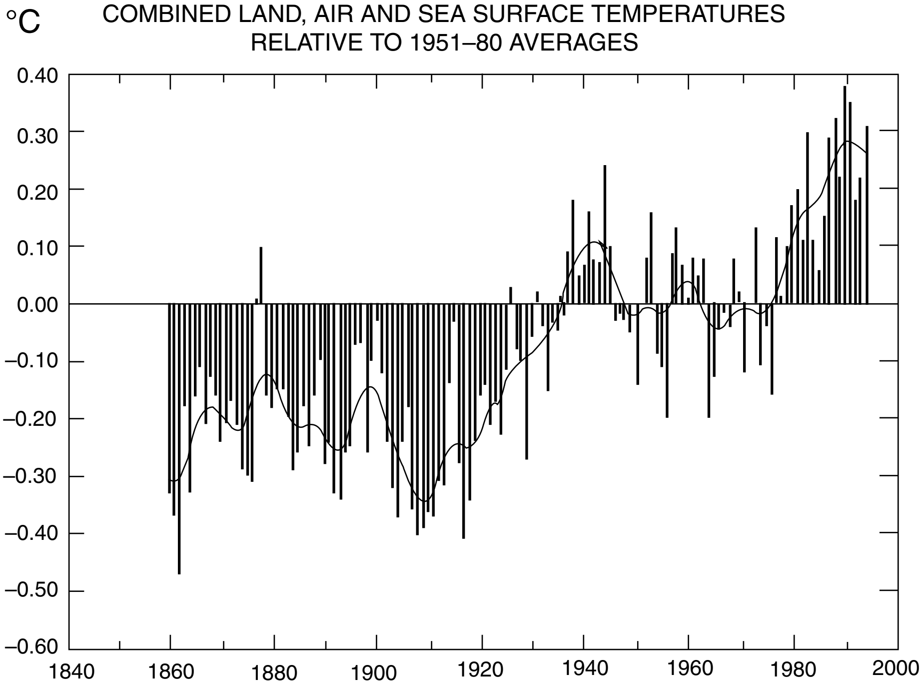


FIGURE 1 Observed changes in the globally-averaged surface temperatures from 1860–1991 relative to the 30-year mean for 1951–1980.

in future emissions and the natural regulation of atmospheric concentrations and is therefore unable to predict *when* the climate changes are likely to happen.

MODEL SIMULATIONS AND PREDICTIONS OF CLIMATE CHANGE

Introduction

Since changes in global and regional climates due to anthropogenic emissions of greenhouse gases will be small, slow and difficult to detect above natural fluctuations during the next 10 to 20 years, we have to rely heavily on model predictions of changes in temperature, rainfall, soil moisture, ice cover, sea level, etc. Indeed, in the absence of convincing direct evidence, concern over an enhanced greenhouse effect is based almost entirely on model predictions, the credibility of which must be largely judged on the ability of the models to simulate the present observed climate and its variability on seasonal, inter-annual, decadal and longer time scales.

Climate models, ranging from simple one-dimensional energy-balance models to enormously complex three-dimensional global models requiring years of scientific development and vast computing power, have been developed during the last 25 years, the most advanced at three centres in the USA and at the UK Meterological Office and, recently, at centers in Canada, France and Germany.

Until very recently, effort was concentrated on developing models (that evolved from weather prediction models) of the global atmosphere coupled to the oceans and cryosphere (sea and land ice) only through prescribing and up-dating surface parameters such as temperature and albedo, from observations. However, realistic predictions of long-term changes in climate, natural or man-made, must involve the atmosphere, ocean play, cryosphere and, eventually, the biosphere, treated as a single, strongly coupled and highly interactive system. The oceans play a major stabilizing role in global climate because of their inertia and heat storage capacity. Moreover, they transport nearly as much heat between the equator and the poles as does the atmosphere. The oceans absorb about half of the carbon dioxide emitted by fossil fuels and also absorb and transport a good deal of the associated additional heat flux and hence will delay warming of the atmosphere.

During the 1980s the UK Meterological Office (UKMO) developed one of the most advanced models of the global atmosphere coupled to a shallow mixed-layer ocean and used this to simulate the present climate and to study the effects of nearly doubling the present level of carbon dioxide to 600 ppmv. A general description of the physical basis, structure and operation of the model, of its simulations and predictions may be found in Mason (1989).

Simulation of the Present Climate

Models of the type just mentioned, the most important computed variables of which are:

E–W and N–S components of the wind
Vertical motion
Air temperatures and humidity

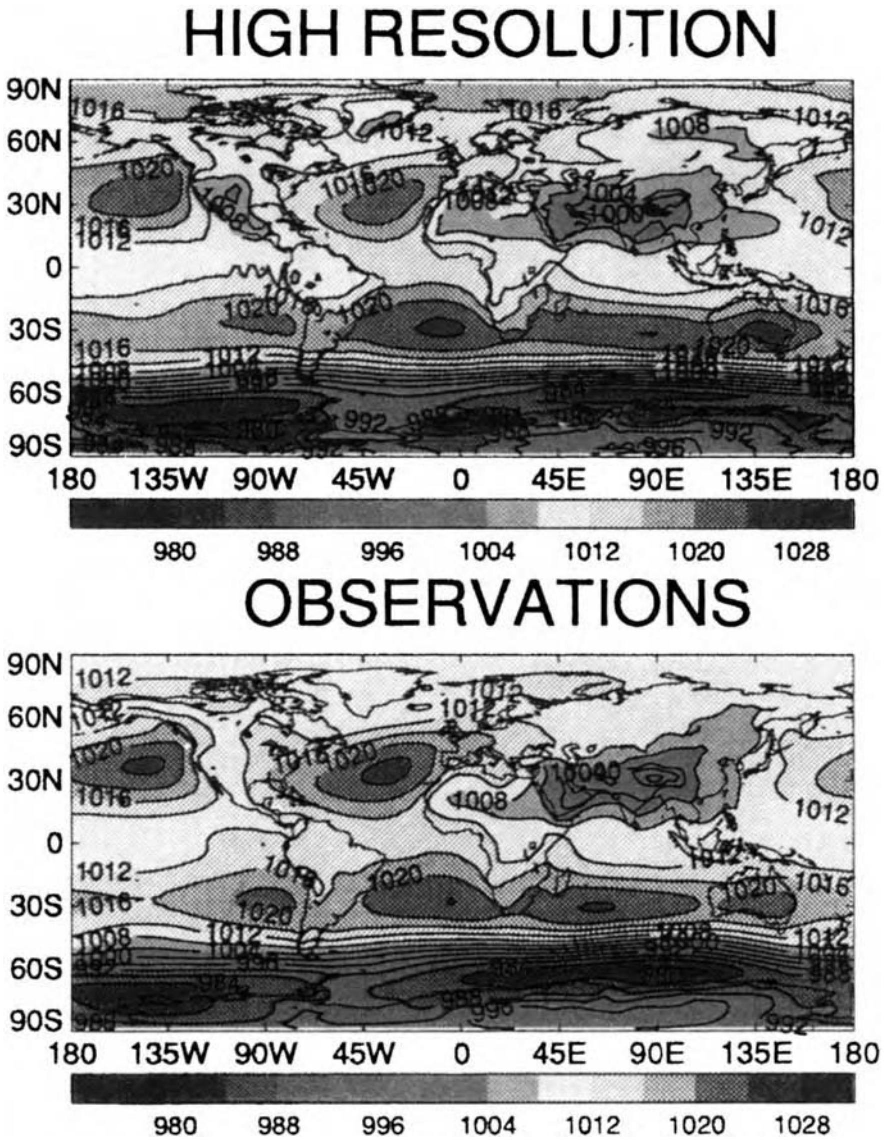


FIGURE 2 Simulation of the global mean surface pressure field for June, July and August by the UKMO climate model compared with observation.

Heights of the 11 specified pressure surfaces
Short- and long-wave radiation fluxes
Cloud amount, height and liquid-water content
Precipitation/rain/snow
Atmospheric pressure at Earth's surface
Land surface temperature
Soil moisture content
Snow cover and depth
Sea-ice cover and depth
Ice-surface temperature
Sea-surface temperature

are remarkably successful in simulating the main features of the present global climate—the distribution of temperature rainfall, winds, etc. and their seasonal and regional variations. They do, however, contain systematic errors, some different in different models, and some common to most. Identification of these errors and biases by comparison with the observed climate is important since these must be taken into account when evaluating predictions. These may not appear to be too serious in making predictions of the effects of a prescribed (e.g., man-made) perturbation since these involve computation of the *differences* between a perturbed and a control (unperturbed) simulation in which the systematic errors may largely cancel. However this linear reasoning may not necessarily be valid for such complex non-linear systems even if the perturbations are small, and the predictions will carry greater credibility if the control runs realistically simulate the observed climate and its variability.

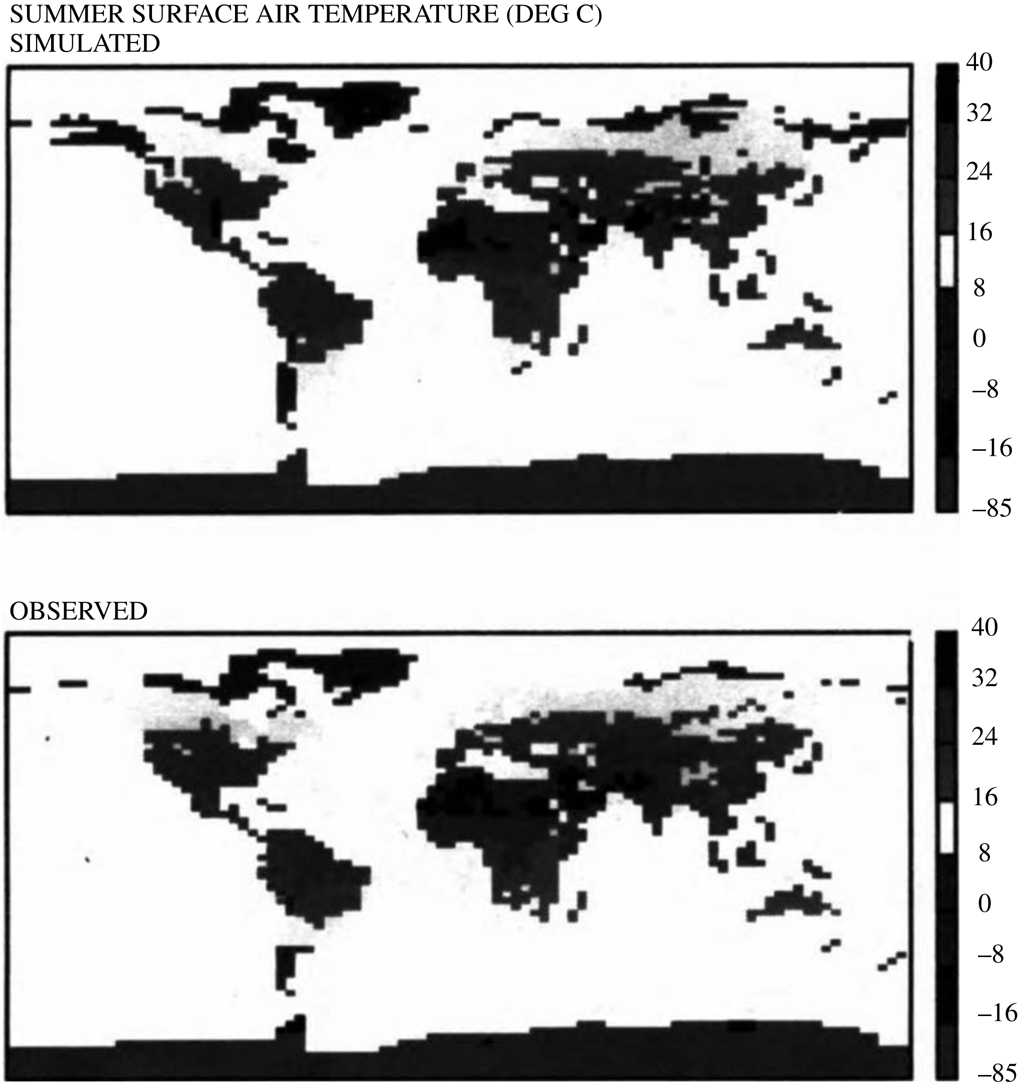


FIGURE 3 Model simulation of the mean near-surface temperatures over land for June, July and August compared with observation.

The main errors in model simulations of the present climate are discussed in IPCC (1992, 1996) and by Mason (2004). Simulations with the best models are close to reality despite the rather low model spatial resolution as illustrated by Figures 2 and 3.

Model Simulations of Ocean Climate

The role of the oceans in influencing climate and climate change is discussed in some detail in Mason (1993). Only the salient facts will be summarised here.

The oceans influence climate change on seasonal, decadal and longer time scales in several important ways. The large-scale transports of heat and fresh water by ocean currents are important climate parameters and affect the overall magnitude, timing and the regional pattern of response of the climate system to external forcing. The circulation and thermal structure of the upper ocean control the penetration of heat into the deeper ocean and hence the time delay which the ocean imposes on the atmospheric response to increases of CO_2 and other greenhouse gases. The vertical and horizontal motions also control the uptake of CO_2 through the sea surface and thus influence the radiative forcing of the atmosphere.

If *ocean models* are to play an effective role in the prediction of climate change, they must simulate realistically the present circulation and water mass distribution and temperature fields and their seasonal variability. Ocean modelling and validation are less advanced than atmospheric modelling, reflecting the greater difficulty of observing the interior of the ocean and of inadequate computer power. They suffer from inadequate spatial resolution, problems in parameterizing sub-grid-scale motions, and in estimating the fluxes of heat, moisture and momentum across the air/sea interface.

When forced with observed surface temperatures, salinities and wind stresses, ocean models have been moderately successful in simulating the observed large-scale circulation and mass distribution, but most models underestimate the meridional heat flux and make the thermocline too deep, diffuse and too warm.

The deeper ocean is also driven, in part, by fluxes of radiant heat, momentum, and of fresh water derived from precipitation, river run-off and melting ice, but measurements of all these are difficult and very sparse at the present time. Different models show considerable differences in their simulations of the deep ocean circulation, but identification of systematic errors is hardly possible because of the paucity of observations. The distribution of temperature and salinity are the primary sources of information for checking model simulations, but it is very difficult to simulate the salinity field because the distribution of sources and sinks of fresh water at the surface is so complex.

Perhaps the most effective way of checking ocean models on decadal time scales is to see how well they simulate the horizontal spread and vertical diffusion of transient tracers such as tritium/He^3 and C^{14} produced in nuclear bomb tests. Current models simulate quite well their shallow penetration in the equatorial ocean and deep penetration in high latitudes but fail to reproduce the deep penetration at 30–50°N, probably because of inadequate resolution of the Gulf Stream and its interaction with the North Atlantic current. The computed poleward transport of heat and the transport across other designated vertical sections can be checked against hydrographic measurements being made from research ships as part of the World Ocean Circulation Experiment, as described in Mason (1993). Some detailed measurements are also being made on the seasonal variation in the depth of the mixed ocean layer and the thermocline that can be compared with the model simulations.

Coupled Atmosphere—Deep Ocean Models

The UKMO has developed a deep global ocean model coupled to its global atmospheric model to carry out long-period climate simulations and to make realistic predictions of climate changes produced by *gradual* increases of atmospheric CO_2 until it reaches double the present value. The results of the first of these enhanced CO_2 experiments, and of similar ones conducted elsewhere, are described in the following section.

Here we summarise the structure and operation of the coupled model, its problems and deficiencies, and the research in progress to overcome them. A more detailed analysis of the first version is given by Murphy (1995). In the latest version, the model atmosphere is divided into 19 layers (20 pressure levels) between the surface and 50 km with 5 levels in the surface boundary layer (lowest 1 km) to allow calculation of the surface fluxes of heat, moisture and momentum. There are also four levels in the soil to calculate the heat flux and hence the surface temperature. The variables listed in the previous section Simulation of the Present Climate are calculated on a spherical grid with mesh 2.5° lat × 3.75° long, about 7,000 points at each level.

The incoming solar radiation is calculated as a function of latitude and season, and diurnal variations are included. Calculations of radiative fluxes at each model level use four wavebands in the solar radiation and six bands in the long-wave infra-red, allowing for absorbtion and emission by water vapour, carbon dioxide, ozone and clouds. Sub-grid-scale convection is represented by a simple cloud model that treats the compensating subsidence and detrainment of air and the evaporation of precipitation. Precipitation is calculated in terms of the water and ice content of the cloud; cooling of the atmosphere by evaporation of precipitation is allowed for. Reduction in wind speed caused by the aerodynamic drag of mountains, oceans waves, and by the breaking of orographically-induced gravity waves are computed. In calculating changes in the extent and thickness of sea ice, drifting of the ice by wind-driven ocean currents is taken into account.

In the land surface model the different soil types and their differing albedos are specified, as are the different types of vegetation, their seasonal changes and their effects on evaporation, albedo, aerodynamic drag.

The ocean model computes the current, potential temperature, salinity, density and the transports of heat and salt at 20 unequally-spaced levels (depths) in the ocean, eight of these being in the top 120 m in order to simulate better the physics and dynamics in the active, well-mixed layer, its seasonal variation, and the surface exchanges of heat, moisture and momentum with the atmosphere. The vertical velocity at the sea floor is computed assuming flow parallel to the slope of the bottom topography specified on a 1° × 1° data set. The horizontal grid, 2.5° × 3.75°, the same as that of the atmospheric model, is too coarse to resolve oceanic meso-scale eddies of scale ~100 km which contain much of the total kinetic energy, but are crudely represented by sub-grid-scale turbulent diffusion and viscosity. The latter has to be kept artificially high to preserve computational stability with the penalty that the simulated currents, such as the Gulf Stream, are too weak. Lateral diffusion of heat and salt take place along ispycnal (constant density) surfaces using diffusion coefficients that decrease exponentially with increasing depth. The coefficients of vertical diffusion are specified as functions of the local Richardson number, which allows for increased mixing when the local current shear is large.

Coupling with the atmosphere is accomplished in three stages. The atmospheric model, starting from an initial state based on observations, is run on its own until it reaches an equilibrium climate. The ocean model, starting from rest and uniform temperature and salinity is also run separately, driven by the wind stresses, heat and fresh-water fluxes provided by the atmospheric model. This spin-up phase of the ocean takes place over 150 years (restricted by available computer time) during which a steady state is achieved in the upper layers of the ocean as they come into equilibrium with the atmospheric forcing. Finally, the ocean is coupled to the atmosphere, sea-ice and land-surface components

TABLE 1
Global mean changes in temperature and precipitation caused by doubling CO_2 in various models in "Equilibrium"

Model	T(°C)	P(%)	Remarks
UKMO (1987)*	5.2	15	
GDFL (1989)	4.0	8	
GISS (1984)*	4.8	13	Very low (8° × 10°) resolution
SUNY (1991)	4.2	8	
CSIRO (1991)	4.8	10	
NCAR (1991)	4.5	5	
Models with Computed Cloud Water/Ice			
UKMO (1989)	3.2	8	Fixed radiative properties
	1.9	3	Variable radiative properties as function of water/ice content

GDFL Geophysical Fluid Dynamics Laboratory, Princeton, USA
GISS Goddard Institute of Space Studies
SUNY State University of New York
SCRIO Commonwealth Scientific and Industrial Research Organization, Australia
NCAR National Center for Atmospheric Research, Boulder, USA

and run in tandem with two-way feedbacks between ocean and atmosphere transmitted at five-day intervals. Thus the atmospheric model is run separately for five days with unchanged sea-surface temperatures and sea-ice extents, accumulating relevant time-averaged surface fluxes, which are then used to drive the corresponding time step of the ocean model, following which the updated sea-surface temperatures and sea-ice cover are fed back to the atmosphere for the next iteration. When an internally consistent balance is obtained between all four main components of the climate system, the final state may be taken as the starting point for perturbation experiments such as the doubling of carbon dioxide.

MODEL PREDICTIONS OF CLIMATE CHANGES CAUSED BY DOUBLING PRESENT CONCENTRATIONS OF CARBON DIOXIDE

Introduction

We recall that atmospheric concentrations of carbon dioxide are likely to double by the second half of this century and that simple radiative calculations, allowing only for feedback from the accompanying increases in water vapour, indicate that this might cause the globally and annually averaged surface air temperature to rise by about 1.5°C. Because, as discussed by Mason (1995), many other feedback processes, both positive and negative, operate within the complex climate system, and because their effects are likely to vary with season, latitude and geographical location, firmer estimates can come only from model experiments in which the climate simulated by a model perturbed by the doubling of CO_2 is compared with that from an unperturbed (control) model, the differences being attributed to the enhanced CO_2.

We now compare and discuss the results of two types of experiments, produced by different models. In one set, involving a global atmosphere coupled to only a shallow ocean, the CO_2 concentration is doubled in one step and the climatic effects are assessed after the system has reached a new equilibrium. In the second set, in which the atmosphere is coupled to a multi-layered deep ocean, the CO_2 is allowed to increase at 1% p.a. compound and so doubles after 70 years.

Prediction of Global Mean Changes in the 'Equilibrium' Experiments

All six models cited in Table 1 comprise a global atmosphere with 9–12 levels in the vertical, coupled to a shallow (50 m deep) ocean with prescribed heat transport. The input solar radiation to all models follows a seasonal cycle, but only those marked with an asterisk include a diurnal cycle. All the models have a rather low horizontal resolution and all the experiments were run for <50 years. Furthermore, all of them *prescribe* the cloud amount and height by empirical formulae that relate cloud to relative humidity and are based on satellite observations of cloud. The radiative properties of the clouds (classified into low, medium and high-level categories) are also prescribed and remain fixed during the model simulation.

The predicted globally and annually-averaged increases in surface air temperature due to doubling of CO_2 are remarkably similar, ranging from 4.2°C to 5.2°C with an average of 4.6°C. This is probably because the sea-surface temperatures and sea-ice cover are constrained to be near observed values by adjusting the advective heat fluxes in the shallow ocean. The predicted increase in precipitation,

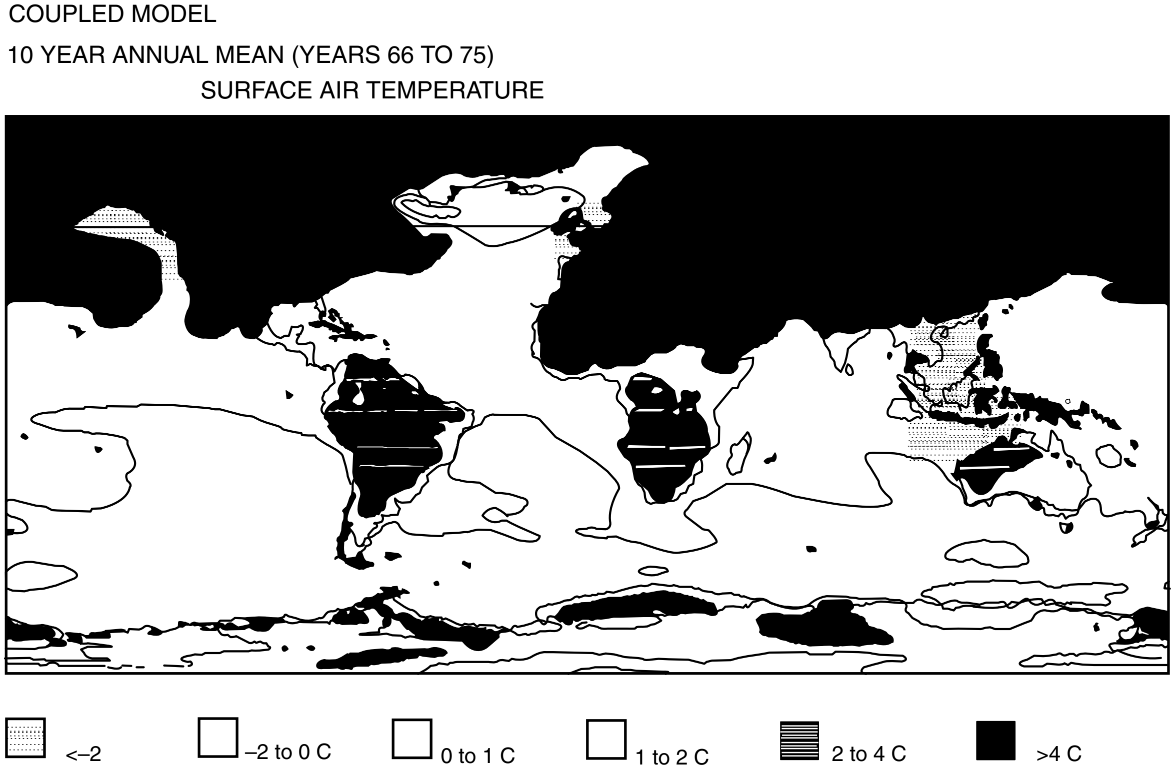


FIGURE 4 Prediction of the UKMO coupled atmosphere—deep ocean model of global warming caused by increasing the concentration of atmospheric carbon dioxide by 1% p.a. compound after 75 years.

not surprisingly, show a greater spread, from 5 to 15% with an average of 10%.

These predictions were not much affected by doubling the horizontal resolution (having the grid spacing). However, they were much more sensitive to the formulation of physical processes, in particular the representation of clouds and their interactions with solar and terrestrial radiation. Model simulations in which the cloud water was *computed* from the model variables and their radiative properties (emissivity, absorptivity and reflectivity) were allowed to vary with the liquid water and ice content produced significantly different results as summarized in Table 1.

The UKMO model, using three progressively more sophisticated and realistic cloud/radiation schemes, has progressively reduced the predicted global warming from 5.2°K to 1.9°K and the corresponding precipitation increases from 15% to 3%. It is important to identify and understand the underlying physical reasons for these results which, if confirmed, are likely to have an important influence on the whole GHW debate.

In the first version of the model, in which cloud cover was related empirically only to relative humidity and the radiative properties were fixed during the whole simulation, enhanced CO_2 produced unrealistic *decreases* in high-, medium- and low-level clouds, except at very high latitudes and, consequently, an exaggerated warming of the atmosphere. Decrease in cloud amount seems inconsistent with the predicted increase in precipitation and suggests that the empirically derived cloud cover was incompatible with the internal dynamics of the model. In the most sophisticated treatment, the cloud water is computed from the dynamical and physical equations; it is transformed progressively from liquid water to ice as the temperature falls below −15°C; rapidly growing ice crystals are allowed to fall out of the cloud; and the radiative properties are varied as a function of the cloud water path and the solar angle for the incoming solar radiation and as a function of the water/ice path for terrestrial long-wave radiation. In this case, enhanced CO_2 leads to a marked *increase* in the extent and optical depth of call clouds, and especially of low clouds in middle and high latitudes, which reflect more of the solar radiation to space and therefore reduce the GHW of the atmosphere to only 1.9°K. The small 3% increase in precipitation is consistent with a 2–3% increase in low cloud cover and a 2% increase in medium-level cloud in the Northern Hemisphere. A more detailed account is given by Senior and Mitchell (1993).

Transient Experiments in Which CO_2 Increases at 1% p.a.

The fact that we now have fully three-dimensional models of the global oceans coupled interactively to the atmosphere, land-surface and sea-ice components of the climate model, enabling

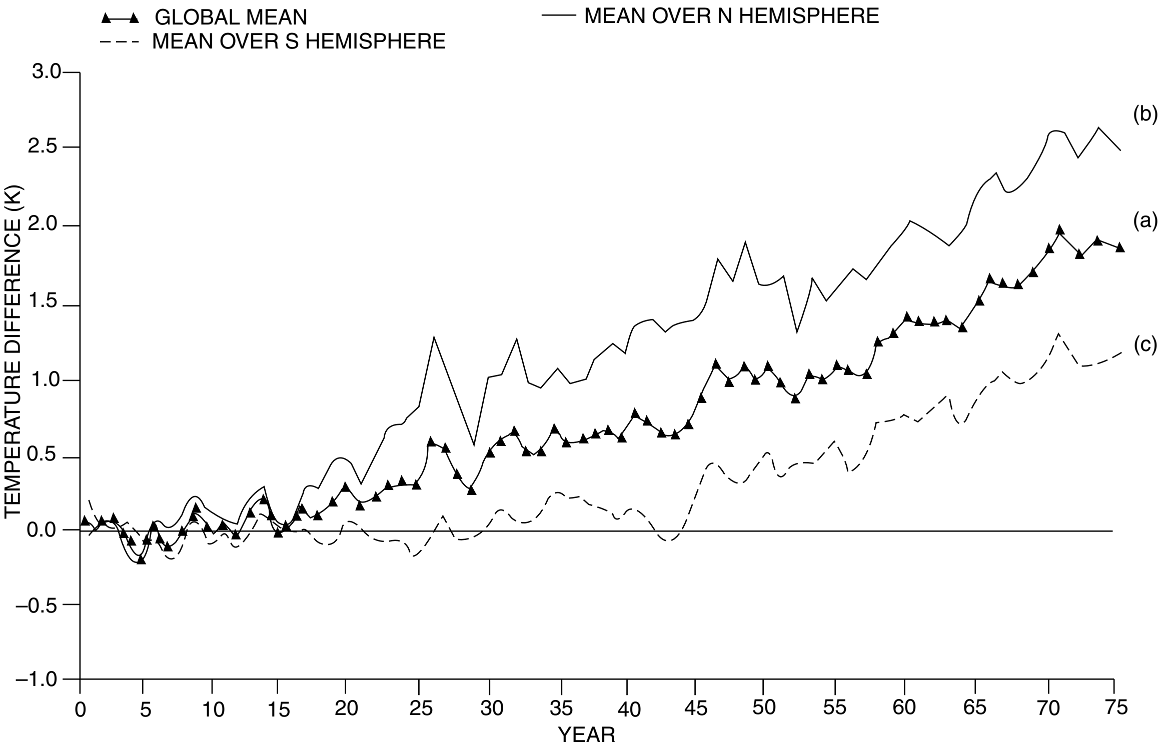


FIGURE 5 Predictions of globally—averaged warming caused by increasing the concentration of carbon dioxide by 1% p.a. compound over 75 years showing the year-to-year changes. The changes for the northern and southern hemispheres are shown separately.

more realistic simulations in which the carbon-dioxide, instead of being doubled in one step, is increased gradually at 1% p.a. compound to double after 70 years. On this time-scale, the atmospheric response will be influenced by changes occurring at depth in the oceans, and especially in the top 1 km.

The first results of such an experiment were published by Manabe *et al.* (1990) from GDFL. The globally and annually averaged increase in surface air temperature was 2.3 K, lower than in earlier models with a shallow ocean. The reduced warming was especially marked in the Southern Hemisphere, which showed little amplification in the Antarctic compared with the Arctic. This is explained by the ocean circulation in the southern oceans having a downward branch at about 65°S, which carries much of the additional 'greenhouse' flux of heat from the surface to depth of >3 km, where it remains for many decades.

Very similar results were produced with the earlier version of the UKMO model by Murphy (1990), Murphy and Mitchell (1995). The annually averaged response in global mean surface temperature to CO_2 increasing 1% p.a. over 75 years is shown in Figure 4, and also in Figure 5, which also shows the results for the hemispheres separately. Averaged over the years '66 –'77, the global mean warming was 1.7°K. The corresponding increase for the Northern Hemisphere was 2.6°K, with warming of >4°K over large areas of the Arctic. The UKMO model, like the GDFL model, shows that the much smaller response of the Southern Hemisphere is due to the transport of heat from the surface to depth in a strong down-welling circulation near 60°S. A similar vertical circulation, caused by melting ice, and penetrating to about 1.5 km depth, occurs at about 60°N in the North Atlantic (see Figure 6). After a slow start, the enhanced global warming settles down at about 0.3 K/decade. Moreover, the model exhibits variability on inter-annual and decadal time-scales; the peak-to-peak variation on the decadal scale being about 0.3°K—of the same magnitude as the predicted signal due to 'greenhouse' warming.

A similar long-term run with a coupled atmosphere—deep ocean model has been carried out at the Max Planck Institute in Hamburg by Cusbasch *et al.* (1992). CO_2 is allowed to increase rather more rapidly to double after 60 years and produces a global mean warming of 1.3°K, the lowest value so far reported.

The transient responses to the doubling of CO_2 by all three models, ranging from 1.3 to 2.3 K, correspond to about 60% of the expected equilibrium response. This implies a lag of about 30 years due largely to the delaying effect of oceans.

The predicted changes in precipitation, though small on average, are far from uniformly distributed. The UKMO model indicates increases in high latitudes of the Northern Hemisphere throughout the year, in middle latitudes

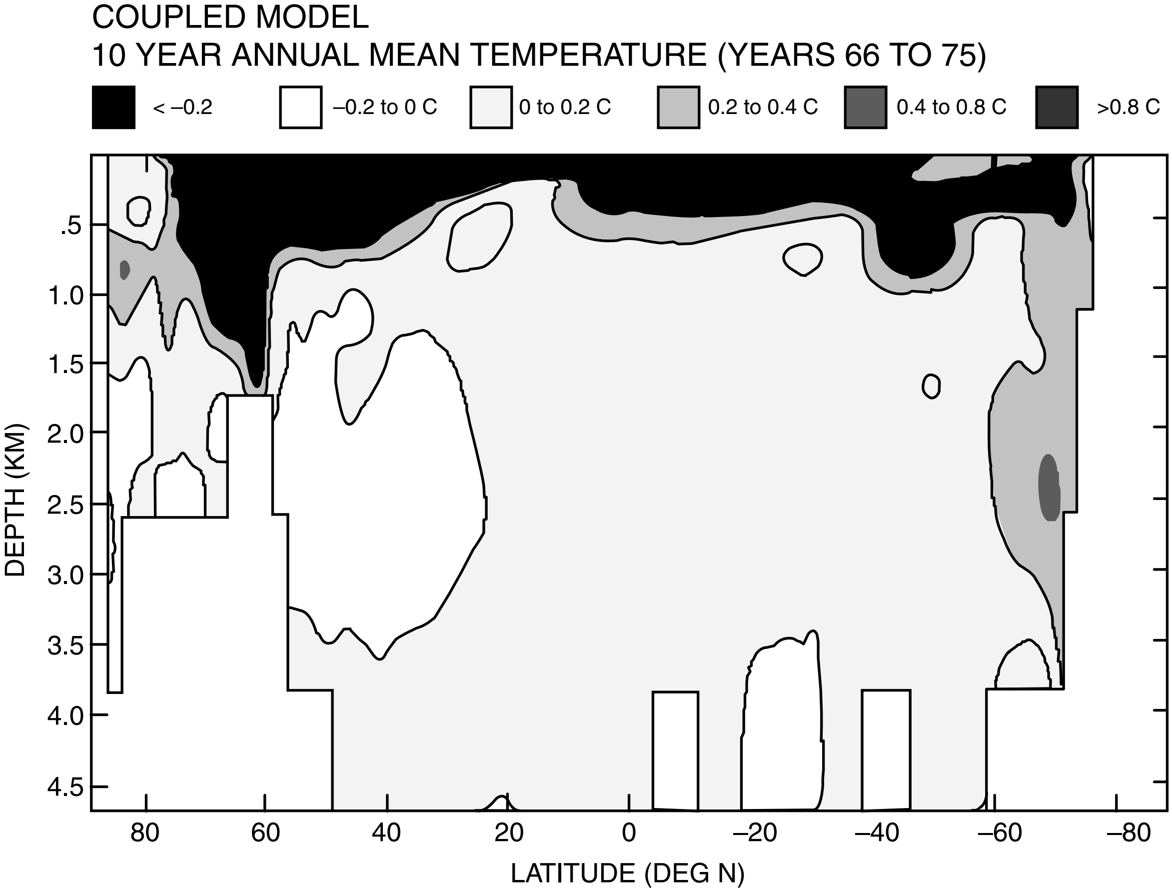


FIGURE 6 Changes in the ocean temperatures averaged around latitude bands and shown as a function of depth after the carbon dioxide has doubled in the model experiment of Figures 4 and 5. These range from about 1°K near the surface to about 0.4 K at 3 km depth near 65°S. (See Color Plate VII)

in winter, and during the S.W. Asian monsoon. In the Southern Hemisphere precipitation increases in the middle-latitude storm tracks throughout the year. Soil moisture is enhanced over the middle latitude continents of the Northern Hemisphere in winter but, in summer, many areas show a deficit mainly because of the earlier retreat of the snow cover under the enhanced temperatures.

Although the four models show broadly similar *global* patterns of response to double CO_2 concentrations, they show marked differences on regional and sub-regional scales, especially in precipitation and soil moisture.

Predictions of globally-averaged changes in temperature, precipitation and soil moisture are of little value in assessing their political, economic and social impact. Although current global models with rather low spatial resolution cannot be expected to provide reliable scenarios in regional and sub-regional scales, the UKMO has been asked to make deductions from its 'transient' CO_2 experiment for Western Europe. The results, which should be treated with caution, may be summarized as follows.

Summer temperatures rise throughout the 70-year experiment, stabilizing at about 0.3°K per decade after year twenty. There is a similar but less steady warming in winter, most pronounced over land. Winter precipitation increases rapidly during the first 30 years (possibly an artefact of an inadequate spin-up period) but thereafter remains rather steady at an average increase of about 0.3 mm/day, the main increases occurring over N. Europe and reductions in S. Europe and the Mediterranean. In summer the precipitation *decreases* by about 0.2 mm/day. The warmer, wetter winters and the slightly warmer drier summers are reflected in the changes of soil moisture.

Since the decadal changes are comparable in magnitude to the decadal variability, the comparable in magnitude to the decadal variability, the confidence in these estimates is low, especially in respect of precipitation and soil moisture changes, which are only marginally significant relative to the variability of the 'control' model, for any single decade.

THE EFFECT OF AEROSOLS

Aerosol particles influence the Earth's radiation balance directly by their scattering and absorption of solar radiation.

They also absorb and emit long-wave radiation but usually with small effect because their opacity decreases at longer wavelengths and they are most abundant in the lower troposphere where the air temperature, which governs emissions, is close to the surface temperature. Aerosols also serve as cloud condensation nuclei and therefore have the potential to alter the microphysical, optical and radiative properties of clouds.

The larger aerosol particles of $d > 0.1$ μm, if produced in large quantities from local sources such as forest fires, volcanoes and desert storms, may significantly influence the radiation balance on local and regional scales, both by scattering and by absorption and emission, especially if they contain carbon particles. However, such particles are rapidly removed from the troposphere by precipitation and are not normally carried long distances. On the global scale, smaller particles of $d < 0.1$ μm are more important, their dominant effect being to cool the atmosphere by scattering solar radiation to space.

Some recent calculations by Charlson *et al.* (1990) of the impact of anthropogenic sulphate particles on the short-wave radiation balance in cloud-free regions conclude that, at current levels, they reduce the radiative forcing over the Northern Hemisphere by about 1 W/m^2 with an uncertainty factor of two. A rather more sophisticated treatment by Kiehl and Briegel (1993) calculated the annually-averaged reductions in radiative forcing due to back-scattering of solar radiation by both natural and anthropogenic sulphate aerosols to be 0.72 W/m^2 in the N. Hemisphere, 0.38 W/m^2 in the southern hemisphere the global value of 0.54 W/m^2 being about half of that calculated by Charlson. However, the high aerosol concentrations over the heavily industrialised regions of the eastern USA, central Europe and South-East Asia produced reduction of >2 W/m^2 that are comparable to the cumulative increases produced by greenhouse gases emitted since the industrial revolution.

In addition to the direct effect on climate, anthropogenic sulphate aerosols may exert an indirect influence by acting as an additional source of effective cloud condensation nuclei, thereby producing higher concentrations of smaller cloud droplets leading to increased reflectivity (albedo) of clouds, especially of low clouds, for solar radiation, which is sensitive to the 'effective' droplet radius

$$r_{eff}\alpha(W/N)^{1/3}$$

where W is the liquid–water concentration of the cloud (in g/m^3) and N is the number concentration of the droplets.

The first calculations of this indirect effect on climate have been made to the UKMO by Jones *et al.* (1994), using their climate model that predicts cloud liquid water and ice content and parameterizes r_{eff} linking it to cloud type, water content and aerosol concentration. The concentration and size distribution of the aerosol, and its spatial distribution are calculated in the same manner as in Kiehl and Briegel but the particles are assumed to consist of ammonium sulphate as being characteristic of aerosol produced in industrially polluted air.

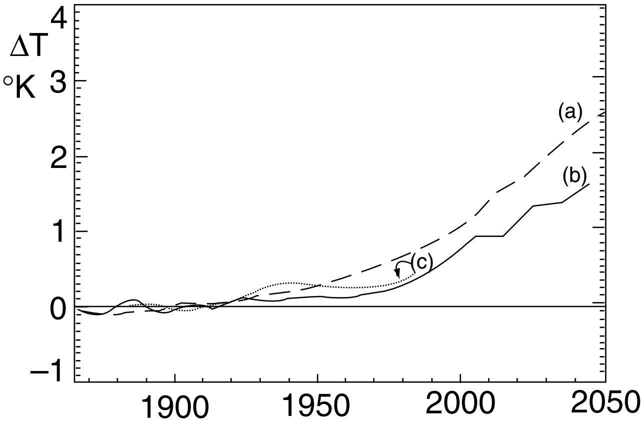


FIGURE 7 Changes in the globally-averaged mean surface temperatures relative to the mean for 1850–1920; dotted curve—observed change since 1880; dashed curve—model computations of the effects of increasing greenhouse gases from 1850–1990 and extrapolated to 2050 AD; solid curve—model predictions of changes caused by both greenhouse gases and aerosols from 1850–2040.

The calculations indicate that the enhanced back-scatter of solar radiation, mainly from low-level clouds in the atmospheric boundary layer, produces an annually-averaged global cooling of 1.3 W/m^2 but that over the highly industrialized regions, where r_{eff} may be reduced by as much as 3 μm, the cooling may exceed 3 W/m^2. However, it must again be emphasized that these calculations contain major uncertainties, probably even larger than those for the direct effect.

Taking them at face value, the calculations of the direct and indirect effects combined, suggest an average global negative forcing of 1.5–2 W/m^2 that may have largely offset the positive forcing of 2.3 W/m^2 by greenhouse gases to-date, and this may be at least part of the reasons for failure to detect a strong greenhouse signal.

The first results of introducing sulphate aerosols into a coupled atmosphere-ocean model come from the UKMO (Mitchell *et al.* 1995). The model, starting from an initial state determined by surface observations in 1860, was run forward to 1990 with no man-made greenhouse gases or aerosols as a control experiment. The model's average global surface temperatures showed realistic inter-annual variations but no overall rise over this period. In the perturbation experiment greenhouse gases were gradually increased from 1860 to reach a 39% equivalent increase in CO_2 by 1990; this resulted in a temperature rise of 1°C compared with an observed rise of only 0.5°C, (Figure 7). The next step was to compute the effects of sulphate aerosols with best estimates of concentration and geographical distribution. The direct effects of increasing the back-scatter of solar radiation was to reduce the warming between 1860 and 1990 to only 0.5°C, very close to the observed, but over and downwind of the highly industrialized regions of North America, Europe and Southern Asia, the aerosols largely nullify the warming caused by the greenhouse gases.

When the coupled model runs were carried forward from 1990 to 2050, increasing the CO_2 by 1% p.a. compound, the effect of aerosols was to reduce the global greenhouse warming from the 0.3°C/decade shown in Figure 5 to only 0.2°C/decade, and to largely offset it in highly polluted regions.

More reliable estimates of the effects of aerosols on climate must await much better observational data on the sources, concentration, size, chemical composition and spatial distribution of both natural and anthropogenic aerosols, including strongly absorbing carbonaceous particles, and dusts, and on the difference between droplet concentrations and sizes in clean maritime and polluted continental clouds. These data will be difficult and expensive to acquire; meanwhile, we are likely to have too many theories and computations chasing too few observations.

When aerosol particles are injected into the stratosphere, they remain there for much longer periods and become much more uniformly distributed than in the troposphere. Concentrations remained consistently high in the 1980s relative to earlier decades, largely due to the El Chichon volcanic eruption in 1982. Recently they were much enhanced by the Mt. Pinatubo eruption in June 1992 which injected about 20 million tons of SO_2 directly into the stratosphere. The consequent reduction in radiative forcing at the top of the atmosphere, estimated at 4 W/m^2, would have required the surface to cool by about 1°C in order to restore equilibrium. The fact that the temperature fell by only 0.3–0.5°C during the following two years may be due partly to absorption of radiation and infra-red emission by the aerosols, to the thermal lag of the oceans and to other negative feedbacks in the system.

SEA-LEVEL RISE

A potentially important consequence of greenhouse warming is the melting of sea-ice and ice sheets on land, only the latter resulting in a rise in sea level. The sea level will also rise as the ocean waters expand in response to the additional warming. Estimates of these consequences involve large uncertainties because of the lack of observations and understanding of the mass balance and dynamics of glaciers and ice sheets. These uncertainties are compounded by the uncertainty in the predicted increases in surface temperature due to greenhouse warming.

Over the past 100 years, the sea level is estimated to have risen about 10 cm. Thermal expansion of the ocean waters has probably been responsible for 4 cm of this rise, melting of mountain glaciers for 4 cm, and melting of the Greenland ice sheet for 2.5 cm. Glaciologists believe that there has been little, if any, overall melting of the Antarctic ice sheet because the air temperatures are too low.

If air and surface temperatures were to increase because of greenhouse warming, thermal expansion of the oceans and melting of mountain glaciers are likely to continue to make the largest contribution to sea-level rise. We have seen that globally coupled atmosphere deep ocean models predict that, when the atmospheric concentrations of carbon dioxide approaches double the present value, the average surface air temperature will increase by 0.3°C/decade. The best estimate of the corresponding rise in seal-level is 4 cm/decade, about half resulting from expansion of ocean waters and half from melting of landbased ice. These estimates, which may conceivably be in error by a factor of two either way, imply serious consequences for low-laying, highly populated areas such as Bangladesh, but they are very much smaller than the wildly exaggerated values that have appeared in the media.

UNCERTAINTIES IN MODEL PREDICTIONS

In summarising the current state of knowledge and understanding of the likely magnitude, timing and impacts of enhanced greenhouse warming, it is virtually certain that the troposphere is warming very slowly in response to the continually increasing concentrations of carbon dioxide and other greenhouse gases, but the signal is as yet barely detectable above the large natural climate variations, probably because it is being delayed by the large thermal inertia of the oceans and has also been masked by the cooling effect of man-made aerosols.

Predictions of the increase in globally averaged temperature that may result from a doubling of carbon dioxide have recently converged towards lower values ranging from 1.3°C to 2.3°C, based on coupled models with a deep ocean. However, this trend may be deceptive because only a small number of 70–100 year simulations have been published and considerable problems and uncertainties remain, both in the atmospheric physics and in the ocean dynamics. These arise largely from the sensitivity of the models to the simulation of clouds and their interaction with the radiation fields, the uncertainty as to how well they simulate the ocean circulation, and the necessity to adjust the ocean surface fluxes in order to ensure that the ocean temperature and salinity remain close to present-day climatology and that the control model climate does not drift during long runs. Long-term drift in the climate of the Southern Hemisphere arises from an imbalance in the heat budget of the Antarctic leading to a spurious slow-melting of the ice. This has now been corrected and changes in the pack-ice are now included. Another important defect of current low resolution ocean models is that they do not capture narrow features such as the Gulf Stream and Kuroshio currents and the regions of strong upwelling off South Africa and South America, all of which play an important role in heat transport. Some tests with higher resolution (1.25° × 1.25°) in the UKMO model improved this situation but only partially. The fact that current models have only limited success in simulating and predicting such a spectacular event as the El Nino is also evidence of defects in the treatment of atmosphere-ocean interactions. There is also a need for an improved representation of atmospheric boundary layer.

Even if the various models agree quite well on the globally-averaged effects, they show larger differences on regional and sub-regional scales, which are politically and economically more relevant. Further improvements in model development will require higher spatial resolution, especially in the oceans, better model physics, much faster computers and, above, all, an adequate supply of global observations from both the atmosphere and the oceans, to feed and validate the models, and to monitor the actual changes in climate that may eventually become evident.

The need for observations from both the surface and the interior of the oceans, and how they might be provided by new and advanced technology, are discussed by Mason (1993).

Despite these uncertainties and the fact that a doubling of CO_2 will cause an increase of only $\backsim$3% in the downward flux of infra-red radiation from the greenhouse gases, future predictions of the resulting globally-averaged temperature rise are unlikely to lie outside the range 1°C to 2.5°C. However, the models provide little guidance as to when these events are likely to occur. Their timing will be determined largely by the very uncertain future global emissions of greenhouse gases and their retention in the atmosphere. We must also realise that no existing climate model incorporates the carbon cycle in which exchanges of CO_2 between the earth's surface and the atmosphere are dominated by terrestrial and especially marine biology, man-made emissions being only about 3% of the natural two-way exchanges. We are always faced with having to compute small differences between large quantities whose magnitudes are uncertain. Given this and the complexity of the models, it is remarkable that they simulate the climate and its variability as well as they do, but there is a tendency to infer more from model predictions than their input data, spatial resolution and simplified physics can justify.

The latest version of the UKMO model has 38 pressure levels in the atmosphere, 20 levels in the ocean, where the horizontal grid spacing is reduced to 1.25° × 1.25°, obviating the need for artificial flux corrections at the ocean/atmosphere interface. Representations of the radiative effects of clouds, of atmospheric convection and of the drag exerted by mountain-induced gravity waves have all been improved. The model now remains stable when run for 1,000 years and shows no long-term drift in the global climate. Changes are calculated at about one million grid points so that computation of one annual cycle of the global climate involves about 10^{15} numerical operations.

When estimates of the radiative effects of greenhouse gases and aerosols prevailing in the years 1860 to 2000 were inserted in the model the predicted increase in the mean global near-surface air temperature T_s was 0.7°C in very good agreement with observation. When the model run was continued to Y2100, during which time CO_2 emissions were assumed to increase from the present value of 6.3 GtC/yr to 13.3 GtC/yr and the atmospheric concentration to almost double from 365 to 620 ppm, T_s was predicted to rise by 3.0°C relative to 1860, or 2.3°C relative to Y2000. The corresponding rise in global sea-level was 34 cm.

If the CO_2 concentration was assumed to increase at 1% pa to double after 70 years to 730 ppm, the predicted rise in T_s was 1.9°C. Similar experiments elsewhere with comparable models gave values between 1.1° and 3.1°C with an average of 1.8°C.

REFERENCES

Charlson, R.J. *et al. Tellus,* **43AB,** 152 (1991).
Charlson, R.J. *et al. Science,* **255,** 423 (1992).
Cusbasch, U. *et al. Climate Dynamics,* **8,** 55 (1992).
IPCC Climate Change. Report of Intergovernmental Panel on Climate Change. Cambridge. University Press Ed. J.T. Houghton *et al.* (1996, 2001).
IPCC Climate Change—Supplementary Report. Cambridge University Press Ed. J.T. Houghton *et al.* (1992).
Jones, A. *et al. Nature,* **370,** 450 (1994).
Kiehl, J.T. and Briegel, P.B. *Science,* **260,** 311 (1993).
Manabe, S. *et al. J. Phys. Oceanography,* **20,** 722 (1990).
Mason, B.J. *Contemporary Physics,* **30,** 417 (1989).
Mason, B.J. *Contemporary Physics,* **34,** 19 (1993).
Mason, B.J. *Contemporary Physics* **36,** 299 (1995).
Mitchell, J. F.B. *et al. Nature,* **346,** 501 (1995).
Murphy, J.M. World Met. Org. Report No. 14 (1990).
Murphy, J.M. *J. Climate,* **8,** 36 (1995).
Murphy, J.M. and Mitchell, J.F.B. *J. Climate,* **8,** 57 (1995).
Senior, C.A. and Mitchell, J.F.B. *J. Climate,* **6,** 700 (1993).

B.J. MASON
Imperial College

HAZARDOUS GASES: see PREVENTION OF TOXIC CHEMICAL RELEASE

GROUNDWATER RESOURCES

Water plays a vital role in sustaining the existence of all living things. The importance of managing the different aspects of the water, including its quantity, quality, transport processes, utilization, and management becomes apparent when the supply of water, especially potable water, is limited. This article is intended to be an introduction of the subject of groundwater. It includes a historical perspective of development, well hydraulics, the quality aspect, and a number of remediation technologies.

It has been estimated that approximately 300 million cubic miles of water exist on the earth at different locations and in different states phases. From Table 1, it can be seen that the ocean forms about 97% of the earth's water as saline water. The remaining 3% is freshwater. This freshwater supply would be quickly depleted if it were not for its endless interaction and exchange with the ocean. This movement of the earth's waters, known as the *hydrologic cycle,* is shown schematically in Figure 1. Due to the complexity of the hydrologic cycle, not all of the components can be determined easily. A detailed discussion on the subject is presented by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) (USSR National Committee for the International Hydrological Decade, 1978). On a long-term average basis, the total freshwater supply is to be considered constant. Therefore, the budget equation for freshwater can be expressed as:

$$[\text{Total precipitation}] = [\text{Evaporation}] + [\text{Transpiration}]$$

Here, the freshwater is found in the atmosphere as water vapor; on the land as water, dew, snow, and ice; and in the ground usually in the form of water. Approximately 99.5% of all the Earth's water is in a form or location unavailable for use (see Table 1), leaving only about 0.5% of the earth's water remaining for consumption. The largest source of available water is groundwater, which forms an important part of the water supply for municipalities, agriculture, and industry. For instance, in the United States, approximately 50% of the total water usage is from groundwater, while over 90% of the world's potable water supply is derived from groundwater. Thus, groundwater is one of the most important natural resources. Groundwater occurs in the saturated zone of water-bearing soils. However, it may also be found in sedimentary deposits, fissures in igneous rocks, limestone caverns, and other water-bearing sources, which may lie near the surface or hundreds of feet beneath the surface. The subject of groundwater is addressed by a variety of disciplines, including hydrology, geology, and hydraulics.

In general, hydrologists and engineers are concerned with the long-range yield from watersheds and aquifers. Yield is a function of watershed storage capacity and the supply, which fluctuates with time. Typical means of groundwater removal are shown in Figure 2. It is of paramount importance to ensure enough supply to last through drought periods. Currently, accurate prediction of drought phenomena is difficult and unavailable, as indicated by the National Research Council. The subject is still an active research in water-resource engineering. Because of the cost of well drilling, it is desirable to have some assurance that the water-supply well will reach a satisfactory aquifer. In order for the hydrologist or engineer to obtain an appreciation of groundwater flow, he/she must understand that the notion of "safe yield" in groundwater is the quantity of groundwater that can be withdrawn without impairing its quality.

HISTORICAL NOTES

A long time passed in the history of man before the concept of the hydrologic cycle was established. Groundwater has been used since ancient times. Plato wrote about rainfall, runoff, and the infiltration of source water as the source for springs. Aristotle recognized the hydrologic cycle, including evaporation, condensation, streamflow, infiltration, and the reappearance of groundwater as springs. Among the early investigations of the hydrologic cycles and groundwater flow, the contributions of Leonardo da Vinci, Bernard Palissy, Claude Perrault, Edme Mariotte, Edmond Halley, and La Metherie are of particular note.

In 1856, studying the movement of water through sand, Henry Darcy developed the celebrated Darcy's Law, that is, the velocity of percolation as a function of permeability and hydraulic gradient. Darcy's Law provides the basis for a quantitative method used in groundwater evaluation. Most of the developments centered in this period on the problems of well hydraulics and well-field design. A detailed account of Darcy's contributions and life history is presented in a recent publication (Brown, Garbrecht, and Hager, 2003). During the late nineteenth and twentieth centuries, Boussinesq, Dupuit, Theim, Forchheimer, and many others laid an analytical foundation for the subject. Their contributions encompassed a large number of analytical solutions for which equations

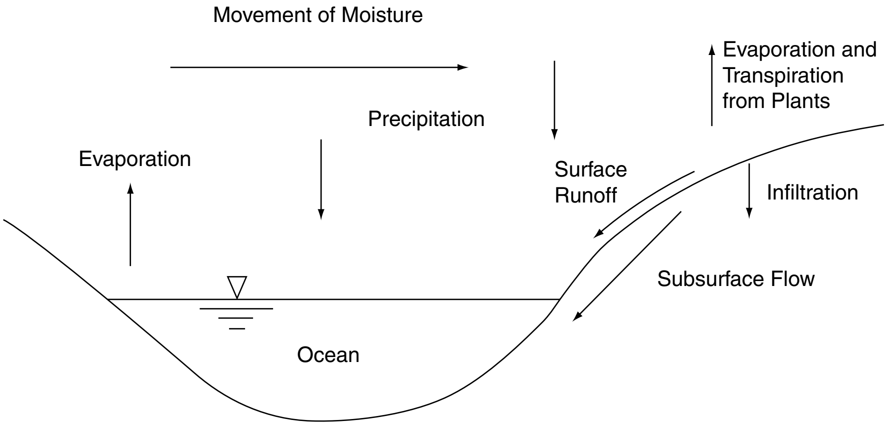


FIGURE 1 The hydrologic cycle.

TABLE 1
World water distribution

Form of Water	Area Covered (km^2)	Volume (km^3)	Share of World Reserves (%)	
			of Total Water Reserves	of Fresh Water Reserves
World ocean	361,300,000	1,338,000,000	96.5	—
Groundwater	134,800,000	23,400,000	1.7	—
Predominantly fresh groundwater	134,800,000	10,530,000	0.76	30.1
Soil moisture	82,000,000	16,500	0.001	0.05
Glaciers and permanent snow cover	16,227,500	24,064,100	1.74	68.7
Antarctica	13,980,000	21,600,000	1.56	61.7
Greenland	1,802,400	2,340,000	0.17	6.68
Artic islands	226,100	83,500	0.006	0.24
Mountain areas	224,000	40,600	0.003	0.12
Ground ice in zones of permafrost strata	21,000,000	300,000	0.022	0.86
Water reserves in lakes	2,058,700	176,400	0.013	—
Freshwater	1,236,400	91,000	0.007	0.26
Saltwater	822,300	85,400	0.006	—
Marsh water	2,682,600	11,470	0.0008	0.03
Water in rivers	148,800,000	2,120	0.0002	0.006
Biological water	510,000,000	1,120	0.0001	0.003
Atmospheric water	510,000,000	12,900	0.001	0.04
Total water	510,000,000	1,385,984, 610	100	—
Total freshwater	148,800,000	35,029,210	2.53	100

bear their names. In the United States, significant works in field, laboratory, and analytical studies were made by Slichter, Meinzer, Theis, Jacob, Hubbert, Hantush, and Muskat, to name a few. The major achievements of the subject, up to the 1950s, have been summarized by Ferris and Sayre (1955).

In the second half of the twentieth century, a considerable body of highly scientific knowledge and sophisticated technology were developed. However, the most important impact is perhaps the invention and proliferation of the digital computer as a tool for performing difficult and tedious computation, which has led to the new computational methods capable of analyzing complex systems. First, an extensive investigation has gone into the different aspects of the chemical processes in groundwater and biodegradation in the subsurface system, among others. Second, a number of solute-transport models have become available, such as the USGS MOC (U.S. Geological Survey method of characteristics) series and the USGS MODFLOW-2000 (modular finite-difference flow model), to name a few. As a result, analysis, planning, and development of large-scale problems can be achieved. Thus far, studies of groundwater resources have made tremendous strides. Major developments include economical use, environmental considerations, and man's effect on the groundwater resource.

WELL HYDRAULICS

Well hydraulics is the most important single element in groundwater-resources engineering. Wells are drilled either for abstraction or observation purposes. Abstraction wells, or pumping wells, are drilled and developed to achieve the optimum output quantity of groundwater, while observation wells are for evaluation of the aquifer characteristics. Equations of well hydraulics under steady and unsteady conditions with specific geologic settings (i.e., confined, unconfined, leaky, and fractured) and their specific boundary conditions can be formulated by fluid-dynamics equations when used in conjunction with Darcy's Law. For practical purposes, most solutions have been reduced to graphical form (Lohman, 1972), tables (Hantush, 1964), or computer programs (Appel and Reilly, 1994). Here, the case of unsteady flow to a well in a confined aquifer (unsteady artesian well) is presented because of its fundamental characteristic and importance in practical applications. A variety of solutions to well hydraulics for other aquifer-pumping and well conditions are available in standard texts (e.g., Todd and Mays, 2005; Bear, 1979).

In the derivation of the nonequilibrium drawdown equation for an aquifer, it is assumed that the release of water occurs instantaneously and that any compaction occurs only in the aquifer. If the aquifer is a limestone formation, such compaction is not too likely to occur. On the other hand, if the aquifer is a clay formation, it will compact quite readily when it loses its moisture content. Thus, before controlled withdrawals are put into effect, the ground levels can subside substantially. For example, the subsidence of Mexico City was caused by the compaction of the clays overlying the limestone formation from which water was drawn. In any case, if compaction does occur, the storage coefficient will vary. Since the variation of the storage coefficient is rather slow, it can be considered a basic property of the soil for most purposes. Assumptions are summarized as follows (see Figure 3):

1. The aquifer is horizontal and infinite in extent in all directions.
2. Pumping starts instantaneously, at a rate Q.
3. Darcy's Law is valid.

The discharge of groundwater flow can be evaluated by Darcy's equation:

$$Q = KiA \tag{1}$$

where Q is the discharge flow rate, ft^3/d; K is the hydraulic conductivity, ft/d; i is the hydraulic gradient; and A is the

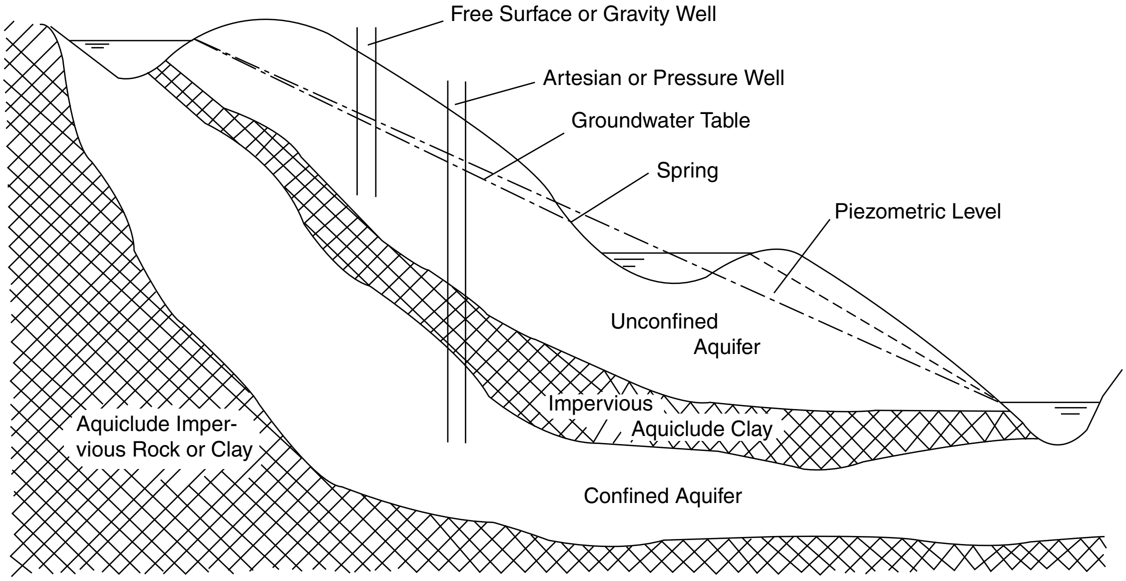


FIGURE 2 Types of groundwater withdrawal.

area through which flow occurs, ft^2. From the basic mass balance equation, the resulting expression of nonequilibrium (transient) flow forward wells can be written as follows:

$$\frac{1}{r}\frac{\partial s}{\partial r}+\frac{\partial^2 s}{\partial r^2}=\frac{S}{T}\frac{\partial s}{\partial t}, \qquad T = Kb \tag{2}$$

where s is drawdown, ft; S is the storage coefficient; T is the transmissibility, ft^3/d per ft; r is the radial coordinator, ft; and t is the time, days. The solution for Equation (2) with the pertinent initial and boundary conditions is given as

$$s=\frac{Q}{4\pi T}\int_{-u}^{\infty}\frac{e^{-u}}{u}\,du=\frac{Q}{4\pi T}W(u), \qquad u=\frac{S}{4T}\frac{r^2}{t} \tag{3}$$

where $W(u)$ is known as the "well function," whose values can be obtained from a standard mathematical handbook. Equation (3) is called the Theis equation and is the expression for the drawdown for radial flow to a well. It is perhaps the most widely used equation for the field determination of transmissibility and the storage coefficient. However, due to the mathematical difficulties involved, several approximate techniques have been developed.

The Theis Method

This is simply the graphical solution of Equation (3), also known as the type-current method. The Theis equation can be rewritten in the form

$$\frac{W(u)}{u}=\frac{16\pi T^2}{QS}\frac{s}{(r^2/t)} \tag{4}$$

By plotting $W(u)$ versus u and s versus r^2/t on log-log scales, respectively, and then overlapping these two curves with

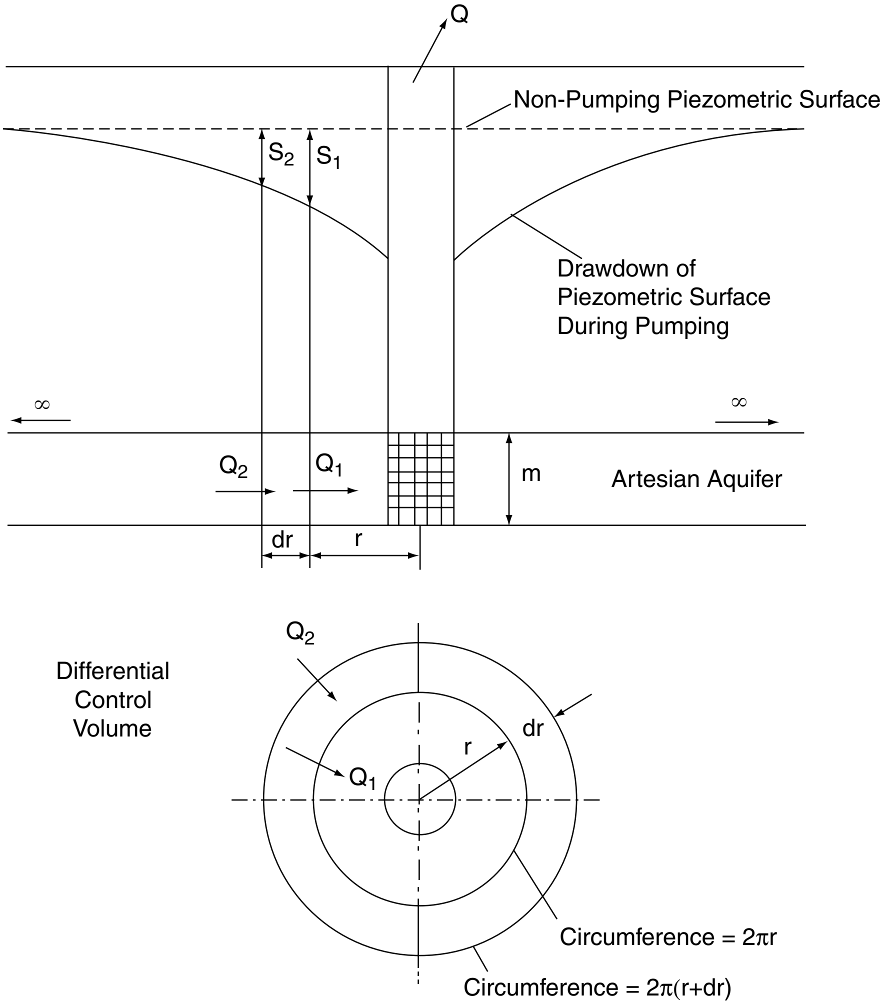


FIGURE 3 Definition sketch for the derivation of the nonequilibrium Theis equation.

their coordinate axes parallel as shown in Figure 4, the values of $W(u)$, u, s, and r^2/t may be obtained for the center of the matching segments. Then, T and S can be calculated as follows:

$$T = \frac{Q}{4\pi s} W(u) \qquad \text{and} \qquad S = \frac{4uT}{(r^2/t)}$$

The Jacobs Method

This method involves a simplification of the well function. Assume that a test well is being pumped at a rate Q with two observation wells located at r_1 and r_2 in a radial direction from the test well. As the test well is pumped, the drawdowns and times of drawdowns are observed for the two observation wells. The difference of drawdowns between the two observation wells can be obtained as follows:

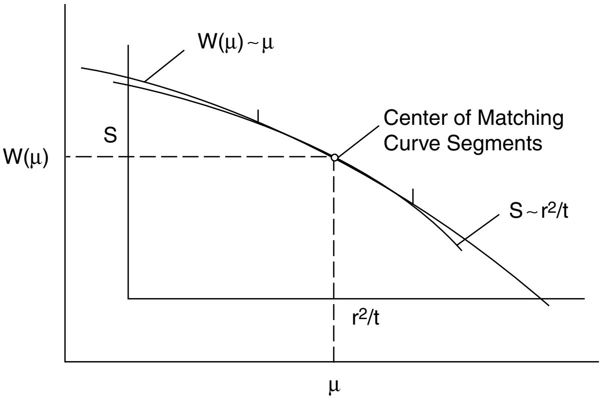


FIGURE 4 Application of Theis method.

$$\begin{aligned} s_2 - s_1 &= \frac{2.3Q}{4\pi T}\left(\log_{10}\frac{0.3Tt_2}{r_2^2 S} - \log_{10}\frac{0.3Tt_1}{r_1^2 S}\right) \\ &= \frac{2.3Q}{4\pi T}\log_{10}\frac{r_1^2/t_1}{r_2^2/t_2} \end{aligned} \tag{5}$$

where the subscripts 1 and 2 refer to the observation wells. Now, if there exists only one observation well (i.e., $r_1 = r_2$), the subscripts now refer to the time at which observations are made and recorded. Then the above equation can be reduced to:

$$s_2 - s_1 = \frac{2.3Q}{4\pi T}\log_{10}\frac{t_2}{t_1} \tag{6}$$

The above equation can be used to evaluate T from field data. For a constant Q, drawdown versus time can be plotted on a semi-log scale as shown in Figure 5. If $t_2/t_1 = 10$, then the transmissivity, T, can be calculated as follows:

$$T = \frac{2.3Q}{4\pi(s_2 - s_1)} \tag{7}$$

If it is assumed the drawdown began at some time, say t_0, the storage coefficient, S, can be determined as

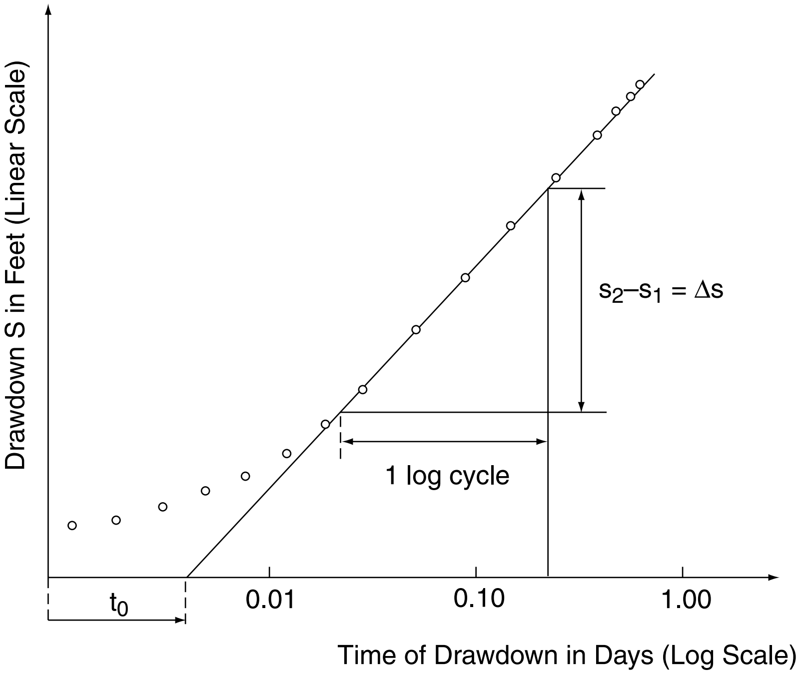


FIGURE 5 Application of semi-log drawdown curve—Jacobs method.

$$S = \frac{0.3Tt_o}{r^2} \tag{8}$$

In order to account for some of the different situations encountered in the field, modification to the basic equations may be necessary by incorporating the well friction losses in the form

$$\text{well friction losses} = kQ^n \tag{9}$$

when $n = 2$ is generally used. The total loss can be expressed as

$$s''_w = s'_w + kQ^2 \tag{10}$$

where

$$s'_w = Cs_w \tag{11}$$

and

$$s_w = \frac{2.3Q}{4\pi T}\log_{10}\frac{0.3Tt}{r_w^2 S} \tag{12}$$

where coefficient C is a gross measurement of the degree of well screen penetration, flow curvature, the radius of the well itself, and the anisotropy of the permeability of the aquifer. It can be evaluated by making field tests at the site.

True equilibrium between the rate of withdrawal and the rate of replenishment is seldom, if ever, achieved. However, a state of quasi-equilibrium can sometimes be attained through several means. For example, replenishment may be by a stream located near the point of withdrawal or, in the case of artesian aquifers, by rain falling on an exposed portion of the aquifer miles away or by rainwater infiltrating to the groundwater in the near vicinity of the well.

QUALITY CONSIDERATION

While the quantity of available groundwater and its movement are still of increasing interest to engineers and scientists, the solutions for well hydraulics represent the quantitative aspect of groundwater resources. It is now recognized that the quality of groundwater is equally important. A major effort has been made by environmental engineers and scientists throughout the world on water quality. In this and the following sections, the quality aspect and remediation considerations are introduced. From a practical standpoint, the quality of any water is a relative term. Water that is of good quality for an industrial-cooling purpose may be of unacceptable quality as a drinking-water source. Generally, water-quality considerations are related to chemical, biological, and radiological content and temperature, and are based on the intended use.

As water moves through the hydrologic cycle, its quality is altered. Rainwater, which was considered a "pure" substance, is not pure at all. When rainfall reaches the ground surface, it contains dissolved gases (CO_2, O_2, N_2, NO_2, NH_3, CO, SO_2, H_2S, etc.) as well as other dissolved materials. The expression "acid rain," for example, is familiar to most people.

As water reaches the soil, its quality changes further. As water percolates through the biologically active soil mantle, organic material, both suspended and dissolved, is removed due to aerobic bacterial action. Filtration occurs, as do ion exchange and adsorption. Thus, by the time water has reached the level of rock, a great deal of the organic material, effectively all of the suspended material, and some of the dissolved inorganic material have been removed. However, inorganic material may also have been added to the water, due to ion-exchange reactions and chemical equilibrium. For example, ammonia dissolved in the rainwater or present in the soil will be oxidized by soil bacteria to nitrate, and the nitrate will be released into the water and carried downward to the water table. As the water moves through rock strata, it picks up inorganic matter as part of dissolution action. This dissolving action is responsible for the relative hardness of most groundwater. It is also responsible for the formation of caves and caverns, especially in limestone ($CaCO_3$) strata. The natural changes in groundwater quality may be seen in Figure 6.

In accordance with their genesis, therefore, there are two different sources of pollutants in the groundwater system: (1) the dissolved chemicals from the geologic formation, and (2) the man-made wastes. The movement of groundwater, as a carrier of the dissolved chemical mass, gives rise to many of the reaction-transport processes that occur in the subsurface. In many geologic processes, groundwater plays a critical role in the formation and dissolution of certain ores and hydrocarbon deposits. For a detailed discussion, the reader is referred to Phillips (1991) and Ingebritsen and Sanford (1998). On the other hand, the dissolved chemicals from man-made waste become a primary source of pollutants in the groundwater system. Groundwater hydrologists are interested in the quality of existing groundwater resources and the effect that human influence will have on the quality of groundwater.

Human influence on the quality of groundwater results primarily from activities that generate wastes. (There are some exceptions, such as nitrates in water due to fertilizers, and saltwater intrusion due to pumping of aquifers.) There are many potential sources of environmental contamination, including agrochemicals, industrial effluents, storage-tank leaks, seepage from disposal sites for toxic substances, and petroleum-product spills. It has been estimated that the amount of hazardous waste generated annually is 264 million metric tons in the United States alone. The enormous amount of hazardous waste that has accumulated over the years makes this figure even more astounding. It is reported that only 10% of the waste generated prior to 1980 was disposed of by practices that would be considered adequate according to current standards. Thus, as much as 90% of the hazardous waste was disposed of at unregulated facilities. These irresponsible disposal practices have created over 22,000 sites containing unregulated hazardous waste in United States. The improper disposal of hazardous waste has caused a number of serious problems that not only result in

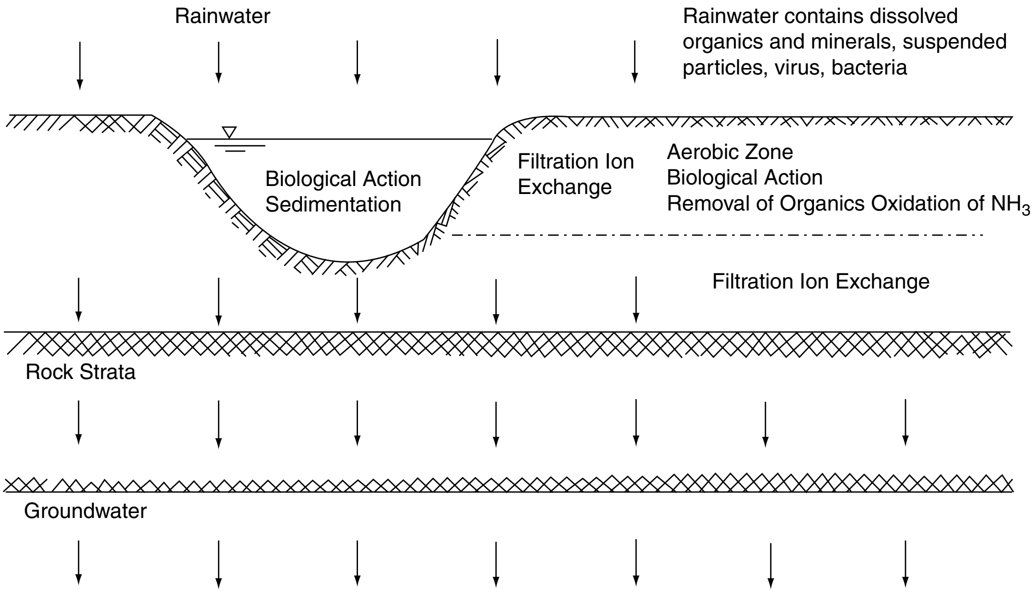


FIGURE 6 Effect of water movement on water quality.

the destruction of the ecological system and natural resources but also present a danger to public health. Pollutants have been broadly classified into eight categories by the U.S. Public Health Service.

1. Oxygen-demanding wastes: These include domestic and industrial wastewaters, leachates from landfill, and any other wastes containing biodegradable organic material. These wastes are stabilized by microorganisms and will remove dissolved oxygen from water.
2. Infective agents: Bacteria, virus, protozoa, and other infectious agents may be present in domestic wastewaters. Those that reach drinking water or even recreational waters represent a health hazard to humans.
3. Plant nutrients: Mineral substances, particularly nitrates and phosphates, stimulate the growth of plants. However, nitrates in groundwater potentially infiltrating through heavily fertilized soil can be a public health hazard.
4. Organic chemicals: Chemicals such as insecticides, pesticides, and detergents are toxic at low concentrations and can cause significant changes in groundwater quality.
5. Other minerals and chemicals: This group consists primarily of materials found in industrial wastes, including chemical residues, petrochemicals, salts, acids, etc. New chemical substances are being created at the rate of 400 to 500 compounds per year. Many of these substances are known to be toxic, although very little is known about their biological action. Almost nothing is known about the possible long-term effects of these chemical compounds.
6. Sediments from land erosion: These materials have little effect on groundwater quality since they would be filtrated out in the first stages of percolation.
7. Radioactive substances: Expected increases in the use of nuclear reactors for power generation indicate that radioactivity may be a serious problem in the future. The extremely long half-lives of some radioactive materials make the disposal of radioactive wastes difficult. Once a radionuclide reaches the groundwater it remains a hazard for a long time. One rule is that a radionuclide is no longer dangerous after nine half-lives.
8. Heat: Heat from industry and power generation is considered a serious pollutant. Its effect on groundwater is limited to reducing the value of a groundwater source for cooling water.

In addition to the eight categories of pollutants used by the U.S. Public Health Service, oil can be considered a ninth category, since oil spills in coastal waters have found their way into groundwater supplies near the coast.

The quality of water required for industrial-process usages varies widely with the type of industry and the type of process used. For example, the recommended total solids concentration in boiler feed waters varies from 50 to 3,000 mg/l depending on the design of the boiler and the pressure of operation. The quality of water used for agricultural operations also varies considerably. Range-cattle-use water

TABLE 2
Drinking-Water Standards

Chemicals	Standards	
	MCLG (mg/l)	MCL (mg/l)
Organics		
Aldicarb	0.001	0.003
Benzene	0	0.0002
Benzo[a]pyrene (PAH)	0	0.0002
Bromodichloromethane	0	0.08
Carbon tetrachloride	0	0.005
Chlordane	0	0.002
Chloroform (THM)	0.07	0.08
Di (2-ethylhexyl) phthalate	0	0.006
Dichloroethane (1,2-)	0	0.005
Dichloroethylene (1,1-)	0.007	0.007
Dichloroethylene (cis-1,2-)	0.07	0.07
Dichloromethane	0	0.005
Dichloropropane (1,2-)	0	0.005
Endrin	0.002	0.002
Heptachor	0	0.0004
Lindane	0.0002	0.0002
Monochlorobenzene	0.1	0.1
Polychlorinated biphenyls (PCBs)	0	0.0005
2,3,7,8-TCDD (Dioxin)	0	3E-08
Tetrachloroethylene	0	0.005
Trichloroethane (1,1,2-)	0.003	0.005
Xylenes	10	10
Inorganics		
Arsenic	0	0.01
Cadmium	0.005	0.005
Mercury (inorganic)	0.002	0.002
Nitrate (as N)	1	1

MCLG—maximum contaminant-level goal, a nonenforceable health goal that is set at a level at which no known or anticipated adverse effect on the health of persons occurs and that allows an adequate margin of safety.

MCL—maximum contaminant level, the highest level of a contaminant that is allowed in drinking water. MCLs are set as close to the MCLG as feasible using the best available analytical and treatment technologies and taking cost into consideration. MCLs are enforceable standards.

containing 5,000 mg/l dissolved solids but low concentrations of sodium, magnesium, and sulfate are desirable. The quality required for irrigation water depends on the soil, composition, climate, type of crops, amount of groundwater used, etc.

Groundwater is primarily used as a water source for public and private drinking-water supply, industrial-process and cooling water, and agricultural supply. The acceptability of a groundwater source for use depends on the standards for quality for that use. The Drinking Water Standards for major chemicals are regulated by the U.S. Environmental Protection Agency (EPA) (2004). These standards are shown in Table 2.

REMEDIATION CONSIDERATION

Pollution Control and Remediation Policy

This section briefly describes the major federal environmental laws that control and regulate groundwater quality. In the United States, the protection of groundwater is accomplished through a set of statutes passed at different times. The goal of these laws is to protect the public health and regulate sources of contaminations.

The first and probably most important modern environmental law was enacted in January of 1970, namely the National Environmental Policy Act (NEPA) of 1969. NEPA is a milestone in man's understanding regarding his own survival and the survival of the total ecology that has supported all life of earth. This act dictated that decision makers approach industrial development by balancing the environmental, economic, and technological factors to protect and enhance public health and welfare. Following NEPA, many important statutes were enacted to restrict the improper handling of hazardous wastes. Currently, there are more than 30 laws concerning environmental issues, among which 2 deal directly with water.

The Clean Water Act (CWA) of 1972, although it does not deal directly with groundwater, is the cornerstone of surface-water-quality protection. It established the basic structure for regulating pollutant discharges into the water system as well as regulatory tools to sharply reduce those pollutant discharges. These tools are employed to achieve the broader goal of restoring and maintaining the chemical, physical, and biological integrity of the water system.

The Safe Drinking Water Act of 1974 was established to protect the quality of drinking water in the United States. It focuses on all waters actually or potentially designed for drinking use, including both surface water and groundwater sources. Since their enactment, these statutes have played a very important role in preventing the creation of new contaminated sites. However, they have had no effect on the hazardous-waste sites already in existence. If a healthy environment is to be established, these long-term dangers cannot be ignored.

The cleanup action started with the enactment of the Comprehensive Environmental Response, Compensation, and Liability Act of 1980 (CERCLA, or Superfund), which is a remedial program for cleaning up the nation's worst hazardous-waste sites created by past industrial disposal practices. CERCLA established a hazardous-substance Superfund as well as regulations controlling inactive hazardous-waste sites. Extensive liability has been brought to owners and waste generators alike. As a consequence, the public is more concerned than ever about groundwater contamination. The implementation of Superfund is challenging, both technically and economically.

Remediation

Groundwater cleanup is very expensive and extremely difficult, if not impossible. An important and necessary step is removal of contaminants from the vadose (unsaturated) zone, which in most cases serves as a source of contamination.

The groundwater remediation can never be complete so long as mobile contaminants remain in the vadose zone. Currently, there are two basic approaches for decontamination, namely *ex situ* and *in situ* techniques. *Ex situ* remediation requires excavation with on-site or off-site treatment, while the *in situ* remediation removes (or treats) the contaminants in place, thus minimizing disturbance to the site. Since excavation may not be feasible in many situations, the *in situ* remediation approach, in general, is technically and economically superior. As a result, a number of *in situ* remedial technologies have been developed to treat contamination in the soil. These remedial actions can be grouped into the following general categories based on their characteristics.

Physical Treatment Physical treatment is a relatively simple and safe approach compared with other treatment methods. Since no foreign materials are introduced into the formation, there is little chance that the contamination situation will be worsened.

The soil vapor extraction (SVE) method, known as soil venting, *in situ* volatilization, enhanced volatilization, or soil vacuum extraction, is one of the most popular physical treatment technologies. In this process, a vacuum is applied through extraction wells to create a pressure gradient that induces transport of chemical volatiles through the soil to extraction wells. This technology has been widely used in volatile organic compound (VOC) removal from the vadose zone, and it can also be useful in decontaminating groundwater, since the lowered VOC vapor pressure in the vadose zone will increase volatilization of contaminants in the groundwater. Principal factors governing the applicability of SVE include contaminant distribution at the site, site hydrogeology, and contaminant properties. In general, SVE is an effective process for highly permeable formations such as sand, but is ineffective for low-permeability formations such as silt and clay, since tight formations restrict the airflow through the porous medium.

Pump-and-treat is a physical treatment technology used for the saturated zone in combination with other technologies. Contaminated groundwater is pumped out of the formation and treated either off-site or on-site. This technology is commonly applied to reduce the rate of plume migration, or to confine the plume to a potentiometric low area. In addition, physical containment systems (cutoff walls, reactive barriers, etc.) have been used. The primary objective is to construct a physical impediment to groundwater flow.

Chemical Treatment In general, chemical methods for detoxification of contaminants in the soil are severely hindered by the difficulties of dispersing chemical amendments into the formation. As such, most attempts of chemical *in situ* treatments have been limited to chemical spills or dump sites where near-surface contaminated soil could be treated. Potential chemical reactions of the treatment reagents with the soils and wastes must be considered with all chemical treatments. Since most hazardous-waste disposal sites contain a mixture of contaminants, a treatment approach that may neutralize one contaminant could render another more toxic or mobile. In addition, the chemical amendment introduced into the soils may create new pollution. The general chemical treatments include solidification-stabilization, neutralization, and oxidation-reduction.

The solidification-stabilization method is designed to make contaminants physically bound or enclosed within a stabilized mass. This method includes inducing chemical reactions between the stabilizing agent and contaminants to reduce their mobility. Neutralization involves injecting dilute acids or bases into the ground to adjust the pH. This pH adjustment can serve as pretreatment prior to oxidation-reduction or biological remediation. The oxidation-reduction method consists of using oxidation-reduction reactions to alter the oxidation state of a compound through loss or gain of electrons, respectively. Such reactions can detoxify and solubilize metals and organics. This technology is a standard wastewater-treatment approach, but its application as *in situ* treatment is limited.

Biological Treatment The ultimate goal of biological treatment is to achieve biodegradation of the organic chemicals. From attempts to utilize bacterial cultures to sophisticated genetic-engineering applications, a myriad of biodegradation technologies have been explored. Since a large portion of the hazardous-waste contamination in the United States stems from petroleum-hydrocarbon products that have been discharged or spilled into the soil at petroleum refineries, airports, and military bases, bioremediation has become an accepted, simple, and effective cleanup method. Basically, there are two approaches: stimulating the growth of “indigenous” microorganism populations and adding new “endigenous” microorganisms. Biological treatments may utilize either one or both of these approaches.

Many toxic organic chemicals can be metabolized or degraded to some degree by indigenous soil microorganisms. This natural process can be accelerated by pumping oxygen and nutrients into the contaminated zone to stimulate the resident microorganisms. However, if the infiltration rate is low, remediation may be a very slow process.

In the endigenous approach, genetically endigenous microorganisms are introduced into the formation to clean the contaminants. There are two major difficulties in this approach. First, the dispersal of introduced organisms throughout the contaminated zone is very difficult, since microorganisms tend to be absorbed by solid particles and become clustered. Second, the newly introduced organisms may destroy the microbial balance once the porous medium has become decontaminated, which creates new ecological problems.

Overall, a significant engineering deficiency with *in situ* bioremediation is the absence of proven methods to introduce degrading populations of microorganisms, nutrients, and other chemicals into the subsurface environment for efficient mixing with microorganisms and the contaminants of concern. In addition, the infiltration rate is an important governing parameter to the effectiveness of the process.

Thermal Treatment

Thermal treatment involves introducing extra energy into the contaminated zone to increase the formation temperature. Two different temperature ranges have been employed: high-temperature treatment is used to destroy chemical structures as well as soil constituents to retard chemical movement, while low-temperature treatment is used to increase chemical mobility and removal rates.

The *in situ* vitrification (ISV) technology is a high-temperature treatment designed to treat soils, sludges, sediments, and mine tailings contaminated with organic, inorganic, and radioactive compounds. Heating, applied via electrodes, is used to melt contaminated soils and sludges, producing a glass and crystalline structure (at about 3000°F or 1600°C) with very low leaching characteristics. The glass and crystalline product will permanently immobilize hazardous substances and retain its physical and chemical integrity for geologic time periods. Since the ISV process is costly, it has mostly been restricted to radioactive or highly toxic wastes. The demand for high energy, specialized equipment, and trained personnel may greatly limit the use of this method.

The thermally enhanced SVE process uses steam/hot-air injection or electric/radio frequency heating to increase the mobility of vapors and facilitate SVE. The temperature in this process is controlled in a low range so that there is no chemical destruction.

Enhancement Technologies

The success of *in situ* remediation technologies depends largely upon the transport efficiency of materials in and out of the contaminated zone. Contaminants must be transported out of the formation, while chemical, biological, and other amendments must be transported in. Without enhancement, most *in situ* remediation methods are effective only in relatively permeable formations and are inadequate for fine-grained soils due to the low natural permeability. Enhancement technologies were developed to improve the transport conditions for the current remediation industry.

The hydraulic-fracturing process begins by using a hydraulic jet to cut a disk-shaped notch on the borehole wall. Water (with or without chemicals) is then injected into the notch until a critical pressure is reached and a fracture is formed. A proppant composed of a granular material (e.g., sand) and a viscous fluid (e.g., guar gum and water mixture) is then pumped into the fracture. As a result, the mobility through difficult soil conditions can be increased. However, since water or another liquid is used in the process, the moisture content of the formation is increased during hydraulic fracturing. This additional water or liquid may block the pathway for gaseous transport and subsequently reduce the removal efficiency from fractured formations.

Pneumatic fracturing is a relatively new enhancement technology. This process involves injection of pressurized air into soil or rock formations to create fractures and increase the permeability. The injection is a quick process (e.g., taking within 10 to 20 seconds), and clean air is the only ingredient of the injection fluid. Thus, the potential chemical hazard or disturbance to the formation's chemical constituents is minimal.

As indicated, groundwater remediation is an expensive and difficult task, which requires a comprehensive understanding of the problem, the identifying of remediation technologies, abilities, and the selecting of an appropriate remediation goal. New engineering remediation technologies are being developed and tested; the interested reader is referred to a document by the EPA (1993).

CONCLUDING REMARKS

It has long been recognized that groundwater is one of the most valuable natural resources. The subject of groundwater is vast and encompasses a great variety of disciplines, including hydrology, hydraulics, geology, and chemistry. This article is intended to serve as an introduction to the subject, and as such, the emphasis is on the underlying principles. Discussion on modeling of solute transport in groundwater is omitted here because of its complexity. For an overview of modeling, the reader is referred to Sun (1996), Van der Heijde and Elnawawy (1993), and others. In short, a great many challenging problems in groundwater await solutions.

REFERENCES

Appel, C.A. and Reilly, T.E. 1994. "Summary of computer programs produced by the U.S. Geological Survey for simulation of ground-water flow and quality." USGS Circular 1104.

Bedient, P.B., Rifai, H.S., and Newell, C.J. 1994. *Ground water contamination.* Prentice Hall, New Jersey.

Brown, G.O., Garbrecht, J.D., and Hager, W.H. 2003. *Henry P. G. Darcy and other pioneers in hydraulics.* ASCE Press.

Carslaw, H.S. and Jaeger, J.C. 1959. *Conduction of heat in solids.* Clarendon Press, Oxford.

Driscoll, F.G. 1995. *Groundwater and wells.* 2nd Edition. Johnson Screens, Minnesota.

Dupont, R.R. 1993. "Fundamentals of bioventing applied to fuel contaminated sites." *Environmental Progress,* 12(1), 45–53.

EPA. 1993. "Guidance for evaluating the technical impracticability of ground water restoration," OSWER, Directive 9234.2-25, Oct.

EPA. 2004. Drinking Water Standards and Health Advisories.

Ferris, J.G. and Sayre, A.N. 1955. "The quantitative approach to groundwater investigations." *Economic Geology,* 50th Anniversary Volume.

Grasso, D. 1993. *Hazardous waste site remediation source control.* CRC Press, New York.

Hantush, M.S. 1964. "Hydraulics of wells." *Adv. Hydrosciences,* Vol. 1. Ed. V. T. Chow. Academic Press, New York.

Ingebritsen, S.E. and Sanford, W.E. 1998. *Groundwater in geologic processes.* Cambridge University Press.

Lohman, S.W. 1972. Groundwater hydraulics. U.S. Geological Survey Professional Paper 708. U.S. Government Printing Office, Washington, D.C.

Phillips, O.M. 1991. *Flow and reactions in permeable rocks.* Cambridge University Press.

Polubarinova-Kochina, P.Ya. 1962. *Theory of groundwater movement.* English translation by R. deWiest. Princeton University Press.

Rouse, H. and Ince, S. 1957. *History of hydraulics.* Dover, New York.

Scheidegger, Adrian E. 1957, 1960, 1974. *The physics of flow through porous media.* University of Toronto Press.

Sun, N.Z. 1996, *Mathematical modeling of groundwater pollution.* Springer-Verlag, New York.

Tolman, C.V. 1937, *Ground water.* McGraw-Hill, New York.

USSR National Committee for the International Hydrological Decade. 1978. *World water balance and water resources of the earth.* English translation. UNESCO, Paris.

PAUL CHAN

YUAN DING
JOHN R. SCHURING, JR.
New Jersey Institute of Technology

H

HAZARDOUS WASTE MANAGEMENT

HISTORICAL OVERVIEW

The development of the Resource Conservation and Recovery Act of 1976 dates to the passage of the Solid Waste Disposal Act of 1965, which first addressed the issues of waste disposal on a nationwide basis. Prior to the 1960s land disposal practices frequently included open burning of wastes to reduce volume, and were controlled only by the general need to avoid creating a public health impact and nuisance, such as a bad smell or visual blight—problems that one could see, smell, taste or touch. At that time, what few landfill controls existed were generally focused only on the basics of sanitation, such as rodent control, and the prevention of fires. The early concept of the "sanitary" landfill was to cover the waste with soil to reduce pests and vermin, create separate chambers of earth to reduce the spread of fire, and control odor and unsightly appearance—the key environmental concerns of the time.

Throughout the '60s and into the '70s, the use of industrial pits, ponds or lagoons on the land were viewed as legitimate treatment systems intended to separate solids from liquids and to dissipate much of the liquids. They were not only intended to store waste, but also to treat it. That is, solids would sink when settling occurred and the liquid could be drained, evaporated, or allowed to percolate into the ground. The accumulated solids ultimately would be landfilled.

Similarly for protection of receiving waters, pollution control laws prior to the mid-1960s were generally concerned with water-borne diseases and nuisances. The concept of water pollution was far more closely linked to the bacterial transmission of disease and physical obstruction or offense than it was to the impact of trace levels of chemicals. Waterways were viewed as natural systems that could handle waste if properly diluted and if the concentrations were within the assimilative capacity of the rivers and streams. The environmental concerns were primarily odor, appearance, oxygen content, and bacterial levels. Individual chemical constituents and compounds, at this time, were not typically regulated in a waterway.

The science of testing for and measuring individual contaminants was unrefined and typically not chemical specific until the 1970s. Water and wastewater analyses were generally limited to indicator parameters, such as Biochemical Oxygen Demand, turbidity, suspended solids, coliform bacteria, dissolved oxygen, nutrients, color, odor and specific heavy metals. Trace levels of individual chemical compounds and hazardous substances as we know them today were not among the parameters regularly analyzed.

"Hazardous waste" became a household word in the late 1970s with the publicity surrounding the Love Canal incident. How much waste has been disposed of is still questionable. Unfortunately, significant amounts were "thrown away" over the past decades and have endured in the environment in drum disposal sites such as "The Valley of the Drums" and in land disposal facilities where they have not degraded.

Throughout the '70s and '80s significant changes were made in the laws governing environmental protection. New laws adopted in the '70s include the Clean Air Act, the Federal Water Pollution Control Act, Safe Drinking Water Act, Resource Conservation and Recovery Act (RCRA), Toxic Substance Control Act, Marine Protection Research and Sanctuaries Act, and in 1980 the "Superfund" (CERCLA) statute. Of all the laws passed in the '70s, RCRA has had the greatest impact on the definition of wastes and the manner in which these wastes were to be managed, treated and handled. RCRA[1] required the US Environmental Protection Agency to establish management procedures for the proper disposal of hazardous wastes. These procedures are part of the Code of Federal Regulations dealing with environmental protection. They cover a "cradle-to-grave" procedure which regulates generators, transporters, storers and disposers of hazardous materials. Regulations for generators and transporters of hazardous wastes may also be found in the Code of Federal Regulations.[2,3]

Subsequent revisions to RCRA in 1984 included the provisions dealing with underground tanks, the restriction of land disposal of a variety of wastes, corrective action requirements for all releases, and the inclusion of a requirement of the EPA to inspect government and privately owned facilities which handle hazardous waste.

Today the law is again being considered for revision, and among the issues that are always under discussion include "how clean is clean" when remediating industrial and landfill sites. The cleanup standards are not consistent among state and federal programs, frequently causing significant discussion among responsible parties and regulators. At this time, risk assessments are used more often in an effort to design remedial programs that are appropriate for the media, and the resources being protected. A risk assessment might provide, for example, the necessary information to set differing groundwater cleanup goals in a sole source aquifer, than in an industrialized area situated above a brackish water-bearing zone where the groundwater will not again be used for potable purposes.

With the preceding paragraphs as general background, the brief discussion which follows on hazardous wastes emphasizes some of the technologies that have been successfully used for the treatment and disposal of hazardous wastes, and remediation of contaminated properties.

HAZARDOUS WASTE DEFINED

Hazardous wastes encompass a wide variety of materials. In 1987, the US EPA estimated that approximately 238 million tons could be classified as hazardous. This number is probably generous but suffice it to say that a great deal of material of a hazardous and dangerous nature is generated and disposed of every year.

The Resource Conservation and Recovery Act defines a hazardous waste as a solid waste that may cause or significantly contribute to serious health or death, or that poses a substantial threat to human health or the environment when improperly managed. Solid waste, under the present guidelines, includes sludges, liquids, and gases in bottles that are disposed of on the land.

From this working definition, a number of wastes have been defined as hazardous. These include materials that are ignitable, corrosive, reactive or explosive or toxic. These characteristic identifiers are further delineated in the regulations.[4] In addition, using these general characteristics and specific tests, the US Environmental Protection Agency has listed materials from processes, such as electroplating, or specific classes of materials, such as chlorinated solvents, or specific materials, such as lead acetate, or classes of compound, such as selenium and its compounds, which must be managed as "hazardous wastes" when they are disposed. This list changes periodically. In many cases disposers have treated materials not on the list as hazardous if they believe them to be so.

Some general classes of materials such as sewage, mining and processing of ore wastes are excluded by law at the present time.

Managing Wastes

Advancements in science and technology have given us opportunities to address environmental contamination issues in ways that are technologically more advanced, and more cost and time efficient than ever before. Technologies that were unknown, unproven and unacceptable to regulatory agencies just a few years ago, now exist and are being implemented at full scale. Regulations have changed, as have government policies governing cleanup and enforcement.

On a technical level, many ideas for hazardous waste treatment and remediation were rejected a few years ago by the engineering, business and regulatory community as being unproven or unreliable. Entrepreneurial scientists and engineers have adapted their knowledge of manufacturing process chemistry and engineering to the sciences of geology and hydrogeology and have refined the necessary equipment and techniques for waste treatment and remediation. Technologies have been tested at bench and pilot scale, and many have proven effective on a large scale. Pressure by the industrial community for engineers and regulators to reach a common ground has driven the process.

Contaminated soil and groundwater remedial techniques have tended toward the "active" end of the spectrum, with the installation of pumps, wells and above ground treatment systems of the capital and labor intensive variety. Progress has been made at the opposite end of the spectrum, ranging from intrinsic bioremediation, which involves no active treatment, to incremental levels of treatment that are far less costly than *ex-situ* pump and treat methods.

Programs like the EPA SITE (Superfund Innovative Technology Evaluation) Program and other Federal test and evaluation facilities, University research organizations and privately sponsored technology incubator and test evaluation facilities have been very successful in testing and establishing new hazardous waste treatment and disposal technologies. Currently, there are several dozen organizations nationally that specifically focus on the development of emerging hazardous waste treatment technologies. The results have been very positive, and many of today's front-edge technologies are the offspring of programs such as these.

On a regulatory/compliance level, the extensive time frame for receipt of approvals led many companies down the path of the traditional treatment and disposal methods, since they were "proven," as well as being approvable by the regulatory agencies. Environmental agencies have become more sophisticated, and cleanup levels are more often based on risk rather than standards set at an earlier data in technical and regulatory development. More than ever, agency personnel are now trained as specialists in the various segments of the environmental industry, including risk assessment, hydrogeology, remediation engineering and personal protection. As a result, the agencies are often more willing to engage in discussions regarding site specific conditions and remedial goals. Further, modifications to state permitting programs have allowed variations on typical operating permits for new and emerging technologies that appear to have promise.

An analysis of Superfund remediation activities indicates that significant progress has been made in the use of innovative technologies for site remediation. The predominant new technologies used at Superfund sites include soil vapor extraction (SVE) and thermal desorption. It is important to note that there are many derivative technologies that will now stand a greater chance of receiving government and industry support as a result.

Remediation technologies that are derived from soil vapor extraction include dual phase extraction and sparing. The two phases are typically a) removal of free product or contaminated groundwater and b) vapor. The *in-situ* addition of certain compounds by sparging into the soil and groundwater has made bioremediation attractive. The addition of the additional components to an earlier technology that was moderately successful has made the modified treatment train much more effective. The new treatment train is therefore more approvable.

On a financial level, methods have been developed for the evaluation of large projects to provide a greater degree of financial assurance. The concept of the "unknown" cost of remediation due to the inability of scientists to accurately see and measure subsurface contamination is diminishing. Probabilistic cost analyses are frequently completed on assignments so that final remediation costs can be predicted within a much narrower range.

Management practices have changed dramatically over the past 20 years at most industries. They have been driven by the improvements in technologies, as well as the laws and regulations. The real estate boom of the 1980s also impacted operating practices, as many properties were bought and sold during this time. The desire of buyers to be assured that they were purchasing "clean" properties, as well as some state environmental property transfer requirements, was the genesis of facility environmental audits as we now know them.

For purposes of discussion, hazardous wastes fall primarily into two categories, organic and inorganic. Some management technologies will apply to both, but in general organic material can be destroyed to relatively innocuous end products while inorganic material can only be immobilized. The key technologies for hazardous waste management include:

- Pollution Prevention
- Recycling and Reuse
- Waste Minimization
- Chemical Treatment and Detoxification
- Destruction
- Stabilization
- Land disposal

Of these, land disposal is the least attractive alternative from the standpoint of long-term liability exposure and environmental impact.

Waste Concentration—A Key Where a waste must be ultimately disposed of, concentration or volume reduction is beneficial. The simplest approach to this is to separate wastes at the source; that is, at the place of origin. This will increase handling costs and effort, but will more than pay dividends in minimizing analytical and disposal costs. First, it will mean that analysis must be done less frequently. Second, waste can be disposed of at the lowest degree of care consistent with the most hazardous contaminant, thus minimizing the volume of waste that must get a greater degree of care because of slight cross-contamination by a more toxic material. This is true whether the material is in the liquid or solid state.

Another method of reducing volume is concentration. For liquids, this generally means distillation or evaporation. Evaporation to date has been acceptable, however, with increased emphasis on the presence of volatile hazardous materials in the atmosphere, evaporation ponds, will, in all probability, no longer meet the necessary standards for waste control and management. In addition, ponds must be permitted under RCRA, which imposes additional financial and operating requirements on the waste concentrator. Double and triple effect evaporators and distillation units will be acceptable but are very energy-expensive. Innovative techniques will be required because of the high energy of the traditional liquid separation systems.

Where a material is dissolved in water or an organic solvent, precipitation may be advisable. The solid can then be separated out from the majority of the liquid by filtration or other liquid/solid separation technology. Typical of this would be the precipitation of lead by the use of a sulfide salt, resulting in lead sulfide which has extremely low solubility. The solid may be suitable for reclamation at present or be stored in a secure landfill in a "non- or less-hazardous form" for eventual reuse.

Pollution Prevention The passage of Pollution Prevention Laws has driven many industries toward better utilization of their resources. Many companies now actively participate in the preparation and update of a pollution prevention program, designed to guide personnel toward goals established to improve waste generation and disposal practices.

Traditional environmental quality and pollution control programs typically focus on an end-of-pipe approach. The pollution prevention plan approach typically begins earlier in the "equation" by reviewing an operation and making modifications that will positively impact a facility. Some examples include reducing harmful chemical purchases, increasing operation efficiencies, and ultimately generating a smaller quantity of waste.

The pollution plan approach will include involvement by a wider range of facility personnel than the traditional environmental management approach. Purchasing, accounting, production and engineering all participate. Proponents suggest that a program is easy to implement, although corporate personnel involved in the effort know that it is an effort which requires broad-based management support, is time consuming, and not necessarily inexpensive to implement. The benefits are potentially significant, as reduced emissions make it easier to comply with discharge standards, and will reduce long-term liabilities.

Recycling and Reuse In many cases, in addition to economically attractive alternatives, a very attractive alternative will be recycling or reuse of hazardous wastes. The economic realities of the regulations, where disposal of a barrel of waste can demand a 5–$10 per gallon, and up to $1,200 per ton or greater fee, may make processing for recycling and/or reuse the best practice. In the present context, we are defining recycling as internal to the plant, and reuse as external to the plant. This is not a legal definition which defines recycling as essentially both internal and external, but it is helpful in this discussion.

Internal recycling will require, in general, high efficiency separation and potential additional processing. Thus, if a solvent is being recycled, impurities such as water, by-products, and other contaminants must be removed. Depending on the volumes involved, this may be done internally to the process or externally on a batch basis.

Reuse involves "selling" the waste to a recycle and reclaimer. The reclaimer then treats the waste streams and recovers value from them. The cleaned-up streams are then his products for sale.

From a regulatory, liability perspective, there are advantages to reuse as the liability for the waste ends when it is successfully delivered to the reclaimer. Because he processes the material, he then assumes responsibility for the products and wastes that are generated. If the material is internally recycled, then the recycler, that is the plant, maintains responsibility for any wastes that are generated as a result of the recycling operation.

In some cases, it may be desirable to dispose of wastes directly to the user. This is particularly true when there are large quantities involved and a beneficial arrangement can be worked out directly. Waste exchanges have been organized to promote this type of industrial activity. Detailed discussions of their mode of operation can be obtained directly from the exchanges.

Waste Minimization The alternative scenario development will be not only site, but substance specific. Two basic approaches to hazardous waste management are:

1) In-process modifications
2) End-of-pipe modifications

Each will have advantages and disadvantages that are processes, substance, and site specific.

In-process alternatives include changing process conditions, changing feedstocks, modifying the process form in some cases, or if necessary eliminating that process and product line.

In-process modification is generally expensive and must be considered on a case-by-case basis. There are some potential process modifications that should be considered to minimize the production of toxic materials as by-products. These include minimization of recycling so side-reaction products do not build up and become significant contributors to the pollution load of a bleed stream. For example, waste must be purged regularly in the chlorination of phenols to avoid the build-up of dioxin. It may also be desirable to optimize the pressure of by-products. For example, phenol is produced and found in condensate water when steam-cracking naphtha to produce ethylene unless pressures and temperatures are kept relatively low.

It may be desirable to change feedstocks in order to eliminate the production of hazardous by-products. For example, cracking ethane instead of naphtha will yield a relatively pure product stream.

Hydrazine, a high energy fuel, was originally produced in a process where dimethylnitrosamine was an intermediate. A very small portion of that nitrosamine ended up in a waste stream from an aqueous/hydrocarbon separation. This waste stream proved to be difficult, if not impossible, to dispose of. A new direct process not involving the intermediate has been substituted with the results that there are no noxious wastes or by-products.

In the ultimate situation, production of a product may be abandoned because either the product or a resulting by-product poses an economic hazard which the corporation is not willing to underwrite. These include cases where extensive testing to meet TSCA (Toxic Substances Control Act) was required. They include the withdrawal of pre-manufacturing notice applications for some phthalate ester processes. However, production of certain herbicides and pesticides was discontinued because a by-product or contaminant was dioxin.

Treatment/Destruction Technology

Chemical Treatment/Detoxification Where hazardous materials can be detoxified by chemical reaction, there the molecule will be altered from one that is hazardous to one or more that are non-hazardous, or at least significantly less hazardous. For example, chlorinated hydrocarbons can be hydro-dechlorinated. The resulting products are either HCl or chlorine gas and nonchlorinated hydrocarbons. A number of these processes are being developed for the detoxification of PCB (polychlorinated biphenols) and are being demonstrated as low concentrations of PCB's in mineral oil. The end products, if concentrated enough, can be useful as feedstocks or the hydrocarbons may be used as fuel.

Cyanide can be detoxified using any number of chemical reactions. These include a reaction with chlorine gas to produce carbonate and chlorine salt. Cyanide can also be converted to cyanate using chlorine gas. In addition, ozone can be utilized to break up the carbon-nitrogen bond and produce CO_2 and nitrogen.

Hexavalent chromium is a toxic material. It can be reduced to trivalent chromium which is considerably less hazardous and can be precipitated in a stable form for reuse or disposal as a non-hazardous material. Chromium reduction can be carried out in the presence of sulfur dioxide to produce chromium sulfate and water. Similar chemistry is utilized to remove mercury from caustic chlorine electrolysis cell effluent, utilizing sodium borohydride.

Lead, in its soluble form, is also a particularly difficult material. Lead can be stabilized to a high insoluble form using sulfur compounds or sulfate compounds, thus removing the hazardous material from the waste stream.

Acids and bases can most readily be converted to nonhazardous materials by neutralizing them with appropriate base or acid. This is probably the simplest chemical treatment of those discussed and is widely applicable; care must be taken, however, to insure that no hazardous precipitates or dissolved solids forms.

Incineration Incineration has been practiced on solid waste for many years. It has not, however, been as widely accepted in the United States as in Europe where incineration with heat recovery has been practiced for at least three decades. Incineration of industrial materials has been practiced only to a limited extent; first, because it was more expensive than land disposal, and second, because of a lack of regulatory guidelines. This has changed because landfills are not acceptable or available, costs for landfilling are becoming extremely high, and regulatory guidance is available. Equipment for incineration of industrial products has been, and is available, however, it must be properly designed and applied.

Incineration is the oxidation of molecules at high temperatures in the presence of oxygen (usually in the form of air) to form carbon dioxide and water, as well as other oxygenated products. In addition, products such as hydrogen chloride are formed during the oxidation process. The oxidation, or breakdown, takes place in the gaseous state, thus requiring vaporization of the material prior to any reaction. The molecules then breakdown into simpler molecules, with the least stable bonds breaking first. This occurs at relatively lower temperatures and shorter times. It is followed by the breakdown of the more stable, and then the most stable bonds to form simple molecules of carbon dioxide, water, hydrogen chloride, nitrogen oxides, and sulfur oxides, as may be appropriate.

Thus, the primary considerations for successful oxidation or destruction are adequate time and temperature. Good air/waste contact is also important. Regulatory guidelines require a destruction and removal efficiency (DRE) of 99.99% thus, time and temperature become all the more important. For the most refractory compounds, such as PCB's, residence times in excess of three seconds and temperatures in excess of 1000°C are required. These temperatures may be reduced in light of special patented processes utilizing oxidation promoters and/or catalysts. As a result of the high required DRE, a test burn is required to demonstrate adequate design.

In addition to time and temperature considerations, there are other important factors which must be considered when designing or choosing equipment to incinerate industrial waste. Most important is adequate emission gas controls. Where materials which contain metals, chlorides, or sulfides are to be incinerated, special provisions must be made to minimize emission of HCl, SO_2, and metal oxides. Usually a scrubber is required, followed by a system to clean up the scrubber-purge water. This system includes neutralization and precipitation of the sulfur and metal oxides. In addition, where high temperature incineration is practiced, control of nitrogen oxides to meet air quality emissions standards must be considered. These substances do not present insurmountable technological challenges, as they have been handled satisfactorily in coal-fired power plant installations, but they do present added economic and operating challenges.

Several types of incineration facilities should be considered. Unfortunately, the standard commercial incinerator utilized or municipal waste will generally not prove adequate for handling industrial waste loads because the temperatures and residence times are inadequate. Municipal incinerators are designed to handle wastes with an energy content below 8000 Btu/pound, while industrial wastes can have heating values as high as 24000 Btu/pound. Municipal incinerators are generally not designed to accept industrial wastes.

A number of incinerator facilities have been built for industrial wastes. Small, compact units, utilizing a single chamber with after-burner, or two-stage, multi-chamber combustion are available. In general, a single-state unit will not suffice unless adequate residence time can be assured.

Rotary kiln incinerators are of particular interest for the disposal of industrial materials. Generally, they are only applicable for large-scale operations, and can handle a large variety of feedstocks, including drums, solids and liquids. Rotary cement kilns have been permitted to accept certain types of organic hazardous materials as a fuel supplement.

Of increasing interest for industrial incineration is the fluid bed incinerator. This has the additional advantage of being able to handle inorganic residues, such as sodium sulfate and sodium chloride. These units provide the additional advantage of long residence time, which may be desirable when the waste is complex (e.g., plastics) or has large organic molecules. On the other hand, gas residence times are short, and an after-burner or off-gas incinerator is often required in order to achieve the necessary DRE.

Incineration has been used successfully for the disposal of heptachlor, DDT, and almost all other commercial chlorinated pesticides. Organo-phosphorous insecticides have also been destroyed, but require a scrubbing system, followed by a mist eliminator, to recover the phosphorous pentoxide that is generated.

Some special incineration applications have been implemented. These include:

- An ammonia plant effluent containing organics and steam is oxidized over a catalyst to form CO_2, water and nitrogen;
- Hydrazine is destroyed in mobile US Air Force trailers which can handle 6 gpm of 100% hydrazine to 100% water solutions, and maintain an emission has which contains less than 0.03 pound/minute of NO*x*;
- Chlorate-phosphorous mixtures from fireworks ammunition are destroyed in a special incinerator which has post-combustion scrubbing to collect NO*x*, P_4O_{10}, HCl, SO_2 and metal oxides;
- Fluid bed incinerators which handle up to 316 tons per day of refinery sludge and 56 tons of caustic are being utilized.

Wet Air Oxidation Although not strictly incineration, wet air oxidation is a related oxidation process. Usually air, and sometimes oxygen, is introduced into a reactor where hazardous material, or industrial waste, is slurried in water at 250° to 750°F.

Operating pressures are as high as 300 psig. Plants have been built to treat wastes from the manufacture of polysulfite rubber and other potentially hazardous materials. Emissions are similar to those obtained in incineration, with the exception that there is liquid and gaseous separation. Careful evaluation of operating conditions and materials of destruction are required.

Pyrolysis Pyrolysis transforms hazardous organic materials by thermal degradation or cracking, in the absence of an oxidant, into gaseous components, liquid, and a solid residue. It typically occurs under pressure and a temperature above 800°F.

To date, the process has found limited commercial application but continues to be one that will eventually be economically attractive, the prime reason being the potential for recovery of valuable starting materials. A great deal of experimentation has been carried out both on municipal and industrial wastes. For example, polyvinyl chloride can be thermally degraded to produce HCl and a variety of hydrocarbon monomers, including ethylene, butylene, and propylene. This is a two-stage degradation process with the HCl coming off at relatively low temperatures (400°C) and the hydrocarbon polymer chain breakdown can be obtained with Polystyrene, with styrene as the main product, and most other polymers. Experimental work carried out in the early 1970s by the US Bureau of Mines, indicates that steel-belted radial tires can be pyrolyzed to reclaim the monomers, as well as gas and fuel oil.

Other target contaminant groups include SVOCs and pesticides. The process is applicable for the treatment of refinery, coal tar, and wood treating wastes and some soils containing hydrocarbons.

Disposal Technology

Land Storage and Disposal Disposal of hazardous materials to the land remains the most common practice. It is highly regulated and a practice which has been limited because of public pressure and federal rules which require the demonstration of alternate means of disposal. The design of secure landfills for the acceptance of hazardous materials must be such that ground waters, as well as local populations are protected. The US Environmental Protection Agency has implemented strict landfills. In practice all landfills accepting hazardous wastes must insure that the wastes stored in close proximity are compatible so that no violent reactions occur should one or more waste leak.

Federal and State regulations prohibit the disposal of liquids in landfills. Of equal importance to the disposal of hazardous wastes, whether solid or semi-solid, is the assurance that material will not leach away from the landfill or impoundment. This assurance is provided by the use of "double-liners" with a leak detection system between the liners, a leachate collection system for each cell, and a leachate treatment system designed and operated for the facility.

In dilute form liquid wastes can be "landfarmed" where microbial action will decompose the compounds over time. This methodology has been utilized over many years for hydrocarbons and has worked well. For highly toxic compounds, such as chlorinated organics, it is less attractive even though decomposition does occur. Land treatment of PCB contaminated soils has been tested with some success.

Stabilization The stabilization of hazardous materials prior to land disposal is frequently practiced. Generally, the stabilization is in the form of fixing the hazardous material with a pozzolanic material, such as fly ash and lime, to produce a solid, non-leachable product which is then placed in land disposal facilities. Typically, this methodology is applicable to inorganic materials. Most of the commercial processes claim that they can handle materials with some organic matter.

Polymer and micro-encapsulation has also been utilized but to a significantly lesser extent than the commercially available process which utilize pozzolanic reactions. Polymers which have been utilized include polyethylene, polyvinylchloride and polyesters.

Grube[9] describes a study of effectiveness of a waste solidification/stabilization process used in a field-scale demonstration which includes collecting samples of treated waste materials and performing laboratory tests. Data from all extraction and leaching tests showed negligible release of contaminants. Physical stability of the solidified material was excellent.

Remediation Technologies

Natural Attenuation and Bioaugmentation The concept of natural attenuation, or intrinsic bioremediation, has gained a greater acceptance by the regulatory community as data presented by the scientific community have demonstrated the results of natural attenuation, and the costs and time frames associated with traditional remedial methods.[1] This approach is most appropriate for the dissolved phase groundwater contamination plume. It is still necessary to remove or remediate the source zone of an affected aquifer, after which natural attenuation may be a reasonable approach to the dissolved phases.

Natural attenuation should not be considered "No Action." It requires a solid understanding of the contaminant, geologic and aquifer characteristics, and a defined plan of action. The action involves demonstrating that the contaminants will breakdown, will not migrate beyond a specified perimeter, and will not impact potential receptors. It may involve the stimulation of microorganisms with nutrients or other chemicals that will enable or enhance their ability to

---

[1] Example of traditional remediation methods are *ex-situ* treatment of soil and groundwater, such as soil excavation/disposal, groundwater pump-and-treat using air stripping and granulated carbon polishing.

degrade contaminants. Some limitations may include inappropriate site hydrogeologic characteristics (including the inability of the geostrata to transport adapted microorganisms) and contaminant toxicity. Monitoring and reporting is required, and a health-based risk assessment may be required by regulators.

Natural attenuation is frequently enhanced by several components, such as the creation of a barrier or the addition of a chemical or biologic additive to assist in the degradation of contaminants.

The overall economics of this approach can be significantly more favorable than the typical pump-and-treat approach. One must be careful to consider, however, that the costs of assessment will equal or exceed that necessary for other methods, and the costs associated with sentinel monitoring will be borne for a longer period of time.

Barriers This has been used in instances where the overall costs of the remedial action is very high, and the geologic features are favorable. It involves the installation of a physical cut-off wall below grade to divert groundwater. The barriers can be placed either upgradient of the plume to limit the movement of clean groundwater through the contaminated media, or downgradient of the plume with openings or "gates" to channel the contaminated groundwater toward a remedial system. This technology has proven to be more efficient and less costly than traditional pump and treat methods, but also requires favorable hydrogeologic conditions. It allows for the return of treated groundwater to the upgradient end of the plume with a continuous "circular" flushing of the soil, rather than allowing the dilution by groundwater moving from the upgradient end of the plume. The result is greater efficiency, and a shorter treatment time period. While the cost of the cutoff wall is significant, it is important to conduct a proper analysis of long-term pump-and-treat costs, including the operation and maintenance of a system that would otherwise be designed to accept a much larger quantity of groundwater.

The creation of a hydraulic barrier to divert upgradient groundwater from entering the contaminant plume allows the pumping of groundwater directly from the affected area and often allows the reinjection of the treated water back into the soils immediately upgradient of the plume. This allows for the efficient treatment of the impacted area, without unnecessary dilution of the contaminated groundwater plume. It does, however, require an accurate assessment of the groundwater regime during the assessment stage. This promising concept is not radical, but its use in connection with natural remediation is growing rapidly.

Passive Treatment Walls Passive treatment walls can be constructed across the flow path of a contaminant plume to allow the groundwater to move through a placed media, such as limestone, iron filings, hydrogen peroxide or microbes. The limestone acts to increase the pH, which can immobilize dissolved metals in the saturated zone. Iron filings can dechlorinate chlorinated compounds. The contaminants will be either degraded or retained in concentrated form by the barrier material.

Physical Chemical Soil Washing Soil is composed of a multitude of substances, with a large variance in size. These substances range from the very fine silts and clays, to the larger sand, gravel and rocks. Contaminants tend to adsorb onto the smallest soil particles, as a result of the larger surface per unit of volume. Although these smaller particles may represent a small portion of the soil volume, they may contain as much as 90% of the contamination.

Soil washing involves the physical separation, or classification, of the soil in order to reduce the volume requiring treatment or off-side disposal. It is based on the particle size separation technology used in the mining industry for many decades. The steps vary, but typically begin with crushing and screening. It is a water-based process, which involves the scrubbing of soil in order to cause it to break up into the smallest particles, and its subsequent screening into various piles. The fraction of the soil with the highest concentration of contamination can be treated using technologies frequently used by industry. The goal is to reduce the quantity of material that must be disposed. The clean soil fractions can often be returned to the site for use as fill material where appropriate.

The use of soil washing technology has some limitations, including a high initial cost for pilot testing and equipment setup. It will be most useful on large projects (requiring remediation of greater than 10,000 cubic yards of soil). Sites with a high degree of soil variability, and a significant percentage of larger particles will show the greatest economic benefit.

Soil Vapor Extraction Soil Vapor Extraction (SVE) is an effective method for the *in-situ* remediation of soils containing volatile compounds. Under the appropriate conditions volatile organic compounds will change from the liquid phase to the vapor phase, and can be drawn from the subsurface using a vacuum pump. There are several factors necessary for the successful use of this technology, including 1) the appropriate properties of the chemicals of concern (they must be adequately volatile to move into a vapor phase), and 2) an appropriate vapor flow rate must be established through the soils.

Air is drawn into the soils via perimeter wells, and through the soils to the vapor extraction well. It is drawn to the surface by a vacuum pump and subsequently through a series of manifolds to a treatment system such as activated carbon or catalytic oxidation.

A concentration gradient is formed, whereby in an effort to reach equilibrium, the liquid phase volatile contaminants change into the vapor phase and are subsequently transported through the soils to the treatment system.

This technology is particularly effective for defined spill areas, with acceptable soils. It is most effective in remediating the soils in the vadose zone, the area that is in contact with the fluctuating groundwater table. Groundwater contaminated with these compounds and similar soil conditions can be remediated using air sparging, a variation of soil vapor extraction.

A variation of this technology is thermal enhanced SVE, using steam/hot air injection or radio frequency heating to increase the mobility of certain compounds.

Air Sparging Air sparging is the further development of soil vapor extraction, wherein that process is extended so that soils and groundwater in the capillary fringe can be effectively treated. Air sparging involves injecting air or oxygen into the aquifer to strip or flush volatile contaminants from the groundwater and saturated soils. As the air channels up through the groundwater, it is captured through separate vapor extraction wells and a vapor extraction system. The entire system essentially acts as an *in-situ* air stripper. Stripped, volatile contaminants usually will be extracted through soil vapor extraction wells and usually require further treatment, such as vapor phase activated carbon or a catalytic oxidation treatment unit. This technology is effective when large quantities of groundwater must be treated, and can provide an efficient and cost-effective means of saturated zone soil and groundwater remediation.

The biological degradation of organic contamination in groundwater and soil is frequently limited by a lack of oxygen. The speed at which these contaminants are degraded can be increased significantly by the addition of oxygen in either solid or liquid form. Air sparging is often combined with *in-situ* groundwater bioremediation, in which nutrients or an oxygen source (such as air or peroxide) are pumped into the aquifer through wells to enhance biodegradation of contaminants in the groundwater.

Oxygen Enhancement/Oxidation In this *in-situ* process, hydrogen peroxide is used as a way of adding oxygen to low or anoxic groundwater, or other oxidative chemicals are added as an oxidant to react with organic material present, yielding primarily carbon dioxide and water. The application of this technology is typically through the subsurface injection of a peroxide compound. It has been injected as a liquid, above the plume, and allowed to migrate downward through the contaminated plume. Alternately, it has been placed as a solid in wells located at the downgradient edge of the plume; in this fashion it can act as a contamination "barrier," limiting the potential for contaminated groundwater to move offsite. As the organic contaminated groundwater moves through the high oxygen zone, the contaminant bonds are either broken, or the increased oxygen aid in the natural biodegradation of the compounds.

The process is exothermic, causing a temperature increase in the soils during the process. This acts to increase the vapor pressure of the volatile organic compounds in the soil, and subsequently increases volatilization of the contaminants. This process can be utilized in connection with a soil vapor extraction and/or sparging system to improve remediation time frames.

It does not act, however, on the soil groundwater vadose zone. This may not be a critical flaw, however, since the strategic placement of the wells may positively impact the contaminant concentrations adequately to meet cleanup standards.

Dual Phase Extraction Dual phase extraction is an effective method of remediating both soils and groundwater in the vadose and saturated zones where groundwater and soil are both contaminated with volatile or nonvolatile compounds. It is frequently used for contaminant plumes with free floating product, combined with known contamination of the vadose zone. This technique allows for the extraction of contaminants simultaneously from both the saturated and unsaturated soils *in-situ.* While there are several variations of this technique, simply put, a vacuum is applied to the well, soil vapor is extracted and groundwater is entrained by the extracted vapors. The extracted vapors are subsequently treated using conventional treatment methods while the vapor stream is typically treated using activated carbon or a catalytic oxidizer.

The process is frequently combined with other technologies, such as air sparging or groundwater pump-and-treat to minimize treatment time and maximize recovery rate.

Chemical Oxidation and Reduction Reduction/oxidation reactions chemically convert hazardous contaminants to nonhazardous or less toxic compounds that are more stable, less mobile and/or inert. The oxidizing agents typically used for treatment of hazardous contaminants are ozone, hydrogen peroxide, hypochlorites, chlorine and chlorine dioxide. These reactions have been used for the disinfection of water, and are being used more frequently for the treatment of contaminated soils.

The target contaminant group for chemical reduction/oxidation reactions is typically inorganics, however hydrogen peroxide has been used successfully in the *in-situ* treatment of groundwater contaminated with light hydrocarbons.

Other Technologies Many other technologies are being applied with increasing frequency. The following is only a very brief description of several that have promise.

- *Surfactant enhanced recovery* Surfactant flushing of non-aqueous phase liquids (NAPL) increases the solubility and mobility of the contaminants in water, so that the NAPL can be biodegraded more easily in the aquifer or recovered for treatment aboveground via pump-and-treat methods.
- *Solvent extraction* Solvent extraction has been successfully used as a means of separating hazardous contaminants from soils, sludges and sediments, and therefore reducing the volume of hazardous materials that must be treated. An organic chemical is typically used as a solvent, and can be combined with other technologies, such as soil washing, which is frequently used to separate, or classify, various soil particles into size categories. The treatment of the concentrated waste fraction is then treated according to its specific characteristics. Frequently, the larger volume of treated material can be returned to the site.
- *Bioremediation using methane injection* The method earlier described for the injection of hydrogen peroxide into wells has also been successfully utilized using methane. It is claimed that this bioremediation process uses microbes which co-metabolize methane with TCE and other chlorinated solvents,

potentially cutting treatment costs and time frames by 30 to 50%.

- *Thermal technologies* The EPA has conducted tests of thermally-based technologies in an evaluation of methods to treat organic contaminants in soil and groundwater. Low temperature thermal desorption is a physical separation process designed to volatilize water and organic contaminants. Typical desorption designs are the rotary dryer and the thermal screw. In each case, material is transported through the heated chamber via either conveyors or augers. The volatilized compounds, and gas entrained particulates are subsequently transported to another treatment system for removal or destruction.

Mobile incineration processes have been developed for use at remedial sites. While permitting is frequently a problem, the economics of transporting large quantities of soil can drive this alternative. One method is a circulating fluidized bed, which uses high-velocity air to circulate and suspend the waste particles in a combustion loop. Another unit uses electrical resistance heating elements or indirect-fired radiant U-tubes to heat the material passing through the chamber. Each requires subsequent treatment of the off gases. Also certain wastes will result in the formation of a bottom ash, requiring treatment and disposal.

In summary, the current business and regulatory climate is positive for the consideration of alternate treatment technologies. The re-evaluation of ongoing projects in light of regulatory and policy changes, as well as new technological developments may allow cost and time savings. The arsenal of techniques and technologies has developed substantially over the years, as has our knowledge of the physical and chemical processes associated with the management of wastes. Effluents and contaminated media are now easier to target with more efficient and cost-effective methods.

BIBLIOGRAPHY

1. Pojasek, R.B. (ed.), *Toxic and Hazardous Waste Disposal*, 1, Processes for Stabilization and Solidification, Ann Arbor Science, Ann Arbor, Michigan, 1979.
2. Merry, A.A. (ed.), *The Handbook of Hazardous Waste Management*, Technomic, Westport, Connecticut, 1980.
3. Overcash, M.R., *Decomposition of Toxic and Nontoxic Organic Compounds in Soils*, Ann Arbor Science, Ann Arbor, Michigan, 1981.
4. *Toxic and Hazardous Industrial Chemicals Safety Manual.* The International Technical Information Institute, Tokyo, 1981.
5. Bertherick, L., *Handbook of Reactive Chemical Hazards*, Butterworths, London, 1979.
6. Hatayma, H.K., *et al.,* A Method of Determining Hazardous Waste Compatibility, USEPA, Cincinnati, 1981.
7. Kaing, Y. and Metry, A.A., *Hazardous Waste Processing Technology*, Ann Arbor Science, Ann Arbor, Michigan, 1982.
8. *Damages and Threats Caused by Hazardous Material Sites*, US EPA/430/9–80/004, USEPA, Washington, 1980.
9. *Management of Uncontrolled Hazardous Waste Sites*—US EPA *Conference Proceedings*, USEPA, 1980.
10. Stoddard, S.K., *et al.*, *Alternatives to the Land Disposal of Hazardous Wastes—An Assessment for California, Office of Appropriate Technology*, State of California, 1981.
11. Grube, W.E., Jr., "Evaluation of Waste Stabilized by the Solid Tech Site Technology," *J. Air Waste Manag. Assoc.* (1990).
12. Evanoff, S.P., *Hazardous Waste Reduction in the Aerospace Industry, Chem. Eng. Prog.*, 86, 4, 51 (1990).
13. Jackson, D.R., *Evaluation of Solidified Residue from Municipal Solid Waste Combustor*, EPA Repot 600/52–89/018 Feb. 1990.
14. *Innovative Hazardous Waste Treatment Technologies: A Developers Guide to Support Services, Third Edition*, EPA Report EPA/542-B-94–012, September 1994.
15. *Hazardous Waste Clean-up Information Database* (*CLU-IN*), US EPA, 1996.
16. *Innovative Treatment Technologies: Annual Status Report* (*Seventh Edition*) *Applications of New Technologies at Hazardous Waste Sites,* USEPA Report EPA-542-R-95–008, Number 7, Revised September 1995.
17. *Remediation Case Studies: Soil Vapor Extraction,* USEPA Report EPA-542-R-95–004, March 1995.
18. *Superfund Innovative Technology Evaluation Program, Technology Profiles Seventh Edition,* USEPA Report, EPA/540/R-94/526, November 1994.
19. *Superfund XV Abstract Book*, Hazardous Materials Control Resources Institute, November 1994.
20. *Remediation Technologies Screening Matrix and Reference Guide*, USEPA Report, EPA 542-B-93–005, July 1993.
21. *Remediation Case Studies: Thermal Desorption, Soil Washing, and In Situ Vitrification,* USEPA Report, EPA-542-R-95–005, March 1995.
22. *Proceedings, Fifth Forum on Innovative Hazardous Waste Treatment Technologies: Domestic and International,* USEPA Report, EPA/540/R-94/503, May 1994.
23. LaGreca, M.D., Buckingham, P.L., Evans, J.C., *Hazardous Waste Management,* McGraw-Hill, Inc., 1994.
24. Freeman, H.M. (ed.), *Standard Handbook of Hazardous Waste Treatment and Disposal,* McGraw-Hill, Inc., 1989.
25. Sell, N.J., *Industrial Pollution Control: Issues and Techniques*, Second Edition, Van Nostrand Reinhold, 1992.
26. Corbitt, R.A. (ed.), *Standard Handbook of Environmental Engineering*, McGraw-Hill, Inc., 1990.
27. Kolluru, R.V. (ed.), *Environmental Strategies Handbook, A Guide to Effective Policies & Practices*, McGraw-Hill, Inc., 1994.

REFERENCES

1. PL 95-580, Resource Conservation and Recovery Act of 1976, 42 USC 6901, 1976.
2. 40 CFR 262.
3. 40 CFR 263.
4. 40 CFR 261.
5. 40 264, 265.
6. SW-968, Permit Applicants' Guidance Manual for the General Facility Standards of 40 CFR 264, Oct. 1983.
7. Lindgren, G.D., "Managing Industrial Hazardous Waste: A Practical Handbook," 350 pp., 1989, Lewis Publ., Boca Raton, FL.
8. Industrial Pollution Prevention Planning, Meeting Requirements Under the New Jersey Pollution Prevention Act, New Jersey Department of Environmental protection, Office of Pollution Prevention, September 1985, Second Edition.
9. Grube, W.E., Jr., "Evaluation of Waste Stabilized by the Solid Tech Site Technology," *J. Air Waste Manag. Assoc.*, 40 310 (1990).

RICHARD T. DEWLING
GREGORY A. PIKUL
Dewling Associates, Inc.

HAZARDOUS WASTES

A hazardous waste is a solid waste than may (1) cause or significantly contribute to an increase in serious irreversible or incapacitating reversible illness, or (2) pose a substantial threat or potential hazard to human health or the environment when improperly treated, stored, transported, or disposed of, or otherwise managed. Hazardous wastes are managed under two categories: (1) waste material currently being generated and subject to control by the Resource Conservation and Recovery Act (RCRA), and (2) wastes that have accumulated at inactive or abandoned sites or wastes resulting from spills that require emergency response. Both are addressed by the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), commonly known as Superfund.

The RCRA identifies hazardous wastes based upon lists of specific wastes and four characteristics. All waste generators are required to determine if any of these four characteristics are exhibited by their wastes.

SOLID WASTES

Hazardous wastes are managed as part of solid wastes; therefore, a discussion of the legal definition of solid waste is necessary. The word "solid," in this case, is misleading, because solid wastes can include liquids and contained gases. Thus, the important term is "waste," which is any material that is discarded—for example, discarded military munitions are considered solid wastes. Recycling has become a major factor in waste management, and the manner in which a waste is recycled determines whether it is to be treated as a solid waste. Material in a waste stream that is recycled by introducing it directly as an ingredient in a production process is not a solid waste, nor is it a solid waste if it is used as a substitute for a commercial product. On the other hand, waste disposed of by burning as a fuel, or in fuels, is a solid waste, as is material containing dioxin. If wastes are stored for recycling but are not 75% recycled in a calendar year, they are solid wastes.

Certain specific types of wastes, such as untreated domestic sewage, industrial wastewater discharges, irrigation return flows, and nuclear materials as defined by the Atomic Energy Act of 1954, are excluded from the definition as solid wastes.

HAZARDOUS WASTES

The growth of industry during World War II was accompanied by a great increase in the generation of hazardous wastes. It is estimated that this increase was accelerated after the war from half a million metric tons per year to almost 280 million metric tons by 1995. This phenomenal rate of growth caused Congress to react by passing the RCRA in 1976 as an amendment to the Solid Waste Disposal Act (SWDA) of 1965. The RCRA itself has been amended several times, most significantly in 1984 by the Hazardous and Solid Waste Amendments (HSWA) of 1984. The management of underground storage tanks (USTs) was included among the amendments.

There are three subtitles in the RCRA that have special significance to hazardous wastes. Subtitle C describes the hazardous-waste program, and Subtitle I describes the underground storage-tank problem. Subtitle D, which pertains to solid wastes, is important because a hazardous waste must first be identifiable as a solid waste.

As directed by Congress, the Environmental Protection Agency (EPA) has defined hazardous wastes by listing certain specific solid wastes as hazardous and by identifying characteristics of a solid waste that make it hazardous.

Is a waste Hazardous?

1. Is the waste excluded? (40 CFR 261.4) Code of Federal Regulations

If not,

2. Is it listed? (40 CFR 261, Subpart D)

F	Non specific sources such as spent solvents.
K	Source specific wastes.
P - U	Discarded products, off-spec material, container or spill residues.

If not,

3. Is it a characteristic waste? (40 CFR 261 Subpart C)

All of the wastes on these lists are assigned an identification number. All listed wastes are presumed to be hazardous regardless of their concentration and must be handled according to Subtitle D of RCRA.

Hazardous-Waste Characteristics

A waste that falls into one of the lists must still be examined to determine if it exhibits one or more of these characteristics: ignitability, corrosivity, reactivity, or toxicity. If it does, it is considered to represent an additional hazard and may necessitate special regulatory precautions.

Ignitability: Liquid wastes with a flash-point test less than 60°C and nonliquids that spontaneously catch fire through friction or by reaction with moisture and materials that meet the Department of Transportation definition of an oxidizer

Corrosivity: Aqueous wastes with a pH greater than or equal to 12.5 or less than or equal to 2, and liquids that can readily corrode or dissolve flesh, metal, or other materials (liquids that corrode steel at greater than 0.25 in. per year at 55°C)

Reactivity: Wastes that readily explode or undergo violent reactions, for example, sodium hydride

Toxicity: Wastes likely to leach chemicals into groundwater when discarded, for example, in a municipal landfill

The EPA designed a procedure called the toxicity characteristic leaching procedure (TCLP) to be used to determine the leachability of a waste. If the leachate contains an amount of any of 40 different toxic chemicals above its regulatory level, the waste is given the waste code associated with that compound or element.

Excluded Wastes (40 CFR 261.4)

Certain wastes have been excluded from hazardous wastes for practical or economic reasons. Wastes are excluded because:

- They fall within 19 categories that are listed as not solid wastes
- They fall within 17 categories that are listed as not hazardous wastes
- They are regulated under other laws, such as the Clean Water Act or the Atomic Energy Act
- They are generated in laboratory experiments or waste-treatability studies
- They are generated in raw material, product storage, or manufacturing units

Examples:

Domestic sewage discharged into a sewer to a municipal wastewater treatment plant, regulated under the Clean Water Act

Certain radioactive materials that are regulated by the Atomic Energy Act

Scrap metal that has been processed to make it easier to handle

Circuit boards that have been shredded and are free of potentially dangerous materials

Household wastes generated by normal household activities that could technically be considered hazardous wastes

Certain wastes from the exploration and production of oil, gas, and geothermal energy

Treated wood containing hazardous materials such as arsenic that is discarded by end users

Dredge materials subject to the Marine Protection, Research, and Sanctuaries Act of 1972

Underground Storage Tanks The greatest potential hazard from underground tanks is leakage that will contaminate groundwater. The EPA estimates that about 25,000 tanks out of 2.1 million regulated tanks contain hazardous materials. In 1984, Congress added Subtitle I to the RCRA to initiate a program to protect the environment from underground tanks leaking petroleum products and other hazardous materials.

For practical and economic reasons, many tanks are excluded from meeting federal requirements (but not necessarily state or local):

- Tanks of 1,100 gallons or less used to hold motor fuels for noncommercial purposes on farms and residences
- Tanks for heating oil used on the premises where tank is stored
- Tanks in basements or tunnels
- Septic tanks and storm and wastewater collectors
- Tanks with a capacity of 110 gallons or less

The nature and complexity of the program led the federal government to allow approved state programs to act in its place. Inasmuch as state regulations can be more stringent, it is important that anyone concerned about underground storage tanks be familiar with state requirements. The Petroleum Equipment Institute Web site presents a map showing which states regulate tanks; however, it is recommended that your local regulatory agency be consulted.

LUST, the program for leaking underground storage tanks, is funded by a 0.1-cent federal tax on every gallon of motor fuel sold in the country.

Accumulated Wastes

Past practices have left huge quantities of hazardous wastes deposited in landfills, ponds, fields, and storage in a manner that threatens our health and environment. Surface and groundwater supplies of drinking water are especially sensitive to contamination by these accumulations, which can take thousands of years to clean themselves. Depending on the type of contamination, treatment may be impractical or prohibitively expensive.

The federal government response to the problem of accumulated wastes and emergencies that release hazardous wastes was CERCLA, commonly referred to as Superfund. This law was enacted by Congress December 11, 1980, and amended by the Superfund Amendments and

TABLE 1

Toxicity-characteristic waste

The D List

If your waste contains one or more of these contaminants at or above the regulatory level, it is a hazardous waste.

Maximum Concentration of Contaminants for the Toxicity Characteristic

EPA Waste #	Contaminant	CAS No.	Regulatory Level (mg/l)
D004	Arsenic	7440-38-2	5.0
D005	Barium	7440-39-3	100.0
D018	Benzene	71-43-2	0.5
D006	Cadmium	7440-43-9	1.0
D019	Carbon tetrachloride	56-23-5	0.5
D020	Chlordane	57-74-9	0.03
D021	Chlorobenzene	108-90-7	100.0
D022	Chloroform	67-66-3	6.0
D007	Chromium	7440-47-3	5.0
D023	o-Cresol	95-48-7	200.0*
D024	m-Cresol	108-39-4	200.0*
D025	p-Cresol	106-44-5	200.0*
D026	Cresol	—	200.0*
D016	2,4-D	94-75-7	10.0
D027	1,4-Dichlorobenzene	106-46-7	7.5
D028	1,2-Dichloroethane	107-06-2	0.5
D029	1,1-Dichloroethylene	75-35-4	0.7
D030	2,4-Dinitrotoluene	121-14-2	0.13
D012	Endrin	72-20-8	0.02
D031	Heptachlor (and its epoxide)	76-44-8	0.008
D032	Hexachlorobenzene	118-74-1	0.13
D033	Hexachlorobutadiene	87-68-3	0.5
D034	Hexachloroethane	67-72-1	3.0
D008	Lead	7439-92-1	5.0
D013	Lindane	58-89-9	0.4
D009	Mercury	7439-97-6	0.2
D014	Methoxychlor	72-43-5	10.0
D035	Methyl ethyl ketone	78-93-3	200.0
D036	Nitrobenzene	98-95-3	2.0
D037	Pentachlorophenol	87-86-5	100.0
D038	Pyridine	110-86-1	5.0
D010	Selenium	7782-49-2	1.0
D011	Silver	7440-22-4	5.0
D039	Tetrachloroethylene	127-18-4	0.7
D015	Toxaphene	8001-35-2	0.5
D040	Trichloroethylene	79-01-6	0.5
D041	2,4,5-Trichlorophenol	95-95-4	400.0
D042	2,4,6-Trichlorophenol	88-06-2	2.0
D017	2,4,5-TP (silvex)	93-72-1	1.0
D043	Vinyl chloride	75-01-4	0.2

[*] If the o-, m-, and p-cresol concentrations cannot be differentiated, the total cresol (D026) concentration is used. The regulatory level of total cresol is 200 mg/l.

Reauthorization Act (SARA) on October 17, 1986, creating a hazardous-waste-site response program and liability scheme that authorizes the government to hold persons who caused or contributed to the release of hazardous substances liable for the cost or the cleanup of affected sites. The president or the delegated agency is authorized to draw funds from a revolving trust fund (Superfund) to respond to releases of hazardous substances: (1) the EPA can take action at the site using Superfund money that it can recover from potentially responsible parties (PRPs), (2) the EPA can order PRPs directly or through a court to clean up a site, and (3) the EPA can enter into settlement agreements with PRPs that require them to clean up or pay for the cleanup of a site.

Superfund money can be used for sites that meet the following criteria:

(1) The site is listed on the National Priority List (NPL).
(2) The state in which the site is located either contributes or provides financial assurances for 10% of any remedial costs incurred.
(3) The remedial action is not inconsistent with the National Oil and Hazardous Pollution Contingency Plan (NCP). The NCP was revised in 1994 to reflect the oil-spill provisions of the Oil Pollution Act of 1990 (OPA).

To establish that a person is liable under CERCLA, the EPA must prove that a hazardous substance was released from a facility that caused the government to incur costs in responding. CERCLA imposes liability on current and former owners and operators of a facility, persons who arranged for treatment or disposal of hazardous substances, and transporters of hazardous substances who selected the disposal site.

Planning Ahead

The Toxic Substances Control Act (TSCA), which became law on October 11, 1976, authorizes the EPA to secure information on all new and existing chemical substances and to control any that were determined to be unreasonable risks to public health or the environment. The Asbestos Hazard Emergency Response Act (AHERA) amended the TSCA on October 22, 1986, and the Radon Reduction Act amended it in October 1988. In 1990 AHERA was amended by the Asbestos School Hazard Abatement Reauthorization Act, which required accreditation for persons conducting asbestos inspection and abatement activities in schools and commercial and public buildings. The TSCA was amended in October 1992 to add the Lead-Based Paint Exposure Reduction Act.

All manufacturers and importers, processors, distributors, and users of chemical substances may be subject to TSCA reporting, record keeping, and testing requirements. Penalties for noncompliance may be up to $27,500 per violation per day. The demand for greater environmental protection at less cost led the EPA to recommend a one-stop approach toward controlling the release of pollutants from different types of facilities in the May 10, 1996, Federal Register. This Public Performance-Based Permitting Program (P3) is oriented toward identifying the actual impact on the environment, the compliance over time, and how well the enforcement agencies perform.

The ultimate approach to the prevention of exposure of the public and contamination of the environment by hazardous wastes is to find ways to produce what we need while minimizing the hazardous side products. The flood of legislation and its costs in money and time to industry has helped to generate a movement sometimes known as “green technology” that introduces environmental consciousness at the early stages of product development. Industry’s aim for a maximum yield is being tempered by the necessity to minimize unwanted and costly-to-dispose-of materials that must be classified as hazardous wastes.

The Control of Exposures

Certain factors must be understood and managed in controlling exposures to hazardous materials:

- Where in the environment have hazardous materials accumulated or are currently being introduced as a result of ignorance, carelessness, or criminal disregard for human health or the welfare of the environment?
- How much hazardous material has accumulated or is being released?
- What are the environmental pathways and mechanisms by which these materials reach human receptors?
- How hazardous are specific materials and how can their risks be estimated?
- How can their generation be minimized?
- How can their impact be minimized?
- How best can the public be informed about the important aspects of hazardous material control?
- How can the credibility of the agents responsible for informing and protecting the public be established?
- How can the priority for control of hazardous materials be determined within the complex structure of society’s problems?
- How can public outrage, often generated by media treatment of environmental problems, be managed so that priorities and resources are directed to where they are most needed?

The Effects of Exposures

The prime motivation for continuing generation of laws to control exposures to hazardous materials has been the fear of cancer; however, it should be noted that there are other health perils inherent in exposures. Environmental exposures pertinent to carcinogenesis include not only hazardous materials but also diets, infectious agents, and even social behavior, as illustrated by the summary prepared in 1981 by the Office of Technology Assessment of the Congress of the United States (Table 2).

Determining the probability of cancer being caused at the concentration levels of a substance encountered in the environment is very difficult. Testing on human beings is not to be considered; therefore, accidental exposures are an important source of information, but the actual exposures and doses are usually poorly documented and existing records can be misleading. As a result, the use of animals as surrogates is the usual alternative, with the size of the doses used and the difference in response between humans and animals as serious complicating factors.

Toxicity and Risk Assessment

Three elements must be present for a situation to have toxicological implications:

1. A chemical or physical agent capable of causing a response
2. A biological system with which the agent interacts to produce a response
3. A response that can be considered deleterious to the biologic system

The deleterious response—that is, injury to health or the environment—should be significant. It is counterproductive to the goal of focusing our available resources on significant toxicological problems to include substances that cause only momentary discomfort or quickly reversible physiological change from the types of exposures that can be reasonably expected in daily life. There are no harmless substances if one does not place a limit on the type or extent of exposure. Nitrogen, oxygen, carbon dioxide, and water can all be harmful or even deadly under different types of exposures. Therefore, judgment must be used in making decisions about substances and the possibility of exposures at levels that can cause harm.

Substances that can cause harm by interfering with or destroying the functions of organs can be distinguished from those whose mode of action is attack on cell growth and reproduction.

Thus, the action of cyanide (which prevents tissues from using the oxygen provided by the blood) classifies it as different from a carcinogen (which causes the abnormal proliferation of cells), a mutagen (which changes the genetic material and thus damages new cells), and a teratogen (which changes the cell framework of an embryo).

Quantification of the risk associated with environmental exposures is a major activity of environmental control. Originally, this effort focused on the effects of carcinogens; noncarcinogens were delegated to minority status. Now there is an increased level of concern about effects other than cancer—for example, the impact of synthetic chemicals on the endocrine system that can result in disruptions of the immune system or behavioral problem. The immunotoxicity, neurotoxicity, teratogenicity, and toxicity to specific body tissues of many compounds have not received significant attention in comparison to their mutagenicity and carcinogenicity. There is danger that once a substance's carcinogenicity or noncarcinogenicity has been declared, its other harmful properties will be neglected.

To evaluate the risk from exposure to a specific hazardous material, it is important to obtain as realistic as possible an estimate of the dose received by a representative individual via inhalation, ingestion of food or drink, and other possible ways that substances can be absorbed. In the case of children in contact with soil and other materials, both indoor and outdoor, the possibility of hand-to-mouth exposure must be considered. As an example, an individual living in the vicinity of an operation that releases dioxin into the atmosphere is affected by both inhalation and by the amount of food and water ingested because of possible contamination from fallout. Obtaining the necessary dose data requires knowledge of emission rates and patterns, stack heights and local meteorological conditions, and

TABLE 2
Summary of cancer-associated environmental factors

Factor	Sites	Range of estimates
Diet	Digestive tract, breast endometrium, ovary	35–50%
Tobacco	Upper respiratory tract, bladder, esophagus, kidney, pancreas	22–30%
Asbestos	Upper respiratory tract, others	3–18%
Occupational	Upper respiratory tract, others	4–38%
Alcohol	Upper digestive tract, larynx, liver	3–5%
Infection	Uterine cervix, prostate, and other sites	1–5%
Sexual development, reproductive patterns, and sexual practices	Breast, endometrium, ovary, cervix, testis	1–13%
Pollution	Lung, bladder, rectum	5%
Medical drugs and radiation	Breast, endometrium, ovary, thyroid, bone, lung, blood	1–4%
Natural radiation	Skin, breast, thyroid, lung, bone, blood	1–3%
Consumer products	Possibly all sites	1–2%
Unknown—(e.g. new chemicals, dumps, stress)	All sites	?

the topography of the land and its uses. The behavior patterns of the individual—that is, the time and places occupied, the food and drink ingested and its sources, and the possibility of direct ingestion of contaminated soil through work or play—must be factored in to obtain a realistic dose estimate.

Estimation of the carcinogenic hazard of a substance has become a very complex procedure; however, the general idea can be presented in a model that calculates the individual or aggregate risk based on linear extrapolation from experimental data to zero dose of the curve relating dose to the probability of cancer.

This assumes that there is no threshold for the incident of cancer, i.e., the only exposure for which there is zero probability of cancer is zero exposure. The experimental data most convincing are those derived from studies of the human population supported by documented exposures and incidences of cancer. In most cases such data are not available; therefore animal-study data are used, with conservative factors introduced to compensate for lack of information about interspecies relationships. The safety factors that have been used have been as large as 10,000—that is, in some cases the effects on humans are assumed to take place at exposures that can be as much as 10,000 times less than those causing similar effects on the surrogate species.

The doses are commonly measured in terms of milligrams of chemicals absorbed in the body each day per kilogram of body weight over a lifetime. For the part that is inhaled, it is customary to convert dosage units into micrograms per cubic meter in the inhaled air. For the purposes of this conversion, a body weight of 70 kg is assumed and the inhaled volume is 20 cubic meters per day.

$$1\frac{\text{mg}}{\text{kg-day}} \times 70\frac{\text{kg}}{\text{person}} \times \frac{1}{20}\frac{\text{person-day}}{\text{m}^3}$$

$$\times 1000\frac{\mu\text{g}}{\text{mg}} = 3500\frac{\mu\text{g}}{\text{m}^3}$$

Therefore, a bodily intake of 1 mg per kilogram of body weight per day is equivalent to the inhalation of air containing 3500 μg of the chemical per cubic meter.

The unit risk value (URV) is the cancer probability corresponding to the inhalation over a lifetime of 70 years of air containing 1 μg per cubic meter of the substance—that is, the risk is the ratio of the probability of cancer to the microgram per cubic meter inhaled:

$$\text{URV} = \frac{\text{Probability of cancer}}{\mu\text{g}/\text{m}^3}$$

The URV multiplied by the average concentration inhaled over a lifetime is the individual lifetime risk of cancer:

$$\text{URV} \times \frac{\mu\text{g}}{\text{m}^3} = \text{individual life time risk}$$

The individual risk multiplied by the number in the population is the aggregate lifetime risk.

A commonly used value to describe a hazardous substance is the atmospheric concentration that over a lifetime will result in one cancer per million inhabitants.

$$\text{URV} \times \frac{\mu\text{g}}{\text{m}^3} \times 10^6 = 1$$

$$\frac{\mu\text{g}}{\text{m}^3} = \frac{1}{\text{URV} \times 10^6}$$

HAZARDOUS-WASTE INFORMATION

Information about Superfund locations and hazardous-waste activities in your zip code, city, county, or state can be obtained over the Internet from the U.S. Environmental Protection Agency's (EPA) Envirofacts Data Warehouse.

REFERENCES

Book of Chemical Lists (CD-ROM), Business and Legal Reports Inc.
CERCLA/Superfund Orientation Manual, 1992 EPA Office of Solid Wastes and Emergency Response
National Center for Environmental Assessment, EPA
RCRA Orientation Manual, Jan. 2003, EPA 530-R-02–016
Toxicology Excellence for Risk Assessment, TERA@Tera.org

EDWARD F. FERRAND
Edward F. Ferrand Associates

HYDROLOGY

THE PURPOSES OF HYDROLOGICAL STUDIES

Hydrology is concerned with all phases of the transport of water between the atmosphere, the land surface and subsurface, and the oceans, and the historical development of an understanding of the hydrological process is in itself a fascinating study.[6] As a science, hydrology encompasses many complex processes, a number of which are only imperfectly understood. It is perhaps helpful in developing an understanding of hydrological theory to focus attention not on the individual physical processes, but on the practical problems which the hydrologist is seeking to solve. By studying hydrology from the problem-solving viewpoint, we shall see the interrelationship of the physical processes and the approximations which are made to represent processes which are either imperfectly understood or too complex for complete physical representation. We shall also see what data is required to make adequate evaluations of given problems.

A prime hydrological problem is the forecasting of stream-flow run-off. Such forecasts may be concerned with daily flows, especially peak flows for flood warning, or a seasonal forecast may be required, where a knowledge of the total volume of run-off is of prime interest. More sophisticated forecast procedures are required for the day-to-day operation of flood control reservoirs, hydropower projects, irrigation and water supply schemes, especially for schemes which are used to serve several purposes simultaneously such as hydropower, flood control, and irrigation.

Hydrologists are also concerned with studying statistical patterns of run-off. A special class of problems is the study of extreme events, such as floods or droughts. Such maximum events provide limiting design data for flood spillways, dyke levels, channel design, etc. Minimum events are important, for example, in irrigation studies and fisheries projects. A more complex example of statistical studies is concerned with sequential patterns of run-off, for either monthly or annual sequences. Such sequences are important when testing the storage capacity of a water resource system, such as an irrigation or hydropower reservoir, when assessing the risk of failing to meet the requirements of a given scheme. A specially challenging example of sequential flow studies concerns the pattern of run-off from several tributary areas of the same river system. In such studies it is necessary to try to maintain not only a sequential pattern but also to model the cross-correlations between the various tributaries.

The question of land use and its influences on run-off occupies a central position in the understanding of hydrological processes. Land use has been studied for its influence on flood control, erosion control, water yield and agriculture, with particular application to irrigation. Perhaps the most marked effect of changed land use and changed run-off characteristics is demonstrated by urbanization of agricultural and forested lands. The paving of large areas and the influence of buildings has a marked effect in increasing run-off rates and volumes, so that sewer systems must be designed to handle the increased flows. Although not so dramatic, and certainly not so easy to document, the influence of trees and crops on soil structure and stability may well prove to be the most far-reaching problem. There is a complex interaction between soil biology, the crop and the hydrological factors such as soil moisture, percolation, run-off, erosion, and evapo-transpiration. Adequate hydrological calculations are a prerequisite for such studies.

A long-term aim of hydrological studies is the clear definition of existing patterns of rainfall and run-off. Such a definition requires the establishment of statistical measures such as the means, variances and probabilities of rate events. From these studies come not only the design data for extreme events but also the determination of any changes in climate which may be either cyclical or a longterm trend. It is being suggested in many quarters that air pollution may have a gradual effect on the Earth's radiation balance. If this is true we should expect to see measurable changes in our climatic patterns. Good hydrological data and its proper analysis will provide one very important means of evaluating such trends and also for measuring the effectiveness of our attempts to correct the balance.

A BRIEF NOTE ON STATISTICAL TECHNIQUES

The hydrologist is constantly handling large quantities of data which may describe precipitation, streamflow, climate, groundwater, evaporation, and many other factors. A reasonable grasp of statistical measures and techniques is invaluable to the hydrologist. Several good basic textbooks are referenced,[1,2,3,8,9] and *Facts from Figures* by Moroney, is particularly recommended for a basic understanding of what statistics is aiming to achieve.

The most important aspect of the nature of data is the question of whether data is independent or dependent. Very often this basic question of dependence or independence is not discussed until after many primary statistical measures have been defined. It is basic to the analysis, to the selection of variables and to the choice of technique to have some idea of whether data is related or independent. For example, it is usually reasonable to assume that annual flood peaks are independent of each other, whereas daily streamflows are usually closely related to preceding and subsequent events: they exhibit what is termed *serial correlation.*

The selection of data for multiple correlation studies is an example where dependence of the data is in conflict with the underlying assumptions of the method. Once the true nature of the data is appreciated it is far less difficult to decide on the correct statistical technique for the job in hand. For example, maximum daily temperatures and incoming radiation are highly correlated and yet are sometimes both used simultaneously to describe snowmelt.

In many hydrological studies it has been demonstrated that the assumption of random processes is not unreasonable. Such an assumption requires an understanding of statistical distribution and probabilities. Real data of different types has been found to approximate such theoretical distributions as the binomial, the Poisson, the normal distribution or certain special extreme value distributions. Especially, in probability analysis, it is important that the correct assumption is made concerning the type of distribution if extrapolated values are being read from the graphs.

Probabilities and return periods are important concepts in design studies and require understanding. The term "return period" can be somewhat misleading unless it is clearly appreciated that a return period is in fact a probability. Therefore when we speak of a return period of 100 years we imply that a magnitude of flow, or some other such event, has a one percent probability of occurring in any given year. It is even more important to realize that the probability of a certain event occurring in a number of years of record is much higher than we might be led to believe from considering only its annual probability or return period. As an example, the 200 year return period flood or drought has an annual probability of 0.5%, but in 50 years of record, the probability that it will occur at least once is 22%. Figure 1 summarizes the probabilities for various return periods to occur at least once as a function of the number of years of record. From such a graph it is somewhat easier to appreciate why design floods for such critical structures as dam spillways have return of 1,000 years or even 10,000 years.

ANALYSIS OF PRECIPITATION DATA

Before analyzing any precipitation data it is advisable to study the method of measurement and the errors inherent in the type of gauge used. Such errors can be considerable (Chow,[1] and Ward[5]).

Precipitation measurements vary in type and precision, and according to whether rain or snow is being measured. Precipitation gauges may be read manually at intervals of a day or part of a day. Alternatively gauges may be automatic and yield records of short-term intensity. Wind and gauge exposure can change the catch efficiency of precipitation gauges and this is especially true for snow measurements. Many snow measurements are made from the depth of new snow and an average specific gravity of 0.10 is assumed when converting to water equivalent.

Precipitation data is analyzed to give mean annual values and also mean monthly values which are useful in assessing seasonal precipitation patterns. Such figures are useful for determining total water supply for domestic, agricultural and hydropower use, etc.

More detailed analysis of precipitation data is given for individual storms and these figures are required for design of drainage systems and flood control works. Analysis shows the

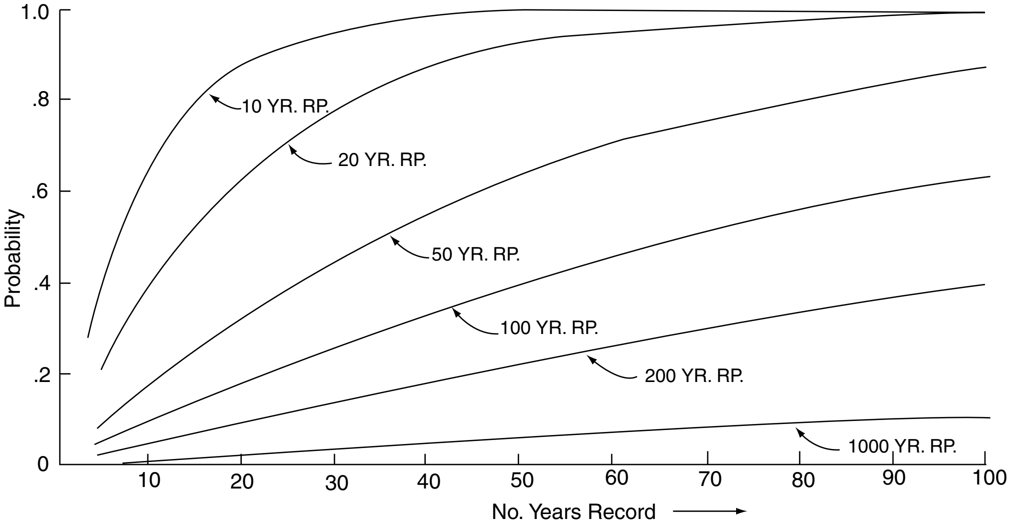


FIGURE 1 Probability of occurrence of various annual return period events as a function of years of record.

relationship between rain intensity (inches per hour) with both duration and area. In general terms, the longer the duration of storm, the lower will be the average intensity of rainfall. Similarly, the larger the area of land being considered, the lower will be the average intensity of rainfall. For example, a small catchment area of, say, four square miles may be subjected to a storm lasting one hour with an average intensity of two inches per hour while a catchment of two hundred square miles would only experience an average intensity of about one inch per hour. Both these storms would have the same return period or probability associated with them. Such data is prepared by weather agencies like the U.S. Weather Bureau and is available in their publications for all areas of the country. Typical data is shown in Figure 2. The use of these data sheets will be discussed further in the section on run-off.

Winter snowpacks represent a large water storage which is mainly released at a variable rate during spring and early summer. In general, the pattern of snowfall is less important than the total accumulation. In the deep mountain snowpacks, snowtube and snowpillow measurements appear to give fairly reliable estimates of accumulated snow which can be used for forecasts of run-off volumes as well as for flood forecasting. On the flat prairie lands, where snow is often quite moderate in amounts, there is considerable redistribution and drifting of snow by wind and it is a considerable problem to obtain good estimates of total snow accumulation.

When estimates of snow accumulation have been made it is a further problem to calculate the rate at which the snow will melt and will contribute to stream run-off. Snow therefore represents twice the problem of rain, because firstly we must measure its distribution and amount and secondly, it may remain as snow for a considerable period before it contributes to snowmelt.

EVAPORATION AND EVAPO-TRANSPIRATION

Of the total precipitation which falls, only a part finally discharges as streamflow to the oceans. The remainder returns to the atmosphere by evaporation. Linsley[2] points out that ten reservoirs like Lake Mead could evaporate an amount equivalent to the annual Colorado flow. Some years ago, studies of Lake Victoria indicated that the increased area resulting from raising the lake level would produce such an increase in evaporation that there would be a net loss of water utilization in the system.

Evaporation varies considerably with climatic zone, latitude and elevation and its magnitude is often difficult to evaluate. Because evaporation is such a significant term in many hydrological situations, its proper evaluation is often a key part of hydrological studies.

Fundamentally, evaporation will occur when the vapor pressure of the evaporating surface is greater than the vapor pressure of the overlying air. Considerable energy is required to sustain evaporation, namely 597 calories per gram of water or 677 calories per gram of snow or ice. Energy may be supplied by incoming radiation or by air temperature, but if this energy supply is inadequate, the water or land surface and the air will cool, thus slowing down the evaporation process. In the long term the total energy supply is a function of the net radiation balance which, in turn, is a function of latitude. There is therefore a tendency for annual evaporation to be only moderately variable and to be a function of latitude, whereas short term evaporation may vary considerably with wind, air temperature, air vapor pressure, net radiation, and surface temperature.

The discussion so far applies mainly to evaporation from a free water surface such as a lake, or to evaporation from a saturated soil surface. Moisture loss from a vegetated land surface is complicated by transpiration. Transpiration is the term used to describe the loss of water to the atmosphere from plant surfaces. This process is very important because the plant's root system can collect water from various depths of the underlying soil layers and transmit it to the atmosphere. In practice it is not usually possible to differentiate between evaporation from the soil surface and transpiration from the plant surface, so it is customary to consider the joint effect and call it evapo-transpiration. This lumping of the two processes has led to thinking of them as being identical, however, we do know that the evaporation rate from a soil surface decreases as the moisture content of the soil gets less, whereas there is evidence to indicate that transpiration may continue at a nearly constant rate until a plant reaches the wilting point.

To understand the usual approach now being taken to the calculation of evapo-transpiration, it is necessary to appreciate what is meant by *potential evapo-transpiration* as opposed to *actual evapo-transpiration.* Potential evapo-transpiration is the moisture loss to the atmosphere which would occur if the soil layers remained saturated. Actual evapo-transpiration cannot exceed the potential rate and gradually reduces to a fraction of the potential rate as the soil moisture decreases. Various formulae exist for estimating potential evapo-transpiration in terms of climatic parameters, such as Thornthwaites method, or Penman or Turk's formulae. Such investigations have shown that a good field measure of potential evapo-transpiration is pan evaporation from a standard evaporation-pan, such as the Class A type, and such measurements are now widely used. To turn these potential estimates into actual evapo-transpiration it is commonly assumed that actual equals potential after the soil has been saturated until some specific amount of moisture has evaporated, say two inches or so depending on the soil and crop. It is then assumed that the actual rate decreases exponentially until it effectively ceases at very low moisture contents. In hydrological modeling an accounting procedure can be used to keep track of incoming precipitation and evaporation so that estimates of evapo-transpiration can be made. The potential evapo-transpiration rate must be estimated from one of the accepted formulae or from pan-evaporation measurements, if available. Details of such procedures are well illustrated in papers by Nash[17] and by Linsley and Crawford[44] in the Stanford IV watershed model.

RUN-OFF: RAIN

It is useful to imagine that we start with a dry catchment, where the groundwater table is low, and the soil moisture

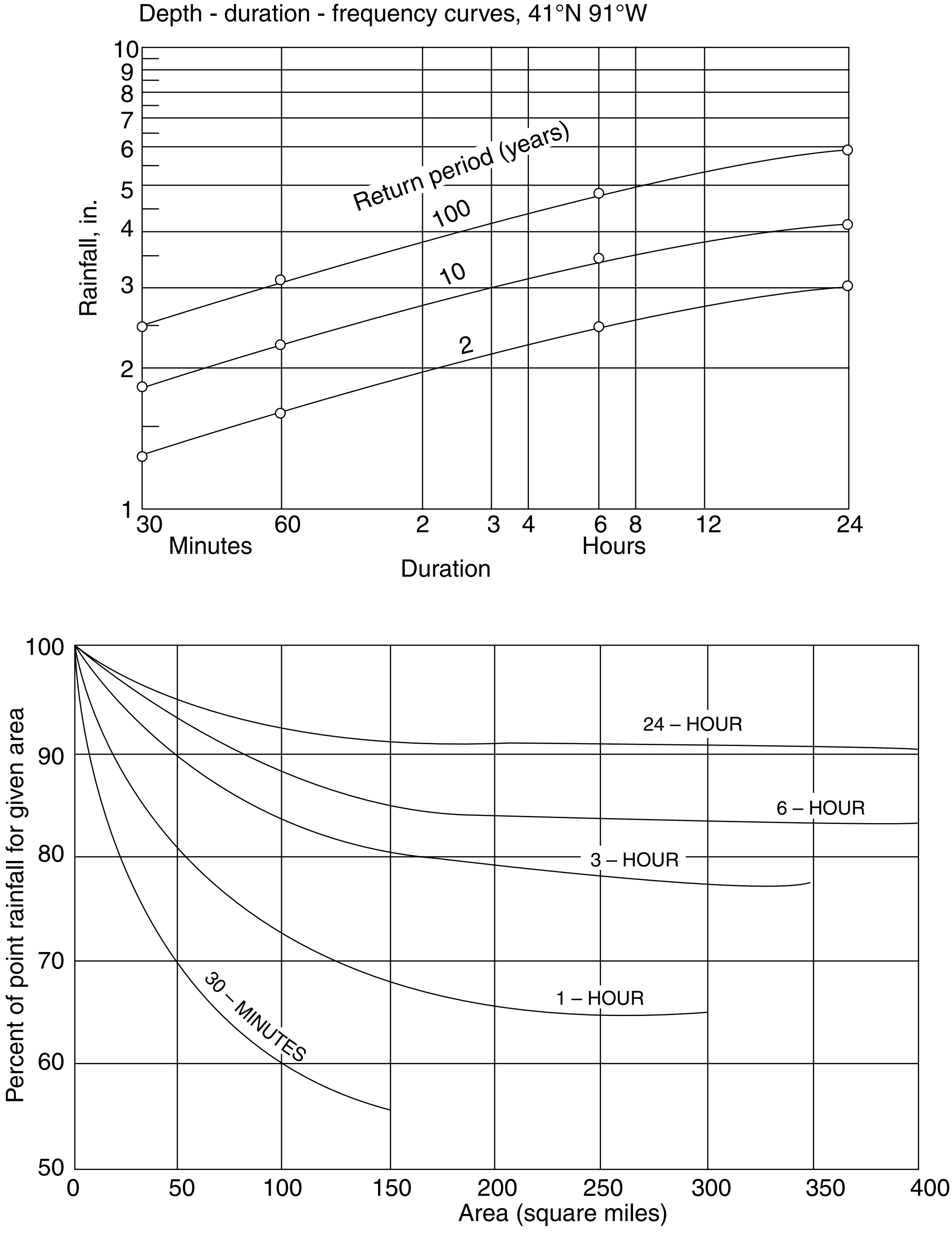


FIGURE 2 Rainfall depth-duration and area-frequency curves (US Weather Bureau, after Chow[1]).

has been greatly reduced, perhaps almost to the point where hygroscopic moisture alone remains. When rain first starts much is *intercepted* by the trees and vegetation and this *interception* storage is lost by evaporation after the storm. Rain reaching the soil *infiltrates* into pervious surfaces and begins to satisfy soil moisture deficits. As soil moisture levels rise, water percolates downward toward the fully saturated water table level. If the rain is heavy enough, the water supply may exceed the vertical percolation rate and water then starts to flow laterally in the superficial soil layers toward the stream channels: this process is termed *interflow* and is much debated because it is so difficult to measure. At very high rainfall rates, the surface infiltration rate may be exceeded and then direct surface run-off will occur. Direct run-off is rare from soil surfaces but does occur from certain impervious soil types, and from paved areas. Much work has been done to evaluate the relative significance of these various processes and is well documented in references (1,2,3).

Such qualitative descriptions of the run-off process are helpful, but are limited because of the extreme complexity and interrelationship of the various processes. Various methods have been developed to by-pass this complexity and to give us usable relationships for hydrologic calculations.

The simplest method is a plot of historical events, showing run-off as a function of the depth of precipitation in a given storm. This method does not allow for any antecedent soil moisture conditions or for the duration of a particular storm.

More complex relationships use some measure of soil moisture deficiency such as cumulative pan-evaporation or the antecedent precipitation index. Storm duration and precipitation amount is also allowed for and is well illustrated by the U.S. Weather Bureau's charts developed for various areas (Figure 2). It is a well to emphasize that the antecedent precipitation index, although based on precipitation, is intended to model the exponential decay of soil moisture between storms, and is expressed by

$$I_N = (I_0 k^M + I_M) k^{(N-M)}$$

where I_0 is the rain on the first day and no more rain occurs until day M, when I_M falls. If k is the recession factor, usually about 0.9, then I_N will be the API for day N. The expression can of course have many more terms according to the number of rain events.

Before computers were readily available such calculations were considered tedious. Now it is possible to use more complex accounting procedures in which soil moisture storage, evapo-transpiration, accumulated basin run-off, percolation, etc. can all be allowed for. These procedures are used in more complex hydrological modeling and are proving very successful.

RUN-OFF: SNOWMELT

As a first step in the calculation of run-off from snow, methods must be found for calculating the rate of snowmelt. This snowmelt can then be treated similarly to a rainfall input. Snowmelt will also be subject to soil moisture storage effects and evapo-transpiration.

The earliest physically-based model to snowmelt was the degree-day method which recognized that, despite the complexity of the process, there appeared to be a good correlation between melt rates and air temperature. Such a relationship is well illustrated by the plots of cumulative degree-days against cumulative downstream flow, a rather frustrating graph because it cannot be used as a forecasting tool. This cumulative degree-day versus flow plot is an excellent example of how a complex day-to-day behavior yields a long-term behavior which appears deceptively simple. Exponential models and unit hydrograph methods have been used to turn the degree-day approach into a workable method and a number of papers are available describing such work (Wilson,[38] Linsley[32]). Arguments are put forward that air temperature is a good index of energy flux, being an integrated result of the complex energy exchanges at the snow surface (Quick[33]).

Light's equation[31] for snowmelt is based on physical reasoning which models the energy input entirely as a turbulent heat transfer process. The equation ignores radiation and considers only wind speed as the stirring mechanism, air temperature at a standard height as the driving gradient for heat flow and, finally, vapour pressure to account for condensation–evaporation heat flux. It is set up for 6 hourly computation and requires correction for the nature of the forest cover and topography. It is interesting to compare Light's equation with the U.S. Crops equation[36] for clear weather to see the magnitude of melt attributed to each term.

By far the most comprehensive studies of snowmelt have been the combined studies by the U.S. Corps of Engineers and the Weather Bureau (U.S. Corps of Engineers[36,37]). They set up three field snow laboratory areas varying in size from 4 to 21 square miles and took measurements for periods ranging from 5 to 8 years. Their laboratory areas were chosen to be representative of certain climatic zones. Their investigation was extensive and comprehensive, ranging from experimental evaluation of snowmelt coefficient in terms of meteorological parameters, to studies of thermal budgets, snow-course and precipitation data reliability, water balances, heat and water transmission in snowpacks, streamflow synthesis, atmospheric circulations, and instrumentation design and development.

A particularly valuable feature of their study appears to have been the lysimeters used, one being 1300 sq.ft. in area and the other being 600 sq.ft. (Hilderbrand and Pagenhart[30]). The results of these lysimeter studies have not received the attention they deserve, considering that they give excellent indication of storage and travel time for water in the pack. It may be useful to focus attention on this aspect of the Corps work because it is not easy to unearth the details from the somewhat ponderous Snow Hydrology report. Before leaving this topic it is worth mentioning that the data from the U.S. studies is all available on microfilm and could be valuable for future analysis. It is perhaps useful at this stage to write down the Light equation and the clear weather equation from the Corps work to compare the resulting terms.

Light's equation[31] (simple form in °F, inches of melt and standard data heights)

$$D = U\left(0.001\,84 T_a \cdot 10^{-0.0000156} + 0.00578\left(e^{-6\cdot 11}\right)\right)$$

where

U = average wind speed (m.p.h.) for 6 hr period
T = air temperature above 32°F for 6 hr period

e = vapor pressure for 6 hr period
h = station elevation (feet)
D = melt in inches per 6 hr period

The U.S. Corps Equation is[36]

$$M = k'(0.00508I_{\mathrm{i}})(1 - a) + (1 - N) \times (0.0212T_{\mathrm{a}} - 0.84) + N(0.029T_{\mathrm{c}}) + k(0.0084U)(0.22T_{\mathrm{a}} + 0.78T\mathrm{d})$$

M = Incident Radiation + incoming clear air longwave + cloud longwave + [Conduction + Condensation]

k' and k are approximately unity.
N = fraction of cloud cover
I_{i} = incident short wave radiation (langleys/day)
a = albedo of snow surface
T_{a} = daily mean temperature °F above 32°F at 10′ level
T_{c} = cloud base temperature
T_{d} = dew point temperature °F above 32°F
U = average wind speed—miles/hour at 50′ level.

Putting in some representative data for a day when the minimum temperature was 32°F and the maximum 70°F, incoming radiation was 700 langleys per day and relative humidity varied from 100% at night to 60% at maximum temperature, the results were:

Light Equation
D = Air temp melt and Condensation melt
$= 1.035 + 0.961$ inches/day
$= 1.996$ inches/day

U.S. Corps Equation
M = incoming shortwave + incoming longwave + air temp. + Condensation
$= 1.424 - 0.44 + 0.351 + 0.59$
$= 1.925$ inches/day.

Note the large amount attributed to radiation which the Light equation splits between air temperature and radiation. It is a worthwhile operation to attempt to manufacture data for these equations and to compare them with real data. The high correlations between air temperature and radiation is immediately apparent, as is the close relationship between diurnal air temperature variation and dewpoint temperature during the snowmelt season. Further comparison of the formulae at lower temperature ranges leave doubts about the influence of low overnight temperatures.

There is enough evidence of discrepancies between real and calculated snowmelt to suggest that further study may not be wasted effort. Perhaps this is best illustrated from some recent statements made at a workshop on Snow and Ice Hydrology.[39] Meier indicates that, using snow survey data, the Columbia forecast error is 8 to 14% and occasionally 40 to 50%. Also these errors occurred in a situation where the average deviation from the long-term mean was only 12 to 20%. For a better comparison of errors it would be interesting to know the standard error of forecast compared with standard "error" of record from the long-term mean. Also, later in the same paper it is indicated that a correct heat exchange calculation for the estimation of snowmelt cannot be made because of our inadequate knowledge of the eddy convection process. At the same workshop the study group on Snow Metamorphism and Melt reported: "we still cannot measure the free water content in any snow cover, much less the flux of the water as no theoretical framework for flow through snow exists."

Although limitations of data often preclude the use of the complex melt equations, various investigators have used the simple degree-day method with good success (Linsley[32] and Quick and Pipes[40,46,47]). There may be reasonable justification for using the degree-day approach for large river basins with extensive snowfields where the air mass tends to reach a dynamic equilibrium with the snowpck so that energy supply and the resulting melt rate may be reasonably well described by air temperature. In fact there seems to be no satisfactory compromise for meteorological forecasting; either we must use the simple degree-day approach or on the other hand we must use the complex radiation balance, vapour exchange and convective heat transfer methods involving sophisticated and exacting data networks.

COMPUTATION OF RUN-OFF—SMALL CATCHMENTS

Total catchment behavior is seen to be made up of a number of complex and interrelated processes. The main processes can be reduced to evapo-transpiration losses, soil moisture and groundwater storage, and flow of water through porous media both as saturated flow and unsaturated flow. To describe this complex system the hydrologist has resorted to a mixture of semi-theoretical and empirical calculation techniques. Whether such techniques are valid is justified by their ability to predict the measured behavior of a catchment from the measured inputs.

The budgeting techniques for calculating evapo-transpiration losses have already been described. From an estimation of evapo-transpiration and soil moisture and measured precipitation we can calculate the residual precipitation which can go to storage in the catchment and run-off in the streams. A method is now required to determine at what rate this effective precipitation, as it is usually called, will appear at some point in the stream drainage system. The most widely used method is the unit hydrograph approach first developed by Sherman in 1932.[16]

To reduce the unit hydrograph idea to its simplest form, consider that four inches of precipitation falls on a catchment in two hours. After allowing for soil moisture deficit and evaporation losses, let us assume that three inches of this precipitation will eventually appear downstream as run-off. Effecitvely this precipitation can be assumed to have fallen on the catchment at the rate of one and a half inches per hour

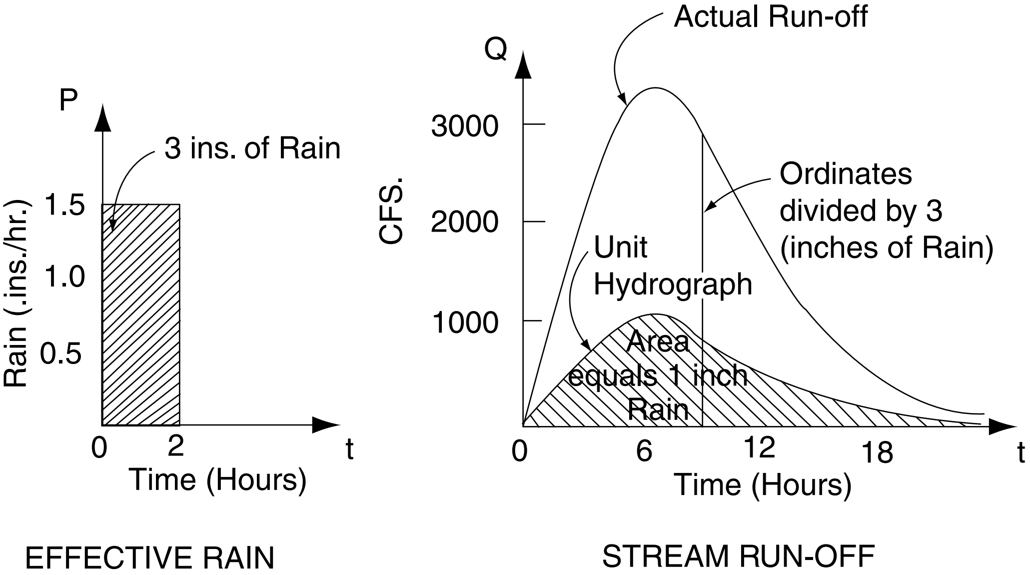


FIGURE 3 Hydrograph and unit hydrograph of run-off from effective rain.

for two hours. This effective precipitation will appear some time later in the stream system, but will now be spread out over a much longer time period and will vary from zero flow, rising gradually to a maximum flow and then slowly decreasing back to zero. Figure 3 shows the block of uniform precipitation and the corresponding outflow in the stream system. The outflow diagram can be reduced to the unit hydrograph for the two hour storm by dividing the ordinates by three. The outflow diagram will then contain the volume of run-off equivalent to one inch of precipitation over the given catchment area. For instance, one inch of precipitation over one hundred square miles will give an area under the unit hydrograph of 2690 c.f.s. days.

When a rainstorm has occurred the hydrologist must first calculate how much will become effective rainfall and will contribute to run-off. This can best be done in the framework of a total hydrological run-off model as will be discussed later. The effective rainfall hydrograph must then be broken down into blocks of rainfall corresponding to the time interval for the unit hydrograph. Each block of rain may contain *P* inches of water and the corresponding outflow hydrograph will have ordinates *P* times as large as the unit hydrograph ordinates. Also, several of these scaled outflow hydrographs will have to be added together. This process is known as convolution and is illustrated in Figure 4 and 5.

The underlying assumption of unit hydrograph theory is that the run-off process is linear, not in the trivial straight line sense, but in the deeper mathematical sense that each incremental run-off event is independent of any other run-off. In the early development, Sherman[16] proposed a unit hydrograph arising from a certain storm duration. Later workers such as Nash[17,23] showed that Laplace transform theory, as already highly developed for electric circuit theory, could be used. This led to the instantaneous unit hydrograph and gave rise to a number of fascinating studies by such workers as Dooge,[18] Singh,[19] and many others. They introduced exponential models which are interpretable in terms of instantaneous unit hydrograph theory. Basically, however, there is no difference in concept and the convolution integral, Eq. (1) can be arrived at by either the unit hydrograph or the instantaneous unit hydrograph approach. The convolution integral can be written as:

$$Q(t) = \int_0^{t<t_0} u(1-\tau)P(\tau)\mathrm{d}\tau \tag{1}$$

Figure 4 shows the definition diagram for the formulation is only useful if both *P,* the precipitation rate, and *u,* the instantaneous unit hydrograph ordinate are expressible as continuous functions of time. In real hydrograph applications it is more useful to proceed to a finite difference from of Eq. (1) in which the integral is replaced by a summation, Eq. (2), and Figure 5.

$$Q_R = \sum_1^M u(m)P(n)\Delta t \tag{2}$$

where *M* is the number of unit hydrograph time increments, and *m, n* and *R* are specified in Figure 2. It should be noted that from Figure 5,

$$m + n = R + 1. \tag{3}$$

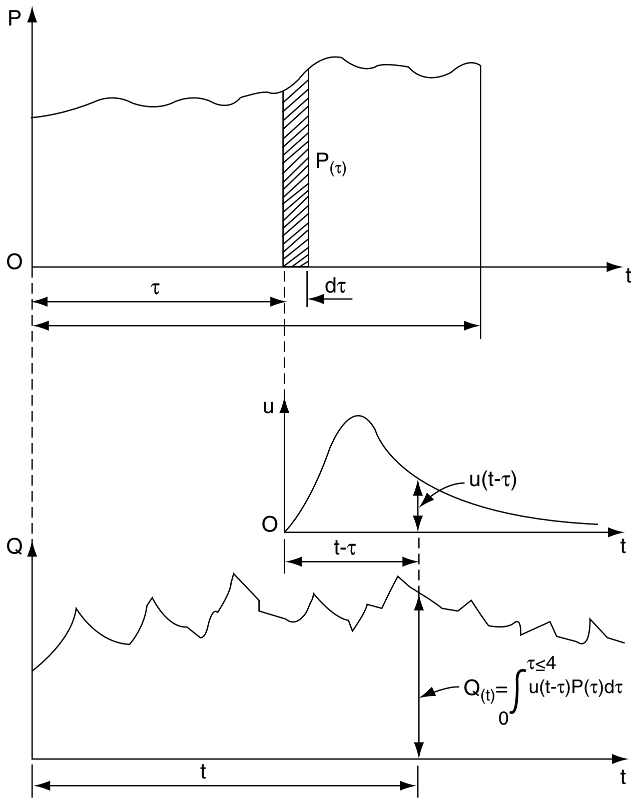


FIGURE 4 Determination of streamflow from precipitation input using an instantaneous unit hydrograph.

Expanding Eq. (2) for a particular value of *R*,

$$Q(10)=u(10)P(1)+u(9)P(2)+\Lambda\ u(1)P(10). \quad (4)$$

The whole family of similar equations for *Q* may be expressed in matrix form (Snyder[20])

$$\begin{pmatrix} Q_1\\ Q_2\\ Q_3\\ .\\ .\\ \\ \\ \\ \\ Q_R \end{pmatrix}=\begin{pmatrix} P_1 & 0 & 0 & . & . & . & . & 0\\ P_2 & P_1 & 0 & & & & & 0\\ P_3 & P_2 & P_1 & 0 & & & & 0\\ . & . & . & & & & . & .\\ . & . & . & . & & & & .\\ P_n & & & & & & & \\ 0 & P_n & & & & & & .\\ 0 & 0 & P_n & & & & & \\ 0 & 0 & 0 & & & & P_1 & 0\\ . & . & . & & & & P_2 & P_1 \end{pmatrix}\begin{pmatrix} u_1\\ u_2\\ u_3\\ .\\ .\\ u_m \end{pmatrix} \quad (5)$$

$\leftarrow m$ columns $\rightarrow$

Or more briefly

$$\{Q\}=[P]\{u\}. \quad (6)$$

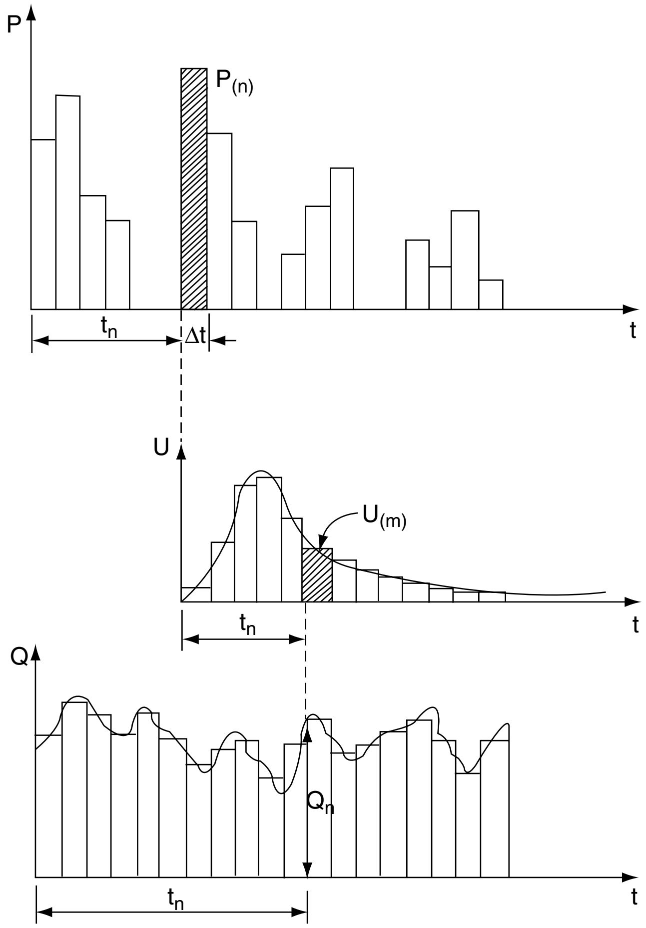


FIGURE 5 Convolution of precipitation by unit hydrograph on a finite difference basis.

Equation (6) specifies the river flow in terms of the precipitation and the unit hydrograph. In practice *Q* and *P* are measured and *u* must be determined. Some workers have guessed a suitable functional form for *u* with one or two unknown parameters and have then sought a best fit with the available data. For instance, Nash's series of reservoirs yields[17,23]

$$u(t)=\frac{1}{(n-1)!}\frac{t^{n-1}}{K^{n}}e^{-t/k} \quad (7)$$

in which there are two parameters, *K* and *n*. Another approach is to solve the matrix Eq. (6) as follows (Synder[21])

$$\{u\}=\left[P^{T}P\right]^{-1}P^{T}\{Q\}. \quad (8)$$

It has already been demonstrated that $R = m + n - 1$ so that there are more equations available than there are unknowns. The solution expressed by Eq. (8) therefore automatically yields the least squares values for *u*. This result will be referred to after the next section.

MULTIPLE REGRESSION AND STREAMFLOW

The similarity of the finite difference unit hydrograph approach to multiple regression analysis is immediately apparent. The flow in terms of precipitation can be written as

$$\begin{aligned} Q_R &= a_1P_t + a_2P_{t+1} + a_3P_{t+2} + \Lambda + a_np_{t+n-1} \\ Q_{R+1} &= a_1P_{t+1} + a_2P_{t+2} + a_3P_{t+3} + \Lambda + a_np_{t+n} \qquad (9) \\ Q_{R+2} &= a_1P_{t+2} + \text{etc.} \end{aligned}$$

Again we can write

$$\{Q\} = [P]\{a\}. \qquad (10)$$

The similarity with Eq. (6) is obvious and may be complete if we have selected the correct precipitation data to correlate precipitation at 6 a.m. with downstream flow at 9 a.m. when we know that there is a 3-hour lag in the system. Therefore, using multiple regression as most hydrologists do, the method can become identical with the unit hydrograph approach.

LAKE, RESERVOIR AND RIVER ROUTING

The run-off calculations of the previous sections enable estimates to be made of the flow in the headwaters of the river system tributaries. The river system consists of reaches of channels, lakes, and perhaps reservoirs. The water travels downstream in the various reaches and through the lakes and reservoirs. Tributaries combine their flows into the main stream flow and also distributed lateral inflows contribute to the total flow. This total channel system influences the flow in two principal ways, first the flow takes time to progress through the system and secondly, some of the flow goes into temporary storage in the system. Channel storage is usually only moderate compared with the total river flow quantities, but lake and reservoir storage can have a considerable influence on the pattern of flow.

Calculation procedures are needed which will allow for this delay of the water as it flows through lakes and channels and for the modifying influence of storage. The problem is correctly and fully described by two physical equations, namely a continuity equation and an equation of motion. Continuity is simply a conversion of mass relationship while the equation of motion relates the mass accelerations to the forces controlling the movement of water in the system. Open channel fluid mechanics deals with the solution of such equations, but at present the solutions have had little application to hydrological work because the solutions demand detailed data which is not usually available and the computations are usually very complex, even with a large computer.

Hydrologists resort to an alternative approach which is empirical; it uses the continuity equation but replaces the equation of motion with a relationship between the storage and the flow in the system. This assumption is not unreasonable and is consistent with the assumption of a stage–discharge relationship which is widely utilized in stream gauging.

RESERVOIR ROUTING

The simplest routing procedure is so-called reservoir routing, which also applies to natural lakes. The continuity equation is usually written as;

$$I - O = \frac{\mathrm{d}S}{dT}, \qquad (11)$$

where

I = Inflow to reservoir or lake
O = Outflow
S = Storage

The second equation relates storage purely to the outflow, which is true for lakes and reservoirs, where the outflow depends only on the lake level. The outflow relationship may be of the form:

$$O = KBH^{3/2} \qquad (12)$$

if the outflow is controlled by a rectangular weir, or:

$$\mathrm{O} = \mathrm{K'H^n} \qquad (13)$$

where K' and n depend on the nature of the outflow channel.

Such relationships can be turned into outflow—storage relationships because storage is a function of H, the lake level. The Eqs. (12) and (13) can then be rewritten in the form

$$O = K''S^{\mathrm{m}}. \qquad (14)$$

Alternatively, there may be no simple functional relationship, but a graphical relationship between O and S can be plotted or stored in the computer. The continuity equation and the outflow storage relationship can then be solved either graphically or numerically, so that, given certain inflows, the outflows can be calculated. Notice the assumption that a lake or reservoir responds very rapidly to an inflow, and the whole lake surface rises uniformly.

During the development of the kinematic routing model described later, a reservoir routing technique was developed which has proved to be very useful. Because reservoir routing is such an important and basic requirement in hydrology, the method will be presented in full.

Reservoir routing can be greatly simplified by recognizing that complex stage–discharge relationships can be linearised for a limited range of flows. It is even more simple to relate stage levels to storage and then to linearise the storage–discharge relationship. The approach described below can then be applied to any lake or reservoir situation, ranging from natural outflow control to the operation of gated spillways and turbine discharge characteristics.

From a logical point of view, it is probably easier to develop the routing relationship by considering storage, or volume changes. In a fixed time interval ΔT, the reservoir inflow volume is $VI(J)$, where J indicates the current time interval. The corresponding outflow volume is $VO(J)$ and the reservoir storage volume is $S(J)$. If the current inflow volume $VI(J)$ were to equal the previous outflow value $VO(J-1)$, then the reservoir would be in a steady state and no change in reservoir storage would occur. Using the hypothetical steady state as a datum for the current time interval, we can define changes in the various flow and storage volumes, where Δ indicates an increment,

$$\Delta VI(J) = VI(J) - VI(J-1) \tag{15}$$

$$\Delta VO(J) = VO(J) - VO(J-1) \tag{16}$$

$$\Delta S(J) = S(J) - S(J-1). \tag{17}$$

To maintain a mass balance for the current time interval,

$$\Delta VI(J) = \Delta S(J) + \Delta VO(J). \tag{18}$$

Using the relationship for a linear reservoir,

$$S(J) = K^{*}QO(J) \tag{19}$$

where $QO(J)$ is the outflow which is equal to $VO(J)/\Delta T$. The corresponding equation for the previous time interval is

$$S(J-1) = K^{*}QO(J-1). \tag{20}$$

Subtracting this equation from Eq. (5) we obtain

$$\Delta S(J) = K^{*}\Delta QO(J). \tag{21}$$

Substituting that $\Delta QO(J) = \Delta VO(J)/\Delta t$,

$$\Delta S(J) = (K/\Delta t)^{*}\,\Delta VO(J). \tag{22}$$

Substituting in equation (4) for $\Delta S(J)$

$$\Delta VI(J) = (K/\Delta t + 1)^{*}\,\Delta VO(J) \tag{23}$$

$$\Delta VO = \frac{1}{1+K/\Delta t} * \Delta VI(J). \tag{24}$$

This equation can be rewritten for flows by substituting $\Delta VO(J)$ equals $\Delta QO(J)^{*}\Delta T$ and $\Delta VI(J)$ equals $\Delta QI(J)^{*}\Delta T$,

i.e.,

$$\Delta QO(J) = \frac{1}{1+K/\Delta t} * \Delta QI(J) \tag{25}$$

where

$$\Delta QI(J) = QI(J) - QO(J-1). \tag{26}$$

Then

$$QO(J) = QO(J-1) + \Delta QO(J). \tag{27}$$

Equation (25) to (27) represent an extremely simple reservoir or lake routing procedure. To achieve this simplicity, the change in inflow, $\Delta QI(J)$, and the change in outflow, $\Delta QO(J)$, must each be changes from the outflow, $QO(J-1)$, in the previous time interval, as defined in a similar manner to Eqs. (15) and (16). The value of K is determined from the storage-discharge relationship, where K is the gradient, dS/dQ. This storage factor, K, which has dimensions of time, can be considered constant for a range of outflows.

When the storage–discharge relationship is non-linear, which is usual, it is necessary to sub-divide into linear segments. The pivotal values of storage, $S(P,N)$, and discharge, $QO(P,N)$, where N refers to the Nth pivot point, are tabulated. Calculations proceed as described until a pivotal value is approached, or is slightly passed. The next value of K is calculated, not from the two new pivotal values, but from the latest outflows $QO(J)$ and from the corresponding storage $S(J)$. The current value of storage is calculated from,

$$\begin{aligned} S(J) &= S(P,N-1) + \Sigma\Delta S \\ &= S(P,N-1) + (QO(J-1) \\ &\qquad - QO(P,N-1)) * K(N-1,N). \end{aligned} \tag{28a}$$

Then the next $K(N,N+1)$ value is calculated,

$$K(N,N+1) = \frac{S(P,N+1) - S(J)}{QO(P,N+1) - QO(J)}. \tag{28b}$$

This procedure maintains continuity for storage and discharge, and is easy to program because no iterations are required.

It will be noted that the routing procedure can be carried out without calculating the latest storage value: only inflows and outflows need be considered. The storage value at any time can be calculated from Eq. (19), which states that there is always a direct and unique relationship between storage, $S(J)$, and outflow, $QO(J)$.

In summary, the factor $1/(1 + K/\Delta t)$ in Eq. (25), represents the proportion of the inflow change, $\Delta QI(J)$ which becomes outflow. The remaining inflow change becomes storage. The process is identical for increasing or decreasing flows: when flows decrease, the changes in outflow and storage are both negative. Eqs. (25) and (27), the heart of the matter, are repeated for emphasis,

$$\Delta QO(J) = \frac{1}{1+K/\Delta t} * \Delta QI(J) \tag{25}$$

$$QO(J) = QO(J-1) + \Delta QO(J). \tag{27}$$

CHANNEL ROUTING

The assumptions of reservoir routing no longer hold for channel calculations. The channel system takes time to respond to an input. Also, storage is a function of conditions at each end of the length of channel being considered, rather than just the conditions at the outflow end.

The simplest channel routing procedure is the so-called Muskingum method developed on the Muskingum River (G.T. McCarthy,[24] Linsley[2]).

Channel routing is also based on continuity, Eq. (11). Before it can be utilized this equation must be rewritten in finite difference form

$$\frac{I_1 + I_2}{2} + \frac{O_1 + O_2}{2} = \frac{S_2 - S_1}{t}. \tag{29}$$

Also, some assumption must be made from which storage can be computed. The Muskingum method linearizes the problem and assumes that storage in a whole channel reach is completely expressible in terms of inflow and outflow from the reach, namely

$$S = K[xI + (1 - x)O]. \tag{30}$$

Substituting from (16) in (15), following Linsley,[2] the result obtained is,

$$O_2 = c_0 I_2 + c_1 I_1 + c_2 O_1 \tag{31}$$

where c_0, c_1 and c_2 are functions of K, x and t and $c_0 + c_1 + c_2 = 1$. The influence of x is illustrated in Figure 6. Extending this equation to N days:

$$\begin{aligned} O_N = {} & c_0 I_N + c_1 I_{N-1} + c_2 c_0 I_{N-1} + c_1 c_2 I_{N-2} \\ & + c_2^2 c_0 I_{N-2} + c_2^2 c_1 I_{N-3} + \Lambda \\ & + c^{N-2} c_1 I_1 + c_2^{N-1} O_1. \end{aligned} \tag{32}$$

Usually, because c_0, c_1 and c_2 are less than one, terms in c_3 and higher are ignored, but whatever the simplification, the result is of the form

$$\begin{aligned} O_N = {} & a_1 I_N + a_2 I_{N-1} + a_3 I_{N-2} + a_4 I_{N-3} \\ & + a_5 I_{N-4} \text{ etc.} \end{aligned} \tag{33}$$

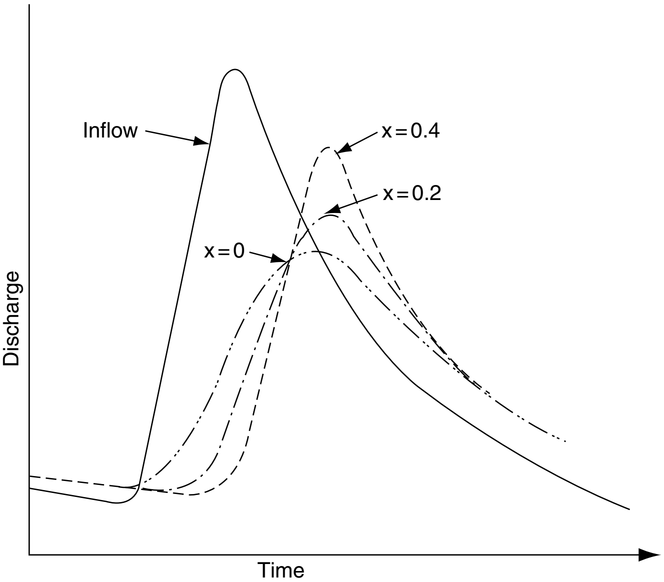


FIGURE 6 Muskingum routing: to illustrate the influence of x on outflow hydrograph (after Linsey[2]).

Once again the set of equations for O_1 to O_R can be written as

$$\{O\} = [I]\{a\}. \tag{34}$$

(N.B. The use of I for inflow is unfortunate and leads to confusion with the identity matrix. Also O is confusable with the null matrix.) It must be admitted that, from the theory, only a_1, a_2 and a_3 are independent, but in practice the precision with which a solution can be obtained for c_0, c_1 and c_2 does not justify calculating a_4, a_5 etc. from the first three. The best approach at this stage is once more to resort to least squares fitting, recasting Eq. (34) in the form of Eq. (8.)

Many hydrologists calculate the routing coefficients by evaluating K and x in the traditional graphical method.

Many assumptions are made in the Muskingum method, such as the linear relationship between storage and discharge, and the implied linear variation of water surface along the reach. In spite of these assumptions the method has proved its value.

Another difficulty which occurs with the Muskingum method usually occurs when the travel time in the reach and the time increments for the data are approximately equal, such as when the travel time is about one day and the data available is mean daily flows. From the strict mathematical viewpoint, the time Δt should be a fraction of the travel time K, otherwise flow gradients such as dO/dt are not well represented on a finite difference basis. However, it is still possible to use time increments Δt greater than K if the flow is only changing slowly, but caution is necessary. The reason for caution is that the C values in Eq. (31) are a function of $\Delta t/K$ and are not constant.

Solving for the C's in terms of K, x and Δt:

$$C_0 = -\frac{x + 0.5(\Delta t/K)}{1 - x + 0.5(\Delta t/K)} \tag{35a}$$

$$C_1 = \frac{x + 0.5(\Delta t/K)}{1 - x + 0.5(\Delta t/K)}. \tag{35b}$$

Hence,

$$C_2 = -\frac{1 - x - 0.5(\Delta t/K)}{1 - x + 0.5(\Delta t/K)} \tag{35c}$$

$$C_0 + C_1 + C_2 = 1. \tag{35d}$$

To illustrate the influence of Δt and K, some synthetic data as used to construct Figure 7. Values for K and x were chosen and when values of C_0, C_1, and C_2 were calculated for different values of Δt. In addition, an assumed inflow was routed using the K and x values and using a time interval Δt which was small compared with K. These resulting inflows and outflows were then reanalyzed for C_0, C_1, and C_2 using time increments five times greater than the original Δt. Such Δt values exceeded K. The new estimates of C_0, C_1, and C_2 are also plotted in Figure 7. Note the general agreement in shape of the C_0 and curves, etc. Exact agreement is lost because of the diminishing accuracy of the d*O*/d*t* terms, etc. as Δt increases. Note also how rapidly the *C* values change when Δt is approximately equal to *K*.

These remarks are not necessarily meant to dissuade hydrologists from using the Muskingum method. The intention is to illustrate some of the pitfalls so that it may be possible to evaluate the probable validity or constancy of the coefficients (Laurenson[22]). The worst situation appears to be when Δt ; *K,* because in real rivers *K* decreases with rising stage and the *C* values are very sensitive to whether Δt is just less than or greater than *K*.

A very real problem in the application of the Muskingum method, and in fact of any channel routing procedure, is the problem of lateral inflow to the channel reach. Given the inflows and outflows for the reach as functions of time, it is necessary to separate out the lateral flows before best fit values of *K* and *x* can be determined. The lateral flows will, in general, bear no relationship to the pattern of main streamflows. Sometimes it is possible to use flow measurements on a local tributary stream as an index of total lateral flow. The cumulative volume of lateral flow can be determined by subtracting summed inflows from summed outflows. The measured tributary flow can then be scaled up to equal the total lateral flow and these flows can then be subtracted from the reach outflows. This residual outflow can be used in the determination of the Muskingum coefficients.

Sometimes it is possible to find periods of record where lateral flows are small or perhaps have a more predictable pattern, such as during recession periods. Also, the routing coefficients can be refined by an iterative procedure and by using various sets of data, although not infrequently it is found necessary to define Muskingum coefficients for different ranges of flow, because real stream flow is not linear as is assumed in the model.

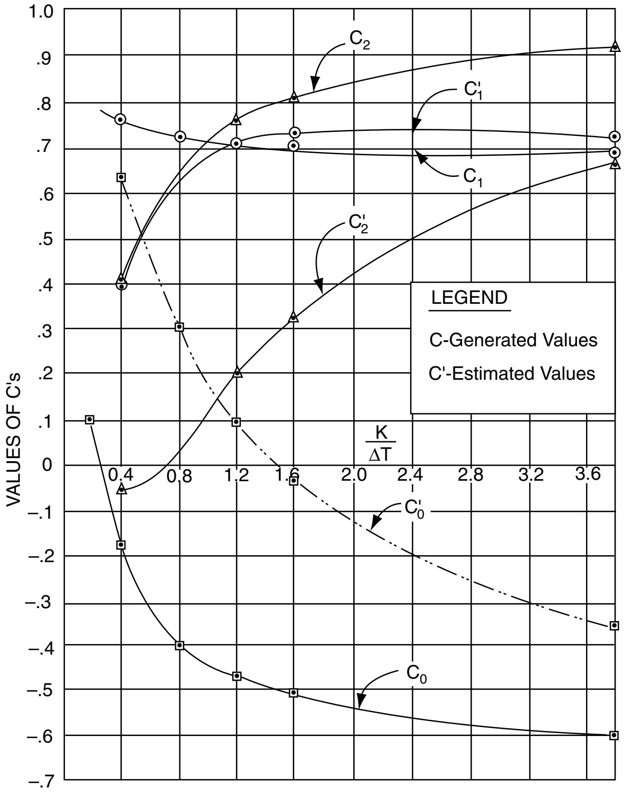


FIGURE 7 Muskinghum coefficients as a function of storage and routing period.

KINEMATIC WAVE THEORY

It has been known for many years that flood wave movement is much slower that would be expected from the shallow water wave theory result of $\sqrt{gy}$. Seddon[25] showed as early as 1900 that flood waves on the Mississippi moved at about 1.5 of river velocity. The theoretical reasons for this slowness of flood wave travel have only been realized during the last twenty years. The theory describing these waves is known either as monoclinic wave theory or kinematic wave theory.

Kinematic wave theory has suffered from neglect because of its usual presentation in the literature. The common approach is to start with the continuity equation and a simplified Manning equation. These equations are combined to yield Seddon's Law that flood waves travel at about 1.5 times the mean flow velocity. Such an approach restricts the method to an unalterable flood wave translating through the channel system. Although such a simple model has the benefit of approximating the real situation, much more useful and general results are available by setting up the two equations in finite difference form and carrying out simultaneous solution on a computer. In particular the method handles lateral inflow with great ease and also calculates discharge against time, eliminating the guess work from the time calculation.

Kinematic wave theory can be seen to be a simplification of the more general unsteady flow theory in open channels, for which the equation of motion and the continuity equation can be written (Henderson and Wooding[4,26])

$$S_0 - \frac{\partial y}{\partial x} - \frac{v}{g}\frac{\partial v}{\partial x} - \frac{1}{g}\frac{\partial v}{\partial t} = \frac{v^2}{c^2 R} \tag{36}$$

Bed-Slope Depth Slope Convective Acceln. Local Acceln. Friction term

$$\frac{\partial Q}{\partial x} + B\frac{\partial y}{\partial t} = \delta i\left(\text{lateral inflow}\right). \tag{37}$$

In different situations various terms in equation (36) become negligible. In tidal flow the friction and bed-slope terms are sometimes omitted, although this simplification is not strictly necessary. In this way the positive and negative characteristics are obtained with propagation velocities of $u \pm \sqrt{gy} = dx/dt$. At the other end of the spectrum is kinematic wave theory in which the bedslope term is assumed to overpower the other three terms on the left hand side, namely, depth slope, convective and local acceleration. Hence kinematic wave are friction controlled. Equation (36) is therefore reduced to

$$S_0 = \frac{v^2}{c^2 R}. \tag{38}$$

To combine Eqs. (36) and (38) into the simplest characteristic form, it is assumed that *R,* the hydraulic radius, may be replaced by *y.* Then Eq. (38) yields

$$Q = vBy = B.Cy^{3/2} S_0^{1/2}. \tag{39}$$

Substituting for in (37) by differentiating Eq. (39)

$$\frac{3}{2} BCS_0^{1/2} y^{1/2} \frac{\partial y}{\partial t} + B\frac{\partial y}{\partial t} = \varepsilon \tag{40}$$

The left hand side is the total differential d*y*/d*t* if the term

$$\frac{3}{2} CS_0^{1/2} y^{1/2} = \frac{\mathrm{d}x}{\mathrm{d}t} = \text{wave velocity.} \tag{41}$$

The important fact that monoclinical waves can only propagate down-stream with a wave velocity given, in the simple case, by Eq. (41), is well known, as is best discussed by Lighthill and Witham.[27] As previously mentioned, as an empirical result, Eq. (41) dates back to Seddon[25] and his study of flood wave movement.

Comparing kinematic wave theory with channel routing, Eq. (37) can be recast in exactly the same form as Eqs. (29) and (30) if the lateral in flow term is ignored. The difference in the two methods depends on the second equation chosen, which is basically a stage–discharge relationship. Channel routing, as illustrated by the Muskingum method, assumes a linear relationship between stage and discharge to arrive at Eq. (30) for storage. Kinematic theory uses a simplified equation of motion which is usually either the Chezy or Manning equation.

In a further development of kinematic theory researchers such as Hyami[29] have shown that the influence of channel storage can be included and this results in an extra term which is effectively a diffusion term:

$$\frac{\partial y}{\partial t} + c\frac{\partial y}{\partial x} = K\frac{\partial^2 y}{\partial t^2} \tag{42}$$

Solutions of this equation are approximately of the form:

$$y = y_0 e^{-ax} \cos(bt - ax) \tag{43}$$

This solution assumes a sinusoidal wave profile, and the solution is seen to be delayed as it moves downstream and also it decays in amplitude. It is seen that effectively an exponential damping term is introduced by the inclusion of diffusion.

Such relationships are complex and cumbersome, demanding detailed data on the channels which often is not available. Inspection of the resulting equations reveals a similarity to the old lag and route method. The author has used a simplified method based on these equations which has proved itself to be very successful.[46] The method uses a variable travel time calculated from the velocity-stage curves for the upstream end of the given channel reach. The kinematic result of approximately 1.5 times channel velocity is used, although a best fit result can be found by optimization. The flow is then routed through a simple linear reservoir, the size of which is related to the channel storage. This reservoir allows a check to be kept on continuity and also it provides an exponential decay term as demanded by the diffusion term in the full kinematic theory. For relatively steep rivers like the Fraser River in British Columbia, one linear reservoir is adequate. Less steep streams may require two reservoirs, but two reservoirs should be adequate in that a two-parameter exponential will fit almost any curve imaginable.

The simple channel routing procedure has been found to be very powerful because it approximates the true physical behavior very closely. It yields a simple non-linear behavior and models the well documented flood wave subsidence. On the Fraser River system the method proved very successful in separating out the very large lateral flows which are characteristic of that particular river system, and this separation was really the key to the modeling of the whole system behavior.

ARTIFICIAL GENERATION OF STREAMFLOW

If streamflows can be considered to be statistically distributed, then it may be possible to establish statistical measures of their distribution. Such statistical measures will be the mean flow, the variance and perhaps some correlation between successive flows. It may also be possible to determine the nature of the distribution, such as normal, or log normal. If such parameters can be found, then these parameters can be used to regenerate data with the same statistical patterns.

The reason for such data regeneration is not apparent until we examine the purposes for which such data can be used. If we are given some 30 years of streamflow data, it is possible to make reasonable estimates of the statistical parameters of, say, monthly or annual flows. The 30 years of record will contain only a very few extreme events, such as floods or prolonged droughts, and will therefore not impose very severe tests on any water resource system which is storage dependent, such as a hydro system or an irrigation project. Generated data could presumably be constructed for very long periods of time and would, if the procedure is justified, reproduce many more extreme flows and consecutive run-off patterns which will test the proposed resource system more thoroughly. The statistical risk of failure can then be evaluated.

This is the aim of streamflow generation studies. Whether the methods currently in use are adequate or valid is still under debate. The aims are sound and the methods promising, if used with caution and an awareness of the intrinsic assumptions.

It is worthwhile comparing the artificial generation methods with the approaches previously used. It has been customary to construct a test period or sequence from the recorded data. As an example, it was not unusual to use the three driest years of record and assume that they occur in successive years. This became the design drought. Without care to evaluate just how extreme an event this might be, such a method might lead to either over or under design. Hopefully, streamflow generation will lead to more balanced designs, especially as experience with the techniques and the techniques themselves improve.

To consider the generation techniques, it is first of all important to appreciate that streamflow forms a time series which can be highly dependent or almost entirely independent according to the time interval chosen. Daily flows are usually highly correlated with each other; monthly flows may still have a reasonable serial correlation, but annual flows may be almost totally independent. Certainly it is usual to assume that annual flood peaks are independent.

The simplest streamflow model assumes that there is no long term trend in the run-off pattern, so that the mean flow for the period of record is a reasonably good estimate of the long term mean. Also it is assumed that the standard deviation from the mean is a good measure of the dispersion of flows around the mean. The model expresses:

Generated flow = Mean flow + Random Component i.e.,

$$Q_i = \bar{Q}_j + t_i \sigma_j \tag{44}$$

i refers to synthetic sequence, *j* to historic sequence.

Q_i = synthetic streamflow for month *i*
= is mean flow for same month
t_i is random normal deviate with mean zero and variance unity (Available in most computer libraries)
σ_j is standard deviation of flows for month *j.*

The random component is seen to be made up of the calculated standard deviation multiplied by a random variable t_i which is generated by a random number generation, and has a statistical distribution which is the same as the normal random error. The variable t_i therefore redistributes σ_j in a random Gaussian fashion.

A more sophisticated model by Thomas and Fiering[42,43] includes the memory that this month's flow has of last month's flow. This memory is described by serial correlation:

Generated Flow = Mean Flow + Serial Component + Random Component
i.e.,

$$Q_{i+1} = \bar{Q}_{j+1} + b_j\left(Q_i - \bar{Q}^j\right) + t_i \sigma_{j+1}\left(1 - r_j^2\right)^{1/2} \tag{45}$$

The additional symbols are

b_j is regression coefficient for serial correlation of *j* and *j* + 1 month
r_j is correlation coefficient between *j* and *j* + 1 month.

It will be noticed that when a serial component is introduced, the random component must be correspondingly decreased by the factor $(1-r_j^2)^{1/2}$ so that there is not a net increase in flow.

Many other synthetic streamflow methods have been tried or suggested but at the present time only the simple models [Eq. (44) and (45)]. Thomas and Fiering's model[43] seem to be in use by practising hydrologists. Also, the reader should appreciate, that the methods are not without their problems, for generated flows may go negative, so that modeling of droughts is questionable. However, a study of the statistical frequency of failure may still be better than the old methods.

So far, generation techniques have been discussed only for a single streamflow record. If a whole river basin with several reservoirs on different tributaries and on the main stream is being studied, then a much more difficult problem arises. It is not difficult to generate artificial data for each tributary or for the main stream. But such data will have none of the cross-correlation which exists in real data. That is to say, with the real data, if flows are high in one tributary, then probably the other tributaries are also reasonably high. This cross-correlation is difficult to preserve in generating techniques. It is discussed at great length by Fiering but the method suggested for preserving cross-correlations is very demanding on computer storage and time.

The writer wonders whether use of a physically based computer model of a river basin, coupled to a random generation of precipitation events both for one area of storm and intensity of storm might not produce data more economically and with the correct statistical interdependence.

HYDROLOGICAL SIMULATION

Simulation is a very broad term covering many types of mathematical and physical procedures. The streamflow generation techniques already discussed are one type of simulation. Analogue computer models and physical flow models are also simulations. Digital computer models for simulation may be based on statistics entirely or on the physical behavior of various components of the system. Some models may be a mixture of both physical and statistical methods.

Although the types of simulation models vary widely, their aims are similar, namely to describe the behavior of a complex system so that predictions of system behavior can be made from some specified input. These predictions may be short-term flood-flow predictions to give flood warning, especially for operation of flood protection schemes. Alternatively, predictions may be needed to evaluate water resource schemes and to examine the influence of storage, diversion, consumptive use, etc. on the system of operation.

STATISTICAL TECHNIQUES

Many simulations, particularly run-off forecasts, have been based on multiple regression analysis. Flow has been treated as the dependent variable and many factors such as precipitation, antecedent precipitation, sunshine hours, etc. have been used as the independent variables.[13] Such methods have had at least moderate success. The difficulties with the method have been that many of the so-called independent variables are in reality somewhat related, for example, rainfall is inversely correlated with sunshine hours. Also, the lack of true functional relationship usually means that coefficients for one year's data vary for other years or for periods of higher or lower flow, and therefore forecast precision in inherently limited and evaluation of changes to the system, such as introduction of reservoirs is not feasible.

The deficiencies of multiple regression analysis have been somewhat overcome, according to some investigators, by the use of multivariate analysis. The principle component techniques select the most significant relationship even if the data is interrelated. However, this type of analysis is still blind to the physics of the system. Examples of this technique and the theory are given in Kendall, Snyder and Wallis[10,14,15] and the reader can consult these for more details.

Monte Carlo methods, of which the stream generation is a simple example, have been proposed and used. Such methods have had staggering success in some fields of physics, such as nuclear studies. The classic example was the original study of critical mass for the atomic bomb. Such success led people to hope for more general applications. But even the keenest exponents of such methods agree that they are no substitute for an understanding of the physical behavior of the system. The writer considers that such statistical techniques are a reasonable approach if there is no physically based alternative. Experience with constructing physically based simulation models shows immediately how valuable it is to incorporate a functional relationship which bears some resemblance to the physical system. An example of this statement is the unit hydrograph approach which can be considered as a simple simulation model. The actual physical behavior of a catchment seems to be well described by a unit hydrograph. As soon as such a unit hydrograph is introduced, it becomes a relatively simple matter to relate cause and effect, which in that case is precipitation and run-off. A similar result can be achieved if the correct data is used in a multiple regression analysis, but such data selection presupposes a knowledge of the system behavior.

PHYSICAL COMPUTER SIMULATION MODELS

Qualitatively, we can describe the behavior of a hydrologic system. The catchment soil layers have storage which decreases and delays run-off and permits evapo-transpiration to occur. The lake and channel system further delays flows and modifies the shape of the outflow hydrograph. Groundwater supplies a highly damped outflow which is significant during dry spells. Quantitatively we may know the precipitation input and the run-off output, although there may be data error, especially in the precipitation. There may also be data from which estimates of potential evapo-transpiration can be made. Presumably data will also be available of such matters as lake areas, stage-discharge relationships, catchment areas, and elevations, and the details of the streamflow network.

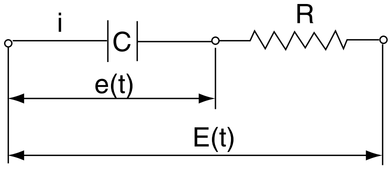
It is important to realize what can be discovered about a system simply from studying the input to the system and the output from the system, as has been very well demonstrated by Nash.[23] He considers a simple electrical system made up of a capacitor and a resistance as shown in Figure 8. Then if $E(t)$ is considered to be the input and $e(t)$ the output it can be shown that

$$e(t) = \frac{1}{1+RCD} E(t) \tag{46}$$

where D is the differential operator. Hence we can solve for $e(t)$ if $E(t)$ is given. Alternatively if $e(t)$ and $E(t)$ are given we can solve for RC as a lumped term, but we can never find the separate values for R and C. Therefore when we construct a simulation model, we may be able to correctly model the output from the input, although the parameters used in the model may be lumped terms describing various factors in the real system.

Hydrological systems are considerably more complex than the above example and it is important to realize that we cannot start to "fit" the model to the data until we have simulated the total system behavior. Also, in general, the system will not be linear like the simple electrical network, so that response becomes a function of flow.

In constructing a simulation model we must first of all decide the factors or processes which should be included to correctly and adequately describe the system. Each factor must then be approximately fitted using any data or knowledge which we may have. It is a basic rule that these factors, such as evaporation, unit hydrograph, soil moisture storage, etc. should be modeled with as few parameters as possible, as long as adequate description of a process is not jeopardized. This minimization of parameters has been well named



$$E(t) = e(t) + Ri(t)$$

$$i(t) = C\frac{de}{dt}$$

$$\therefore e(t) = \frac{1}{1+CRD} E(t)$$

FIGURE 8 Simple system demonstrating non-separability of *R* and *C* (after Nash).

as the "principle of parsimony." The final optimization of the model parameters is done when the whole model is complete. One parameter at a time is adjusted in a step by step process until best fit values are obtained.

To illustrate the process we shall consider a simple model, Quick and Pipes.[47]

WATERSHED MODELING IN MOUNTAIN CATCHMENTS

The rain and snowmelt run-off processes in rugged mountain catchments appear, at first sight, to be highly complex. Rain, snow, temperature, and soil and rock composition of the watershed all are highly variable at different elevation levels in the watershed.

In reality this apparent complexity is simplified by the orographic influences. The strong orographic gradients of behavior impose a useful discipline on the various processes. In particular, there tends to be a greater areal uniformity of precipitation and temperature within each elevation zone. This uniformity of behavior by elevation zone, if proved to be real, offers a great simplification of the most difficult of hydrologic behavior, namely the variation of rain and snow and temperature over an area.

The UBC Watershed model was designed to take advantage of the strong orographic discipline which exists in a mountain watershed. The primary purpose in designing the model was to provide a tool for forecasting flood run-off, but in reality, its most useful purpose has become the day to day forecasting of run-off for hydropower production. The model can also be used as a research tool to investigate the total system behavior of a mountain catchment. Such research investigations are highly dependent on the accuracy and distribution of the meteorological and hydrological data base.

Design of the UBC Watershed Model

There are five main subdivisions of the total model namely meteorological data processing, snowmelt calculation, soil moisture budget, routing of flow components to channel outflow point, and statistical evaluation of model performance. These subdivisions will be discussed briefly. A complete description of model design and use is given in Quick and Pipes.[49]

Meteorological data processing Meteorological data is available only at discrete points. In many situations, only one data station may exist and that station may be in the valley and perhaps not even in the watershed itself. Even when more than one data station exists, there is still the technical problem of distributing the point data to all elevations and regions of the watershed.

It is assumed that the orographic influences are the strongest and that these orographic effects are modified by the moisture content of the air mass. Usually there is no direct information on vapor pressure values, and therefore indirect indications of moisture status are used. For example, temperature range during the day is used as an indicator of relative humidity. A low daily temperature range is associated with humid conditions when the temperature lapse rate will be approximately equal to the saturated adiabatic rate. A high daily temperature range is assumed to indicate that maximum temperature will vary at the dry adiabatic rate, but minimum temperature will tend to show a very low lapse rate. A simple functional relationship is therefore defined between temperature lapse rate and the daily temperature range.

Orographic precipitation gradients are made functionally dependent on the rates of saturated dry adiabatic lapse rates corresponding to the observed temperatures. The larger the difference between saturated, L_S, and dry, L_D, lapse rates, the greater is the convective instability of the air mass, so that convective instability is measured by

$$\frac{L_D - L_S}{L_D}. \tag{47}$$

If the air mass is already very unstable, then the orographic effects will be moderate. The orographic opportunity is therefore expressed as

$$\text{orographic opportunity} = 1 - \frac{L_D - L_S}{L_D} \equiv \frac{L_S}{L_D}. \tag{48}$$

When this ratio approaches 1, the orographic effect on precipitation is high, whereas when the ratio decreases, the orographic effect is weaker. It will therefore be appreciated that orographic effects will be large during the winter, when temperatures are low, and L_S is more nearly equal to L_D. On the other hand, in the warm summer weather, because L_S is then much lower than L_D, the orographic effects will be much weaker.

This simple functional variation of precipitation gradients has proved to be a very valuable tool in distributing meteorological data and has resulted in considerable improvement in forecast accuracy.

Snowmelt calculation The strong elevation dependence of snowmelt supports the use of temperature indices for the calculation of snowmelt. The simple degree day method probably accounts for 80% of the snowmelt process, but there are periods of extreme melt when radiation and condensation melt become important additional factors. The snowmelt algorithms used in the UBC Watershed model have been discussed in a previous paper, but, in summary, additional temperature terms are added to account for radiation and condensation, Quick and Pipes.[46] Radiation is represented by the daily temperature range and condensation is handled by using the minimum temperature as an approximation for the dew point temperature. Under warm summer conditions this dew point assumption is usually good and this is the only time when the extra terms are significant. The snowmelt equation, for forested areas, when slightly simplified, is

$$BM = \left\{ TM + \left(\frac{TX - TN}{k_1} + TN \right) * \frac{TN}{k_2} \right\} * PTM \tag{49}$$

where

BM = Daily melt in an elevation band
TM = Mean daily temperature
TX = Maximum daily temperature
TN = Minimum daily temperature
PTM = Point melt factor (Assumed value 2.1 mm/°C/Day)

k_1, k_2 are constants (assumed value, $k_1 = 8$, $k_2 = 10°C$).

A modified formulation is used for open areas, where additional weighting is given to the maximum temperature and this equation is also used for glaciated regions

$$BM = \left(TX + \frac{TN}{k_2} * TN \right) * PTM \tag{50}$$

A negative melt budget is maintained during ripening periods and following cold periods.

The soil moisture budget When designing a watershed model, there is considerable computational advantage if all the non-linear watershed behavior can be concentrated in one section of the model. This advantage is even greater if the non-linearities are handled completely in each time step, rather than being distributed over the many time steps of the routing procedure.

In the UBC watershed model, non-linearities have been confined to the soil moisture budget section of the model. The soil moisture budget section subdivides the total rain and snowmelt inputs into fast, medium, slow and very slow components of run-off. This subdivision of the total run-off depends on the present status of each section of the soil moisture and groundwater components, and so the subdivision process is non-linear. The degree of non-linearity is in the hands of the model designer and his concept of the various hydrologic processes. For example, in the UBC model, the non-linearities are greatest at the soil surface, where the subdivision between fast and medium run-off is determined by soil moisture deficit conditions. In contrast, the deep groundwater zone has no non-linear behavior within itself; it simply accepts what finally arrives. This gradation of non-linear behavior, which is maximum at the surface and decreases with depth, is probably realistic in mountain conditions, but might require redesign for some soil conditions in flat terrain.

Once the flow has been subdivided into components of run-off, each component can be routed separately to the watershed outflow point. Some models might assume that water can be exchanged between the various run-off components as the water migrates towards the stream system. Although it is probably quite realistic to make such an assumption, it also introduces considerable complexity and extra data handling. More complex versions of Sugawara's[51] tank model could be designed to operate in this fashion.

The routing process itself can also be linear or non-linear. If a linear routing procedure is used, then each flow increment can be routed independently of any other flow, and the routing procedure is computationally simple. Non-linear routing, on the other hand, can become quite complex.

The simplicity of linear routing procedures is attractive to the model designer, but such simplicity is only of value if it is also a realistic representation of actual run-off processes. In the following section, some brief consideration will be given to the physical nature of the various run-off processes and to their mathematical representation.

Two major types of routing procedure are widely referred to in the literature: unit hydrograph routing, Nash[17] and kinematic routing, Hayami.[29] More complex procedures, usually referred to as hydraulic routing, are not usually used within watershed models, but are occasionally used in some more detailed channel routing procedures. The unit hydrograph is defined as a linear process Sherman,[16] yet, some workers make changes to hydrograph shape for extreme rainfall events. This non-linearing of the unit hydrograph can be avoided by recognising that total watershed response is made up of several different unit hydrograph responses, as shown schematically in the figure. Soil moisture budgeting switches varying amounts of water to the various run-off components, so that the composite hydrograph can be more peaked or less peaked according to the severity of the storm. Kinematic routing can be linear or non-linear according to the assumed functional form between incoming volume and flow rate. In channels, this relationship is usually expressed as a velocity-depth or a discharge-area relationship.

THE UNIT HYDROGRAPH

For the unit hydrograph to be truly linear, then, as Nash[23] has pointed out, storage and outflow must be linearly related. A cascade of storages will then yield various possible hydrograph shapes. This simple model assumes a uniform instantaneous input of run-off volume over an δA in time δt. A real watershed receives a non-uniform input of run-off volume over an area, A, in a finite time period, t. Strictly speaking, a watershed should be subdivided into regions where uniform input is a reasonable approximation. With such a subdivision, the final outflow hydrograph would become not only a function of watershed shape and varying response characteristics of sub-areas, but also of rain and snowmelt distribution across the watershed. For example, there should be a recognizable difference between rainfall response and snowmelt response. Snowmelt, which is highly temperature dependent, is greatest at low elevations and least at the top of the watershed. Rain, on the other hand, undergoes orographic enhancement, so that it is greatest at the highest parts of the watershed and least in the valley portions. This orographic effect is always present to some extent, but, as a general rule, the warmer the air mass, the less is the orographic enhancement. In addition to this complete reversal of elevation dependent distribution of run-off, there is also a considerable contrast in the intensities of water input from rain and snow events. Snowmelt, even under quite extreme snowmelt conditions, rarely exceeds 80 mm of water equivalent per day, and some 75% of this melt occurs during the daylight period. Rainfall daily volumes can be considerably greater and can occur over much shorter time periods. These contrasting areal distribution and time distribution effects are partly self-compensating. Snowmelt is greatest at low elevations which tends to make hydrograph response more peaked, but the lower intensity of snowmelt tends to produce a flatter response. Rain, on the other hand, is of greater intensity, producing a higher peak outflow, but is greatest in the more remote regions of the watershed, which has a tendency to reduce the peak flow. There seems to be little or no conclusive data to test this contrast between rain and snow response. This lack of conclusive data may be that the difference between rain and snow events are small, or because the data base is not sufficiently detailed and accurate to demonstrate any difference.

Returning to the question of the watershed routing, some simple theoretical models will be examined to explore the representativeness of linear routing models.

The linear routing process is usually assumed to represent a relationship between the outflow, *QO,* and storage *S,* of a linear reservoir, with a storage constant *K,*

$$S = K(QO). \tag{51}$$

Cascades of such linear reservoirs are used to produce a range of unit hydrograph responses to inputs of rain or snowmelt. A linear Weir is not a structure which can easily be imagined as part of a watershed system, because it is much more realistic to imagine flow through various interstices and flow paths in the soil layers and to assume such flow to be friction controlled. Such a conceptualisation leads to examination of the groundwater flow equations, for example, the Darcy equation,

$$Q = KA\frac{\mathrm{d}y}{\mathrm{d}x}. \tag{52}$$

The area of flow, *A,* can be represented by a width, *B,* and a saturated depth of flow, *y.* Assuming *B* to be constant, then *y* is a linear measure of storage. The equation can be linearized if it is assumed that d*y*/d*x* is constant for all flows, an assumption which might be reasonable for steep mountain catchments where the ground slope dominates the flow process. Eq. (6) might therefore be reduced to the linear flow-storage equation, like Eq. (5).

A single storage-flow relationship produces an instantaneous unit hydrograph response which is a simple exponential decay, as given by the Nash single reservoir result. Even such a simple model, when convoluted with the time distributed input of effective rainfall, will produce a typical time distribution of outflow, modeling the rising and falling limbs of a unit hydrograph. It is not even strictly necessary to resort to a cascading of linear storages, although, from a modeling viewpoint, a cascade of storages offers a flexible method of controlling hydrograph shape.

Other frictionless and friction-controlled equations can also be shown to imply relationships between storage and flow, but many of the relationships are non-linear, for example, a Weir formula and the Manning equation,

Weir: $$Q = KBH^n \tag{53}$$

Manning: $$Q = \frac{A}{n}R_h^{2/3}S_0^{1/2}. \tag{54}$$

In these equations *H* and R_h can be considered as expressions of storage.

A more interesting relationship can be derived from consideration of kinematic wave behavior. Following Lighthill and Whitham's work and later work by Henderson and Wooding, many authors have proposed that kinematic wave behavior is very representative of hydrologic run-off processes, Lighthill and Whitham,[27] Quick and Pipes,[48] either from watersheds or in channels. The starting point of the kinematic approach is the continuity equation

$$\frac{\partial Q}{\partial x} + B\frac{\partial y}{\partial t} = 0. \tag{55}$$

Re-writing and comparing with the total differential of *y,*

$$\frac{\partial Q}{\partial y}\frac{\partial y}{\partial x} + B\frac{\partial y}{\partial t} = 0 \tag{56}$$

and

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}x}{\mathrm{d}t}\frac{\partial y}{\partial x} + \frac{\partial y}{\partial t} = 0. \tag{57}$$

These equations are equivalent if,

$$\frac{1}{B}\frac{\partial Q}{\partial y} = \frac{\mathrm{d}x}{\mathrm{d}t} = C. \tag{58}$$

So far these results only have the restriction that *Q* should be a function of *y* only. Consider the special case when *Q* is a linear function *y,* as previously argued for the Darcy groundwater equation,
i.e.,

$$Q = ky. \tag{59}$$

Then, solving for *C,*

$$C = \frac{1}{B}\frac{\mathrm{d}Q}{\mathrm{d}y} = \frac{k}{B} = \text{constant}. \tag{60}$$

The wave travel time, *T,* for a catchment dimension or channel length, *L,* is

$$T = \frac{L}{C} = \frac{LB}{k} = \text{constant}. \tag{61}$$

Therefore we have a system in which any and all discharges, *Q,* travel through the catchment or channel in exactly the same travel time. The outflow response for any level of flow is therefore only dependent on *y,* which is a measure of storage in the system. Following Henderson and Wooding's analysis of rainfall run-off,[26] the resulting run-off unit hydrograph would have a constant shape, which for the simplified situation is a simple triangle. Convolution of this triangular hydrograph with the effective rainfall will again yield a typical outflow hydrograph. Very accurate data would be needed to distinguish between this final outflow hydrograph and the result from the linear reservoir discussed earlier.

It therefore appears that the use of linear routing is not too unreasonable. Further justification will depend on the results of watershed modeling which are discussed in the next section.

MODEL CALIBRATION AND ASSESSMENT OF PERFORMANCE

Some examples of model calibration will be given for a region in the Upper Columbia River basin (Figure 9). The watersheds, such as the Jordan River and Spillimacheen, selected for calibration purposes are deliberately of moderate size, some 200 to 500 km^2. Too small a watershed responds too quickly and requires a data base time increment of a few hours, whereas the size suggested above can be modeled using daily data. The calibration for these smaller watersheds forms an excellent basis for modeling the larger watershed regions of several thousand square kilometers used in forecasting mainstream flows.

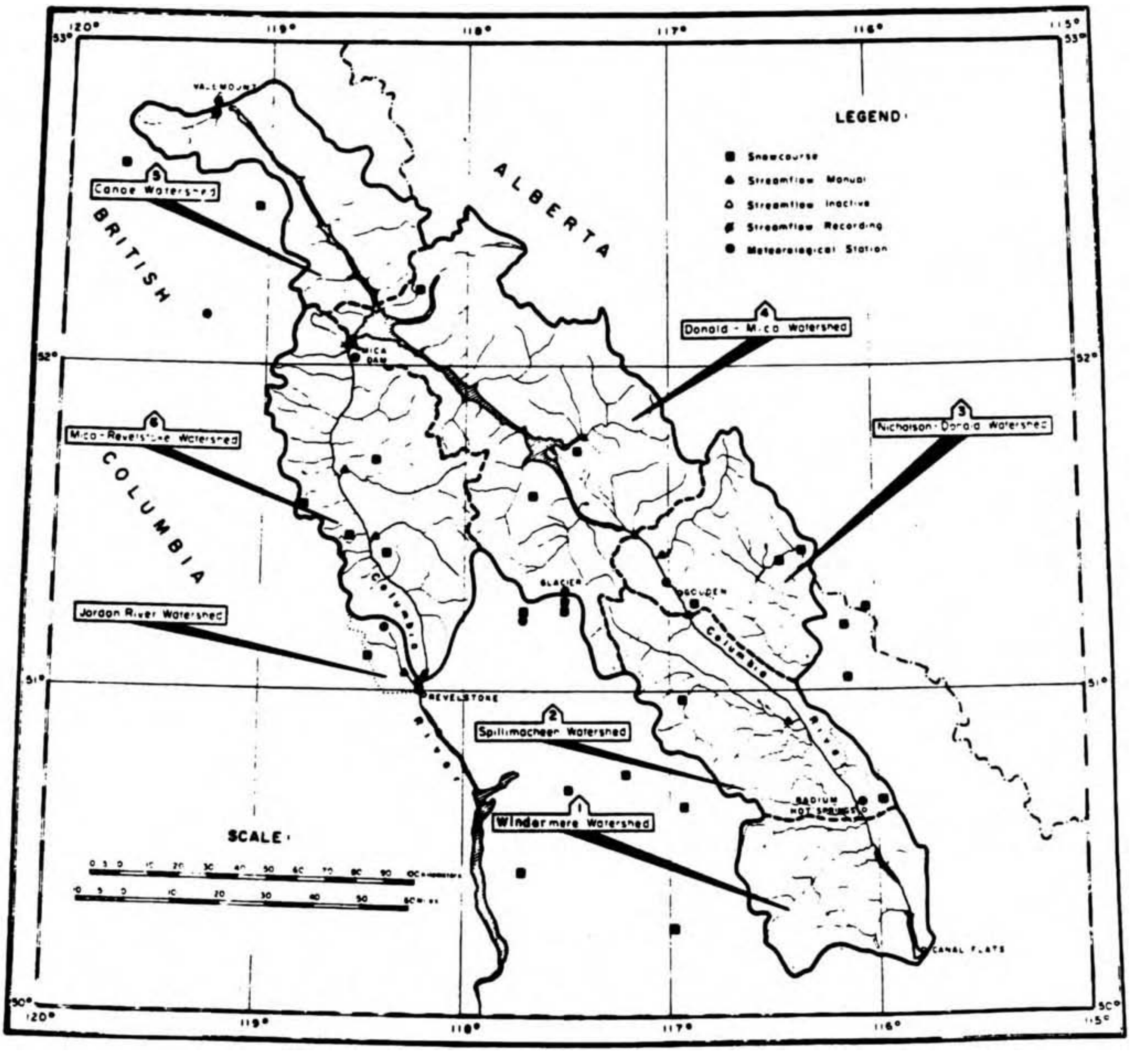
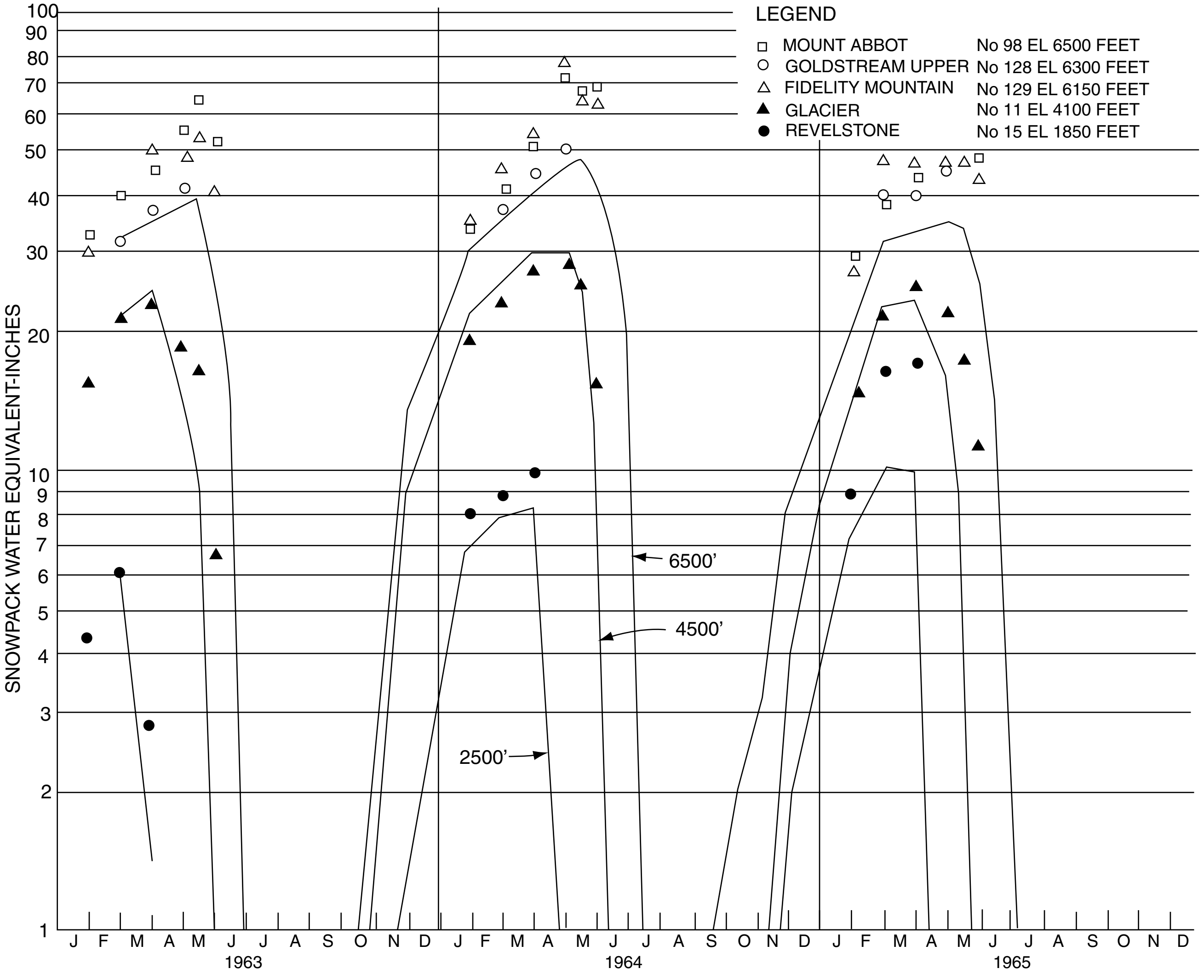


FIGURE 9 Map of Upper Columbia basin (indicate Spill & Jordan)—Remove Goldstream.

One of the first steps in calibrating a watershed is to study the orographic precipitation gradients. The model algorithm describing the variation of precipitation with elevation and temperature is calibrated by comparing calculated snowpacks at various time intervals with the snowcourse measurements, as shown in Figure 10. The complete comparison during accumulation and depletion of snowpacks requires estimation of precipitation by elevation, the form of precipitation, rain or snow, and the snowmelt occurring at various times. The temperature lapse rate is therefore an important ingredient in this part of the analysis, influencing both melt rates and orographic gradients.

Reconstitution of outflow hydrographs is the next stage of model calibration and two examples which have been presented in detail previously, Quick and Pipes[48,50] will be briefly outlined. The Jordan River example, Figure 11, illustrates the considerable improvement which is achieved by using temperature based snowmelt equations which account for additional radiation melt and condensation melt components. For comparison, the melt calculated from only the mean air temperature is plotted. In periods of extreme snowmelt, the melt rate can be approximately double the rate that might be estimated from mean temperature alone. For the example given, plots of temperature and precipitation at a base station are presented for comparison with the outflow hydrograph.

In another example illustrated in Figure 12 for the Spillimacheen River, an earlier paper, Quick and Pipes (1977) discussed and compared the accuracy attainable using different



UBC Watershed Model

Reconstitution of Seasonal Snowpacks (1963–1965)

Goldstream River Basin

FIGURE 10 Comparison of computed snowpacks and snowcourse data.

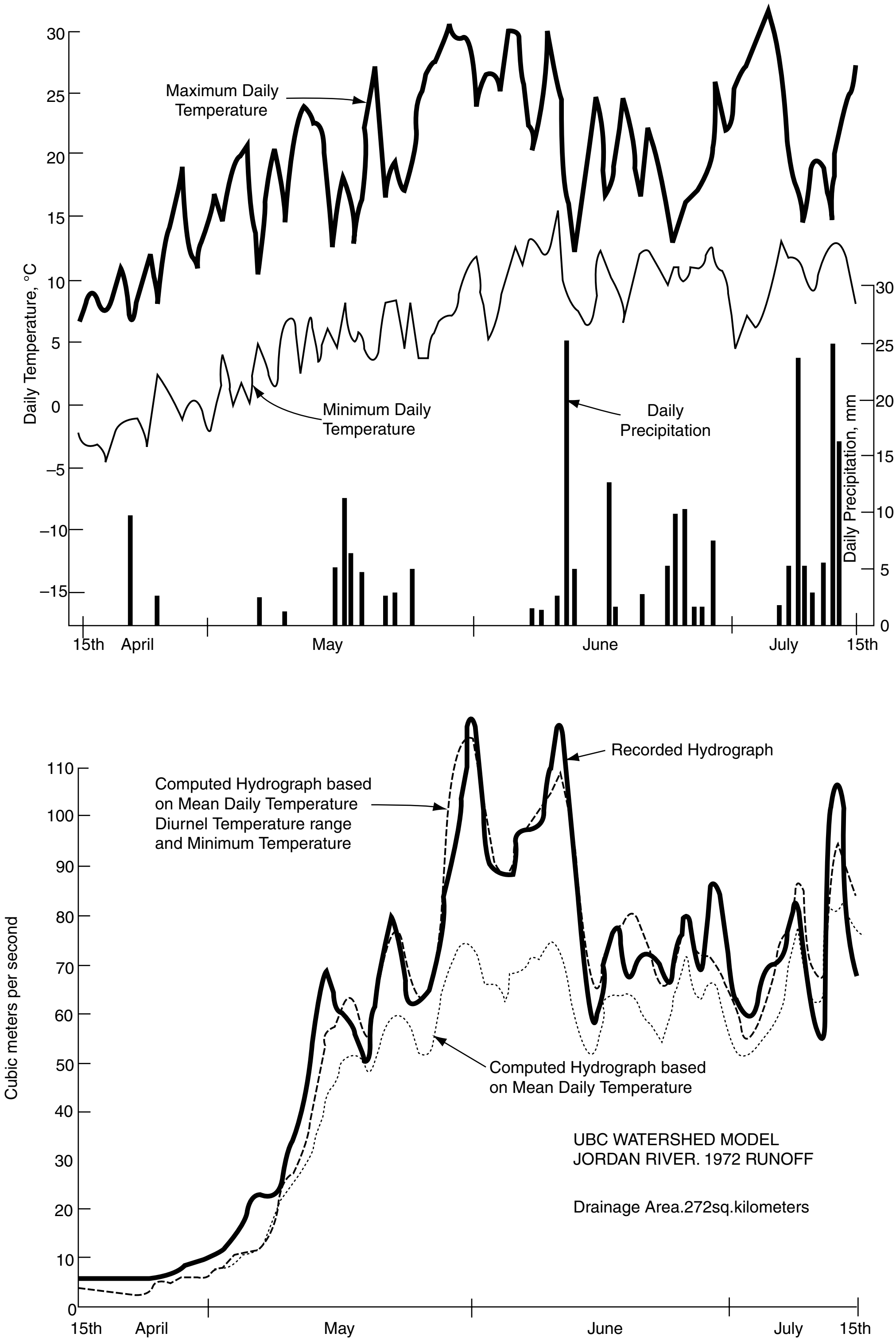


FIGURE 11 Jordan River hydrograph.

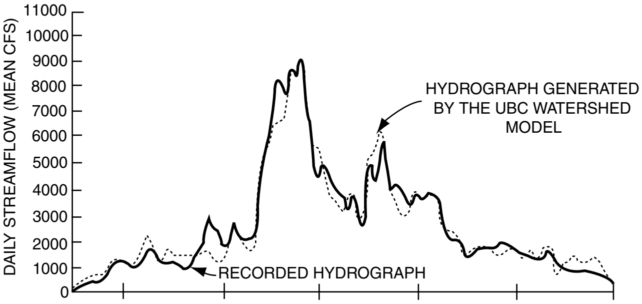


FIGURE 12 Spillimacheen River hydrograph.

single data station and combinations of data stations. None of the data stations were in the watershed. One valley data station exhibited far less precipitation than the mountain stations, varying in frequency, duration, and amount. The study showed that the nearest station, which was also near the mid-elevation of the basin and in a similar climatic zone, gave the best results, superior even to a combination of stations. Such a mid-elevation station reduced data extrapolation errors and is more representative of amount, duration, and frequency of precipitation and of the actual basin temperature regime. For the data tested, the errors of maximum peak flow, monthly volumes and hydrograph shape, measured by residual variance, were all less than 5%.

In summary, simulation techniques should be kept as simple as possible, provided that representation of reality is not jeopardized. As much precalibration as possible for each element of the model should be done during the model construction, using good common sense. The value of graphical fitting should be utilized because, by use of graphical representation, the mind is able to handle large quantities of data simultaneously. Final fitting of the model is done when the main simulation blocks have been assembled, and at this point it is useful to have a simple quantitative measure of the goodness of fit. Several measures have been proposed, but Nash's[23] is as good as any. This measure is calculated by subtracting the recorded output from the simulated output for each time increment. These differences are squared and summed to give a *residual variance.* The optimization can then proceed to minimize this residual variance.

REPRESENTATIVE DESIGN APPLICATIONS

Minor Design Problems

Such structures as small dam spillways, road culverts, small bridges, etc. require estimation of a design flood. The approach by which such a design flood can be estimated is shown diagrammatically in the flow diagram Figure 13.

First it must be decided whether rain floods or snowmelt floods are likely to be limiting and it may be necessary to calculate each before a conclusion can be reached.

Secondly, the response time of the catchment must be considered and this is typically characterized by the unit hydrograph, especially the time from the start of run-off to the peak.

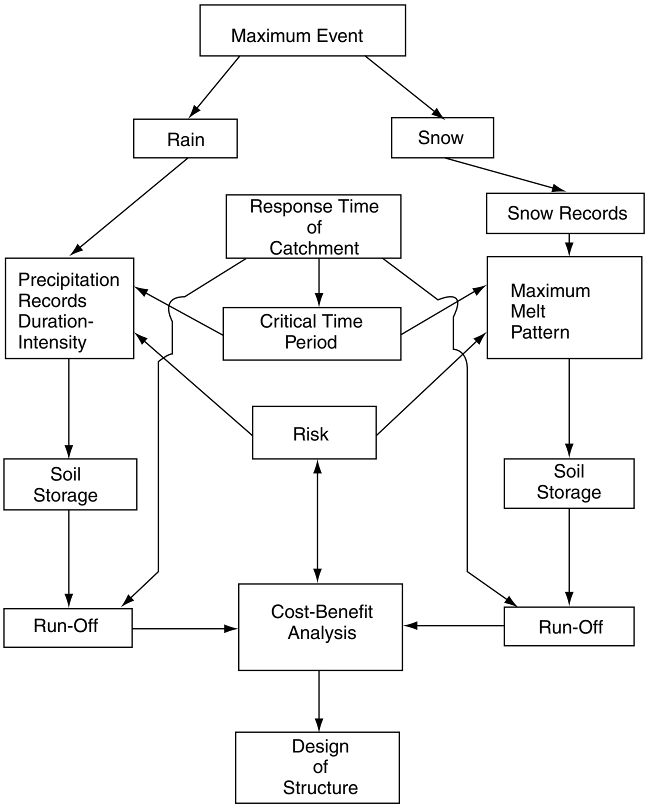


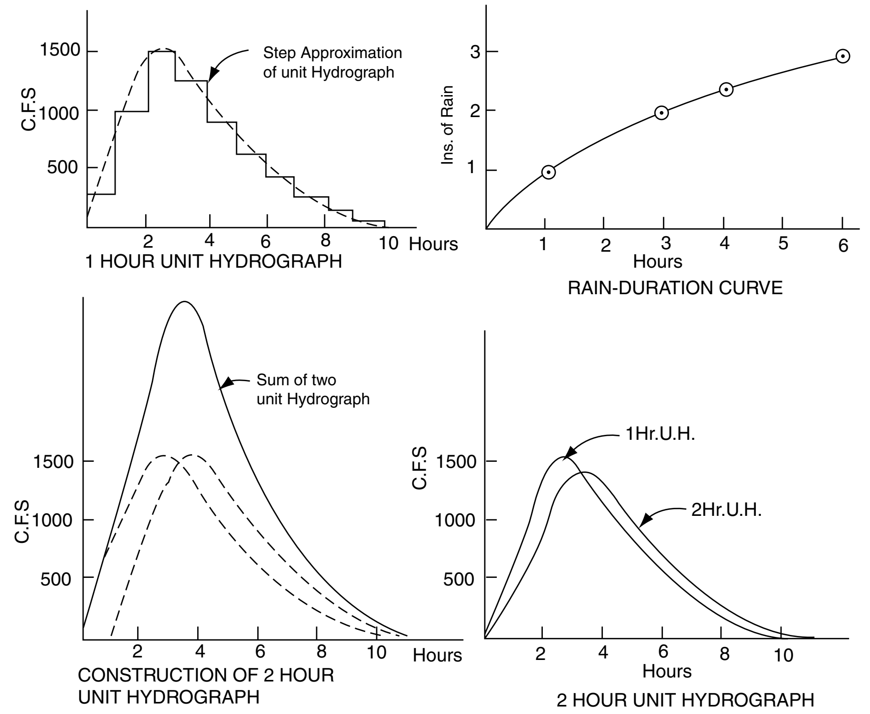
FIGURE 13 Estimation of design flood for a small spillway or road culvert, etc.

Thirdly, the precipitation records for rain events must be examined. In many instances data for the catchment may not exist, but representative data from adjacent areas may be available. Such data must be analyzed to yield duration-intensity plots for given return periods, or probabilities and for the catchment area in question.

The unit hydrograph data and the precipitation data must now be combined to determine the critical period of run-off. As an example, starting with the 1 hour unit hydrograph, it is easy to superimpose several 1 hour unit hydrographs to obtain 2 hour, 3 hour, 4 hour, etc. unit hydrographs. These longer duration unit hydrographs have relatively lower peaks as is illustrated in Figure 14. However, the greater volume of rain in longer duration storms may give a higher resultant peak and such a situation is illustrated in the figure.

A similar calculation for snowmelt would yield a maximum flow for run-off. Note that, for small catchments, the diurnal temperature variations impose limitations on the maximum snowmelt run-off because the high temperatures only last for a limited number of hours each day.[30,31,32]

Once the peak flow has been determined for a given rain or snow event probability, a first design of the structure should be made. The costs can be estimated and also damages resulting from higher floods can be at least approximated. At this point the cost of added protection can be considered and the assumed risk can be reevaluated.



PEAK RUN-OFF	RAIN	U.H.PEAK	RUN-OFF-RAIN x U.H.PEAK
1 Hour Storm	1.0 inch	1540	1540 C.F.S
2	1.7	1290	2190
3	2.0	1290	2580
4	2.5	1125	$\boxed{2810}$ Max.
5	2.75	990	2720

Critical Duration is 4 Hour Storm of 2.5 ins. of Rain, which produces maximum Run–off of 2810 C.F.S.

FIGURE 14 Determination of critical storm duration for a given catchment.

Experience of similar structures is very useful at this stage, especially in evaluating such added constraints as possible log jams or the influence of siltation or erosion (Bureau of Reclamation[54]).

Major Spillway Design Floods

Large dams are a particularly challenging design problem from many viewpoints. In particular, the consequences of failure due to inadequate spillway design have given rise to numerous extensive hydrological studies and the development of suitable techniques. For such a major structure it is usual to calculate design floods by several methods, as discussed in such references as Kuiper and Linsley.[52,53]

The first method is based on extreme probability theory as previously discussed in the statistics section. The analysis of extreme independent time series is based on the work of Fisher and Tipper[11] and is well illustrated by Gumbel.[12] The big question is what level of risk to assume, and probabilities of 1:1000 to 1:10,000 are not uncommon for such projects. The economic arguments reveal that the spillway cost is often small compared with the havoc that could occur if the dam failed. Hence it is quite usual for spillways to be designed for much more rare events than the design flood used for the rest of the system. Clearly the determination of the spillway design flood involves the incorporation of a large safety factor, but it is still necessary to make as good an estimate as possible of the true probability of such an event.

The second method for estimation of design floods is the *maximum probable flood*[1] approach. The physical processes involved in producing a flood are studied. In some areas, such as the North-Western part of North America, snowmelt is the principal factor for large rivers. Elsewhere rain floods may be limiting. Probable maximum precipitation can be estimated by storm maximization,[55] or by statistical methods. A simple approach was developed by Hershfield[56] and has been called the poor man's probable maximum precipitation, the PMPMP. It is necessary to study the weather patterns which produce either high snowmelt or high rainfall intensities. Storm patterns, durations, and movements can be studied. The storm movement can be crucial for a large river for it is not unknown for storms to move downstream at a rate similar to the flood wave movement, so that considerable reinforcement of the flood can take place. On northern rivers like the Mackenzie, which flows to the north, snowmelt and ice breakup occur from south to north, resulting in build up of ice jams and floods. There is also the question of combination of rare events such as rain on snow or several consecutive storms. Concerning rain on snow, it is noticeable that the snow-covered areas are the ones which mainly contribute to run-off, probably because the snowmelt has already primed the soil and also because the 100% humidity of the warm air during rain can produce considerable condensation melt.

The calculation of flood run-off from such storm or melt patterns requires a run-off simulation calculation of the type previously discussed. The resulting flood flow may be modified by the reservoir storage until maximum permissible reservoir stage is reached. From then on the spillway must be capable of passing the total flood, less any flow passed through the turbines or diverted to irrigation.

A third approach, which is often of comfort to the designer, is to compare the design flood with that used on other similar rivers. Careful comparison and cross-correlation may help to confirm the reasonableness of the final design values.

Standard project floods[1] can be determined by similar processes and used for the design of other aspects of a scheme, such as downstream diking and channel protection schemes. In this case the resulting damage from failure may not be so widespread and may not endanger life. It is not unusual for such design floods to be only a fraction of the value of a spillway design flood, perhaps half.

Complex Systems

Usually a complex water resource system involves the operation of storage or even several storages. Such operation aims to minimize flood risk at one extreme and drought risk at the other end of the scale. The scheme must satisfy various complex demands for such uses as hydro-power, irrigation, fisheries requirements, water supply, water quality, and navigation. Recreation is also now a demand of increasing importance and imposes limitations regarding water levels on beaches or in marinas. Water intakes and wastewater effluent lines set additional limitations. Water quality also sets demands on minimum flows, water temperature, lake levels, etc.

An important aspect of river basin behavior which has long been neglected is the effect of the alteration of run-off pattern on the ecological balance and erosion patterns downstream. Hydrological studies can supply the biological ecologists with information on how flood levels and seasonal run-off patterns will change with given schemes so that harmful effects can be recognized before damage is done.

Added to this complexity is the question of run-off forecast accuracy which can greatly influence the efficiency of water utilization. In the evaluation of such schemes hydrology is seen in its most necessary and challenging role. The hydrologist is called on to develop accurate forecasts of run-off. He also must simulate the total system behavior and then subject the system to sequences of run-off patterns generated from the characteristics of the recorded data. The resulting performance of the system is then analyzed from the economic viewpoint and, frequently, projections must be made into the future to estimate eventual demands and trends and system capability. In this type of problem the whole technical capability of the hydrologist is utilized and as techniques and data improve so the conflicting requirements of water use, water quality, recreation, and flood control can be better reconciled.

REFERENCES

Textbooks and Reference Books

1. Chow, V.T., *Handbook of Hydrology*, McGraw-Hill.
2. Linsley, Kohler, and Paulhus, *Hydrology for Engineers*, McGraw-Hill.
3. Bruce, and Clark, *Introduction to Hydrometeorology*, Pergamon.
4. Henderson, *Open Channel Flow*, MacMillan.
5. Viessman, Knapp, Lewis and Harbaugh, *Introduction to Hydrology*, 2nd Edition, IEP Series, Dun-Donnelley, 1977.
6. Meinzer (Ed.), *Hydrology*, Dover.
7. Maas *et al., Design of Water Resource Systems*, Harvard.

Statistics

8. Moroney, M.J., *Facts from Figures*, Pelican.
9. Parl, B., *Basic Statistics*, Doubleday College Courses, Refs. 1,2,3.
10. Kendall, M.G., *A Course in Multivariate Analysis.*
11. Fisher, and Tippett, Limiting forms of the frequency distribution, *Cambridge Phil. Soc.*, Vol 24, 1927–28.
12. Gumbel, Applications of extreme statistics, lecture 4, applied mathematics series 33, U.S. Nat. Bureau of Standards.
13. Matalas, N.C., and M.A. Benson, Effect of interstation correlation on regression analysis. *J. of Geophys. Res.,* **66**, No. 10, pp. 3285–3293, Oct. 1961.
14. Snyder, W.M., Some possibilities for multivariate analysis in hydrological studies, *J. of Geophys. Res.*, **67**, No. 2, pp. 721–729, Feb. 1962.
15. Wallis, J.R., Multivariate statistical methods in hydrology—A comparison using data of known functional relationship. *Water Resources Research,* **1**, No. 4, pp. 447–461, 1965.

Runoff

16. Sherman, L.K., Streamflow from rainfall by the unit-graph method, *Eng. News. Rec.,* **108**, pp. 501–505, April 7, 1932.
17. Nash, J.E., The form of the instantaneous unit hydrograph, *Intern. Assoc. of Sci. Hydrology*, Pub. 45, **3**, pp. 114–121, 1957.
18. Dooge, J.C.I., A general theory of the unit hydrograph, *J. Geophys. Research*, **64**, No. 1, pp. 241–256, 1959.
19. Singh, K.P., A non-linear approach to the instantaneous unit hydrograph, Ph.D. thesis, University of Illinois, 1962.
20. Snyder, W.M., Hydrologic Studies by electronic computers in TVA, *Proc. A.S.C.E.* HY 2, pp. 1–10, Febr. 1960.
21. Snyder, W.M., Matrix operations in hydrograph computations, Tennessee Valley Authority, Res. Paper No. 1, 10 pages, Knoxville, Tennessee, Dec. 1961.
22. Laurenson, E.M., Storage analysis and flood routing in long river reaches, *J. Geophys. Res.*, **64**, No. 12, pp. 2423–31, Dec. 1969.

23. Nash, J.E., Lectures on Parametric or Analytical Hydrology, Ontario Committee for the Hydrologic Decade, **PR** 28, 1969.
24. McCarthy, G.T., The Unit Hydrograph and Flood Routing, Conference of North Atlantic Division, U.S. Corps of Engineers, June 1938.
25. Seddon, J.A., River Hydraulics, *Trans. A.S.C.E.*, **43**, p. 179, 1900.
26. Henderson, F.M., and R.A. Wooding, Overland Flow and Interflow From Limited Rainfall of Finite Duration. *J. Geophys. Res.*, **69**, No. 8, p. 1531, April 1964.
27. Lighthill, and Whitham, On kinematic waves: I-Flood movement in long rivers, *Proc. Roy. Soc.* (London) A., **229**, No. 1178, p. 281, May 1955.
28. Dooge, and Harley: Linear Routing in Uniform Open Channels, Intern. Hydrology Symposium, Fort Collins, Sept, 1967, 8.1, pp. 57–63.
29. Hyami, S., On the propagation of flood waves. Bulletin No. 1 Disaster Preventino Research Institute, Kyoto University, Japan (Dec. 1951) (Summarized by Henderson [4] p. 384).

Snow Hydrology

30. Hildebrand, C.E., and T.H. Pagenhart (1955) Lysimeter studies of snowmelt. U.S. Corps Research Note.
31. Light, P. (1941), Analysis of high rates of snowmelting. *Trans. Am. Geophys. Union* Pt. I., **22**, pp. 195–205.
32. Linsley, R.K. (1943), A simple procedure for the day to day forecasting of run-off from snowmelt. *Trans. Am. Geophys. Union* Pt. III, pp. 62–67.
33. Quick, M.C. (1965), River flood flows, *Journal of the Hydraulics Division, Proc. A.S.C.E.*, No. HY 3, **91**, pp. 1–18.
34. SIPRE (1959), Report No. 12, Annotated Bibliography of Snow and Ice.
35. Snow and Ice Hydrology, Proceedings of Workshop, 1969, Colorado State University.
36. U.S. Corps of Engineers (1956), Report No. PB 151660 *Snow Hydrology,* U.S. Corps of Engineers, Dept. of Commerce, Washington, D.C.
37. U.S. Corps of Engineers (1960), Run-off from snowmelt, EM 1110–2-1406, U.S. Govt. Printing Office, Washington, D.C.
38. Wilson, W.T. (1941), An outline of the thermodynamics of snowmelt. *Trans. Am. Geophys. Union* Pt. I, pp. 182–195.
39. Snow and Ice Hydrology, Proceedings of Workshop 1969, Colorado State University.
40. Quick, M.C., Experiments with physical snowmelt models, National Research Council Hydrology Symposium No. 8: Runoff from snow and ice, Quebec City, May 1971.
41. Quick, M.C. *Forecasting Runoff-Operational Practices*, invited theme paper: International Symposium on the role of snow and ice in hydrology, Banff, 1972. UNESCO, WMO, IHD, Published by National Research Council, Ottawa.

Artificial Streamflow Generation

42. Maass *et al., Design of Water Resource System,* Chapter 12, Harvard.
43. Fiering, M.B., *Streamflow Synthesis.*

Simulation

44. Crawford, N.J., and R.K. Linsley, *Digital Simulation in Hydrology: Stanford Watershed Model IV,* Tech. Report No. 39, Dept, Civil Eng., Stanford University, Stanford, California, 1966.
45. O'Connell, Nash, and Farrell, River flow forecasting through conceptual models II, *Journal of Hydrology,* 1970, p. 322.
46. Quick, M.C., and A. Pipes, (1976), Combined snowmelt and rainfall runoff model, *Canadian Journal of Civil Engineering,* **3**, No. 3, pp. 449–460.
47. Quick, M.C., and A. Pipes, *Daily and Seasonal Forecasting with a water Budget Model,* Proceedings of the International Symposium on the role of snow and ice in hydrology. Banff 1972, UNESCO, WMO and IHD, National Research Council, Ottawa.
48. Quick, M.C., and A. Pipes, (1975), Nonlinear channel routing by computer, *Journal of the Hydraulics Division, Proc. A.S.C.E.,* **101**, No. HY6.
49. Quick, M.C., and A. Pipes, (1977), Snowmelt floods in mountain catchments, *Proceedings of the Canadian Hydrology Symposium 77 Floods*, Environment Canada, Ottawa, Ontario, KIA OE7.
50. Quick, M.C., and A. Pipes. (1977), *UBC Watershed Model: Users Manual,* Dept. of Civil Engineering. University of B.C.
51. Sugawara, Ozaki, Katsuyama and Watanabe, (1975), *Tank Model, Symposium on the Application of Math Models in Hydrology,* IAHS, IHP, UNESCO, WMO, Bratisla6va.

Designing

52. Kuiper, *Water Resources Development,* Butterworths.
53. Linsley, and Franzini, *Water Resources Engineering,* McGraw-Hill.
54. Bureau of Reclamation: Design of Small Dams U.S. Govt. Printing Office, Washington, 1965.
55. World Meteorological Organization, *Manual for Estimation of Probable Maximum Precipitation Operational Hydrology,* Report No. 1, WMO—No. 332. 1973. (Also obtain current listing of WMO Hydrological Publications.)
56. Blench, T., *Mechanics of plains rivers,* University of Alberta, Edmonton (1986).

MICHAEL C. QUICK
University of British Columbia

HEALTH PHYSICS: see MANAGEMENT OF RADIOACTIVE WASTES; INDUSTRIAL HYGIENE ENGINEERING; RADIATION ECOLOGY

I

INDOOR AIR POLLUTION

PART 1

Laboratory work and chemical testing involves procedures that could contaminate the air inside occupied spaces. The nature of the contaminants varies widely: high humidity from steam baths, odors from hydrogen sulfide analyses, corrosion capabilities of alkalies and acids, solubility of acetone, explosive properties of perchloric acid, health hazards of bacteriological aerosols, and poisonous properties of nickel carbonyl. Ideally, best procedure is *not to emit*; but the next best is to remove or exhaust directly and as close to the point of origin for safety of laboratory personnel and protection of property.

To achieve this end, the accepted methods used for containment and removal of contaminants is by restricting the contaminant procedures to within an enclosure or hood. Simultaneously the air is drawn across the hood face to capture and remove the contaminants before escaping into the room.

In the design of a fume exhaust system utilizing hoods the following factors must be analyzed and evaluated: Capture velocities, Fume hood design, Seven basic hood designs, Makeup air source, Air distribution, Exhaust system, Exhaust duct materials, Exhaust air treatment, Special systems.

CAPTURE VELOCITIES

Air flow rates required for hood exhaust systems are based on a number of factors, the most important of which is *capture velocity.* For most applications these will range from 50 to 200 fpm. The lower figure is used to control contaminants released at low speed into relatively quiet room air (15 to 25 fpm). The higher figure is used to control contaminants released at high rates. Under special conditions hood face velocities as low as 25 fpm have been used with industrial type hoods.

Conclusions regarding optimum face velocity selection are rather mixed. In conceptual design of a lab facility this is given much thought and argument, especially when air conditioning is to be included. For every 1000 cfm of air exhausted through hoods, 3 to 4 tons of refrigeration are required to be added to system capacity for makeup air. At $1,000 per ton of refrigeration the cost of exhausting 1000 cfm could range from $3,000 to $4,000. This certainly adds to hood burden and capital outlay.

Some design emphatically forbid hood face velocities less than 100 fpm. Attempts have been made to relate face velocity to hood service by compromising fume hood usage with the added responsibility of supervision by laboratory personnel to insure that fume hood usage is restricted to the type contaminant for which face velocities were selected. To this end, Brief (1963) offers a method of hood classification as a step toward economy of design and operation. He classified type “S” hoods for highly toxic contaminants (threshold limit values less than 0.1 ppm) as having face velocities from 150 to 130 fpm. Type “A” hoods for moderately toxic contaminants (TLV’s of less than 100 ppm) can be sized for face velocities of 100 to 80 fpm. Hoods for non-toxic contaminants, Type “B” (TLV’s above 100 ppm), are sized for a face velocity from 60 to 50 fpm.

It should be emphasized that TLV’s should be used with care and not as sole criteria since they represent airborne concentrations that most workers may be exposed to repeatedly during a normal work day of 8 hours duration for a working lifetime.

Fume hood efficiency depends on the amount of air exhausted and hood design. To assure flexibility of operation and maximum safety to lab personnel, a fume hood should be designed for exhaust air rates ample for complete removal of all contaminants. This may be a logical step when only one hoods or two are involved in a single facility. However, with more than two, generous exhaust through all hoods can impose a heavy initial and operating cost penalty on the air conditioning system. From actual experience with laboratory design, it is difficult to select a one-hood design that will satisfy all situations.

FUME HOOD DESIGN

The function of a hood exhaust system is to protect the lab technician from exposure. Thus, the heart of the system is the hood and the design begins with the hood, which is, at best, a compromise between the ideal and the practical. Basically, a hood is a simple box, Figure 1(a). Without the necessary indraft shown for the basic ventilated hood, Figure 1(b), the material inside the hood can become airborne and be emitted into the room by one or a combination of the following normal laboratory operations: thermal action and convection currents, mechanical agitation, aspirating action by cross currents of the air outside the box. Material can escape from the basic hood only through the door or opening in front. However, in the simple ventilated hood, contaminants are kept inside by the action of the air flowing into the opening. To contain and keep the material from escaping, sufficient air must be exhausted to create and maintain an indraft through the face of the hood opening.

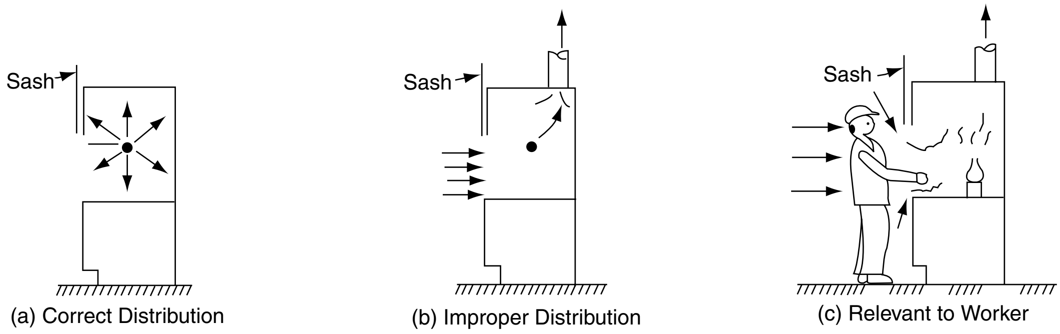
Hoods should control contaminated air so that the contaminant does not reach the breathing zone of the lab technician in significant quantities.

Nearly all hood designs presently in use attempt to provide protection in three ways: a mechanical shield, direction of air movement, dilution of contaminant by mixing with large volumes of air inside the hood. The mechanical shield comprises the hood sash. When an experiment is being set up the sash is in the raised position. In many experiments, the sash is lowered two-thirds the way down or even closed off entirely while an unattended experiment is being carried out. Only the occasional visit by the technician is needed. Care should be exercised not to lower the sash to a level that can cause too high an indraft velocity with attendant overcooling or snuffing out of a burner flame. Protection is provided by the direction of air flow across the back of the worker and into the hood proper, past the equipment within the hood and thence into the exhaust system, Figure 1(c). Lastly, because large amounts of air are being moved through the hood, dilution of the contaminated air takes place readily and further reduces the hazard of breathing hood air.

SEVEN BASIC HOOD DESIGNS

Seven basic hood designs are in use, all as shown in Figures 2a–2g.

1) *Conventional hood* (Figure 2a)
All exhausted air taken from the room. This is the simplest, low in initial cost and effective. However, high exhaust air rates place a heavy burden on air conditioning capital cost and operation.
2) *Conventional hood with reduced face velocity* (Figure 2b)
An attempt to compromise hood effectiveness to reduce air conditioning load chargeable to the hoods. Although low in relative cost, it does reduce air conditioning load but its effectiveness in removing fumes generated within the hood is weakened.
3) *Conventional hood with use factor* (Figure 2c)
Exhaust hoods may be needed at random intervals and it is not likely that they would be simultaneously. As with other types of air conditioning loads, there is a usage or diversity factor that is apparent, yet difficult to define precisely. This factor depends upon judgment, experience, and logic. For example, a large number of hoods in a laboratory room does not necessarily mean all hoods will be operating at one time since the number of lab personnel will be limited and thus reflect on the number of hoods in operation. On the other hand, it is the policy of some laboratories to keep *all* hoods in operation 24 hours a day, even though they are used intermittently. So much depends on the management of the facility and it behooves the designer to explore the total operation with the ultimate user.
4) *Internally supplied hood* (Figure 2d)
Required makeup air is fed directly inside the hood without affecting the overall room air conditioning. This air need not be cooled in summer but merely tempered in winter. Although an additional air handling system is required, the saving on the air conditioning load can offset



(a) Correct Distribution (b) Improper Distribution (c) Relevant to Worker

FIGURE 1 Flow directions through hoods.

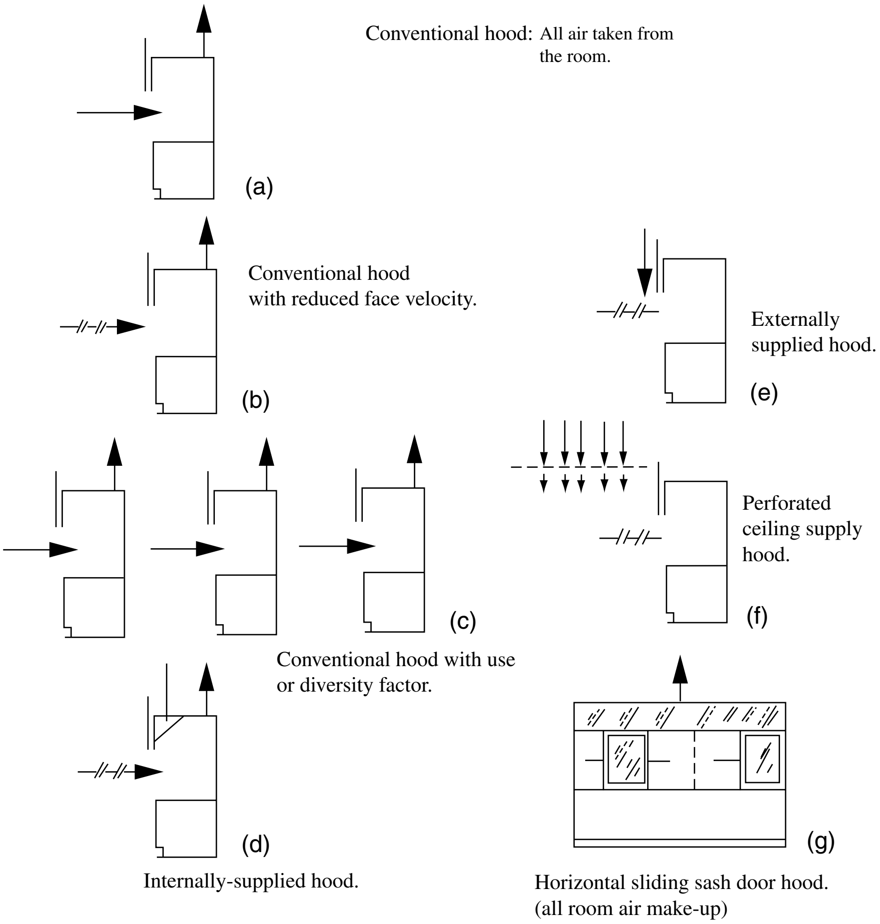


FIGURE 2 Hood designs.

this. Cost of hood runs medium to high but unless carefully designed and balanced fume removal effectiveness can be poor.

5) *Externally supplied hood* (Figure 2e)
Because of the additional duct system required such a system is relatively more expensive, relatively low cost effect on air conditioning, and because air is being exhausted across the hood face, fume removal effectiveness is good.
6) *Perforated ceiling supply hood* (Figure 2f)
This allows ample opportunity for the conditioned air to mix with room air and it becomes often necessary to sensibly cool but not dehumidify this auxiliary supply. Because air is exhausted across the hood face, fume removal effectiveness is good.
7) *Horizontal sliding sash hood* (Figure 2g)
Compared to the conventional hood with its vertical sliding door, the horizontal sliding ash unit presents much less area to be exhausted and total exhaust is thereby reduced. Relative cost of the hood is low and since less air is exhausted, air conditioning costs are low. Air conditioning and fume removal effectiveness are good.

To keep hood face disturbances to a minimum high velocity streams from the air conditioning system should not be permitted to disturb the even, smooth flow of air across the hood face.

MAKEUP AIR SOURCE

Makeup air to balance that exhausted is the most essential design feature of any hood exhaust system. When a fume hood is operating poorly, closer analysis will most often show inadequate makeup air supply. There is no air for the hood to "breathe" and an improperly sized makeup system will starve the fume hood and restrict its intended operation.

Some designs depend on air drawn from adjoining corridors and office spaces. Introduction of makeup air by indirect means is an economical approach. However, such a system can lead to balancing problems and cross-contamination between laboratory spaces. Positive introduction of air from corridors and office spaces by use of transfer fans can improve this. It has been found that the most reliable, flexible, and easily maintained system arrangement is that in which an adequate supply of outside conditioned air as makeup is supplied to the laboratory space to balance the air being exhausted. It is good practice to supply a little less makeup air this way than that being exhausted. A slight negative pressure will be maintained, drawing air through door louvers from corridors or adjacent offices.

Air exhausted from a hood is *never* recirculated so that hood burden goes up. Operating costs can be reduced by supplying makeup air from an auxiliary source instead from the cooling system. The air handled is filtered and tempered in winter only. Of the seven basic hood designs, numbers 4d, 4e, and 4f make use of the auxiliary system. An auxiliary system can be either a central unit or unitary type with an outside air inlet for each laboratory. Correct selection of the type of makeup air system can be made only by an engineering analysis and flow sheet of the hood exhaust system.

One of the most important characteristics of an exhaust system is that at some point the system must end and discharge to atmosphere. Unfortunately, while the exhaust system has ended at this point, the problems associated with that exhaust system may have just begun. If too much air discharged from an exhaust system is recirculated through the supply system not much good has been accomplished. If by poor design the exhaust air is not properly located with respect to the intakes of other supply systems, potentially disastrous results can be attained. Many poor designs are commonplace. The real cure for this type of problem is not higher exhaust velocities, higher stacks, better weather caps, better separation of discharge and intake openings, or other, although one or more of these can contribute to the cure. The real remedy must start back at the source of contamination itself.

Because the pattern of natural air flow around buildings is not predictable, contamination by the location of vent effluents and air intakes is difficult to put to practice. Halitsky[2] (1963) and Clarke[3] (1967) have advanced theoretical knowledge and rule of thumb that aid greatly in the solution of such problems.

AIR DISTRIBUTION

In review, air movement within each room of a laboratory complex must be such that a definite flow pattern will be maintained throughout the building along with flow from non-contaminated to potentially contaminated areas. To bring about this differential flow pattern, the nature barriers between the various classes of rooms will assist. The pattern will also be assisted by supplying outside clean air to the non-contaminated and semi-contaminated areas and by exhausting air only from the moderately and extremely contaminated areas. In general, supply fans should take suction from the upper portions of the building. Also, the exhaust fans should discharge to the outdoors through stacks of varying heights depending on adjacent structures. To help, the building should be maintained at a slight positive pressure with respect to the outdoors. Laboratory rooms should be maintained at a negative pressure with respect to the surrounding rooms.

Only an adequate supply of makeup air to satisfy exhaust needs will keep the building in balance. This certainly implies there must be an excess of supply over exhaust needs. In actual installations, experience shows that when two fans are exhausting from the same space with no provision for makeup air, the stronger fan will take command and outside air will enter the room through the weaker fan system. When there are multiple exhaust hoods and no makeup air, with one hood off, outside air can downdraft through the idle fan. When a fan must exhaust from a room without makeup, fan capacity will be reduced from design and will result in less control at the hood.

EXHAUST SYSTEM

The exhaust system being under negative pressure will cause leakage flow to be drawn into the system and contamination will be confined. Best location for an exhaust fan serving hoods is on the roof. Then all exhaust ductwork will be on the suction side of the fan and indoors. But this is not always possible. If the fan location must be indoors, say just above the hood, then careful attention must be paid to duct tightness on the discharge side. When flammable material is handled, mounting fan on roof is a distinct advantage because explosion-proof construction may not be required of the fan motor. However, fan wheel should be non-ferrous and inside casing should be epoxy coated for corrosion protection.

EXHAUST DUCT MATERIALS

In many buildings ductwork is often concealed in ceilings or inside walls, making duct inspection and replacement a major problem. Where this condition exists it is reasonable to use ductwork with long life expectancy. For chemicals used in laboratories, galvanized iron and black iron ductwork are highly susceptible to corrosion. Stainless steel, transite, polyvinyl chloride-coated steel or fiberglass-reinforced plastic (FRP) ductwork will not require early replacement for such corrosive service but are costly. Actually, selection of materials will depend on the nature and concentration of contaminants or chemical reagents, space conditions, cost, accessibility. Whatever materials are selected, duct joints must be leaktight and the ductwork should have ample supports. For best service life all longitudinal duct seams should be run along the top panel. An extensive duct system should have inspection and cleaning facilities. Ducts that could develop condensation loading should pitch toward a pocket in the bottom of the run and be provided with a trapped drain.

Type 316 passive stainless steel may be used for bacteriological, radiological, perchloric acid and other general chemical purposes. 316 stainless steel is easy to work but is not suitable for chemical hoods handling concentrated hydrochloric and sulfuric acids.

EXHAUST AIR TREATMENT

Gases that are bubbled through reaction mixtures and then discharge to the hood are generally, by their nature, reactive enough to be completely eliminated by a scrubber of some design. For materials that are acidic, a simple caustic scrubber is all that is necessary to assure essentially complete control. Similarly, for materials of a basic nature, an acid scrubber may be used to advantage. For those materials that do no react rapidly with either caustic or acidic solutions, a column filled with activated charcoal will always provide the desired control.

Perchloric acid is highly soluble in water and hoods have been developed with packed sections built into the hood superstructure and provided with water wash rings in the ductwork downstream of the scrubber to prevent buildup of perchlorates which are explosive on contact. Fume hoods handling highly radioactive materials should have HEPA filters upstream and downstream of the hood. For highly hazardous bacteriological experiments safety can be achieved only by incineration of the exhaust air stream, which is heated to about 650°F to destroy the bacteria.

SPECIAL SYSTEMS

Lowered Sash Operation

A hood exhaust fan maintains proper capture velocity when the sash is wide open, but the exhaust's hood's vertically sliding sashes are sometimes lowered to within a few inches of the work surface when the hood is in operation. A method in use to reduce waste of conditioned air and also to achieve a more constant face velocity over the range of sash positions is the use of a 2-speed fan for each hood. When the sash is pushed up the fan runs at high speed. A micro-switch mounted in the hood is tripped by the sash when it is lowered below a predetermined position. The volume of air the fan will pull on low speed is adequate to maintain desired face velocity for the smaller cross-sectional area. The proper placement of the switch setting can be 50 to 60% of the vertical face opening, i.e., the fan would go on low speed when the sash is lowered to 50 to 60% of opening. This holds for all exhaust hoods despite differences in hood dimensions and other variations in exhaust systems. It has been found to apply equally as well to hoods with minimum face velocities of 80, 100, and 125 fpm. The volume of conditioned air that is normally lost is reduced by about one-third when the sash is below the set point.

In a conventional hood with a single speed fan, the excessively high face velocities experienced at low sash settings and the cooling effect on the backs of lab personnel using the hood has an overall adverse effect. Further, still, when the laboratory technician stands in front of a hood in operation his body presents an obstruction to the flow of air into the hood. Thus, a low pressure area develops in the space between the man and the hood. Under certain conditions, the resulting low pressure area can cause fumes to be aspirated from the hood and out into the room. The reduction in face velocity using the 2-speed fan reduces the probability of such a hazardous condition developing.

Type "S" hoods should be provided with fan speeds so that at no point across the hood face should a velocity greater than 250 fpm exist. Another way to control this velocity is to provide by-pass dampers in the exhaust duct just downstream of the hood itself. By-pass hoods are made to accomplish this effect by providing this feature in the hood structure itself.

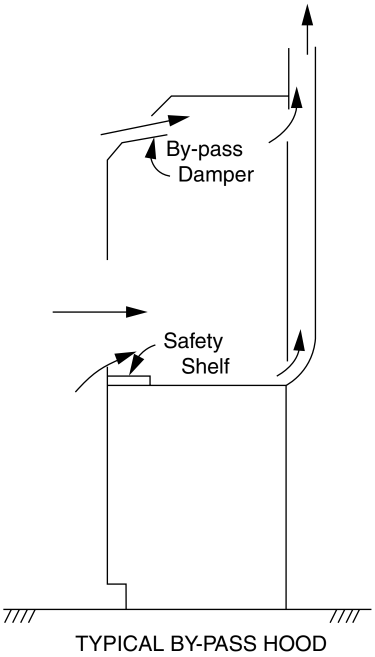
By-Pass Hoods

These provide for a constant rate of room exhaust and uniform face velocities at any door position. They stabilize the room exhaust and the room they supply. The by-pass may be an integral part of the hood itself. As the hood door begins to close, the damper starts to open. Another important aspect and advantage of the by-pass hood is that the hood interior is continuously being purged of fumes even while the door is closed tight. For the by-pass hood see Figure 3.

Supply Air Hoods

Two types are commercially available. The first has auxiliary air introduced outside and in front of the sash, normally from the overhead position. In this design the auxiliary air supply is drawn into the sash opening as a part of the room air. Relative cost of this type compared to the conventional is high. Relative cost of air conditioning is low because amount of room air exhausted is reduced. Air conditioning effectiveness, fume removal effectiveness, and convenience to lab personnel are good. However, acceptability to local authorities should be investigated. See Figure 4(a).

In the second type, auxiliary air is fed directly into the hood on the inside. Relative cost is high, cost of air conditioning is low, air conditioning effectiveness is good, but fume removal effectiveness is poor. Because effective face velocities can drop



TYPICAL BY-PASS HOOD

FIGURE 3

below the safe value needed to prevent leakage of fumes, its use is discouraged by many health authorities. See Figure 4(b).

Induction Venturi

For many fume exhaust applications such as those involving hazardous fumes or gases, the conventional exhaust method of passing gases through the fan casing could be potentially hazardous. With exhaust from perchloric acid fume hoods in particular, a build-up of crystals can occur on duct walls and fan. This crystalline growth is explosive under normal conditions and special treatment of such a system is mandatory.

To overcome this, there are commercially available induction venturi systems with water wash facilities. Since perchloric acid crystals are highly soluble this system is provided with spray rings or nozzles and are washed down internally at regular intervals. Drainage is provided to a trough attached to the back of the hood table (see Figure 5). System operation is accomplished by introducing a high velocity air stream jet inside a specially designed venturi. This in turn induces a flow of gas into the venturi inlet. This induced flow can then be used to exhaust the hood without any of the gas having to pass through the fan. Venturi is usually of stainless steel (316L). Blower is mild steel. Such a system used to exhaust 12000 cfm against ½″ w.g. required a primary flow of clean air of 500 cfm and ¾ hp fan motor.

Other perchloric acid fume exhaust systems use fans of PVC construction but wash the gas stream upstream of the fan. Its construction is also PVC. Each hood should be provided with its own exhaust system; no combinations should be manifolded. Organic compounds must be avoided in the construction of the system as well as the chemical used in testing inside the hood.

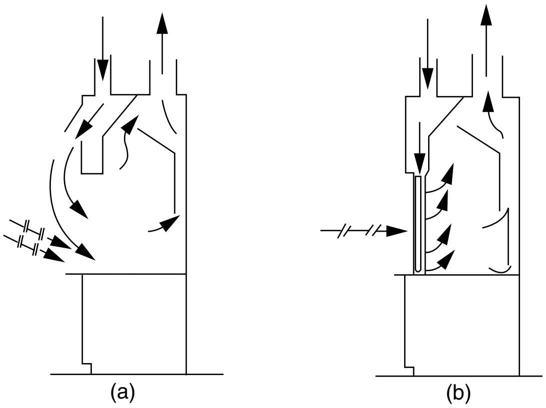
Multihood Single Fan System

Should each fume hood be provided with its own exhaust fan or should several hoods be serviced by one fan common to all? A common exhaust duct and fan system may be used if the facility handles similar and compatible chemical reagents. In the consideration of exhaust systems for a chemical research facility, where the chemical nature of the reagents to be used cannot be predicted in advance, or cannot be controlled, safest procedure is to use separate and individual exhaust fans and ducts.

DESIGN PROCEDURE

For a most economical design and the use of the various criteria outlined herein the following procedure is suggested:

1) Set inside conditions of dry bulb temperature and relative humidity in the upper range of the comfort zone. Since relative humidity is critical to operating costs, place greater emphasis on this aspect.
2) Select a hood face velocity sufficiently high to control the type hazard, using the recommendations outlined in reference 1. Review hood



(a)

(b)

FIGURE 4 Auxiliary air supply schematic.

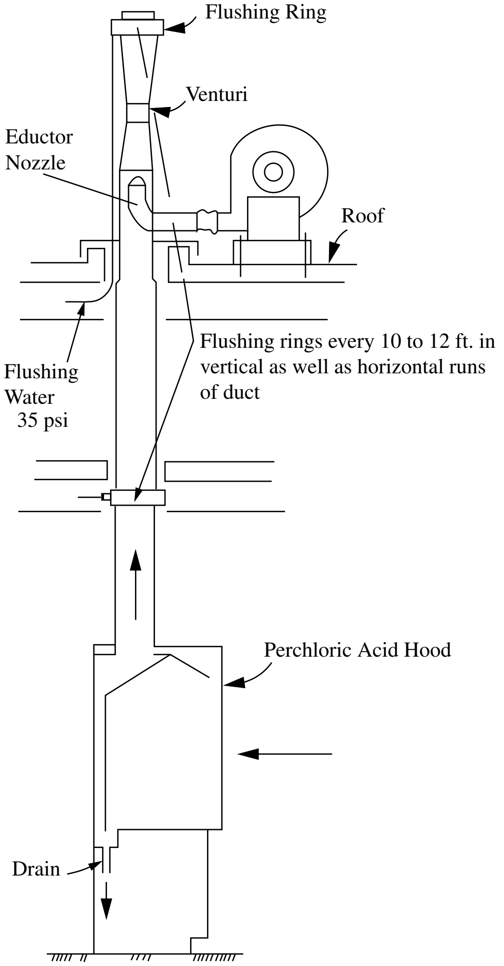


FIGURE 5 Induction venturi system.

operation carefully since not all hoods require same face velocities.

3) In cooperation with laboratory management determine the minimum number of hoods requiring continuous operation. Determine if a hood or hoods can operate intermittently or a minimum and estimate if its exhaust flow can be eliminated insofar as its effect on air conditioning load is concerned.
4) Avoid the use of hoods to store material and merely provide local exhaust.
5) Determine the acceptability of face screens or shields or horizontal sliding panels.
6) Locate hoods so that they are set clear of doorways and frequently traveled aisles.
7) Determine if laboratory management is willing to take a "slip" in room conditions when more air is exhausted than is originally planned.
8) Consider use of perforated ceiling supply hood arrangement with conditioned air supply through ceiling diffusers for spot cooling effect.

General

Exhaust stack should be vertical and straight and discharge up; no weather caps should be used. Brief[1] suggests when open-face velocities exceed 125 fpm, install an atmospheric damper downstream of the hood just before the exhauster to prevent excessive indraft velocities when conventional hoods are used. At high face velocities, laboratory equipment placed within the hood should be set so that points of release of contaminant are at least 6 inches back of the hood face. This can be ensured by placing a ¼ inch thick edging 6 inches wide on the bench top near the hood entrance face. Brief[1] found that concentrated head loads within the hood proper, exceeding 1000 watts per foot of hood width created thermal vectors that require higher face velocities for control. Obstruction of hood face by large objects is discouraged; blockage causes control problems.

PART 2

Factors to Be Considered in Fume Hood Selection

In the selection of a fume hood the following factors should always be considered:

Space

- What actual space requirements will be required?
- What are the future requirements?
- What physical space is available?

Function

- What chemicals and procedures will be involved in this application? (Highly corrosive, TLV, etc.)
- High heat procedures?
- Extremely volatile?

Location

- Are your present ventilation capabilities adequate and will they be taxed by the new hood installation?
- Is the area where the hood will be installed adequately suited to the new installation? For instance, high traffic areas give rise to undesirable crosscurrents and cause materials to be drawn from hoods. Hoods should not be installed next to doors but preferably in corners.
- Is the operation such that the use of an auxiliary air system might compromise the safety of the operator? Safety is paramount in any hood application.

Hood Construction Materials

Although basic hood design has changed very little, many advances have been made in the materials from which hoods are constructed. Here are some of the basic materials and their more distinctive features.

Wood

- Generally poor chemical resistance.
- Inexpensive to fabricate and modify in the field.
- Can present a fire hazard in applications involving heat and flame.
- Poor light reflectivity causes a dark hood interior.

Sheet Metal (Cold rolled steel or aluminum)

- Requires secondary treatment for chemical resistance.
- Demands extreme care to avoid damaging the coating since corrosion can occur in damaged areas.
- "Oil canning" due to thin-gauge metal causes noise in operation.
- Relatively inexpensive.
- Usually heavy and cumbersome to install.

Fiberglass

- Excellent chemical resistance.
- Lightweight for ease of installation or relocation.
- Easily modified in field with readily available tools.
- Sound-dampening because of physical construction.
- Some inexpensive grades can cause fire hazards and are not chemically resistant.
- Available with good light reflective properties for a light and bright work space.
- Shapes are limited to tooled mold configurations, and can be moulded with covered interiors.

Cement/Asbestos (Transite)

- Excellent chemical resistance.
- Has inherent sound dampening qualities.
- Excellent fire resistance.
- Heavy and difficult to install.

- Extremely brittle, requiring care in handling to avoid breakage.
- Poor light reflectivity.
- Stains badly when exposed to many acids, etc.
- Easily modified in field with only minor tooling difficulties.
- Inexpensive.

Stainless Steel

- Better general chemical resistance than cold rolled steel.
- Not well suited to many acid applications.
- Generally provided in type 316 for specific applications to which it is well suited such as perchloric acid.
- Heavy and expensive.
- Difficult to modify in field.
- Excellent fire resistance.

Polyvinyl Chloride

- Excellent chemical resistance except for some solvents.
- Good fire-retardant properties.
- Particularly well suited to acid digestion applications such as sulfuric and hydrofluoric.
- Easily modified in field.
- Generally not available in molded configurations.
- Expensive.
- Distorts when exposed to intense direct heat.

Stone

- Excellent chemical resistance.
- Excellent fire resistance.
- Difficult and extremely heavy to install.
- Extremely difficult to field modify.
- Expensive.

WALK-IN HOOD

This type of hood was not mentioned in Part I but will be now included. The walk-in hood is a standard hood whose walls extend to the floor, thus providing sufficient space to accommodate a more elaborate experimental setup requiring additional height. Such hoods have double or triple hung sashes, which may be raised and lowered to provide access to any part of the setup while the remaining space is enclosed to contain fumes. The back baffle of such a hood extends over the full height of the hood and is equipped with at least three adjustable slots to regulate the amount of air passing over various parts of the setup.

SPECIAL PURPOSE FUME HOODS

Perchloric Acid Fume Hood

Due to the potential explosion hazard of perchloric acid in contact with organic materials, this type hood must be used for perchloric digestion. It must be constructed of relatively inert materials such as type 316 stainless steel, Alberene stone, or ceramic coated material. Wash-down features are desirable since the hood and duct system must be thoroughly rinsed after each use to prevent the accumulation of explosive residue. Air flow monitoring systems are recommended to assure 150 fpm open face velocity operation. An additional monitoring system for the wash-down facilities is also recommended.

Radiological Fume Hoods

Hoods used for radioactive applications should have integral bottoms and covered interiors to facilitate decontamination. These units should also be strong enough to support lead shielding bricks in case they are required. They should also be constructed to facilitate the use of HEPA filters.

Canopy Fume Hoods

Canopy fume hoods are a type of local exhauster which normally has limited application in a laboratory. Their main disadvantage is the large amount of air required to provide an effective capture velocity. Since the contaminant is drawn across the operator's breathing zone, toxic materials can be quite dangerous. A canopy hood can, however provide a local exhaust for heat or steam.

INTEGRAL MOTOR-BLOWERS

Many hoods are available with motors and blowers built directly into the hood superstructure. From the standpoint of convenience, the hood is relatively portable and can be installed easily. A built-in motor-blower should not be used for highly toxic applications since it causes a positive pressure in the exhaust system ductwork and any leaks in the duct could spill the effluent into the lab area. There may be more noise associated with this type hood since the motor-blower is closer to the operator.

Fume Discharge

Each individual exhaust fan on the roof should have its own discharge duct to convey the fumes vertically upward at a high velocity as far above the topmost adjacent roof as possible. Failure in this will result in potential recirculation of fumes into building air intakes and will be particularly hazardous to personnel who use the roof for maintenance, research, or relaxation.

As the wind blows over the leading edge of a roof parapet, as shown in Figure 6, a disturbance is created that sweeps from the edge of the parapet up over the top of the building. Above the boundary of this disturbance, wind flow is undisturbed. Below the boundary, the influence of the sharp edge of the building creates eddy currents that can pocket fumes released at the roof. This is known as the wake cavity. Unless fumes are discharged into the undisturbed air stream above the boundary, where they can be carried away, they will remain relatively undisturbed and undiluted on the roof and in the lee of the building, where they can enter the building air intakes either on the roof or at ground level. When this happens, all the care taken in the design of a good fume exhaust system

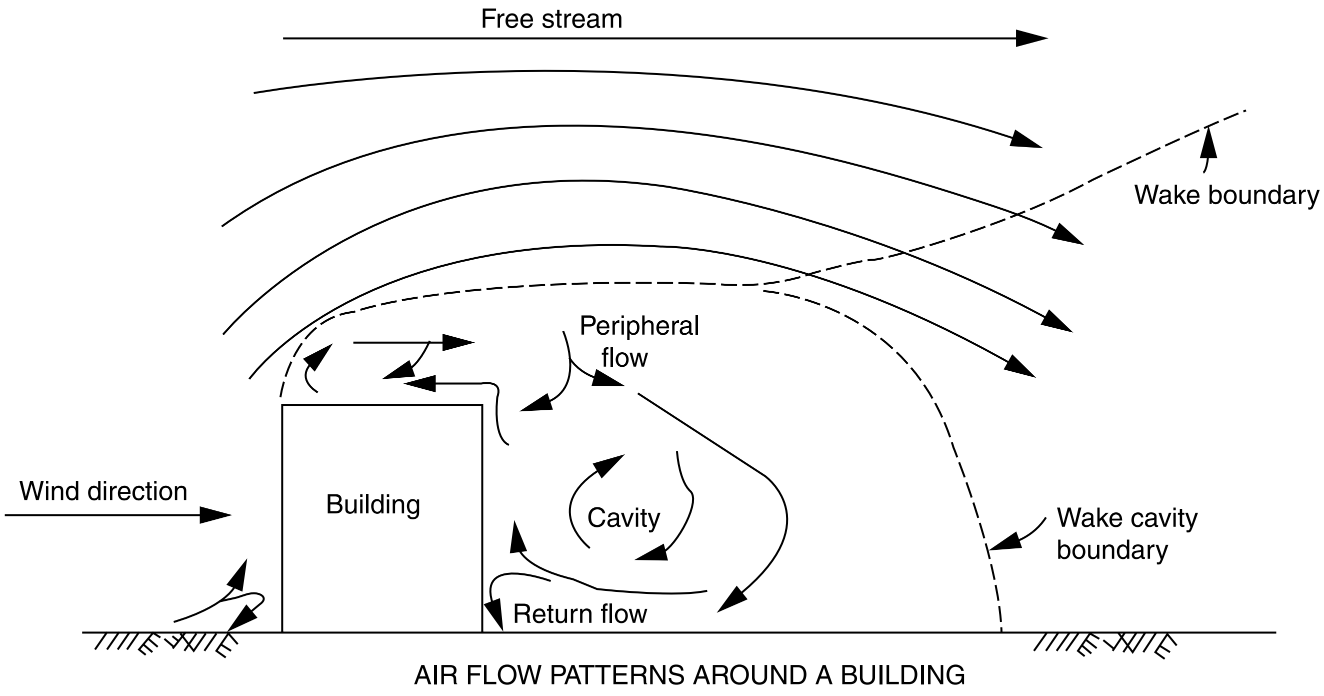


FIGURE 6

may be nullified. And with the present concern over air pollution, failure to disperse the fumes may give rise to legal action against the building owner.

Fume absorbers such as charcoal have been proposed to relieve the fume disposal problem, so have air washers and catalysts. These devices have not been used because the kinds and amounts of fumes released are constantly changing in research and are therefore unpredictable.

Despite the number of warnings in the literature, rain caps, cone shaped covers or hoods fastened to the tops of vertical stacks—are still being used to prevent rain from entering exhaust stacks. It is important that their use be avoided completely. There are several simple stack arrangements that will prevent entry of rain into exhaust stacks when fans are not operating. One such arrangement is shown in Figure 7.

BUILDING AIR INTAKES

In high-rise research buildings, mechanical equipment is frequently installed in the penthouse and in the basement. Because of the possibility of recirculating fumes released from or near the roof, outdoor air is often taken at the second floor level on the prevailing wind side of the building, and away from fume exhausts. Assistance in determining the prevailing wind direction at the building site may be obtained from the local weather bureau.

BASIC PERFORMANCE CRITERIA

The following may be used as a general guide for the selection of hood blower systems that will provide optimum average face velocities for various exhaust materials. Tables listing the TLV for various chemical compounds may be obtained from the American Conference of Governmental Industrial Hygienists.

Very low toxicity level materials	
Noxious odours, nuisance dusts and fumes	80 fpm
General lab use	
Corrosive materials	
Moderate toxicity level materials (TLV of 10–1000 ppm)	100 fpm
Tracer quantities of radioisotopes	
Higher toxicity level materials (TLV less than 10 ppm)	125–150 fpm
Pathogenic microorganisms	
High alpha or beta emitters	
Very high toxicity level materials	(TLV less than 0.01 ppm) An enclosed glove box should be used

HOW TO CUT AIR CONDITIONING COSTS

As a rule of thumb, each 300 cfm of air exhausted through hoods requires one ton of refrigeration. Current operating costs are about 50 to 60 dollars per ton of air conditioning for a four month period. Installed equipment averages about $1,000 per ton. So, a hood exhausting at 900 cfm would require about three tons of air conditioning at a capital expense of $150 to $180 per season. However, if the same hood had

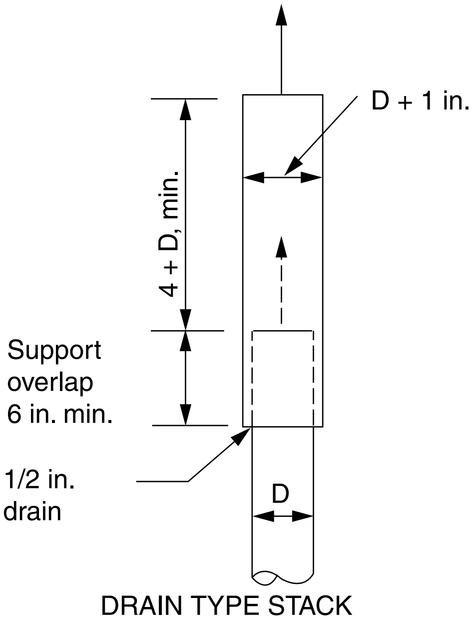


FIGURE 7

the Add-Air feature supplying 50% untempered air, $1,500 would be saved in capital equipment and $75 to $90 to annual operating costs. (Figures are in 1992 $.)

MAINTENANCE AND TESTING

Since the hood performance may be affected by the cleanliness of the exhaust system and the direction of rotation of the exhaust fan, it is important to provide a maintenance schedule of inspections and performance testing throughout the year to make certain that the fume hoods are operating safely and efficiently.

If filters are used to remove dust and other particulates from the exhaust air, they must be periodically inspected and replaced if necessary. Corrosion of ductwork and damper mechanisms should be watched and debris should be removed from inside the ducts, especially at startup time. Excessive corrosion of ducts may cause leakage of air into the system or the failure of balancing dampers that will affect capture velocities well below their design figures. Remember to check fan rotation since this most often causes poor exhaust performance.

PERFORMANCE TESTING

Two performance tests should be conducted periodically on all hoods. One for fume leakage and the other for face velocity. The test for fume leakage consists of releasing odorous fumes such as ammonia or hydrogen sulfide within the hood. If fumes are detected outside the hood, especially around the face opening, the capture velocity at the sash opening may be inadequate, or there may be an interfering air disturbance. Cleaning the exhaust system, adjusting the air flow damper, or increasing the fan speed may improve the performance if low face velocity seems to be the problem. If, on the other hand, leakage seems to be caused by interference from an auxiliary air supply stream or other velocity near the sash, the nature of the interference may be investigated as follows: placing liquid titanium tetrachloride on masking tape around the periphery of the sash opening. Observations can then be made of the path of visible fumes to determine where there is spillage into the room. Smoke bombs have also been used to determine flow patterns at sash openings and to identify interference.

A hot wire anemometer is usually used to measure actual face velocity. This is done as a traverse over the entire sash opening, including especially all edges and corners. The overall face velocity average is obtained by averaging the velocity readings at prescribed positions of the traverse.

These testing procedures are difficult to standardize and are dependent on subjective observations. Thus, they are considered to be unadaptable and inadequate. The *American Society of Heating, Refrigeration, and Air Conditioning Engineers* (ASHRAE) has set up a research project for developing fume hood performance criteria and new test procedures for such laboratory equipment.

SAFETY FEATURES

Interconnection of Hoods

If two or more hoods independently serve a single room or an interconnecting suite of rooms, all of the hoods in these rooms should be interconnected so that the operation of one will require the operation of all. If this is not done, there is a strong possibility that fumes will be drawn from a hood that is not operating to makeup air demands of those in operation.

Alarm for Hood Malfunction

All hoods should be equipped with safety devices such as sail switches to warn personnel that the air volume exhausted from the hood has dropped to a point where it will not provide sufficient capture velocity for safe operation.

Fire Dampers

Most building codes require fire dampers in all ducts that pass through fire walls and floors. However, it is important *not* to install them in fume exhaust systems. Should a fire occur in a hood, or if heat from a fire nearby such a damper should cause the damper to close, the fume backup into the facility would prove disastrous.

EXHAUST FROM LABORATORIES

A laboratory should exhaust 100% of the air fed to it. If the materials that are being handled or tested in the laboratory are hazardous enough to need a hood, the presence of these materials in itself should dictate 100% exhaust. An accidental spill or accidental release of materials at a bench or hood can result in recirculation throughout the entire building. Accidental recirculation is a serious hazard and should be guarded against.

GLOSSARY OF TERMS RELATED TO FUME HOOD SELECTION

Baffle: An air director mounted off the hood's inner surface which causes air to move in specific patterns.

Blower: An air moving device utilizing a rotating impeller within a housing to exhaust air.

BTU: (British Thermal Unit) The amount of heat required to raise one pound of water one degree Fahrenheit.

Capture Velocity: Air velocity at the hood opening necessary to overcome opposing air currents and cause contaminants to flow into the hood.

CFM: (Cubic Feet per Minute) A volume of air moved per minute.

Duct: A pipe system used to convey and constrain a moving air stream.

Ejector: An air moving system which consists of a high pressure air source passing through a Venturi nozzle, creating a suction at the nozzle entry.

Face Velocity: The speed of air measured in feet per minute across the fume hood sash opening perpendicular to the sash.

HEPA Filter: A High Efficiency Particulate Air Filter rated 99.97% effective on particles 0.3 micron or larger.

Inches of Water: A unit of pressure equal to the pressure exerted by a column of water one-inch high at standard temperature.

Manometer: An instrument for measuring pressure. It is essentially a U-tube filled with a liquid, normally water or mercury.

Negative Pressure: Pressure within a system below that of atmosphere, causing an inward flow of air.

Plenum: An air compartment maintained under pressure which serves as a reservoir for a distribution duct.

Positive Pressure: Pressure within a system above that of atmosphere, causing an outward flow of air.

Scrubber: A device used to wash effluent air streams for removing contaminants.

Static Pressure: The pressure exerted in all directions when air moves through a duct system creating a resistance to air flow. Measured in inches of water.

TLV: (Threshold Limit Value) The amount of air-borne toxic materials that represents the maximum concentration to which an average person may be exposed for 8-hours a day with no adverse effects. (Usually expressed in parts per million.)

Transport Velocity: The minimum velocity required to move particulates in the air stream.

LABORATORY SAFETY GUIDELINES

Safety rules should be followed at all times. A shortcut may save you a few minutes, but cost you your life. Here are a few guidelines that may make your laboratory safer.

Hoods: Chemical hoods effectively remove toxic and flammable vapors. Equipment in use should be completely enclosed in a hood with adequate room allowed for experimental procedures. When the apparatus is too large to be housed in a hood and there is no possibility of toxic or flammable materials being released, anchored shields of safety or wired glass should envelope the equipment. Hoods are not designed to be used as storage areas. Remove unused equipment and chemicals and store them in their proper places.

Emergency Equipment and Procedures: Well-equipped chemical laboratories have eye-wash fountains, deluge safety showers, fire blankets, fire extinguishers, and emergency exits. This equipment should be tested periodically. In addition, being familiar with the locations and uses of the equipment may save you needed time during an emergency.

Personal Protection: Rubber aprons, asbestos gloves, safety glasses, full face shields, and approved respirators protect you from spills, burns, spattering chemicals, flying fragments, and irritating fumes. In addition, the laminated safety glass doors on chemical fume hoods protect you from mishaps in the hood.

Health Monitoring: When biological agents or carcinogens are used in the laboratory, special medical control programs are necessary to monitor the workers' health. If radioactive materials or radiation-producing equipment such as an X-ray diffraction unit are used, dosimeters or film badges should be worn to monitor exposures.

Labeling: Chemicals must be prominently and accurately labelled. When a small quantity of material is removed from a large container, immediately label the smaller container. Containers for hazardous chemicals should have precautions such as "poison" or "flammable"; indicated under the label. After you have completed working with a particular material, return the container to storage or dispose of the material. Nothing should be left in open containers.

Equipment should be labeled U.L. or C.S.A. listed and meet all federal and local electrical codes.

Eating, Drinking, and Smoking: Food, beverages, and cigarettes, pipes, or cigars should not be permitted in the chemical laboratory under any circumstances. Chemical glassware should never be used to hold food. Always wash your hands well before eating, drinking, or smoking.

Pipetting: Never pipette toxic, corrosive, or radioactive chemicals by mouth; always use a rubber bulb or syringe.

Glassware: Cracked or chipped glassware should be discarded to prevent cuts or scratches which can cause further complications if chemicals contact the injury.

Always place a towel or cloth over glass tubing being cut or broken. Fire polish sharp ends. When inserting a rod or piece of glass tubing through a perforated stopper, wrap a towel around your hand for protection.

Waste Disposal: Disposal of hazardous waste materials requires special handling.

Place all broken glass in specially marked metal containers—never in waste baskets or containers used for paper or rags.

Flush dilute acids and alkalies down the drain with large quantities of water.

Never pour flammable liquids not miscible with water, compounds that give off toxic vapors, or corrosive materials down the drain. Special disposal containers are needed for each of these wastes.

Storage: An efficiently placed storage room is essential for everyone's safety. Chemical storage rooms should be equipped with fire doors, safety lights, fire extinguishers, as well as good ventilation and sprinkler systems.

Remember

Carefully group liquid reagents to prevent hazardous combinations which may produce fumes, fire, or explosion.

Segregate incompatible materials.

Keep volatile liquids away from ignition sources such as heat, flames, or electric sparks.

Store large or heavy containers and apparatus near the floor.

Store all solvents in safety cans.

Store and frequently vent drummed chemicals according to supplier's instructions.

Secure compressed gas cylinders.

Replace valve caps when not in use.

Smaller laboratories that don't have separate storage rooms should have noncombustible storage cabinets. Large quantities of flammable solvents should be placed outside in ventilated, noncombustible buildings.

Housekeeping: Good housekeeping is essential for safe laboratory operation. All passages, exits, safety showers, fire extinguishers, electrical controls, and stairways must be kept clear of equipment and obstructions. Remove unused equipment or chemicals from work spaces. Clean up spilled chemicals immediately to prevent dangerous chemical combinations, burns, or slips and falls.

FUTURE TRENDS

We may expect stricter enforcement of existing local and Federal regulations for the safe handling of toxic materials. New regulations may dictate high fume hood face velocities and increased exhaust volumes which place an increased load on air tempering and exhaust systems. Because of this, work involving small apparatus will probably be relegated to glove box enclosures. Larger hood enclosures will more than likely be fitted with horizontal sliding doors or sashes rather than the current vertical rising sash. This type of hood would provide full access to the larger hood interior, but would require opening only the areas needed by the operator. This would result in lower exhaust volumes even at the increased face velocities and provide an integral safety shield if required.

Reported measurements of airborne bacteria and fungi have been sparse (see Institute of Medicine, 2004, for a summary). Viable bacteria concentrations, found in homes in the U.S.A. ranged from 2220–4006 CFU/m^3 (i.e., colony forming units per cubic meter of air). In Finland, homes and day care centers with moisture problems and winter conditions had unit concentrations as high as 35,000 CFU/m^3 (Hyvarinen et al., 2001).

REFERENCES

1. Brief, Richard S. *et al., Design and Selection of Laboratory Hoods,* Air Engineering, Oct. 1963.
2. Halitsky, James, *Gas Diffusion Near Building,* ASHRAE Transactions, **69**, pp. 464–485, 1963.
3. Clarke, John H., *Air Flow Around Buildings,* Heating, Piping and Air Conditioning, May 1967.
4. Schulte, H.F. *et al., Evaluation of Laboratory Fume Hoods,* American Industrial Hygiene Association Quarterly, Sept. 1954.
5. *Industrial Ventilation,* A Manual of Recommended Practice, American Conference of Governmental and Industrial Hygienists, 11th Ed., 1970.
6. Sax, Irving, N., *Dangerous Properties of Industrial Materials,* 9th Ed., 1996, Reinhold, New York.
7. Young, J.A., *Improving Safety in the Chemical Laboratory,* 1987, Wiley-Interscience, New York.
8. Di Berardinis *et al., Guidelines for Laboratory Design,* 1987, Wiley-Interscience, New York.
9. Dux, James P. and Stalzer, Robert F., Managing Safety in the Chemical Laboratory (1989), Van Nostrand-Reinhold, New York.
10. Slote, Lawrence, Editor, *Handbook of Occupational Safety and Health,* 1987, Wiley-Interscience, New York.
11. Pipitone, David A., *Safe Storage of Laboratory Chemicals,* 1984, Wiley-Interscience, New York.
12. National Fire Protection Association, *Fire Protection Guide on Hazardous Materials,* 9th ed., 1986.
13. Institute of Medicine, Committee on *Damp Indoor Spaces and Health,* Publ. National Academies Press, Wash. D.C., 2004.
14. Hyvarinen A., Reponen T., Husman T., Nevalainen A., Central European Jnl. of Public Health, 9(3): 133–139 (2001).

JOHN D. CONSTANCE, P.E. (DECEASED)
Cliffside Park, N.J.

INCINERATION: see MANAGEMENT OF SOLID WASTE

INDUSTRIAL ECOLOGY

INTRODUCTION

Industrial ecology is an emerging field of study that deals with sustainability. The essence of industrial ecology was defined in the first textbook of the field in this way:

> Industrial ecology is the means by which humanity can deliberately and rationally approach and maintain sustainability, given continued economic, cultural, and technological evolution. The concept requires that an industrial system be viewed not in isolation from its surrounding systems, but in concert with them. It is a systems view in which one seeks to optimize the total materials cycle from virgin material, to finished material, to component, to product, to obsolete product, and to ultimate disposal. Factors to be optimized include resources, energy, and capital. (Graedel and Allenby, 2003, 18)

Industrial ecology is industrial and technological in the sense that it focuses on industrial processes and related issues, including the supply and use of materials and energy, adoption of technologies, and study of technological environmental impacts. Although social, cultural, political, and psychological topics arise in an industrial-ecology context, they are often regarded as ancillary fields, not central to industrial ecology itself (Allenby, 1999).

Industrial ecology's emphasis on industries and technologies can be explained with the "master equation" of industrial ecology. Originating from the IPAT equation (impact, population, affluence, and technology; Ehrlich and Holdren, 1971; Commoner, 1972), the master equation expresses the relationship between technology, humanity, and the environment in the following form:

$$\begin{aligned} \textit{Environmental impact} &= \textit{Population} \times \frac{\textit{GDP}}{\textit{Person}} \\ &\quad \times \frac{\textit{Environmental impact}}{\textit{Unit of GDP}} \end{aligned} \tag{1}$$

where GDP is a countrys or region's gross domestic product, the measure of industrial and economic activity (Graedel and Allenby, 2003, pp. 5–7; Chertow, 2000a).

In this equation, the *population* term, a social and demographic one, has shown a rapid increase in the past several decades, and continues to increase. The second term, *per-capita GDP,* is an economic indicator of the present population's wealth and living standards. Its general trend is rising as well, although there are wide variations among countries and over time. These trends make it clear that the only hope of maintaining environmental interactions in the next few decades at an acceptable level is to reduce the third term, *environmental impacts per unit of GDP,* to a greater degree than is the product of the increases in the first two terms—a substantial challenge! This third term is mainly technological and is a central focus of industrial ecology.

The name "industrial ecology," combining two normally divergent words, relates to a radical hypothesis—the "biological analogy." This vision holds that an industrial system is a part of the natural system and may ideally mimic it. Because biological ecology is defined as the study of the distribution and abundance of living organisms and the interactions between those organisms and their environment, industrial ecology may be regarded as the study of metabolisms of technological organisms, their use of resources, their potential environmental impacts, and their interactions with the natural world.

The typology of ecosystems has been characterized as three patterns (Figure 1a–c). A Type I system is a linear and open system that relies totally on external energy and materials. In biology, this mode of action is represented by Earth's earliest life forms. A Type II system is quasi-cyclic, with much greater efficiency than Type I. However, it is not sustainable on a planetary scale, because resource flows retain a partially linear character. Only a Type III system possesses a real cyclic pattern, with optimum resource loops and external reliance only on solar energy. This is how the natural biosphere behaves from a very long-term perspective.

The evolutionary path from Type I to Type III taken by nature (from open to cyclic, from unsustainable) provides perspective on the evolution of industrial ecosystems. Historically, the industrial system has mimicked the Type I pattern, with little concern about resource constraints. The best of today's industries come close to Type II (Figure 1d), and a Type III industrial system is a vision of a possible sustainable future for industrial ecosystems.

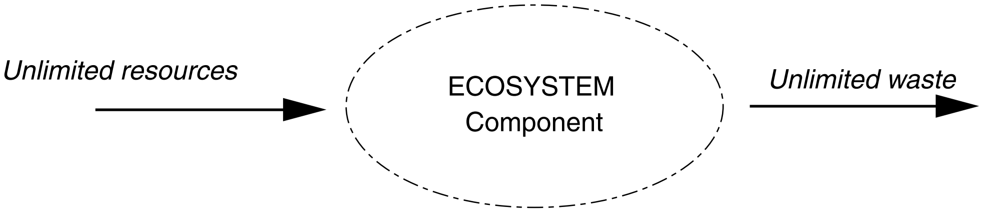
The biological analogy has been explored in other ways as well. From a metaphysical perspective, industrial ecology's philosophy might be labeled as: "nature as model," "learning from nature," and "orientation by nature" (Isenmann, 2002).

In this context, industrial firm-to-firm interactions have been examined by ecological food-web theory (Hardy and Graedel, 2002), and the theoretical approaches of thermodynamics and self-organization have also been applied to these systems (Ayres, 1988).

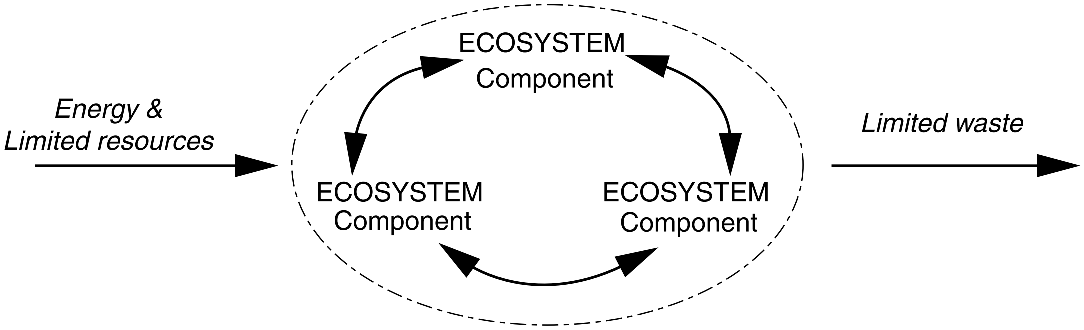
The interaction between the worlds of industry and ecology emphasizes that industrial ecology is a systems science that places emphasis on the interactions among the components of the systems being studied. This systems orientation is manifested in several of the research topics of the field, including life-cycle analysis, industrial metabolism, system models and scenarios, and sustainability assessment (Lifset and Graedel, 2002), topics that are discussed below.

THE ORIGINS OF INDUSTRIAL ECOLOGY

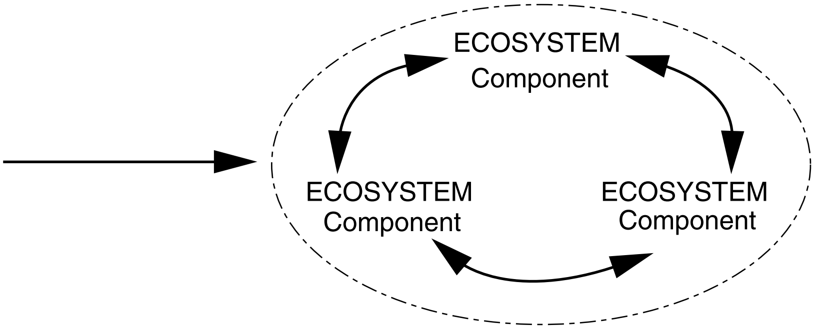
For many thousands of years, nature dominated the human-nature relationship. This dominance was reversed by the growth of agriculture and especially by the industrial revolution of the 1800s. The implications for nature of this



(a) Type I: linear material flows

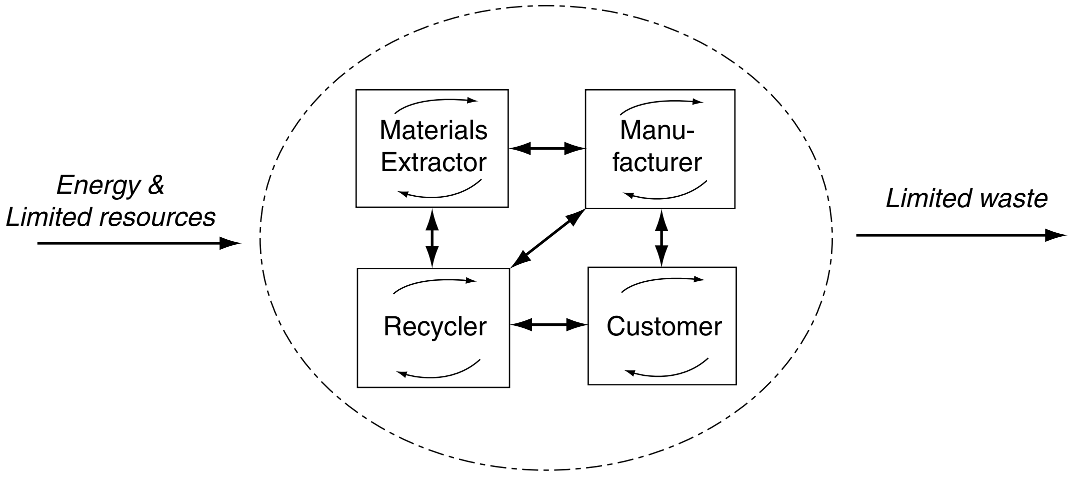


(b) Type II: quasi-cyclic material flows



(c) Type III: cyclic material flows

FIGURE 1 Typology of ecosystems. From Graedel and Allenby, 2003; Lifset and Graedel, 2002. (a) Type I: linear material flows; (b) Type II: quasi-cyclic material flows; (c) Type III: cyclic material flows; (d) Type II industrial ecosystem.



(d) Type II industrial ecosystem

FIGURE 1 (*continued*)

transformation were called out in the third quarter of the twentieth century by several seminal environmental thinkers (Carson, 1962; Lovelock, 1988; Ward *et al.*, 1972). The publication of the Club of Rome's report *The Limits to Growth* also received considerable public attention (Meadows *et al.*, 1972). That report predicted that economic growth could not continue indefinitely because of Earth's limited availability of natural resources, as well as its limited capacity to assimilate pollution of various types. Most of the Club of Rome's dire projections about resource exhaustion have not thus far come to pass. Nonetheless, the issue of the sustainability of human civilization has become a concern of global scope and reach.

The concept of industrial ecology, in which the technology–environmental linkage is explicitly recognized and addressed, can be traced to the early 1920s (Erkman, 1997, 2002). However, 1989 is generally viewed as the formal year of birth of the field (Figure 2). In that year, R. Frosch, then vice president of the General Motors Research Laboratories, and his colleague N. Gallopoulos developed the concept of industrial ecosystems in their seminal article "Strategies for Manufacturing" (Frosch and Gallopoulos, 1989). Their view was that an ideal industrial system would function in a way analogous to its biological counterparts. In such an industrial ecosystem, the waste produced by one process would be used as a resource for another process. No waste would therefore be emitted from the system, and the negative impacts to the natural environment would be minimized or eliminated. This analogy between biological and industrial systems was the conceptual contribution that led ultimately to the new field of industrial ecology.

Industrial ecology's growth since the early 1990s has been marked by a series of institutional milestones, including the first textbook (*Industrial Ecology;* Graedel and Allenby, 1995), the first university degree program (created by the Norwegian University of Science and Technology [NTNU] in 1996), T. E. Graedel's appointment as the first professor of industrial ecology in 1997, the birth of the *Journal of Industrial Ecology* in 1997, and the foundation of the International Society for Industrial Ecology (ISIE) in 2001. As a consequence of these activities, an academic community of industrial ecologists has been formed, research methodologies are being developed and refined, and industrial ecology is being practiced all over the world.

INDUSTRIAL ECOLOGY'S TOOLBOX

Given an evolving field with a wide and evolving scope, industrial ecology's toolbox has become equipped with a variety of methods of approaching the concepts and practices of interest. Three of the most common tools, material-flow analysis (MFA), life-cycle assessment (LCA), and input-output analysis (IOA), are discussed below from a methodological point of view.

Material-Flow Analysis

MFA is "the systematic assessment of the flows and stocks of materials within a system defined in space and time. It connects the sources, the pathways, and the intermediate and final sinks of a material" (Brunner and Rechberger, 2004, p. 3), thus providing information on the systemic utilization of the material within the given boundaries.

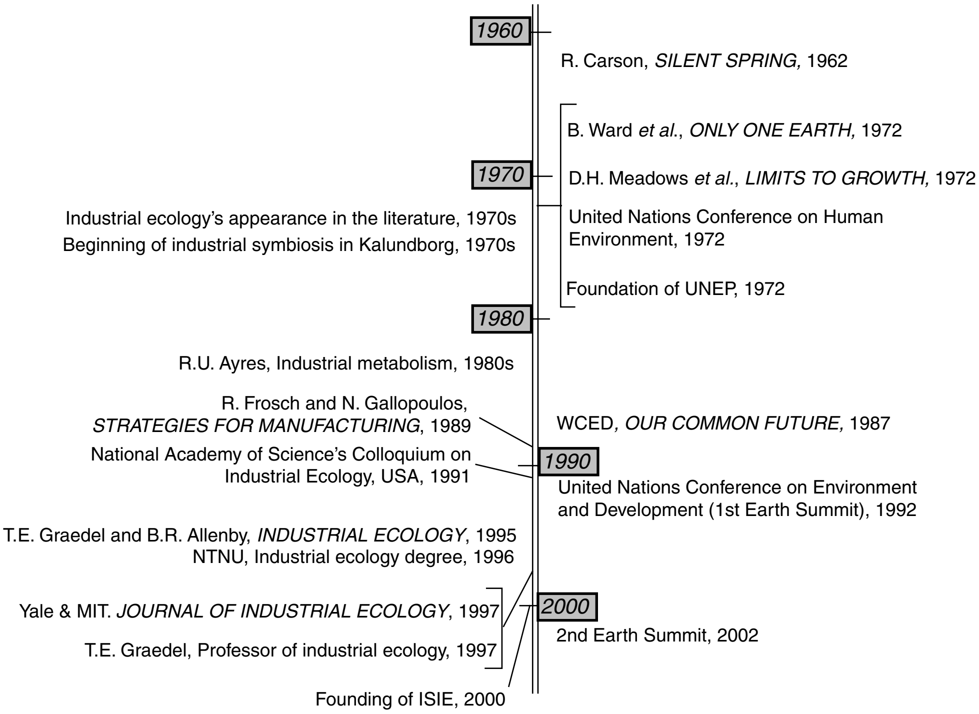


FIGURE 2 Industrial ecology and sustainable development: time line of events.

The principal terminology used in MFA studies is as follows (Graedel and Allenby, 2003, pp. 284–289; Brunner and Rechberger, 2004, pp. 34–40):

Substance: Any (chemical) element or compound composed of uniform units
Material: Substances and combinations thereof, both uniform and nonuniform
Goods: Entities of matter with a positive or negative economic value, comprised of one or more substances
Process: The operation of transforming or transporting materials
Flux: The rate at which an entity enters or leaves a process
Budget: An accounting of the receipts, disbursements, and reserves of a substance or material
Cycle: A system of connected processes that transfer and conserve substances or materials

The central principle upon which MFA is based is that of mass balance, which states that the mass of all inputs into a process equals the sum of the mass of all outputs and any mass accumulation (or depletion) that occurs within. This renders the results of MFA useful for studies of resource availability, recycling potential, environmental loss, energy analysis, and policy studies. MFA may be performed on a local scale and from a technical engineering perspective (as in Type A in Table 1), or, on a broader scale, associated with a geopolitical or socioeconomic dimension (as in Type B in Table 1; Bringezu and Moriguchi, 2002). In each case there is the potential for achieving a better understanding of the materials aspects of the process or entity under study, as well as identifying opportunities for achieving improvements.

Life-Cycle Assessment

LCA is a tool broadly used by industrial ecologists to identify and quantify the environmental impacts associated with a product, progress, service, or system across its “cradle-to-grave” life stages. Unlike the more targeted examination of a product or process in order to understand and quantify its direct environmental impacts, the use of a life-cycle perspective enables one to examine the direct and indirect environmental effects of an object through the stages of extraction of raw materials; various manufacturing, fabrication, and transportation steps; use; and disposal or recycling.

LCA began in the United States in 1969, in an effort to compare several types of beverage containers and determine which of them produced the lesser effect on natural resources and the environment (Levy, 1994; U.S. EPA, 2004). Since the 1990s, the Society for Environmental Toxicology and Chemistry in North America and Europe and the U.S. Environmental Protection Agency (EPA) have worked to promote consensus on a framework for conducting life-cycle inventory analysis and impact assessment. In 1993, the International Organization for Standardization

TABLE 1
Types of material-flow-related analysis

Type of analysis	A		
	a	b	c
Objects of primary interests	Specific environmental problems related to certain impacts per unit flow of:		
	substances	materials	products
	e.g., Cd, Cl, Pb, Zn, Hg, N, P, C, CO_2, CFC	e.g., wooden products, energy carriers, excavation, biomass, plastics	e.g., diapers, batteries, cars
	Within certain firms, sectors, regions		
	B		
	a	b	c
	Problems of environmental concern related to the throughput of:		
	firms	sectors	regions
	e.g., single plants, medium and large companies	e.g., production sectors, chemical industry, construction	e.g., total or main throughput, mass flow balance, total material requirement
	associated with substances, materials, products		

Source: Bringezu and Moriguchi, 2002. (With permission)

(ISO) included LCA in its ISO 14000 environmental certification process. As a result of these efforts, an overall LCA framework and a well-defined inventory methodology have been created.

LCA consists of three phases (Udo de Haes, 2002):

Goal and scope definition: A phase to set the purposes and boundaries of a study, such as geographic scope, impact categories, chemicals of concern, and data-availability issues

Life-cycle inventory analysis: The most objective and time-consuming process, in which the energy, water, and natural resources used to extract, produce, and distribute the product, and the resulting air emissions, water effluents, and solid wastes, are quantified

Life-cycle impact assessment: An evaluation of the ecological, human-health, and other effects of the environmental loadings identified in the inventory

These three phases are usually being followed by an *interpretation* phase in which the results from the above processes are tracked and possibilities for improvement are discussed.

Data availability and uncertainty are continuing concerns of LCA, as are the time and expense required. As a result, there have been efforts to streamline, or simplify, LCA to make it more feasible while retaining its key features (e.g., Curran, 1996).

Input-Output Analysis

IOA is a technique of quantitative economics introduced by Leontief in 1936 (Leontief *et al.*, 1983, p. 20; Polenske, 2004). In this approach, an input-output table is constructed to provide a systematic picture of the flow of goods and services among all producing and consuming sectors of an economy. IOA also registers the flow of goods and services into and out of a given region. The mathematical structure of the basic input-output models is simple:

$$\mathbf{x} - \mathbf{A}\mathbf{x} = \mathbf{y} \tag{2}$$

where **x** is a vector of outputs of industrial sectors and **y** is a vector of deliveries by the industries to final demand. **A** is a square matrix of input-output coefficients; each element a_{ij} represents the amount of sector *i*'s output purchased by sector *j* per unit of *j*'s output (Leontief *et al.*, 1983, p. 23).

IOA approaches material cycles by replacing the monetary flows with material ones. Its initial demonstration was a projection of U.S. nonfuel-minerals scenarios, completed by the creator of the input-output method in the early 1980s (Leontief *et al.*, 1983, pp. 33–205). The analogous approach for physical flows is termed a "physical input-output table" (PIOT). It is the product of the efforts of scholars from various disciplines between the 1970s and 1990s, and has been applied to establish the material accounting system of several countries (Strassert, 2001, 2002). Duchin did the pioneer work to bring the IOA approach to industrial ecology (Duchin, 1992). More recently, the IOA approach has been linked with LCA to produce a new method: economic input-output LCA (EIO-LCA; Matthews and Mitchell, 2000).

INDUSTRIAL ECOLOGY IN PRACTICE

Micro-Level Practice

Micro industrial ecology's practices are mostly centered on firms and their products and processes. Firms are the most important agents for technological innovation in market economies. The persistent supply of greener products from greener processes in facilities constitutes the microfoundations of world environmental improvement. In addition, a present firm is not a sole "policy taker" any more. To overcome the low efficiency of command-and-control environmental regulation, many firms have become "policy makers," so far as the relationship between technology and the environment is concerned.

Pollution prevention (P2), also termed "cleaner production," is industry's primary attempt to improve upon passive compliance with environmental regulations. In P2, attention is turned to reducing the generation of pollution at its source, by minimizing the use of, and optimizing the reuse or recycling of, all materials, especially hazardous ones. The pioneer of this approach is 3M's Pollution Prevention Pays (3P) program in 1975. It succeeded in avoiding 1 billion pounds of pollutant emissions and saved over $500 million for the company from 1975 to 1992. Many companies were spurred to learn 3M's approach, and according to a recent survey, pollution prevention has become an importance operational element for more than 85% of manufacturing companies (Graedel and Howard-Grenville, 2005).

While pollution prevention addresses a manufacturing facility as it finds it, design for environment (DfE) is transformational: it attempts to redesign products and processes so as to optimize environmentally related characteristics. Often used in concert with LCA, DfE enables design teams to consider issues related to the entire life cycle of products or processes, including materials selection, process design, energy efficiency, product delivery, use, and reincarnation. DfE practices are currently being implemented by many firms, large and small.

DfE is mainly a technological approach. It can address a wide range of environmental issues throughout a product's life cycle. However, its capability to address some environmental impacts, especially in disposal of end-of-life products, is limited: it can facilitate, but cannot ensure, recycling. However, the approach designated "extended producer responsibility" (EPR) complements the firm-level practice from the perspective of policy. In this regard, most Organization of Economic Cooperation and Development countries encourage manufacturers to take greater responsibility for their products in use, especially in postconsumer stages. EPR follows the "polluter pays principle," transferring the costs of waste management from local authorities to those producers with greater influence on the characteristics of products (Gertsakis et al., 2002).

It is foreseeable that the acceptance of EPR will, in turn, intensify DfE activities in many firms. We thus begin to see a sequence of environmentally related steps by responsible industrial firms. The first is pollution prevention, which is centered within a facility. The invention and adoption of LCA next expands a company's perspective to include the upstream and downstream life stages of its products. Later on, a core issue—sustainability—is brought to the table. Some assessment methods have been developed to quantify a facility's sustainability, although this remains a work in progress as of this writing.

Meso-Level Practice

Most interfirm practices of industrial ecology relate to the concept of industrial symbiosis and its realization in the form of eco-industrial parks (EIPs). As Chertow (2000b) puts it: "Industrial symbiosis engages traditionally separate industries in a collective approach to competitive advantage involving physical exchange of materials, energy, water, and/or by-products. The keys to industrial symbiosis are collaboration and the synergistic possibilities offered by geographic proximity" (314).

The classic example of industrial symbiosis is Kalundborg, a small Danish industrial area located about 100 km west of Copenhagen. Its industrial symbiosis began in the 1970s as several core partners (a power station, a refinery, and a pharmaceutical firm) sought innovative ways of managing waste materials (Cohen-Rosenthal *et al*., 2000).

Over time, many other industries and organizations have become involved; the result is a very substantial sharing of resources and a larger reduction in waste (Figure 3).

Industrial symbiosis thinking is implemented by but not confined to EIPs. Chertow (2000b) has proposed a taxonomy of five different material-exchange types of industrial symbiosis:

1. Through waste exchanges (e.g., businesses that recycle or sell recovered materials through a third party)
2. Within a facility, firm, or organization
3. Among firms co-located in a defined EIP
4. Among local firms that are not co-located
5. Among firms organized "virtually" across a broader region

Only Type 3 can be viewed as a traditional EIP. No matter which type, or on what scale, industrial symbiosis has proven to be beneficial both to industries and to the environment.

Macro-Level Practice

At macro scales (e.g., a city, a country, or even the planet), MFA has proven to be an important tool for considering the relationships between the use of materials and energy use,

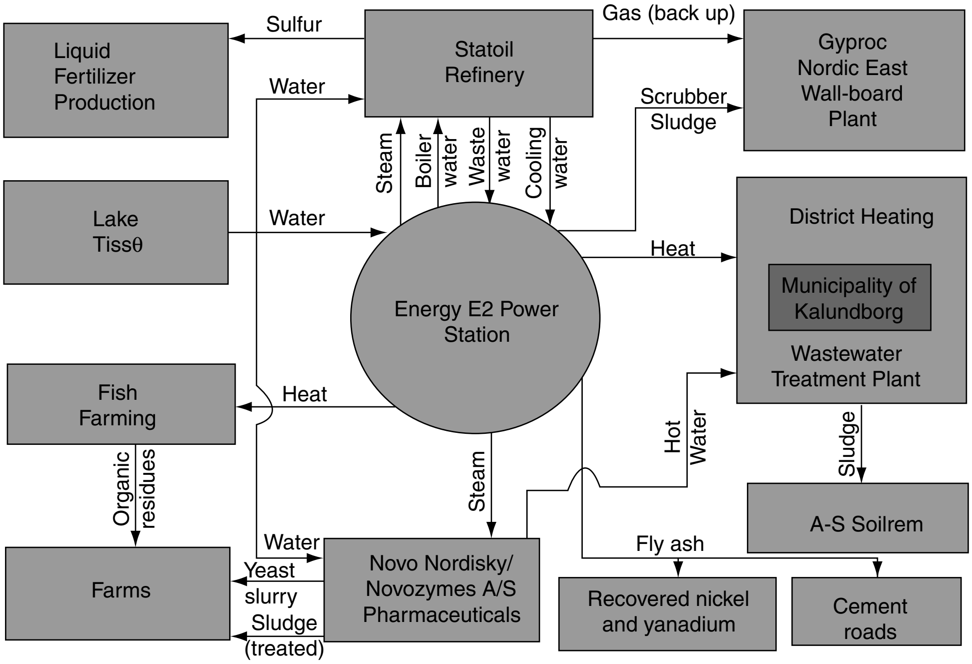


FIGURE 3 Industrial symbiosis at Kalundborg, Denmark. From Chertow, 2000b; updated by M. Chertow.

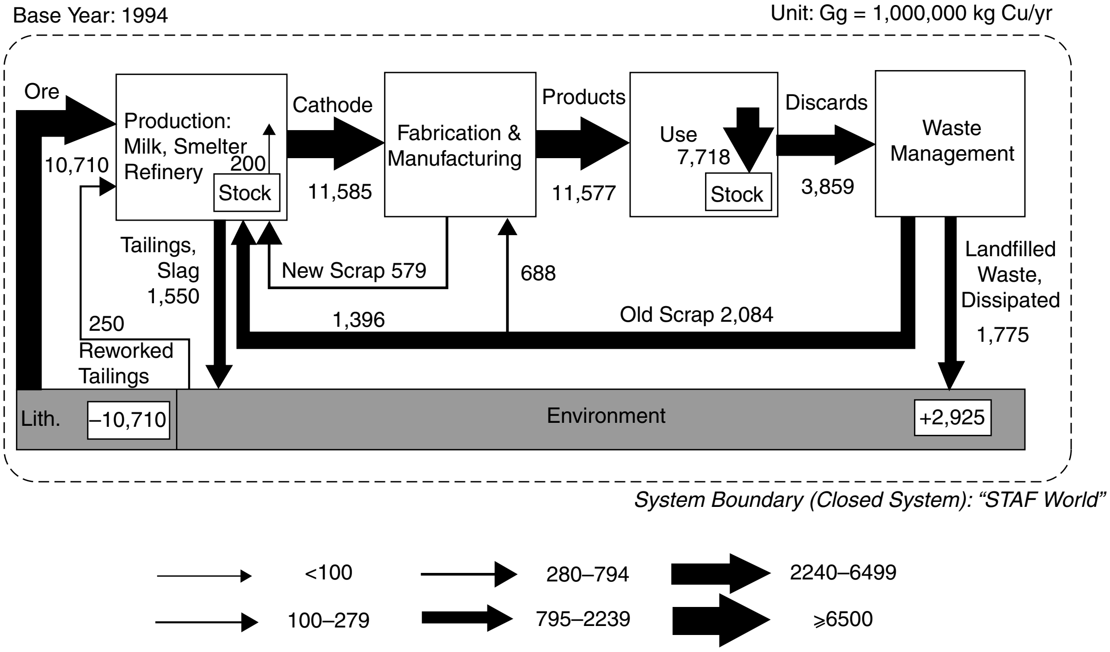


FIGURE 4 Global anthropogenic copper cycle in 1994. From Graedel *et al.*, 2004.

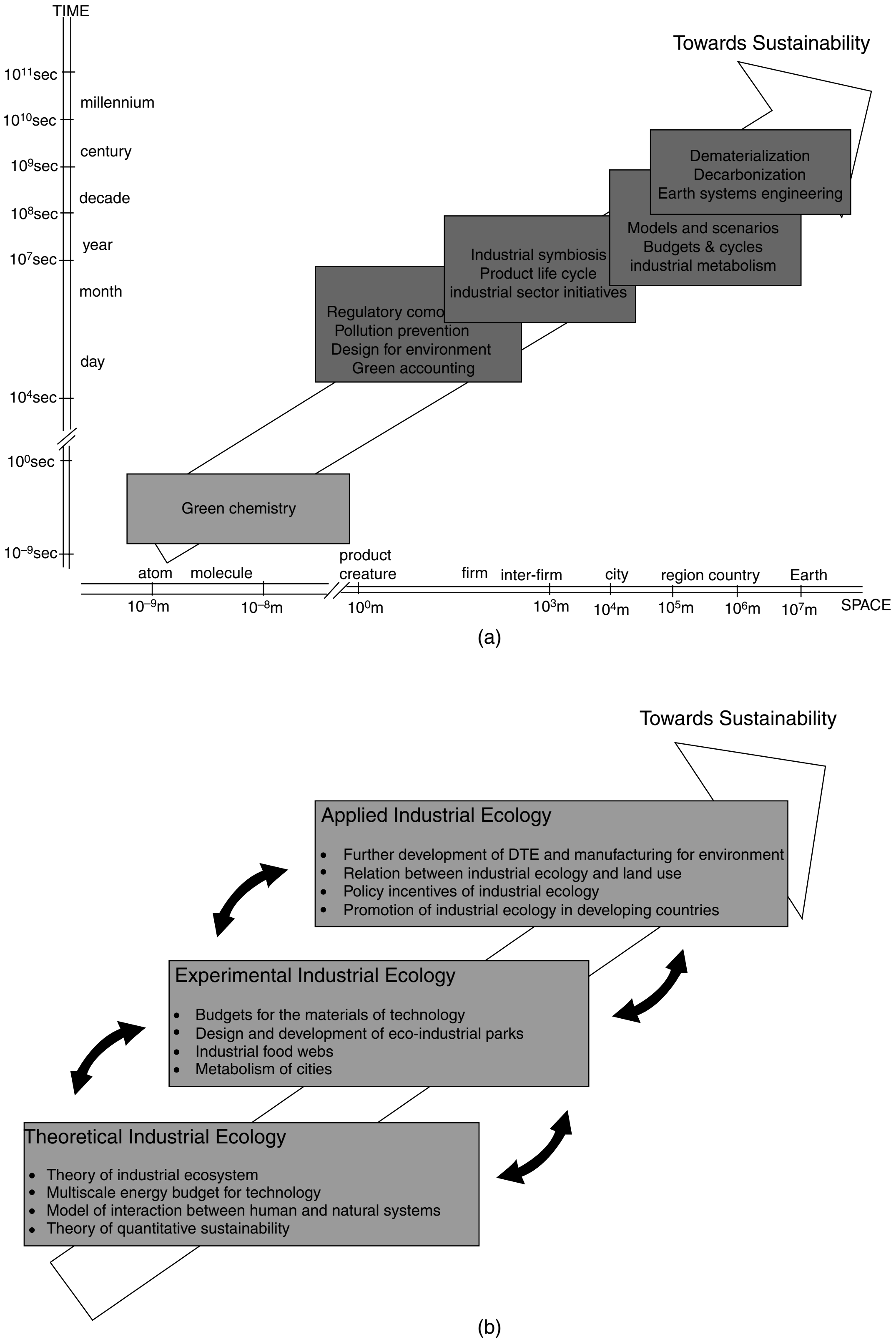


FIGURE 5 A graphical framework of industrial ecology. (a) The spacetime of industrial-ecology tools and methods; (b) An industrial-ecology roadmap.

environmental impact, and public policy. An example, the global copper cycle in 1994, is shown in Figure 4. During 1994, global copper inputs to production were about 83% ore, 11% old scrap, 4% new scrap, and 2% reworked tailings. About 12 Tg of copper entered into use, while nearly 4 Tg were discarded, giving a net addition to in-use copper stock of 7–8 Tg.

Some 53% of the copper that was discarded in various forms was recovered and reused or recycled through waste management. The total environmental loss, including tailings, slag, and landfills, was more than 3 Tg and equaled one third the rate of natural extraction. All of this information provides perspectives impossible to achieve from a less comprehensive analysis.

Material-flow studies can address another macro issue of industrial ecology—dematerialization, which is the reduction in material use per unit of service output. Dematerialization can contribute to environmental sustainability in two ways: by ameliorating material-scarcity constraints to economic development, and by reducing waste and pollution. Dematerialization may occur naturally as a consequence of new technologies (e.g., the transistor replacing the vacuum tube), but can also result from a more efficient provisioning of services, thus minimizing the number of identical products needed to provide a given service to a large population.

SUMMARY

It is difficult to provide a holistic and systematic picture of a young field with its evolving metaphors, concepts, methods, and applications. We attempt to do so graphically, however, in the "spacetime" display of Figure 5a. In this figure, the tools and methods of industrial ecology are located dimensionally, with time and space increasing from the bottom left to the upper right, as does complexity. The figure demonstrates that industrial ecology operates over very large ranges of space and time, and that its tools and methods provide a conceptual roadmap to sustainability.

As an emerging field, industrial ecology has a long list of areas where research and development are needed (Figure 5b). The urgent theoretical needs are to develop general theories for industrial-ecosystem organization and function, and to relate technology more rigorously to sustainability. Experimental industrial ecology needs to complete a set of analytical tools for the design of EIPs, the dynamics of industrial food webs, and the metabolism of cities. Finally, applied objectives can be fulfilled through maintaining the progress of DfE, developing the policy-related aspects of industrial ecology, and promoting industrial ecology in developing countries. The tasks are substantial, but carrying them out is likely to provide a crucial framework for society in the next few decades, as we seek to reconcile our use of Earth's resources with the ultimate sustainability of the planet and its inhabitants, human and otherwise.

REFERENCES

Allenby, B.R. 1999. *Industrial Ecology: Policy Framework and Implementation.* Upper Saddle River, NJ: Prentice-Hall.

Ayres, R.U. 1988. Self organization in biology and economics. *International Journal on the Unity of the Sciences*, 1(3): pp. 10–18.

Bringezu, S. and Y. Moriguchi. 2002. Material flow analysis. In: *A Handbook of Industrial Ecology,* R.U. Ayres and L.W. Ayres (eds.), pp. 79–90. Cheltenham, UK: Edward Elgar Publishers.

Brunner, P.H. and H. Rechberger. 2004. *Practical Handbook of Material Flow Analysis.* Boca Raton, FL: Lewis Publishers.

Carson, R. 1962. *Silent Spring.* Boston, MA: Houghton Mifflin.

Chertow, M.R. 2000a. The IPAT equation and its variants: changing views of technology of environmental impact. *Journal of Industrial Ecology,* 4(4): pp. 13–29.

Chertow, M.R. 2000b. Industrial symbiosis: literature and taxonomy. *Annual Reviews of Energy and Environment,* 25: pp. 313–337.

Cohen-Rosenthal, E. *et al. Designing eco-industrial parks: the North American experience.* http://www.cfe.cornell.edu/wei/design.htm. Accessed in 2000.

Commoner, B. 1972. The environmental cost of economic growth. In: *Population, Resources and the Environment,* R.G. Ridker (ed.), pp. 339–363. Washington, DC: Commission on Population Growth and the American Future.

Curran, M.A. (ed.). 1996. *Environmental Life-Cycle Assessment.* New York: McGraw-Hill.

Duchin, F. 1992. *Industrial input-output analysis: implications for industrial ecology.* Proceedings of the National Academy of Sciences of the United States of America, 89(3): pp. 851–855.

Ehrlich, P.T. and J.P. Holdren. 1971. Impact of population growth. *Science,* 171(3977): pp. 1212–1217.

Erkman, S. 1997. Industrial ecology: an historical view. *Journal of Cleaner Production*, 5(1–2): pp. 1–10.

Erkman, S. 2002. The recent history of industrial ecology. In: *A Handbook of Industrial Ecology,* R.U. Ayres and L.W. Ayres (eds.), pp. 27–35. Cheltenham, UK: Edward Elgar Publishers.

Frosch, R. and N. Gallopoulos. 1989. Strategies for manufacturing. *Scientific American,* 261(3): pp. 144–152.

Gertsakis, J. 2002. Industrial ecology and extended producer responsibility. In: *A Handbook of Industrial Ecology,* R.U. Ayres and L.W. Ayres (eds.), pp. 27–35. Cheltenham, UK: Edward Elgar Publishers.

Graedel, T.E. and B.R. Allenby. 1995. *Industrial Ecology.* Englewood Cliffs, NJ: Prentice-Hall.

Graedel, T.E. and B.R. Allenby. 2003. *Industrial Ecology,* 2nd edition. Upper Saddle River, NJ: Prentice-Hall.

Graedel, T.E. and J. Howard-Grenville. 2005. *Greening the Industrial Facility.* New York: Kluwer.

Graedel, T.E. *et al.* 2004. Multilevel cycle of anthropogenic copper. *Environmental Science and Technology,* 38: pp. 1242–1252.

Hardy, C. and T.E. Graedel. 2002. Industrial ecosystems as food webs. *Journal of Industrial Ecology,* 6(1): pp. 29–38.

Isenmann, R. 2002. Further efforts to clarify industrial ecology's hidden philosophy of nature. *Journal of Industrial Ecology,* 6(3–4): pp. 27–48.

Leontief, W. *et al.* 1983. *The Future of Nonfuel Minerals in the U.S. and World Economy.* Lexington, MA: Lexington Books.

Levy, M.H. 1994. Life cycle analysis. In: *Encyclopedia of the Environment,* R.A. Eblen and W.R. Eblen (eds.), pp. 424–426. Boston, MA: Houghton Mifflin.

Lifset, R. and T.E. Graedel. 2002. Industrial ecology: goals and definitions. In: *A Handbook of Industrial Ecology,* R.U. Ayres and L.W. Ayres (eds.), pp. 3–15. Cheltenham, UK: Edward Elgar Publishers.

Lovelock, J. 1988. *The Ages of Gaia: A Biography of Our Living Earth.* New York: Norton.

Matthews, H.S. and J.S. Mitchell. 2000. Extending the boundaries of life-cycle assessment through environmental economic input-output models. *Journal of Industrial Ecology,* 4(3): pp. 7–10.

Meadows, D.H. *et al.* 1972. *The Limits to Growth.* New York: Universe Books.

Polenske, K.R. 2004. Leontief's "magnificent machine" and other contributions to applied economics. In: *Wassily Leontief and Input-Output Economics,* E. Dietzenbacher and M.L. Lahr (eds.), pp. 9–29. Cambridge, UK: Cambridge University Press.

Strassert, G. 2001. Interindustry linkages: the flow network of a physical input-output table (PIOT): theory and application for Germany. In: *Input-Output Analysis: Frontiers and Extensions,* M.L. Lahr and E. Dietzenbacher (eds.), pp. 35–53. New York: Palgrave Publishers.

Strassert, G. 2002. Physical input-output accounting. In: *A Handbook of Industrial Ecology,* R.U. Ayres and L.W. Ayres (eds.), pp. 102–113. Cheltenham, UK: Edward Elgar Publishers.

Udo de Haes, H.A. 2002. Industrial ecology and life cycle assessment. In: *A Handbook of Industrial Ecology,* R.U. Ayres and L.W. Ayres (eds.), pp. 138–148. Cheltenham, UK: Edward Elgar Publishers.

U.S. EPA. Why LCA? http://www.epa.gov/ORD/NRMRL/lcaccess/whylca.htm. Accessed in 2004.

Ward, B. *et al*. 1972. *Only One Earth: The Care and Maintenance of a Small Planet.* New York: Norton.

TAO WANG
T. E. GRAEDEL
Yale University

INDUSTRIAL HYGIENE ENGINEERING

INTRODUCTION

It can be said that the industrial or occupational physician and nurse are in the front line of industrial health and the contribution they make with regard to diagnosis, establishment of causes, treatment and rehabilitation are of great importance to the health of the workers. But once the potential cause of ill health or discomfort is established it is usually the engineer who is required to design out the problems. Thus the engineer has a most positive contribution to make in the maintenance of good health amongst the working population. The engineer is in the forefront of industrial medicine and as such can make as great a contribution as the medical profession because he or she is in the position to modify the industrial process, or to apply some environmental control system to the workplace, such that workers' exposure to physical, chemical or biological insults can be minimised.

When discussing workplace pollution from airborne materials such as dusts, fumes, gases, vapours and mists, it is useful to first define the outcome of various degrees of overexposure which may result from ineffective, faulty or non-existent pollution controls. In simple terms these may be considered under the categories shown below:

1) The stochastic effects, where the probability of developing a disease increases with dose, as with carcinogens or sensitizers.
2) The acute effects, characterized by rapid absorption and effect due to sudden exposure, typical of asphyxiants.
3) The subacute effects, resulting from repeated and extended exposures over a period of hours or days, an example of which would be the early symptoms of narcosis.
4) The chronic effects, resulting from continued absorption over a long period of time, e.g., materials leading to fibrosis of the lung.
5) The combination of any of the above.

The effects of exposure of workers to heat, noise and radiation also fall into the above categories. Not only will the degree of control depend upon the potential hazard but the requirement for back up and monitoring systems will, to a great extent, be determined by the potential consequences of failure of the control measures. This chapter will discuss in further detail the engineering considerations in designing control measures for workplace environments.

In designing control systems it is necessary to define the objectives such that the alternative engineering options can be seen in perspective. The basic techniques of workplace control are: substitution of a less hazardous material or process, segregation or enclosure of the operation, suppression of emission, ventilation extraction at source, ventilation dilution to reduce concentration in the general atmosphere.

The success of the selected control method will be judged by its ability to reduce personal risk by minimizing worker exposure and absorption. Furthermore the control method should remain effective and maintain the same degree of protection over the working life of the process.

Unfortunately, failure and ineffectiveness of control systems are only too well known and examples are common: enclosures that are left open: ventilation systems that fall into disrepair or are not used due to shopfloor objections to noise, or cold draughts produced as a result of careless or incompetent design; work procedures that are impossible to follow; personal protection that cannot be worn for any length of time. Many more examples could be cited but to the unfortunate recipient failure to protect means a risk to health irrespective of the reason.

Potential risk should be assessed by examining the possible results of failure of the control system. Where chronic hazards exist an occasional transient failure may not be too serious but in the case of sensitizers and carcinogens any overexposure could have serious implications, and, in the case of asphyxiants, failure could be fatal. Thus, in the case of acute hazards the control system should be adequate to prevent even momentary peaks above the control limit. On the other hand, where there are no materials with serious acute hazards, short excursions above a chosen standard level may be permissible provided that the average dose is controlled. Where potential failure could have serious consequences then the engineering software specifications need to be more stringent, i.e., there may be a need for built-in redundancy and back-up controls. The interpretation of control standards and specifications should only be undertaken by a competent industrial hygienist.

As the achievement of comfort is also an objective, physiological and psychological factors should not be overlooked. To allow for human factors in designing controls it is necessary to move outside the realm of pure engineering.

A successful control technique can only be designed by a systematic approach, looking at both hardware and software requirements. Each aspect of the problem can be seen as a link in a chain, if one link fails the worker can become over-exposed to the hazard.

Using the philosophy of preferring a safe place to a safe person, it is usual to examine first the source of the hazard then working backwards to those at risk. When examining the suitability of any strategy it is important to be clear about the degree of protection offered and the consequence of any failure of the control systems.

It is usually best to consider the choice of engineering control systems in the order in which the problem evolves, that is: first tackle the source, then the means of transmission and lastly the exposed population.

The Source

The purpose of dealing with the source is to reduce the potential for emission and this can be achieved in a variety of ways depending upon the nature of the operation.

1) The process: the problem may be eliminated by changing the process so that no hazard is created.
2) Substitution: by replacing the toxic material for one with a lower hazard potential.
3) Enclosure: plant and layouts can be modified to isolate the source of emission or where possible enclosures can be built around the sources.
4) Extract ventilation: where toxic materials are being released into the atmosphere extraction at the source can capture any emitted substances.
5) Suppression at source: wet methods can be used to prevent dust being emitted provided means are available to handle the wetted dust and drying out is prevented.

The Means of Transmission

Once the material is released the means of transmission is usually the air which is eventually breathed by the worker. Thus the hazard can be reduced by one of the following means:

1) Isolation: working systems can be devised to ensure that at times of noxious emission the area is unoccupied or workers are located at sufficient distance to reduce exposure to an acceptable level.
2) Ventilation: in the case of airborne pollutants their concentration can be reduced by dilution ventilation or the workers can be bathed in a supply of cleaned fresh uncontaminated air; in the case of heat the air may be treated via an air conditioning system.
3) Shielding: certain radiations can be blocked by a shield between the source and the worker.

The Exposed Population

Finally, the worker's exposure can be minimized by examining his position in relation to the hazard as follows:

1) Enclosure: the worker can be enclosed in a suitably treated and conditioned refuge.
2) Job Rotation: the duration of exposure and hence the absorbed dose can be reduced by suitable job rotation involving a larger number of workers over a shorter period.
3) Education, Training and Method Study: the workers' awareness of potential hazards can be heightened by training linked to a careful examination of the work methods that are being adopted.
4) Personal Protection: respirators, resistant clothing, hearing defenders, can be employed to reduce contact and to shield exposed and vulnerable parts of the body.
5) Automation: use of automation to either eliminate or reduce the exposure time of the worker is a costly but sometimes necessary alternative.

THE HARDWARE

Control at Source

By far the best approach to minimizing emission at source is to modify the process or the material, provided care is taken to ensure that alternative hazards are not created. The substitution of a safer material for a toxic one has commonly been adopted, typical examples being in the case of paint solvents where xylene and toluene were substituted for the carcinogenic benzene and in foundry work where shot blasting was substituted for silicosis inducing sand blasting.

More commonly it is possible to rearrange the work procedure and methods so that emissions are minimized. Where powders and friable materials such as minerals are transported within a process, the greatest emissions of dust usually occur at the points of transfer from one container to another or from one mode of transport to another. Bags of raw material are emptied into hoppers, hoppers feed on to conveyors, conveyors tip onto other conveyors or to bins; powders are shovelled or blended by hand into open vessels, each stage of the process releasing dust into the air. Where volatile liquids are used they are often poured from one vessel to another or left standing uncovered for periods of time. Where surface coatings are applied, volatile solvents have to evaporate or cure before the coating is complete. This also applies to quick drying printing inks. In many cases the coated surfaces are allowed to dry in the open air of the workplace.

In most cases the emission of dust and vapours can be minimized by enclosure or by redesigning the process so that the escape of pollutants is reduced. The enclosures can be fitted with extract ventilation but often it is just sufficient to enclose as much as possible allowing minimal openings for access. If it is necessary to fit extraction then the smaller the cross-sectional area of the openings the lower the ventilation rate need be. Unfortunately, enclosure obscures visibility, thus where it is necessary to observe the process inside, provision has to be made for simple viewing panels of plate glass or transparent plastic. Video monitors or electronic indicators can also be used.

Redesign of processes can be as simple as fitting special covers on containers and matching discharge ports to entry holes or as complicated as total automatic handling throughout. With materials handling dramatic improvements on hygiene can often be achieved by using sealed transfer systems, eliminating the potential for inhalation and skin contact. Where liquids are transferred the simple use of a discharge nozzle extending below the liquid surface can reduce the evaporation. It is wise to think through every stage of the process noting places where emission is highest. An experienced outsider can often notice features which are missed by the people who are in closest contact. This should be done before the costly expedient of extract ventilation is adopted.

When redesigning new plant it is important to involve experienced health, hygiene and safety professionals in the early stages as often the environmental aspect of production processes are considered too late to have any effect on the installation. Often plant is installed and commissioned before environmental hazards are realized and then there is a great reluctance to make modifications. It should be borne in mind that the high apparent cost of process redesign can often be offset against the reduced capital and running costs of supplementary control, monitoring and maintenance which is required of an inferior process. Also the relation between workforce and management is greatly improved with obviously safer processes.

Extract Ventilation If it is difficult to adopt process redesign or satisfactory enclosure then pollutants can be captured before they are generally released by means of some type of hood, slot or enclosure inside which a negative air pressure has been generated. Air will thus flow in by virtue of the difference in pressure between the mouth of the device and the surround. If the contamination from the process can be moved into the hood by means of the airflow, then some degree of control will be achieved. The air in the hood is removed by a fan via ducting to an outlet or air cleaning system. Any contaminant not drawn in will remain free to disperse within the workroom where it may be inhaled by the operator or other personnel. It is therefore important to consider the patterns of the airstream in the vicinity of the work where the success or failure of the extract system is often determined.

Providing no obstructions exist around a hood, then air will flow in from all directions in a zone of influence which is approximately spherical in shape with the mouth of the hood at its center (Figure 1). The air velocity at any point in that sphere being approximately related to the hood entry velocity by the inverse square of the distance from the hood (Dallavale, 1952).[10] By careful selection and location of screens the shape of the airstreams can be adjusted to reduce unwanted airflow from uncontaminated areas thus increasing the flow of air drawn through the zone of pollution emission.

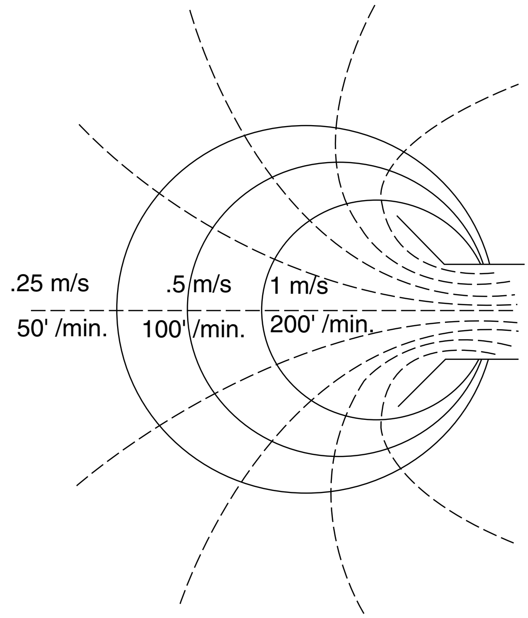


FIGURE 1 Airflow patterns and velocity contours at the mouth of a suction inlet.

For most operations a minimum velocity (capture velocity) is defined as that velocity at which the energy of motion of the contaminant can be overcome thus drawing it into the hood. In the case of the release of welding fumes for example a velocity of at least 0.5 m/s (100 ft/min) is usually required to overcome the thermal motion of the plume (ACGIH). In processes where the velocity of emission of the contaminants is high as with grinding or where high cross draughts exist then higher capture velocities may be necessary. Conversely in certain work, such as paint finishes and gas shielded welding, too high an airflow may affect the finished product. Thus a control envelope bounded by upper and lower velocity contours may have to be defined. That is, the hood should be sited so that the source of pollution lies between chosen velocity contours and preferably on the centerline of the slot.

Centreline velocities can be predicted in relation to the distance from the mouth of the opening using the expressions given below (Fletcher, 1977).[12]

$$\frac{VA}{Q} = \frac{V}{V_0} = \frac{1}{0.93 + 8.58\alpha^2}$$

where

$$-\alpha = xA^{-1/2}\left(\frac{W}{L}\right)^{-\beta}$$

and

$$\beta = 0.2\left(xA^{-1/2}\right)^{-1/3}$$

and

V = centreline velocity at distance x from the hood

V_0 = mean velocity at face of hood

L = length of hood

W = width of hood

$A = L \times W$

Q = required volume flow rate.

The solution of these expressions can be made simpler by the use of the nomogram in Figure 2.

In the use of the above expressions the dimensions L and W will depend upon the size and shape of the point or points of emission of the pollutant, V should be the capture velocity and Q will be the calculated volume flow rate of air to provide the required capture velocity at distance x from the hood.

The relative location of the operator, hood and source of contaminant can also have considerable bearing on the effectiveness of control. Whilst a system utilising the flow energy of the plume to assist the airflow may require less air volume, this will be ineffective if the operator is located between the source and the hood. It should be borne in mind that the operator is acting as a bluff body in a streamline airflow, creating turbulence which can bring back contaminants into the breathing zone. Exposure levels will also be contributed to by secondary emission from contaminants that are not captured. The shape of the workpiece can also influence airflow patterns at the point of emission directing contaminants to remote areas away from the mouth of the hood. In welding, for example, whilst stance and the use of a headshield can help to minimize these effects (Akbarckhanzadeh, 1979) there will still be some residual fume and gases in the breathing zone when using simple extraction hoods or slots.

When designing a hood for ventilation control, it is better to start with an idea of providing maximum enclosure of the source of contaminant. Thus it is prudent to design a booth or enclosure first then progressively remove sections of it to make a practical design. Unfortunately as the size of the openings increase so does the required volume flow rate of air and effectiveness of the extraction may decrease. The use of cardboard and a portable fan to produce a full scale "mock up" model can be an invaluable aid to empirical design, giving an opportunity to judge control effectiveness and ergonomic performance.

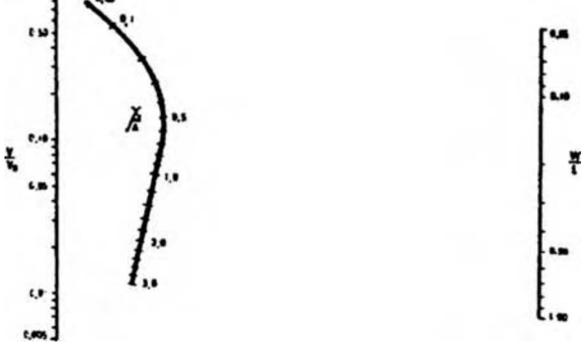


FIGURE 2 Nomogram for the solution of hood entry equations.

Difficulties arise where work processes require access on all sides to facilitate production techniques or where the source of contamination is moving. In the case of benchwork requiring all-round access, narrow slots pulling over a short distance or assisted by supply jets (Hughes, 1988)[16] of air can be used, as can perforated or slotted bench tops through which air is drawn, thus the ventilation hardware occupies the minimum of space. Where moving sources of pollution occur it is possible to use hoods with flexible ducting supported on cantilever arms or crane type supports allowing flexibility over a wide area. It may be necessary to provide clamps or other devices to ensure that the hood is located to achieve good capture and will move with the source. Care must be taken to ensure that the use of the extract system does not hinder the operation or it may fall into disuse to operator resistance. Operationally portable hoods often fail to achieve the degree of control required due to the practical imposition of regularly having to accurately reposition the device, thus it is important to ensure that such devices are lightweight, well balanced and quick and simple for the operator to use.

The hood may also be designed and built into the tool as with portable sanders, grinders and saws. This technique known as low volume high velocity (LVHV) exhaust uses small volumes of air at relatively high velocities. The hood can be an annular slot or multi-hole type operating on a vacuum of 20 kPa (150 mm Hg) at a volume of 16 liter/sec (35 ft^3/min). When these are accurately positioned much of the contaminant can be controlled. However, although this system has the advantage of low volume requirements and portability there are several disadvantages which often preclude its use. The LVHV head has a limited envelope of control due to the low volume of air extracted and workpiece shape, misalignment of exhaust orifice or lack of maintenance can seriously disrupt the efficiency of collection. The weight of the tool plus its extract also presents problems of manoeeverability and acceptability to workers.

Transmission

Once the pollutant has escaped from its source and is airborne or is broadcast many workers in the vicinity can be at risk. With wide area sources capture may be impossible.

Segregation of the worker from the source of pollution by carefully planned production routines can often be used to reduce the potential for exposure. For example processes which give rise to airborne materials can be carried out in separate areas or at times when the building is otherwise unoccupied, operators remaining in the areas can be provided with protective clothing. Local extraction, as described in the previous section can be employed but its effectiveness is limited because of the wide area now involved. Therefore it may be necessary to resort to diluting the airborne material to a safe level by providing quantities of uncontaminated air sufficient for the purpose.

General Ventilation Although control efficiency is considerably less than with a well designed extract system the technique is usually easy to operate and is less susceptible to mechanical or software failure. But because of its limitation it should only be applied to places where the pollutant is of low toxicity and such that localized clouds of high concentrations cannot occur.

Dilution ventilation is based on the fact that a cloud of contaminant mixes with the surrounding air thus reducing the concentration in the resulting mix.

Assuming complete mixing the concentration of a pollutant emitted at a rate of R m^3/s into an airstream flowing at a rate of Q m^3/s is given by:

concentration (in m^3 of pollutant per m^3 of air)

$$= \frac{R}{Q}$$

but is more commonly expressed as a percentage.

$$\text{Concentration (per cent)} = \frac{R}{Q} \times 100$$

or a parts per million

$$\text{concentration (ppm)} = \frac{R}{Q} \times 10^6.$$

These concentrations are expressed as volume for volume but if a mass per volume concentration is required then the density of the substance should be used as shown below:

$$\text{mass concentration (in kg/m}^3\text{)} = \frac{R}{Q} \times D$$

where D is the density of the airborne substance in kg/m^3

$$\text{or mass concentration (in mg/m}^3\text{)} = \frac{R}{Q} \times D \times 10^6.$$

It must be noted that in the case of airborne solvent vapors the density of the vapor is not the same as the density of the liquid but can be calculated from:

$$\text{Vapor density at STP} = \frac{M}{22.4}\ \text{Kg/m}^3$$

where M is the molecular mass of the vapor.

With dilution ventilation it is necessary to calculate the required volume flow rate of air to dilute the pollutant to a desired concentration i.e. to the TLV or below. The calculation can be summarized in the following expressions:—

$$\text{Required flow rate } Q = \frac{R \times 10^6 \times K}{C}$$

where C is the chosen concentration of the pollutant in parts per million and K is a safety factor. The units of R and Q must be consistent, that is, if R is in m^3/s then Q will be the same.

Where the solvent vapors are concerned Q can be found by:—

$$Q = \frac{r \times 22.4 \times 10^6 \times K}{M \times \mathrm{C}}\ \text{m}^3/\text{s}$$

where r is the rate of evaporation of the solvent in kg/s.

The determination of the safety factor K is a matter of judgment and experience and depends upon the relationship between the source of pollution and the airflow distribution of the diluting air, the uniformity of emission and the toxicity of the pollutant. Where there is good mixing between the diluting air and the pollutant and rates of emission are uniform and the pollutant is of a low toxicity then K could be as low as 2 or 3.

The figures calculated from the above formulae should be used with great caution as dilution on its own has little effect on the source of contaminant, serving only to reduce the workers' exposure by minimizing the dose accumulated from the general working environment. If the pattern of airflow in the vicinity of the operator can be modified then a greater effect on the breathing zone concentration can be achieved (see section, *Displacement*).

In practice, airflow patterns in buildings are complex, the energy of motion of the contaminant often giving rise to a plume effect which can remain in a concentrated cloud for some time. Considerable energy in the form of a high air velocity would be required to disperse it. In the case of hot fume the plume can rise and form a layer in the upper parts of the building and some gases in their concentrated form can also do this if they are less dense than air. Likewise concentrated clouds of denser gas released into still air can form layers in the floor or pour into sumps and trenches. When dispersing such accumulations air velocities in excess of 2 m/s (400 ft/min) are required but the whole operation must be undertaken with care to prevent spread into other occupied areas.

However if there is no risk of a contaminant with acute hazard potential accumulating in areas which may become occupied, then the motion of a concentrated plume may be used to good effect. Careful location of screening and exhaust points can be used to remove the concentrated material without the need to distribute and dilute. As a result much greater effectiveness can be achieved at much lower air volumes. Here empirical studies of contaminant flow patterns are required to assess ventilation requirement, the previous formulae being inappropriate.

The Exposed Population

Where the hazard cannot be controlled at source or is not effectively controlled during the transmission stage then it is necessary to examine the exposed population in order to engineer an environment around them which will minimize their exposure.

This can be done in three ways:

1) released into still air by air displacement so that they are bathed in an uncontaminated airstream,
2) by the use of conditioned refuges or enclosures,
3) by the provision of personal protective clothing.

Displacement Correct selection of airflow patterns can serve to reduce the flow of airborne contamination in the breathing zone. Displacement depends upon the creation of a diffused clean air flow over the worker and towards the work, thus carrying away the products of the work, preferably towards some form of extraction system. The provision of work station supply ventilation also has corresponding economies in the volumes of air required, as compared to a system ventilating the whole building. Thus, displacement at the work station can help to reduce the high cost of heating replacement air during winter months. However, care must be taken to ensure that local turbulence is minimized so that the effectiveness of the control is maintained. This technique is most suited to well defined work stations. Working posture and stance also needs to be carefully considered to ensure success.

Thermal comfort of the worker must be considered so that the combination of air temperature and velocity is such that cold draughts are not experienced at the work station. To this end the air supply diffusers must be carefully chosen and the supply air temperature accurately controlled to suit the air velocity at the worker. This is particularly important where air is discharged from above and behind the worker as the back, the neck and the back of the head are parts of the body most sensitive to cold draughts. It should be noted that air at a temperature above normal room temperature can feel cold if it is flowing at a sufficiently high velocity. Charts showing the relationship between air velocity and temperature and the part of the body affected (Houghton *et al.*, 1938) in a room at 24°C are shown in Figure 3.

The relationship is expressed as the number of people complaining of discomfort as a percentage of the number tested. As an example from Figure 3 it can be seen that a draught velocity of 0.2 m/s at a temperature of 1.0°C below room ambient of 24°C blowing on occupants' necks will give a feeling of coolness and will not be acceptable to 20% of the room occupants whereas the same draught blowing on the ankle region will discomfort only about 9% of the occupants.

Conditioned Refuges Isolating the worker from an uncongenial environment is a technique which is often adopted where the process being worked is too large or too expensive to control locally. Isolation cubicles can be used to protect from noise, ionizing radiation, heat and cold as well as from airborne toxins. in most cases the enclosure will require ventilating and possibly air conditioning and the amounts of air to be supplied will require some calculation. As a general rule each person enclosed will require 10 litre/sec of fresh air but this amount can be varied depending upon the size of the enclosure and whether smoking is permitted or not.

If air conditioning is required the airflow rate will generally be in excess of that required for fresh air alone and some recirculation will be normal. The volume of air required for air conditioning will depend upon the heat gain to the space and can be calculated from the expression:—

$$Q = \frac{S}{1.2\Delta t}\ \text{m}^3/\text{s}$$

where

S = the sensible heat gain to the room in kW
Δt = the difference in temperature between the room air and the supplied conditioned air °C.

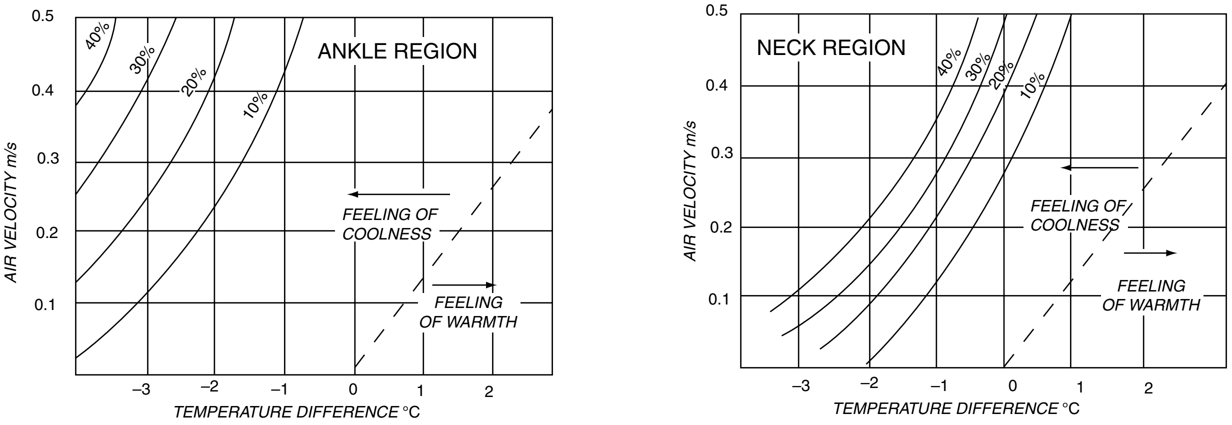


FIGURE 3 Percentage of room occupants objecting to draughts.

Thus a balance has to be made between Q and Δt and this will be influenced by draughts as illustrated in Figure 3.

Personal protection Often considered to be the last resort in reducing worker dose it has been shown (Crockford, 1976)[8] that personal protection can play a major role in an "integrated control programme". It can be used as a single control item as a tool to assist in the smooth operation of a process or it can be used as a "fall back" feature as part of a system with built in redundancy. It has most application where the area of emission is too large to be effectively controlled by other means or in operations where the hazard only occurs for periods of short duration and in a predictable manner and where the alternative control measures are too costly to be a reasonable alternative.

Hitherto most respiratory protective equipment has been uncomfortable to wear for long periods and workers have been reluctant to make full use of it but with the introduction of the self-contained ventilated helmets much of the worker resistance has disappeared. The development of such helmets has put a new light on protection in hitherto traditionally dusty industries.

Finally, aspects of personal hygiene can be included in this section. Absorption by routes other than inhalation can be minimized by frequent washing, the provision of clean work clothing daily, the prevention of eating, drinking and smoking in the workplace.

Ventilation System Design

From the formulae and diagrams already given, having determined the volume flow rate of air to undertake a specific ventilation task and, in the case of extraction, having established the dimensions of the inlet it is necessary to design the remainder of the hardware that make up the ventilation system. Textbooks abound in this topic (Hemeon, 1963; Alden and Kane, 1970; McDermot, 1976) so only an outline of the procedures involved will be given.

Hoods etc. In the case of the extract hood many configurations are given (ACGIH, 1995) to suit a variety of workplace situations. But as every workplace and building varies it is necessary to adapt the designs to fit. This should be done in conjunction with a production engineer so that the flow of the work is not hindered by the presence of the ventilation hardware which tends to impede the transport of raw materials, finished products and can restrict the mobility and visibility of the workforce.

Ducts With regard to ducting to transport air from one part of the workplace to another it is necessary to sketch on a plan and elevation of the workshop the best position for it to be sited. It is usually more convenient to site in the roof, ceiling or spaces above the workplace but the accessibility of overhead cranage must be considered in the layout.

Fittings such as bends, changes of section, branch pieces and dampers are necessary to accommodate multiple intakes and to avoid obstructions forming part of the building fabric or production equipment. Thus the installation will consist of lengths of straight duct plus fittings. As more fittings are added and the duct runs become longer and more complex so the energy required to transport the air increases.

Having sketched the layout it is then necessary to determine the dimensions of each item. For identification purposes it is useful to label each section where there is a change of dimension or direction of airflow. Thus each component can be labelled as AB, BC, etc. and, as such, is suitable for entering into a computer ducting program (Malchaire, 1981; Clapp *et al.*, 1982; Betts, 1984).

With regard to straight ducts the cross-sectional shape can be either rectangular or circular but cross-sectional area is important.

For a given airflow rate the narrower the duct the higher the average air velocity as given by the simple expression:

$$\text{Volume flow rate, } Q = VA$$

where A is the cross-sectional area of the duct at the point where V is known. Where airborne particles are to be conveyed via the ducting then a minimum air velocity must be maintained that will prevent the particles from settling out inside. This velocity is known as the *Transport Velocity* and will vary with the type of particle being transported but as a guide a velocity of 20 m/s (4000 ft/min) will ensure that most particles remain airborne.

Thus having established a volume flow rate and a transport velocity, the cross-sectional area is determined from the above formula. In this situation a circular cross-section is more suitable.

Where transport velocities are not important then a compromise has to be made between the higher energy consumption of narrow ducting and the higher capital cost of wide ducting. Also the passage of air through ducts creates noise which can become obtrusive in quiet situations. As a general rule air-speeds in excess of 5 m/s (1000 ft/min) can create noticeable noise. The cross-sectional shape of the duct is influenced by the size of the space into which it is to be fitted. A circular section will always use less material than a rectangular for a given area but often it is easier to fit rectangular ducts into areas where space is limited as in false ceilings of buildings.

The energy required to move air through the ducting is expressed as a pressure which is necessary to overcome fractional resistance and turbulence caused by the materials of manufacture. Ducting is commonly made of galvanized sheet steel but other materials such as brick, concrete and various plastics can be used each of which provides a different frictional resistance. Charts are published (ACGIH, 1988; CIBS, 1970) giving pressure losses for various diameters of duct for a wide range of airflow rates. If other materials are used the ducting can be assumed to be of sheet steel as above but the resulting pressure losses must be multiplied by a factor depending upon the nature of the material. These factors are published for the commoner materials (CIBS, 1970) or are available from the manufacturer.

Pressure losses in the fittings can either be calculated as an equivalent length of straight duct (ACGIH, 1988) or by a factor which must be multiplied by the velocity pressure in the ductwork (CIBS, 1970; Daley, 1978).

In order to establish the type and size of fan to provide the motive power to move the air it is necessary to establish the total pressure loss throughout the system by adding the losses for each item from one end to the other including any filters and discharge louvres. The chosen volume flow rate plus the total pressure required will identify the fan which can be chosen from catalogues. If a multi-branched system having several extract or supply points is designed then the total volume flow of the fan is the sum of the flows in each of the terminals but the total pressure loss is that from the furthest end to the other and must not include pressures lost in intermediate branches.

Multi-branch duct systems need to be properly balanced to ensure that the correct flow rate is passing through each branch. This can be done either by inserting dampers in each branch and adjusting the dampers to provide the correct flow or by inherently balancing the system by careful selection of duct sizes in each branch, also constant volume flow devices can be inserted to automatically set the branch to the required flow. Where dust is to be carried then any obstruction inside the duct such as a damper becomes a place on which dust can be trapped leading to possible deposition and alteration of flow characteristics, therefore inherent balancing is to be preferred.

Also damper balanced systems have the disadvantage that individual control dampers can be tampered with and adjusted to the detriment of all the other branches in the system. Once a system is balanced then the dampers should be locked in their set positions.

Fans Fan types vary widely and are named: propeller, axial flow, cross flow, mixed flow and centrifugal depending upon their geometric shape. Some have their electric motors in the airstream and some have them outside. If dirty, hot or corrosive air is handled then it is important to ensure that the motor is out of the airstream. Also many atmospheres contain inflammable gases or vapours then flameproof motors and switchgear are required to minimize explosion risks. In some countries it is a legal requirement that where handling flammable gas mixtures, fan motors must also be out of the airstream.

Fan impellers and casings are made of various materials to suit the type of air and contaminant that is being handled. Corrosive contaminants will quickly destroy mild steel fans; thus plastic or stainless steel is required, this would also apply to the materials of the ductwork and fittings. Also impellers accidentally striking the casing as they rotate can give rise to sparking thus where there is an explosion or fire risk materials that produce high temperature or high energy sparks such as aluminium or magnesium or any alloys of those elements must not be used.

A fan performs according to a characteristic curve of pressure against volume flow rate which depends upon its geometric shape and speed of rotation of the impeller (Figures 4–6). The position of the duty point on this curve is governed by the resistance to airflow of the ventilation system against which it has to work. The resistance of the system can be represented by a curve whose shape is a parabola passing through the origin of the form: resistance pressure proportional to volume flow rate squared ($p \propto Q^2$); see Figure 7. The fan efficiency also varies over its range of duties; thus, it is important to choose one whose duty point on the system lies on the most efficient part of the characteristic curve. The manufacturer will advise on this and most fan catalogues indicate efficiencies.

The majority of fans are driven by electric motors although other sources of motive power can be used. Fan drives can either be direct, where the impeller is keyed on to the shaft of the motor or through some drive arrangements such as V-belts and pulleys, gearboxes or variable speed couplings. It is unusual for direct driven fans to be variable in speed unless the electric motor itself has the facility of speech change. Changes of speed can be arranged with V-belt drives

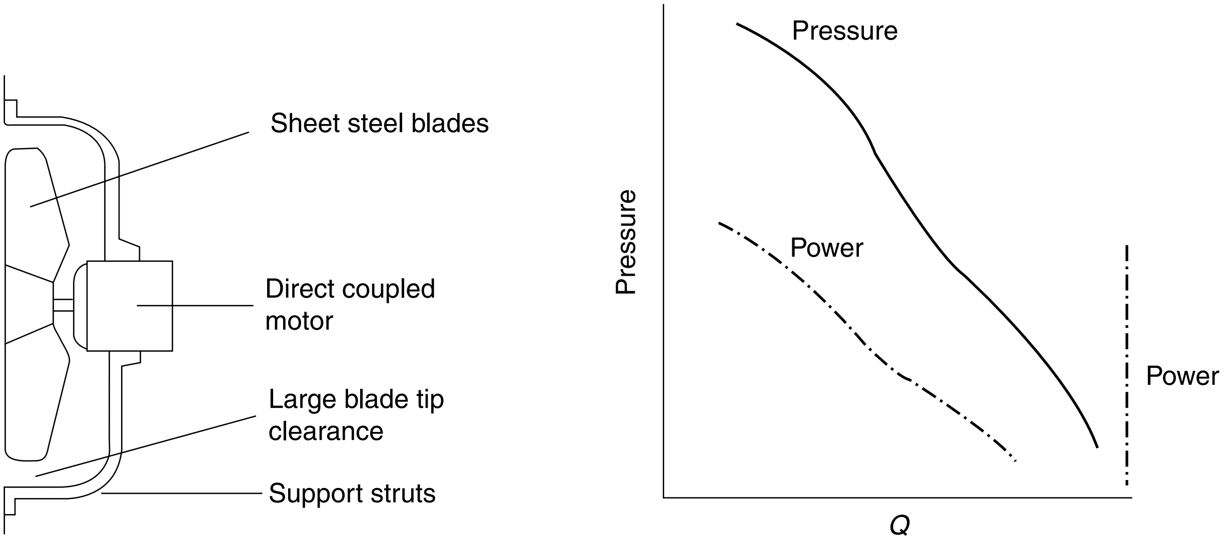


FIGURE 4 Propeller fan and graph showing its characteristic curves.

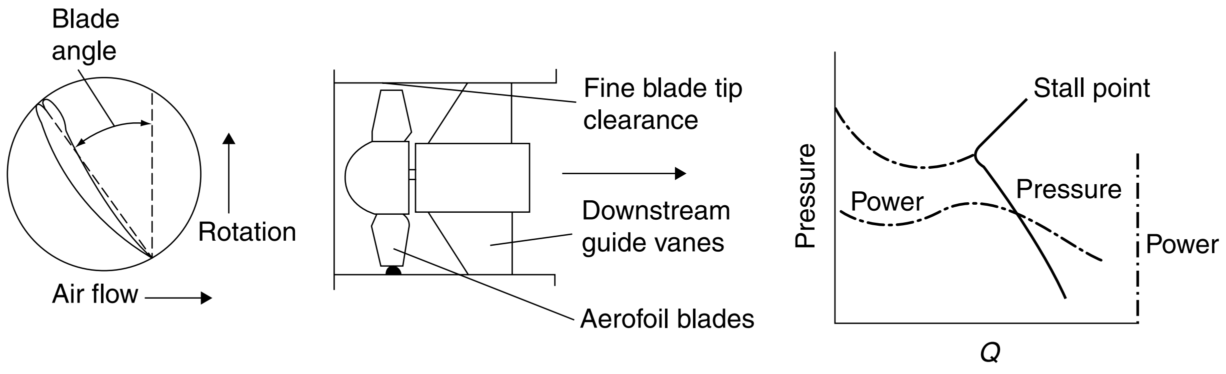


FIGURE 5 Axial flow fan and graph showing its characteristic curves.

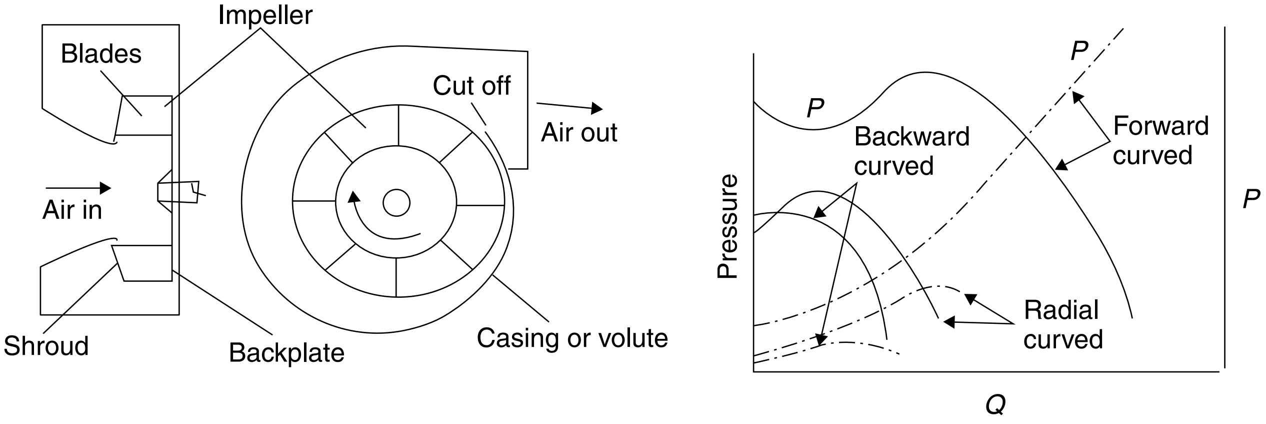


FIGURE 6 Centrifugal fan and graph showing its characteristic curves.

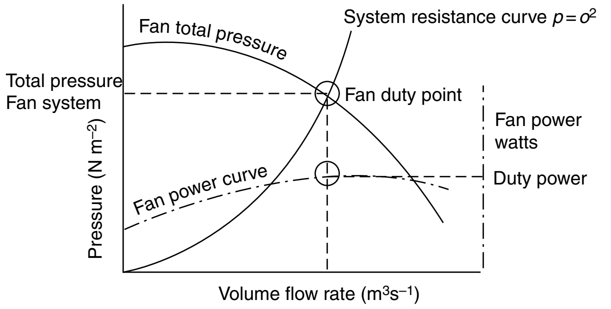


FIGURE 7 Matching of fan and system.

by changing the diameter of the drive and driven pulleys and with gearboxes by changing the spur wheels. If increases in fan speed are contemplated they should be done in consultation with the manufacturer as motor and shaft loadings, impeller materials and bearing sizes are designed for a particular maximum duty and power in mind and any increase over and above the manufacturers' maximum can lead to failure of one of the above components and could cause a serious accident.

Air Cleaning and Filtration This topic can be divided into two fields of speciality depending upon the nature of the dust and its concentration in the airstream. Where air is supplied to a workplace from outside then its particulate concentration is low but of a wide range of sizes and materials. Filtration is required to clean it to the standard required of that workplace, which in some cases can be very stringent, for example, in clean room atmospheres.

In the case of captured air from a dust emission source the loadings can be very high but often of a specified material and of a limited size range often much larger in aerodynamic diameter than with external air pollution. This requires the use of different techniques to provide the necessary air quality standard.

In the former case filtration is by inertial separation using the particles' changes of direction as it zig-zags its way through dry fabric filter media. The particle is caught on the surface of the material which is sometimes coated with a viscous medium to assist capture and to improve loading capacity. The size of the particle captured will depend upon the matrix size of the medium. Filter materials consist of cotton, wool, nylon, viscose fibers, paper, glass fiber, asbestos, cellulose and many others and are either packed in disposable cells or are on a continuous roll which is automatically wound on as the resistance increases.

Where high concentration particulate cleaning is involved mechanical devices imparting centrifugal action are used to collect the larger particles. Cyclone separators and other dry centrifugal cleaners can be used but where particles are sticky or in a moist atmosphere wet methods such as venturi scrubbers, water cabinets, and wet centrifugal devices are employed. These techniques are more effective for large particles rather than small which require large area dry fabric devices such as bag and sleeve filters with reverse jet purging. Where higher collection efficiencies are required for smaller particles then multi-stage filtration using as the last stage inertial filters, may also be used. Cleaning air for gaseous pollutants is more difficult involving a variety of tower scrubbers, washers and absorbers (Dorman, 1974).

Energy and Cost Implications of Extract Ventilation The energy required to provide extraction ventilation can be considered in two parts: that which is required to move the air through the system to overcome the friction losses and that which is required to heat incoming air from outside to replace that which is discarded.

Friction losses The power required to overcome the friction losses can be calculated from the simple expression:

$$\text{Power} = \frac{pQ}{\eta}\,\text{Watt}\,(\text{Joule/sec})$$

where

p = total pressure loss through the system in Pa (N/m^2)
Q = total volume flow rate flowing in m^3/s
η = overall efficiency of the fan and prime mover, usually the fan drive and motor.

Thus, if the system is run continuously for a period of time, the total energy required can be calculated in kJ or kW hours and converted to a monetary equivalent based upon the fuel tariff in operation.

Replacement Air Whenever air is thrown away to outside, replacement air enters the building. This may enter via a properly designed make-up air system or it can find its way in via breaks in the building fabric such as doors, windows, discontinuities in the building. No matter which way it enters, if the building is to be maintained at comfort condition, then entering air requires tempering, usually in the form of heating in the winter and in some places cooling in the summer. The theoretical power required can be calculated from the formula

$$\text{Power} = \rho V \cdot C \Delta t \; Watt$$

where

V is the volume flow rate of air entering m^3/s
ρ is the density of that air in kg/m^3
C is the specific heat of the air in kJ/kg°C
Δt is the difference in temperature between inside air and outside.

For most engineering applications this formula can be approximated to power $= 1.2V\,\Delta t$ Watt. Figure 8 shows a chart for calculating power required for heating and friction losses.

When the duration of operation is taken into account the overall theoretical cost of heating or cooling can be calculated, but it must be remembered that the type of fuel and the efficiency of conversion and tariff in operation at the time must be considered in order to obtain a realistic costing.

In most situations at most times of the year the cost of tempering the incoming air far outweighs the friction costs (Gill, 1980)[13].

Difficulties with Extract Systems

Many systems are unpopular or unsuccessful because they allow the pollutants to escape in sufficient volumes to hazard workers health or comfort. Faulty design, installation, maintenance and procedures can all lead to difficulties. Some of the commoner faults found under these headings are listed below.

Faulty Design

1) The point of release for the pollutant is too far away from the suction inlet,
2) insufficient enclosure has been provided to prevent the escape of pollutants,
3) the airflow may be restricted by a poor backup ductwork system,
4) the wrong fan has been chosen,
5) the fan performance may be restricted due to bad inlet conditions,
6) outside wind and weather conditions are affecting the system,
7) if multi-branched, the system may be out of balance,
8) no provision made for heated "make-up" air to enter,
9) discharges adjacent to air inlets, allowing recirculation of contaminant.

Faulty Installation

1) Fans wired wrongly affecting their direction of rotation leading to reversed flow,
2) with two stage fans, one stage installed the wrong way round,
3) ductwork badly jointed allowing leakages of air,
4) obstructions left in ductwork during installation,
5) controlling dampers may not be functioning correctly or properly anchored.

Faulty Maintenance

1) Ducts can be restricted or blocked due to a build-up of deposits inside,

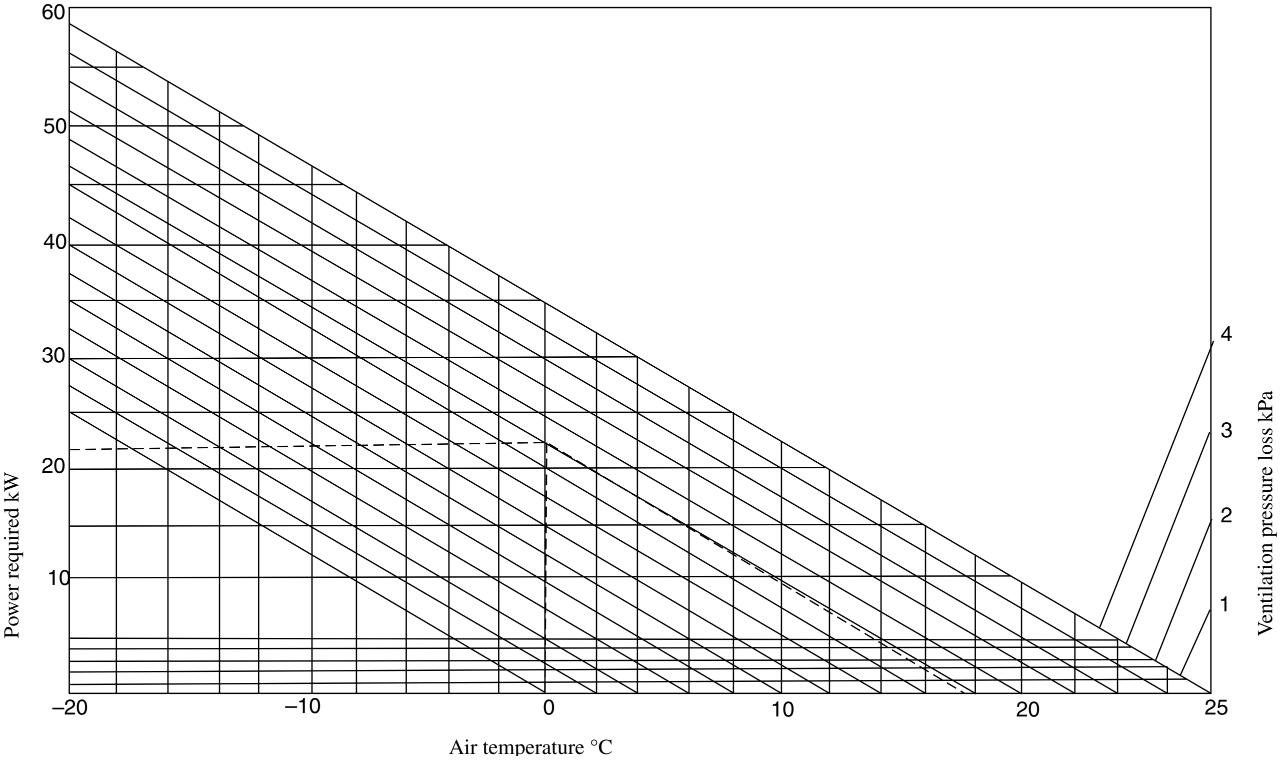


FIGURE 8 Chart for estimating energy requirements for heating.

2) Ducts can be damaged allowing leakage or restriction,
3) Fan blades can be deposited with dirt,
4) Fan blades can be corroded or worn away,
5) Fan motors can be seized or corroded due to being in an uncongenital airstream,
6) Dampers can be corroded, seized or deposited with dirt,
7) Filters or air cleaners can be blocked or missing,
8) Weather protection devices may be corroded or damaged.

Faulty Procedures Ventilation systems are often left to the shop floor operators to control either by an on/off switch or by opening and closing dampers when required. There is a danger that the system is not turned on because of unpopularity or that the opening and shutting of dampers unbalances the system particularly if all dampers are left fully open. Arguments can occur on the shop floor because of unpopularity often as a result of faulty design and systems are left switched off. No ventilation system should be controlled thus but it should be interlocked electrically with the process so that it is running when the process is. Multibranch systems must be either inherently balanced or have their controlling dampers locked in position.

Air Conditioning

Air conditioning plant is normally designed to take air at outside conditions which varies in temperature and moisture content from the very cold and dry in the depth of winter to warm and moist in the height of summer. The plant either heats and humidifies or cools and dehumidifies the air to supply it to the spaces to be ventilated at such a temperature and moisture content to deal with the heat gains and losses and the moisture gains and losses in that space.

Room heat gains and losses vary in intensity with time over the day and year due to variations in the use of the space such as changes in occupancy and/or the use of machinery, plant and equipment, external influences such as movement of the sun and variations in outside air temperature.

Room moisture gains occur due to occupancy and the evaporation of water from plant and equipment in use. Moisture losses are rare but do occur in certain textile and paper manufacturing plants.

If several rooms have similar patterns of heat and moisture gains and losses, they can be grouped together into a zone which can have the same supply of air, conditioned by common controls and plant.

Because the supplied air is the vehicle for heating and cooling, larger volume flows are normally required just for providing fresh air, therefore, for reasons of economy, whenever possible, some extracted air is recirculated and only that amount is discarded which is sufficient to balance the amount of fresh air brought in. The ratio of fresh air to be recirculated can be fixed the same all the year round or variable to take advantage of outside conditions. Programmed outside thermostats controlling motorized dampers achieve this. Some partial air conditioning systems use 100% recirculation and rely on natural infiltration of air to provide the necessary refreshment.

Processes The air conditioning processes for which items of plant are required are as follows:

1) sensible heating, requiring a heater battery supplied with either hot water, steam or refrigerant or electrically powered,
2) humidification requiring either a water spray or a controlled jet of steam,
3) cooling and dehumidification requiring a cooler battery supplied with either chilled water or refrigerant, the process involving cooling the air to below its dew point so that water condenses out.

Plant items are assembled in the correct combination to suit the particular design problem. The systems divide themselves into three basic types:

—The air systems where the required supply conditions are generated centrally, the air being ducted to the rooms. Hot water is supplied from a central boiler plant and air cooling is provided directly from the evaporator of a refrigeration plant or by means of chilled water from a center refrigeration plant.

—The air and water systems in which only sufficient air required for ventilation purposes is handled centrally and ducted to the rooms, the remaining heat transfer being achieved in room units containing finned coils supplied with heated or chilled water over which some room air is continually being passed. Heated and chilled water is provided from a central boiler and water chilling plant.

—The unitary of self-contained systems where the complete air handling and heat transfer plant is sited in the room, the necessary ventilation air being drawn from outside local to the room usually via a hole in the wall or window.

There are many variations of these basic types of systems.

Possible Problems Faults in air conditioning systems usually result in discomfort for the room occupants and give rise to vociferous complaints, which take one or more of the following forms: too hot, too cold, cold draughts, dry throats, too humid, a feeling of stuffiness, unpleasant odours, too noisy. Most of these can be measured or sensed by the investigator. Faults can be divided into: inherent design faults, installation deficiencies, maintenance deficiencies and malfunction of components.

Inherent design faults include:

1) badly sited air inlets and outlets
2) undersized heat transfer components
3) omission of humidification or de-humidification facilities
4) bad siting of thermostats
5) insufficient provision of fresh air supply.

Installation deficiencies include:

1) badly or wrongly connected ducts
2) air leakage
3) badly sited thermostats
4) control dampers and valves functioning incorrectly.

Maintenance deficiencies include:

1) blocked filters
2) no filters
3) dirty heat transfer surfaces and fan blades
4) control dampers and valves out of adjustment
5) thermostats and their associated controlling mechanisms and circuits out of adjustment
6) heating or cooling plant in poor condition.

The breakdown or malfunction of any item of plant or control systems will lead to the creation of an unsatisfactory environment.

An air-conditioning system operates as an integrated design, an alteration of any one item will unbalance the others thus plant should not be tampered with by inexperienced or unqualified personnel. In most cases the identification of faults and the provision of remedial action should be left to a qualified and experienced environmental/heating and ventilation/air-conditioning engineer.

SOFTWARE CONSIDERATIONS

The range of options available for the control of the working environment is diverse and it is all too easy to make a choice based upon what is the easiest to install and implement, but the control must be capable of providing protection over the operational life of the process. Thus monitoring and maintenance must be built in to the operational side of the system. Continued effectiveness depends upon the awareness of the workforce and their supervisors and managers. Regular checks on the working environment and the performance of the control measure must be made and a procedure laid down and adhered to to record the results of the checks and to implement remedial measures if the system is failing in any way. This can be made easier if the control system is seen as an integral part of the production process warranting the same attention as a production machine.

A planned education and training programme may be required to introduce the operators and management to the importance of the control measure and to explain how it works and what could go wrong and what the consequences are of a malfunction.

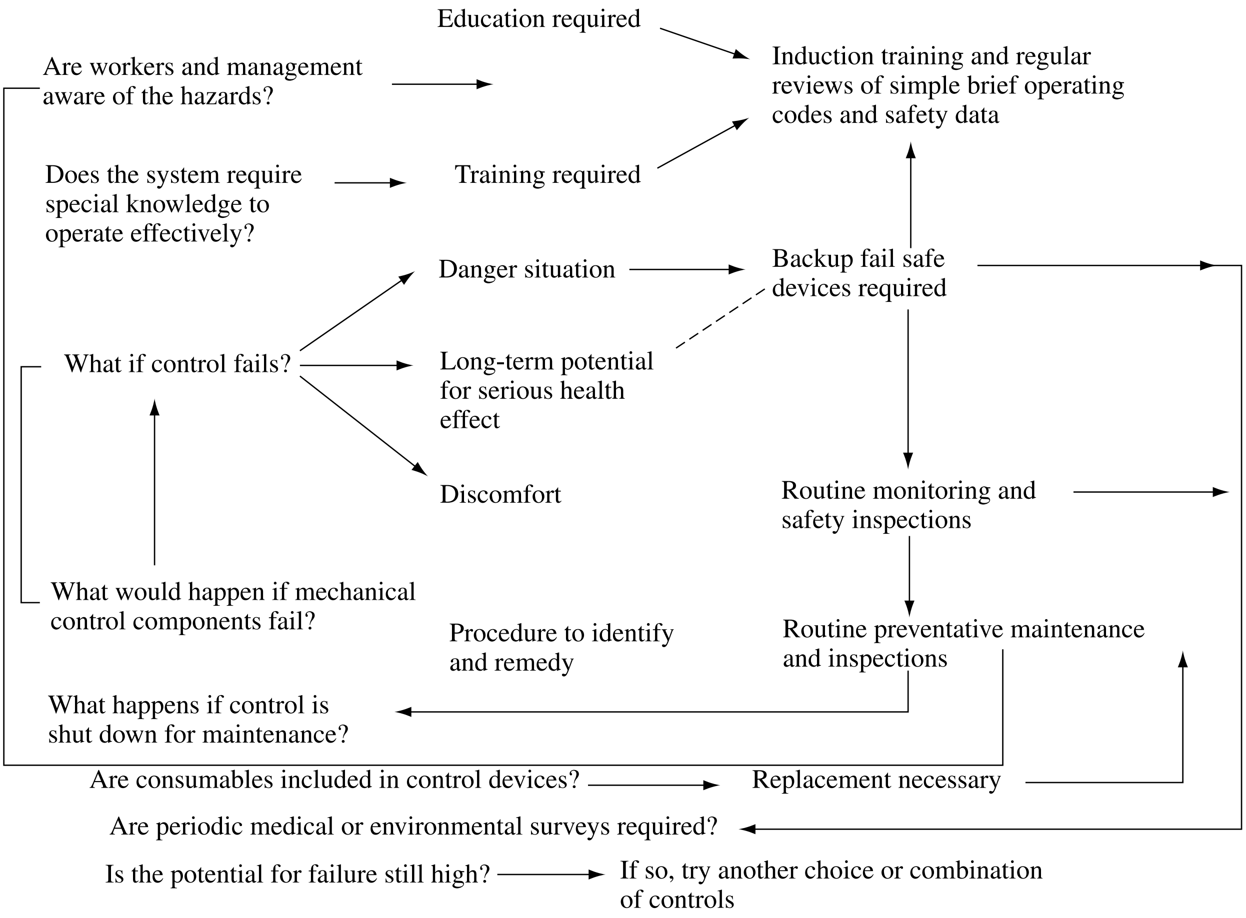


FIGURE 9 Control choice "software" considerations.

To monitor effectiveness of the control may require routine medical and biological checks of workers and environmental monitoring by competent persons and it may be necessary to provide an independent review and situation report to ensure worker acceptability.

It may be useful to introduce a software algorithm (Alesbury, 1980) posing certain questions and providing courses of action following the replies. One such algorithm is offered in Figure 9.

All the above can best be achieved by a genuine desire to make the workplace safe and directions for this can only come from senior management with full consultation and cooperation of the workforce.

REFERENCES

ACGIH, *Industrial Ventilation,* American Conference of Government Industrial Hygienists, Michigan (1995).

Alden, J.L. and Kane, J.M., *Design of Industrial Exhaust Systems,* Industrial Press (1970), New York.

Alesbury, R.J., Health and Safety at Work, **3,** Nov. 1980.

Akbackhanzadeh, F., Metal Construction, **11,** 9, 1979.

Clapp, D.E., Groh, D.S., and Nenadic, C.M., Ventilation design by microcomputer, *Am. Ind. Hyg.,* J37 546 (1982).

CIBS, *IHVE Guide Book C,* Chartered Institution of Building Services 1970, London.

BOHS, Technical Guide Series, A computer program for the design of industrial exhaust systems, British Occupational Hygiene Society, London, 1984.

Crockford, G.W., *Ann. Occ. Hygiene,* **19,** 345, 1976.

Daley, B.B., *Woods Practical Guide to Fan Engineering,* Woods of Colchester Ltd., 1978.

Dallavale, J.M., *Exhaust hoods,* 2nd Ed. (1952), Industrial Press, New York.

Dorman, R.G., Dust Control and Air Cleaning, Pergamon, 1974.

Fletcher, B., *Ann. Occ. Hygiene,* **20,** 141, 1977.

Gill, F.S., *Ann. Occ. Hygiene,* **23,** 1980.

Hemeon, W.C.L., *Plant and Process Ventilation,* Industrial Press (1963), New York.

Houghton, F.C., Gutberlet, C., and Witkowski, E., ASHVE Transactions, **44,** 289, 1938.

Hughes, R.T., Push-pull local exhaust ventilation, Proc. conf. Ventilation 88, London, 1988.

McDermot, H.J., Handbook of Ventilation for Contamination Control, Ann Arbor Science, 1976.

Malchaire, J.B., *Ann. Occ. Hygiene,* **24,** 217, 1981.

FURTHER READING

ACGIH, Ventilation 91, American Conference of Governmental Industrial Hygienists, Cincinnati (1992).

Burgess, W.A. *et al.,* Ventilation for the Control of the Work Environment, Wiley Interscience, New York (1989).

Cralley, L.V. and Cralley, L.A., Plant Practices for Job Related Health Hazard Control, Wiley Interscience, New York 1989).

Garrison, R.P., *Am. Ind. Hyg. Assn. J.,* 44 (1983).

Harrington, J.M. and Gardiner, K., Occupational Hygiene, Blackwell Science, Oxford (1995).

FRANK S. GILL
Consultant
ROGER J. ALESBURY
British Petroleum

INDUSTRIAL WASTE MANAGEMENT

INTRODUCTION

Industries produce large volumes of wastes that may include a wide variety of chemicals containing most toxic pollutants. It has been estimated that there are over 300,000 water-using factories in the United States.[1] As the population grows, the need for manufactured goods will also increase. As a result, the volume of industrial wastes is expected to grow faster than that of municipal wastes. Some of the industrial wastes can be treated jointly in municipal wastewater treatment plants, but others must be pretreated at the source.

In recent years, industrial wastewater management has undergone vast changes. Under legislative mandates and technological advancement, industries are recognizing the need and benefits of using water for several different purposes in descending order of required cleanliness before final treatment and release to the environment. This multiple, cascade, or sequential reuse of water minimizes the need for new water supplies, and reduces and concentrates wastes. A benefit of water and residual reuse may be an economical closed-loop, a zero discharge system that requires minimum make-up water to function.

Presented here is an overview of industrial pollution control legislation and standards, Standard Industrial Classification, industrial waste survey and monitoring, and wastewater treatment systems for selected industries.

INDUSTRIAL POLLUTION CONTROL LEGISLATION AND STANDARDS

Water Pollution Control Legislation

Although water pollution legislative history in the United States had its beginning with the Refuse Act of 1899, the Federal Water Pollution Control Act Amendments of 1972 (PL 92-500) marked the greatest commitment to eliminating pollutants in the nation's lakes, rivers, and streams.[2] The following is a summary of the last 40 years of water pollution control legislation enacted by the Congress of the Unites States.

- Federal Water Pollution Control Act of 1948
- Federal Water Pollution Control Act of 1956
- Federal Water Pollution Control Act Amendments of 1961
- Water Quality Act of 1965
- Clean Water Restoration Act of 1966
- Water Quality Improvement Act of 1970
- Federal Water Pollution Control Act Amendments of 1972
- Clean Water Act of 1977
- Clean Water Act Amendments of 1980
- Clean Water Act Amendments of 1981
- Water Quality Act of 1987

The 1972 and 1977 laws are collectively known as the Clean Water Acts.[3,4] They clearly express a serious national interest in water quality and reflect a strong public commitment to end water pollution. This legislation establishes deadlines for terminating pollution, enforcement provisions by federal, state, and local governments, and a greater federal degree of control over the quality of the nation's waters. The discharge of pollutants into the nation's waters is prohibited unless a permit is obtained under the *National Pollutant Discharge Elimination System* (*NPDES*). There are about 50,000 industrial and 16,000 municipal NPDES permits at the present time.[5] These permits are issued by the states and must be renewed every 5 years. The NPDES permit records effluent limits and spells out requirements for monitoring and recording. Restrictions on amounts of specific pollutants that a given facility may discharge into surface waters are, in general, based on national effluent guidelines. Industry has responded positively to the mandates of the law in meeting the discharge limits based on the *best available* treatment technologies as defined by EPA. Under the NPDES permitting program effluent limitations are established for toxic pollutants for various industrial categories. In a similar manner, *industrial pretreatment standards* are being developed for all pollutants that are discharged into the *publicly owned treatment works* (POTW). Under the pretreatment regulations, two types of federal pretreatment standards are established: (1) prohibited discharges and (2) categorical standards.

Prohibited discharges to sewers or POTWs are those that cause a fire or explosion hazard, corrosion, obstruction, and slug discharges and heat discharges. Categorical standards are developed for those pollutants that are incompatible, that is, those that interfere with the operation of or pass through POTWs, or contaminate the sludge and other residues from POTWs.

Substances considered for categorical standards are those for which there is substantial evidence of carcinogenicity, mutagenicity, and/or teratogenicity; substances structurally similar to aforementioned compounds; and substances known to have toxic effects on human beings or aquatic organisms at sufficiently high concentrations and which are present in the industrial effluents. There are many specific elements or compounds that have been identified as priority pollutants. These include metals, organics, cyanides, and asbestos.

To provide incentives, the Clean Water Act offered federal cost sharing to cover 75 percent of treatment plant construction. In the 1987 amendments, Congress set up a mechanism for states to develop revolving loan funds to pay for future pollution control facilities.[6]

Toxic Substance Control Act

The Toxic Substance Control Act of 1976 (TSCA), was enacted to regulate the introduction and use of new hazardous chemicals.[7] Under TSCA regulations, industry must furnish data on the anticipated production, usage, and health effects of all new chemical substances and mixtures before they are manufactured for commercial distribution. TSCA also regulates the manufacture, processing, use, and final disposal of all chemical substances.

The Resources Conservation and Recovery Act of 1976 (RCRA) restricts the disposal of hazardous waste into the atmosphere, bodies of water, or on land.[8] Thus RCRA was enacted to protect the quality of groundwater, surface water, the land, and the air from contamination by solid wastes.

The Hazardous and Solid Waste Amendments of 1984 (HSWA) emphasized the protection of groundwater through the use of leachate collection (double liners), and monitoring of underground tanks; upgraded criteria for disposing of municipal solid wastes in landfills; and established new requirements for the management and treatment of small quantities of hazardous wastes.[9]

The Comprehensive Environmental Response Compensation and Liabilities Act of 1980 (CERLA), the so-called Superfund legislation, was designed primarily to address the problem or the financial cleaning of abandoned or illegal hazardous waste sites.[10] The Superfund Amendments and Reauthorization Act of 1986 (SARA) provided funds and a timetable and guidance for cleanup standards.[11]

The Pollution Prevention Act of 1990 restates the waste reduction mandate in the 1984 HSWA amendments to RCRA. A pollution prevention office within the EPA was also created by this law along with the expansion of the reporting requirements of the SARA, emergency planning, and community right-to-know provisions. Organizations subject to these provision are required to report the steps taken to achieve waste reduction.[12]

STANDARD INDUSTRIAL CLASSIFICATION

The U.S. Government Standard Industrial Classification (SIC) lists 20 broad industry types bearing two-digit identification numbers. More specific classification of industry is achieved through three and four digit systems. As an example, Standard Industrial Classification 20 is for Food and Kindred Products; 202 for Dairy Products; and 2021, 2022 and 2023 for Creamery Butter, Cheese, and Condensed and Evaporated Milk.[13]

MAJOR CONTAMINANTS—INDUSTRIAL SOURCES AND EFFECTS

The initial step in the rational development of industrial pollution control is to identify and characterize the major industrial contaminants. Generally pollutants can be classified in three basic categories. These categories and parameters are listed below:

(A) Physical Properties

Temperature
Insoluble components
Floating, settleable and suspended matters
Color
Odor
Foamability
Corrosiveness
Radioactivity

(B) Chemical Composition

Organic matter
Inorganic matter
Total dissolved solids
Acid-Base
pH
Acidity
Alkalinity
Nutrients
Nitrogen
Phosphorus
Extractables
Oil
Grease
Reaction
Oxidizing
Reducing
Chlorine demand

(C) Biological Effects

Decomposition
Biodegradable
Nonbiodegradable
Toxic
Organic—phenols, pesticides, polychlorinated biphenyls (PCB), benzidine, etc.
Inorganic—heavy metals and cyanide
Pathogens

Industrial sources as well as major water quality effects of some of these materials are summarized in Table 1.[14]

TABLE 1
Major contaminants—industrial source and effect

Type of contaminant	Industrial source (some examples)	Some major effects
Inorganic salts	Oil refinery, desalination plants, munitions manufacturing and pickle curing	Interferes with industrial usage municipal (drinking water), and agriculture (irrigation water)
Acids and/or alkalies	Chemical manufacturing, tanneries	Corrosion of pipelines and equipment, kills fish
Organic matter	Tanneries, canneries, textile mills etc.	Food for bacteria and thus depletes oxygen
Suspended matter	Paper mills, canneries, etc.	Suffocates fish eggs, degrades stream appearance
Floating solids and liquids	Slaughterhouse, oil refinery	Unsightly, odorous, interferes with oxygen transfer
Heated water	Cooling waters from most industries and power plants	Accelerates bacterial action, lowers total oxygen saturation level
Color	Textile, tanneries, metal finishing, and chemical plants	Objectionable appearance
Toxic chemicals	Munitions manufacturing, metal plating, steel mills, petrochemicals, etc.	Alters stream biota and animal diversity
Microorganism	Pharmaceutical, combined municipal—industrial plants	Unsafe for drinking and swimming
Radioactivity	Nuclear power plants, chemical laboratories	Concentrates in fish, harmful for drinking
Foam-producing matter	Glue manufacturing, slaughterhouse, detergent manufacturing	Aesthetically objectionable

Source: Adapted from Ref (14).

INDUSTRIAL WASTE SURVEY

Industrial waste surveys are conducted to develop knowledge of the waste streams from a specific industry. A survey provides information on sewer lines, waste routing, and material balance. Since an industrial waste survey provides an understanding of the waste flow through the plant, and the potential for water and residual reuse, a majority of the industries find that the survey expenditures yield an excellent return.[15,16]

Sewer Map

Of prime importance for any industry in operation at a site is the development of an up-to-date sewer map showing water, wastewater, and sanitary and storm drains. Locating sewer lines and establishing the manufacturing sources responsible for different waste streams becomes a time-consuming and complex problem in older facilities. Piping diagrams are seldom updated as changes are made over the years. The sewer map should include details such as pipe size; location and type of water supply and drain connections to each processing unit; direction of flow; location of roof and floor drains; manholes; catch basins and control points. To develop a realistic sewer map tracer, studies may be needed. Commonly used tracers are dyes, floats (wood chips, cork floats, and stoppered bottles), and smoke.

Flow Sheet

A flow sheet is prepared for each operation in the entire plant. It should show all raw materials, additives and products, by-products, and liquid and solid wastes. All primary discharges from each process, and the type, period, and duration of each operation should also be indicated on the flow sheet.

Mass Balance

After developing the flow sheet the next step is to obtain the amounts of raw materials, additives, products, and wastes for each operation. From the material balance, the extent of solid and liquid waste characteristics may be determined. This mass balance acts as a check on the waste quantities determined in the preliminary sampling and analysis. It also provides preliminary estimates of flows and parameters to be measured.

Location of Sampling Stations

Sufficient sampling stations should be established to determine the waste load at all of the major processes which contribute wastes. A desirable feature of the sampling station is that the flow be known. If flow is not known, it may be established by use of a flow measurement device. The sampling stations should be easily accessible with adequate safeguards, and wastewater should be well mixed.

Coordination with Production Staff

It is necessary that the efforts of the production staff be fully coordinated during the industrial waste survey. The input and information supplied by the production staff is valuable in identifying the frequency of batch dumps, spills, overflows, and continuous discharges.

SAMPLING AND MONITORING

The basis for industrial pollution abatement programs rests upon information obtained by sampling and monitoring of various waste streams. Serious problems or inaccurate information may result if sampling procedures and test parameters are selected in a careless or naive manner.

Flow Measurement

An essential part of the wastewater sampling and monitoring program is the collection of flow data. A knowledge of flow rate, flow variability, and total flow is essential. A variety of flow measuring devices are available. the selection of the proper measuring method or device will depend on factors such as cost, accessibility, type of flow, and character of waste. A list of many different types of methods and devices commonly used for wastewater flow measurement are given in Table 2.[2]

Sampling

It is essential that waste stream samples be truly representative of the waste discharge. Waste conditions may vary both in magnitude and composition over a 24-hour period. Therefore, care should be taken in selecting the method of sampling, frequency and duration, sample handling, and parameters to be measured. All these items are briefly discussed below.

Method of Sampling

Two most common methods of sampling are known as *grab samples* and *composite samples.* Either may be obtained manually or automatically. Grab samples are single batch samples taken at a given time. Composite or integrated samples are taken at constant time intervals usually over a 24-hour period, then mixed in proportion to flow at the time of sampling to obtain one representative sample of the total flow for the day.

Grab samples provide valuable information at low cost. They are recommended where: (1) condition or quality remains relatively uniform over long periods; (2) the effect of slug loads are desired; (3) concentration of certain constituents are needed for adjustment of chemical feed; (4) flow is intermittent; and (5) the sample requires immediate analysis due to instability of the constituents.

Composite sampling is done to obtain the average quality data for the day. Continuous samplers that take a sample volume in proportion to the flow are desirable. Another type of composite sampler removes samples on the hour and deposits it into different bottles. A volume of sample from each bottle is mixed manually in proportion to flow to obtain the composite

TABLE 2
Types of flow measurement devices commonly used for measuring wastewater discharges

Flow Measurement Devices	Principle of Flow Measurement
1. For pressure pipes	
a. Venturi meter	Differential pressure is measured
b. Flow nozzle meter	Differential pressure is measured
c. Orifice meter	Differential pressure is measured
d. Electromagnetic meter	Magnetic field is induced and voltage is measured
e. Turbine meter	Uses a velocity driven rotational element (turbine, vane, wheel)
f. Acoustic meter	Sound waves are used to measure the velocity
2. For open channels	
a. Flumes (Parshall, Palmer-Bowlus)	Critical depth is measured at the flume
b. Weirs	Head is measured over a barrier (weir)
c. Depth measurement	Float is used to obtain the depth of flow in the sewer, and velocity is calculated from slope
d. Acoustic meter	Uses sound waves to measure velocity and depth
3. Computing flow from freely discharging pipes	
1. Pipes flowing full	
a. Nozzles and orifices	Water jet data is recorded
b. Vertical open-end flow	Vertical height of water jet is recorded
2. Pipe partly flowing full	
a. Horizontal open-end pipe	Dimensions of free falling water jet are obtained
b. Open flow nozzle (Kennison nozzle or California pipe method)	Depth of flow at free falling end is determined
4. Miscellaneous methods	
a. Dilution method	A Constant flow of a dye tracer is used
b. Bucket and stopwatch	A calibrated bucket is used and time to fill is recorded
c. Pumping rate	Constant pumping rate and pumping duration are recorded

average sample for the day. Individual hourly samples are the grab samples representing the condition at that instant.

Frequency and Duration of Sampling

The frequency of sampling depends on the flow rate, wastewater characteristics, and variability in quality and volume. The expected range in flow rate and waste concentration should be determined by a preliminary survey. Although most of the time the frequency is one sample per hour, the frequency for highly variable waste streams could be as high as one sample every three minutes.

An intensive plant survey will generally last between five to ten days of normal plant operation. Since the treatment facilities must be designed to treat the highest pollution load expected, it is important to consider seasonal variations (if applicable).

Sample Handling

In order to obtain a representative sample it is necessary for the sampling point to have sufficient hydraulic turbulence. Sufficient volume of the sample must be obtained to perform all analyses planned. The minimum volume of grab sample should be between one to two liters. Sample containers and sampling device should be clean and uncontaminated. Before the sample is taken, the container should be rinsed several times with the wastewater. Each sample should be labeled with an identification card containing the following information: date and time, sample location, method of sampling (grab or composite), and notation of information obtained from field analyses of parameters that may change before laboratory analyses are made (temperature, pH, appearance).

The sample should be analyzed as quickly as possible. Storage should be in a manner that insures that the characteristics to be analyzed are not altered. Refrigeration in most instances is necessary. In some cases a special chemical may have to be added to prevent changes in chemical or biological characteristics.

Parameters Measured

A major item in any industrial monitoring program is the cost of analytical measurement. It is important that the parameters be properly selected to represent inplant waste streams; and waste characterization, treatment and reuse requirements. In some cases the necessary analyses will be time-consuming and relatively expensive. In many cases, an alternative analytical technique may be used that are less expensive. Major constituents of interest in industrial pollution monitoring are listed in Table 1.

Analytical Considerations

Good analytical procedures are of the utmost importance in a monitoring program. The basic references for wastewater analytical procedures and techniques are EPA publications, Standard Methods for Examination of Water and Wastewater, and ASTM Standards.[17–20] The analysis may fall into several major categories including routine wet chemistry, selective ion electrodes, automated wet chemistry, and bioassary tests.

Data Analysis

Data obtained through a well-planned and executed monitoring program will provide valuable information for waste process selection, plant design and operation, and assist the industry in evaluating the manufacturing process. The monitoring program may result from changes in chemical use and/or industrial process, inplant spills or dumping of baths. Variability of the parameters may be random or cyclic. The data should be analyzed to establish the fluctuations with time, location, work shifts, and type of operation.

Statistical techniques should be used to develop relationships such as average or mean, standard deviation and extreme conditions, and regression coefficients. With a knowledge of basic probability theory and the use of statistical techniques, such as least squares, curve fitting, analysis of variance, regression and correlation analysis, chi-squared goodness of fit, and others, it is possible to construct mathematical models and curves for almost any level of precision desired. Such techniques help to evaluate information having wide variations, so that an estimate of the best value of the parameter being measured can be developed. For more information the reader should consult several excellent references on statistical methods.[21–24]

TREATMENT PROCESSES

Industrial wastewater treatment utilizes a number of unit operations and processes to achieve the desired degree of treatment. The collective treatment schematic is called a flow scheme, flow diagram, flow sheet, process train, or flow schematic. Basic considerations for developing a flow scheme include: (1) the manufacturing process and its flow scheme; (2) characteristics of wastewater streams and degree of treatment; (3) requirements of the regulatory agency; (4) construction cost; (5) level of expertise of treatment plant operation personnel; and (6) operation and maintenance costs.[2]

The wastewater treatment facilities are designed to process liquids and solids or sludges. It is essential that in-plant waste management and control techniques be utilized if cost-effective methods of liquid and solids treatment are to be attained. Some application of in-plant controls include waste reduction, waste segregation, water conservation and recycle, and process modifications. Generally, controlling wastes within the plant is much more cost-effective than implementation of treatment. Certain innovative processes that reduce waste production have resulted in improved process yield or by-product recovery and utilization.

Waste Reduction

A basic materials balance procedure should be developed that accounts for all materials that enter and leave an individual department or processing area. Some specific waste reduction measures include: (1) materials balance accounting; (2) improved process control; (3) more effective cleanup procedures; (4) regular preventive maintenance; and (5) by-product recovery. Poor maintenance and cleanup procedures in many industrial plants are a major source of wasteload generation.

Product spills should be avoided, and when they occur, the waste should be handled in a manner that will minimize the contribution of wasteload to the plant. Proper housekeeping procedures should be developed and implemented.

Waste Segregation

It is important to segregate all wastes in accordance with the physical and chemical properties of the contaminants, potential to react, and their treatability. Important properties are suspended solids, acidity, basicity, organic matter and biodegradability, volatility, and toxic organics and inorganics. Efforts should be made to minimize dilution of waste streams prior to treatment and/or recovery of residuals. Diluted streams should be handled separately.

Water Conservation and Recycle

Water conservation and recycle consist of minimizing the raw water supply and maximizing the amount or wastewater reuse within the plant. The net effect of water conservation and recycle is the reduction of wastewater volume and to concentration of wasteloads.

Each process should be investigated to minimize water supply requirements and the potential for substituting recycled or reclaimed wastewater for fresh water. Water quality requirements for most industrial processes will govern the feasibility or extent of water reuse. However, in every industry using water for washing, rinsing, or cooling, some form of countercurrent flow or recycling can probably be implemented for maximum reuse. Some specific recommendations for water conservation and recycle include: (1) dry (nonwater) cleaning method, (2) water meters at each department to make operators conscious of usage, (3) automatic valves that close when water is no longer required, (4) regular preventive maintenance programs that include leak surveys, and (5) cleanup with high-pressure, low-volume rinse sprays.[16]

Process Modification

In-plant process modifications, though more costly to implement than simple operational changes, may be very effective in controlling wasteload generation. Process modifications may consist of individual unit process changes, or changes to a complete process line. The cost-effectiveness of such modifications will depend on the relative reduction in wasteloads. In older industrial plants, required process modifications may be accomplished during the installation of newer and more efficient equipment. In the design of new plants, each process should be evaluated with respect to use of more efficient equipment for maximum water conservation and minimum wasteload generation.

Liquid Treatment Systems

The removal of various contaminants from the liquid depends on the nature of the impurities and their concentrations. Coarse and settleable inorganic and organic solids are generally removed in sedimentation facilities. Oil and grease is removed by skimmers. The removal of dissolved organics is readily achieved in biological or chemical or physicochemical treatment processes. For example, chemical oxidation-reduction reactions and precipitation are used for removal of heavy metals. Carbon adsorption can be used to remove refractory organics. Ion exchange and membrane processes are suitable for demineralization and byproduct recovery. Major physical, chemical, and biological treatment processes used for liquid treatment are summarized in Table 3. For extensive presentations on these processes the readers should refer to several excellent publications that provide detailed discussions on the theory and design of wastewater treatment processes.[2, 25–31]

Sludge and Brine Processing and Disposal

Safe handling and disposal of residues and brines produced in various treatment units are of equal importance. Solids portions include screenings, grit, scum, organic and inorganic sludges. Brines (high mineral concentrates) are produced from ion exchange, reverse osmosis, and electrodialysis units. In general, the sludge processing and disposal methods include thickening, stabilization, and dewatering or evaporation of liquid. Several of the unit operations and processes used for sludge and brine handling and disposal are illustrated in Figure 1.

MAJOR INDUSTRIAL WASTES AND TREATMENT

It should be recognized that to truly understand the waste problem of any industry, a lifetime serious effort on the part of a qualified waste engineer may be required. Production methods of each industry are normally different. As a result, the treatment process train needed for the waste streams should be individually designed. The purpose of this section is to review the origin and characteristics of wastewater and methods of wastewater treatment for six major industries. In the following discussion, the reader will find a brief summary of the major liquid wastes, their origin, characteristics and current methods of treatment for food, paper and allied products, chemical, petroleum, metals, and power generation industries. For a detailed study on these topics the readers should refer to several references on the subject included in the list of references.[32–35]

Food Industry

There are approximately 50,000 food processing plants in the total food processing industry in the United States.

Food processing is a low-profit margin, highly competitive industry. Water is required for washing, blanching,

TABLE 3

Major physical, chemical, and biological treatment unit operations and processes used for liquid treatment

Unit Operations and Processes	Principal Applications
Screening	Racks or bar screens are the first step in wastewater treatment. They are used to remove large objects.
Flow equilization	Used to dampen the flow rate and mass loading.
Skimming	Used for oil separation where free oil is floated to the surface of a tank and then skimmed off. The oil separator specified by *American Petroleum Institute* is commonly used.
Dissolved Air Flotation (DAF)	Process used to remove suspended solids, and oil and grease from the waste stream. Uses dissolved air to produce line bubble that float the solids.
Foam separation	Process used to remove minute concentrations or refractory organics and heavy metal ions. Contaminants are carried up by rising bubbles to the pool surface where they are deposited as the bubbles exit.
Sedimentation	Use to remove settleable solids.
Neutralization	Process used to neutralize acidic or alkaline waste streams prior to chemical and/or biological treatment.
Precipitation	Dissolved solids in solution are chemically transformed into insoluble form. Process is extensively used to precipitate phosphorous and heavy metals. Heavy metals are generally precipitated as hydroxide through the addition of time or caustic to a pH of minimum solubility.
Chemical Oxidation	Used to oxidize pollutants to terminal end products. Common oxidants are Cl_2, O_3, H_2O_2, and $KMnO_4$. Oxidation of iron and manganese is used for precipitation. Cyanides, sulfides and many organics arc destroyed by oxidation.
Chemical Reduction	Used to precipitate certain ions from solutions, e.g., hexavalent chromium. Also many oxidizing agents are destroyed by reduction. Common reducing agents are ferrous sulfate, sodium metabisulfite, and sulfur dioxide.
Coagulation	Used to agglomerate suspended materials so that efficient settling can take place. Commonly used chemicals are alum, iron salts, and polymers. The process consists of chemical feed, flash mix and flocculation basins.
Suspended growth biological reactor	Biological process using suspended biomass to remove dissolved organics. Principal variation is an activated sludge process.
Attached growth biological reactor	Biological process using attached biomass to remove dissolved organics. Principal variations are a trickling filter and a rotating biological contactor (RBC).
Anaerobic Processes	Attached or suspended microbial process operated in the absence of oxygen. Normally used to treat high strength organic wastes. Methane is produced as an energy source. Sludge (biomass) production is small.
Oxidation Pond	Large basins (normally earthern) used for waste storage and treatment. Treatment is achieved by natural processes of settling and biological decomposition.
Aerated Lagoon	Basins (normally earthern and frequently with plastic liners) with aeration equipment where waste are aerated over long periods of time. The waste characteristics are altered by biological oxidation.
Land Treatment	Organic waste is applied over land for treatment. Common application methods are slowrate irrigation, rapid infiltration, and overland flow.
Nitrification	Process is used to convert ammonia nitrogen to nitrate nitrogen. It can be achieved in suspended or attached growth biological reactors.
Denitrification	Nitrite nitrogen and nitrate nitrogen are reduced to nitrogen gas by microorganisms. Denitrification is achieved under anaerobic condition in suspended or attached growth reactors. An organic source such as methanol is needed.
Disinfection	Process used to reduce the number of water borne pathogens in water. Normally applied to effluents of wastewater treatment plants containing microbes. Chlorination and ozonation are the most common methods.
Ammonia stripping	Ammonia gas is air stripped from the wastewater, normally by using stripping tower. High pH values are required.
Breakpoint Chlorination	Ammonia nitrogen is oxidized to nitrogen gas by breakpoint chlorination in a mixing basin.
Filtration	Used to polish the effluent by removing total suspended solids and turbidity. Biological activity in the filter bed may reduce some of the biochemical oxygen demand (BOD).
Carbon adsorption	Used to remove soluble refractory organics from wastewater effluent.
Ion Exchange	A demineralization process in which the cations and anions in wastewater are selectively exchanged for the ions in an insoluble resin bed. When the resin capacity is used up it is regenerated by using high concentrations of the original ions that are exchanged for the polluted ions attached to the resin. Flow from regeneration is composed of highly concentrated brine.
Reverse osmosis or ultrafiltration	A demineralization process applicable to production of high-quality water from effluent. The water permeates through semipermeable membrane at high pressure, producing high-quality water in one stream and a high concentration of mineral ions in another.
Electrodialysis	A demineralization process where electrical potential is used to transfer the ions through ion-selective membranes. Produces two streams; one high in mineral content and the other free of minerals.

THICKENING	STABILIZATION	CONDITIONING	DEWATERING	DISPOSAL
Concentrates solids	Reduces pathogens, and eliminate offensive odors	Enhances water removal	Removes moisture and produces sludge cake	Used for ultimate disposal of residues
1. Gravity	1. Chlorine oxidation	1. Chemical	1. Vacuum filter	1. Evaporation of brine
2. Flotation	2. Line stabilization	2. Elutriation	2. Filter press	2. Incineration
3. Centrifugation	3. Heat treatment	3. Heat treatment	3. Horizontal belt filter	3. Wet Oxidation
	4. Aerobic digestion		4. Centrifugation	4. Pyrolysis
	5. Anaerobic digestion		5. Drying beds	5. Composting
				6. Land filling
				7. Deep well injection

FIGURE 1 Alternative unit operations and processes for sludge and brine processing and disposal.

pasteurization, cleaning process, equipment, and cooling of final product. The waste is generally characterized by high BOD and suspended solids. The estimated total wastes from food industry for the United States in 1968 was:

320 million cubic meters of wastewater discharged
360 million kg of BOD generated
180 million kg of suspended solids generated
8 billion kg of solids residuals

Table 4 is a brief summary of major liquid wastes, their origin, characteristics and applicable methods of treatment for several major industries in the food industrial category.

Paper and Allied Products Industry

Pulp is produced by mechanically or chemically processing wood or other vegetative materials to extract usable cellulosic fibers as an aqueous slurry. The pulp slurry may be used directly in paper making or it may be shipped elsewhere for processing into paper products. The pulp and paper industry is the ninth largest industry in the United States. The major group of the industries in the pulp and paper industry are pulp mills, paper mills, paperboard mills, miscellaneous converted paper products, paperboard containers and boxes, and building paper and board mills.

The fundamental industrial operations are divided into two major categories: pulp mill and paper mill. The pulp mill operation includes wood preparation, pulping, deinking, pulp washing, screening and thickening, and bleaching. Paper mill operations include stock preparation, paper machine operation, and finishing. Table 5 provides a summary of these operations, origin of major wastes, major characteristics, and treatment methods.

Chemical Industry

The chemical industry is highly diversified and supplies products for virtually every other industry. The number of synthetic compounds manufactured is estimated to range between 500,000 and 600,000, and a host of new products is introduced every year. The chemical products may be used as primary, intermediate, or finished products.

The chemical processing industry has a variety of special pollution problems due to the vast number of products manufactured. The treatment processes combine processing, concentration, separation, extraction, by-product recovery, destruction, and reduction in concentration. The wastes may originate from solvent extraction, acid and caustic wastes, overflows, spills, mechanical losses, etc. Origin of major wastes, characteristics, and treatment and disposal of wastes in several major chemical industries are summarized in Table 6.

Petroleum Industry

The petroleum industry is one of the most important manufacturing industries in the country. It is a complex industry utilizing complex combination of interdependent operations engaged in the storage and transportation, separation of crude molecular constituents molecular cracking, molecular rebuilding, and solvent finishing to produce petrochemical products. Each process is responsible for production of many waste streams containing oil, chemical oxygen demand (COD), phenol, sulfide, chloride, and others. Treatment may involve oil separation, precipitation, adsorption, and biological treatment. The refining operations may be divided into many major categories. Wastewater characteristics, origin of major wastes, characteristics, and treatment and disposal methods from several major processes are summarized in Table 7.

Metals Industry

Primary metal processing and fabricated metal products manufacturing comprise the metals industries. The most important end uses of the products of the metals industries are automobiles, machinery, appliances, electrical equipment, structures,

TABLE 4
Summary of industrial wastes from major food industry

Industries Producing Waste	Origin of Major Wastes	Major Characteristics	Major Treatment and Disposal Methods
Canned goods	Trimming, culling, juicing and blancing of fruits and vegetables	High in suspended solids, colloidal and dissolved organic matter	Screening, lagooning, soil absorption by spray irrigation
Dairy products	Dilutions of whole milk, buttermilk, and whey	High in dissolved organic matter, mainly protein, fat, and lactose	Biological treatment aeration, RBC, trickling filtration, activate sludge, whey recovery
Brewed and distilled beverages	Steeping and pressing of grain; residue from distillation of alcohol; condensate from stillage evaporation	High in dissolved organic solids, containing nitrogen and fermented starches or their products	Recovery, concentration by centrifugation and evaporation, RBC; trickling filtration; use in feeds; anaerobic biological treatment followed by aerobic biological treatment
Meat and poultry products	Stockyards; slaughtering of animals; rendering of bones and fats; residues in condensates; grease and wash water; picking chickens	Organic matter, blood, other proteins, and fats and oils	Screening, settling and/or floatation, trickling filtration, RBC, activated sludge, recovery for animal food products
Animal feedlots	Excreta from animals	High in organic suspended solids and BOD	Land disposal and anaerobic lagoons
Beet sugar	Transfer, screening, and juicing waters; drainings from lime sludge; condensates after evaporator; juice and extracted sugar	High in dissolved and suspended organic matter, containing sugar and protein	Reuse of wastes, coagulation, lagooning, activated sludge, trickling filter, anaerobic process
Cane sugar	Spillage from extraction, clarification, evaporation cooling, and condenser waters	Variable pH, soluble organic matter with relatively high BOD of carbonaceous nature	Neutralization, recirculation, chemical treatment, some selected aerobic oxidation

TABLE 5
Summary of industrial wastes from several major operations in pulp and paper industry

Major Industrial Operations	Origin of Major Wastes	Major Characteristics	Major Treatment and Disposal Methods
Pulp mill			
Wood preparation	Log transportation and storage, debarking and chipping	Solid wastes, hydraulic debanking uses water	Incineration, water is recycled through lagoons
Pulping	Mechanical pulping, kraft pulping, sulfide pulping	Suspended solid, BOD, liquors containing high BOD, sulfide, mercaptans, high pH	Sedimentation, recovery if sodium hydroxide and sodium sulfide, aeration
Deinking	Removal of ink	Contains dirt and clay fillers, chemicals and usually alkaline	Sedimentation and coagulation
Pulp washing, screening, and thickening	Washing action, centrifugal cleaning, gravity and vacuum thickening	Liquor contains pulp fibres	Pulp fibres are recorded and recycled
Bleaching	Bleaching of pulp fibers by chlorine solution	Contains chlorine, low pH, high TDS	Dechlorination, neutralization
Paper mill			
Stock preparation	Addition of filter, colors, chemicals, and screening	Color, clay fillers, chemicals, high TDS	Coagulation and reuse
Paper machine	Removal of free water	Contains fibres and fillers. Known as white water. TSS, TDS and BOD	Fibers reclaimed, liquid wastes treated by coagulation, activated sludge
Finishing and converting	Surface improving, size cutting. Mostly dry process	Little or no liquid waste, mainly solid waste	Recycled in pulping operation, incineration

furniture, and containers. There are approximately 35,000 establishments in the United States. The industry uses approximately 32 billion cubic meters of water per year. Four percent of the plants use 92 percent of the water. These industries generate wastewaters which vary in quantity and quality. Table 8 provides the reader with a brief summary of the major liquid wastes, their origin, characteristics, and methods of treatment in four major metals industries.

TABLE 6
Summary of industrial wastes from several major chemical industries

Industries Productng Wastes	Origin of Major Wastes	Major Characteristics	Major Treatment and Disposal Methods
Acids and Alkalies	Dilute wash waters; many varied dilute acids and bases	Low or high pH, low organic content	Unflow or straight neutralization, coagulation and sedimentation
Detergents	Washing and purifying soaps and detergents	High in BOD and saponified soaps	Flotation and skimming, precipitation with $CaCl_2$
Explosives	Washing TNT and guncotton for purification, washing and pickling of cartridges	TNT, colored, acid, odorous, and contains organic acids and alcohol from powder and cotton, metals, acid, oils, and soaps	Flotation, chemical Precipitation, biological treatment, aeration, chlorination of TNT, neutralization, adsorption
Pesticides	Washing and purification of products	High organic matter, benzene ring structure, toxic to bacteria and fish, acid	Dilution, storage, activated-carbon adsorption, alkaline chlorination
Phospate and phosporus	Washing, screening, floating rock, condenser bleed-off from phospate reduction plant	Clays, slimes and oils, low pH, high suspended solids, phosphorus, silica and fluoride	Lagooning, mechanical clarification, coagulation and settling of refined waste
Formaldehyde	Residues from manufacturing synthetic resins and from dyeing synthetic fibers	Normally high BOD and formaldehyde, toxic to bacteria in high concentrations	Trickling filtration, absorption on activated charcoal
Plastics and resins	Unit operations from polymer preparation and use; spills and equipment washdowns	Acids, caustic, dissolved organic matter such as phenols, formaldehyde, etc.	Discharge to municipal sewer, reuse, controlled-discharge
Fertilizer	Chemical reactions of basic elements. Spills, cooling waters, washing of products, boiler blowdown	Sulfuric, phosphoric, and nitric acids; minerals elements, P, S, N, K, Al, NH_3, NO_3	Neutralization, detain for reuse, sedimentation, air stripping of NH_3, lime precipitation
Toxic chemicals	Leaks, accidental spills, and refining of chemicals	Various toxic dissolved elements and compounds such as Hg and PCBs	Retention and reuse, change in production, neutralization and precipitation, carbon adsorption

TABLE 7
Summary of wastes generated from various oil refinery operations

Major Industrial Operations	Origin of Major Wastes	Major Characteristics	Major Treatment and Disposal Methods
Crude oil and product storage	Primary fractionation of oil and water, spills and leakages	High concentrations of emulsified oil, COD, TSS	API separation, DAF, settling, aeration
Crude desalting	Chemical desalting, heating and gravity separation of oil	Emulsified and free oil, ammonia, phenol, sulfide, TSS, high BOD and COD	API separation, DAF, activated sludge, carbon adsorption
Cracking	Thermal cracking or catalytic cracking, fractionation, steam stripping, and overhead accumulators or fractionators	BODs, COD, ammonia, phenol, sulfides, cyanides, and alkalinity	Chemical oxidation, biological treatment carbon adsorption
Polymerization	Catalytic reaction, acid, removal action, and gas stabilizer	Alkaline waste stream, high in sulfide, mecaptans, and ammonia	Acid catalysts recycled, carbon adsorption
Alkylation	Catalytic reaction caustic and water wastes, neutralization of hydrocarbon streams	Oil, sulfides, TSS, fluoride	Neutralization, chemical oxidation, sedimentation

TABLE 8

Origin, characteristics, and treatment of wastes in four major metal industries

Industries Productng Wastes	Origin of Major Wastes	Major Characteristics	Major Treatment and Disposal Methods
Steel	Coking of coal, washing of blast-furnace flue gases, and pickling of steel	Low pH, acids, phenol, ore, coke, limestone, alkali, oils, mill scale, and fine suspended solids	Neutralization, recovery and reuse, chemical coagulation
Iron-foundry products	Wasting of used sand by hydraulic discharge	High suspended solids, mainly sand; some clay and coal	Selective screening, drying of reclaimed sand
Metal container	Cutting and lubricating metals, cleaning can surface	Metal fines, lubricating oils, variable pH, surfactants, dissolved metals	Oil separation, chemical precipitation, collection and reuse, lagoon storage. Final carbon adsorption
Metal-plated products	Stripping of oxides, cleaning and plating of metals	Acid, metals, toxic, low volume, mainly mineral matter	Alkaline chlorination of cyanide; reduction and precipitation of chromium; lime precipitation of other metals

Power Generation Industry

The power generation industry uses large volumes of water for cooling purposes. It is estimated that total cooling water demand for power generation is in excess of 400 billion cubic meters per year. This is approximately one-fifth of the total runoff in the United States.[33] Direct discharge of cooling water in streams, lakes, and estuaries has an impact on water quality and the ecosystem in three ways: it pollutes the water thermally, it subjects aquatic organisms to a variety of physical stresses as it is pumped through the plant, and it pollutes the water chemically.

Types of Cooling Systems

Basically, there are three ways of dissipating heat: once through cooling systems, cooling ponds, and cooling towers. The average condenser water requirement and consumptive use for different cooling systems are summarized below:[36]

Condenser requirement	120 liters per KWH
Consumptive use:	
once through	0.7 liter per KWH
cooling ponds	1.0 liter per KWH
cooling tower	1.4 liter per KWH

A general discussion on three types of heat dissipating systems is given below:[37]

Once Through Cooling The most common and least expensive method for removing heat from power plant condensers is a once-through-cooling system. The system requires large intake structures and has the largest environmental impact. Aquatic organisms are forced through the condenser tubes resulting in high mortality. The 2 to 5°C warmer discharge upsets stratification, lowers dissolved oxygen, increases the metabolic rate of aquatic animals, and increases undesired bacteria.

Cooling Ponds The cooling water is circulated through cooling ponds in a closed-loop. Cooling occurs by evaporation, conduction, and long-wave radiation. Land area of the ponds are generally 0.5–1 hectare/MW.

Cooling Towers Cooling towers are used to eliminate discharge of heated water into natural bodies of water. There are two types of cooling towers: natural-draft and mechanical-draft. Natural-draft cooling towers rely on natural air circulation to dissipate waste heat to the atmosphere. The mechanical-draft cooling towers use induced drafts to promote evaporative heat transfer. Both require heat transfer between water and air. Dry cooling towers use a temperature driving force between the water inside the tubes and air flowing past the tubes.

Wastewater Treatment

Chemical discharges from the power-generation industry result from makeup water treatment; chemicals added to plant cooling water or steam generator; products of corrosion, erosion, wear, or chemical reaction from plant system; and residuals from pollution control equipment. The chemicals generally controlled in effluent discharge permits are: pH, PCBs, TSS, oil and grease, heavy metals, free and total residual chlorine, phosphorus, and corrosion inhibitors.[34] Treatment methods generally used for removal of contaminants include: neutralization, coagulation, and precipitation of corrosion inhibitors, TSS, heavy metals and phosphorus; flotation or American Petroleum Institute (API) separator for oil and grease; ion-exchange or reverse osmosis for TDS control; and evaporation for concentration and residue recovery and disposal.

REFERENCES

1. Henry Lee, *Legislation and Standards: Status, Trends Significance,* Chapter 1, “Industrial Wastewater Management Handbook”, H.S. Azad (Editor), McGraw-Hill Book Co., New York, 1976.
2. Qasim, S.R., Wastewater Treatment Plant: Planning Design and Operation, Technomic Publishing Company, Inc. Lancaster, PA, 1994.
3. *Federal Water Pollution: Control Act Amendment of 1972* (PL 92-500), 92nd U.S. Congress, October 18, 1972.
4. *Clean Water Act* (PL 95-217), 95th U.S. Congress December 27, 1977.
5. Wentz, C.A., *Hazardous Waste Management,* McGraw-Hill Book Co., New York, 1995.

6. *Water Quality Act of 1987* (PL 100-4), 100th Congress, February 4, 1987.
7. *Toxic Substance Control Act* (PL 99-519), 99th U.S. Congress, October 22, 1986.
8. *The Resource Conservation and Recovery Act* (PL 89-272), 89th Congress, 1976.
9. *The Hazardous and Solid Waste Amendments of 1984* (PL 98-616), 98th U.S. Congress, November 8, 1984.
10. *The Comprehensive Environmental Response Compensation and Liabilities Act of 1980* (CERLA), (PL 96-510), 96th U.S. Congress, 1980.
11. *The Superfund Amendmenls and Reauthorization Act of 1986* (PL 99-499), 99th U.S. Congress, October 17, 1986.
12. *Pollution Prevention Act of 1990* (PL 100-389), 100th U.S. Congress, 1990.
13. "Standard Industrial Classification Manual", Executive Office of the President, Office or Management and Budge, NTIS, PB 87-100012, 1987.
14. Nemerow, N.L., *Liquid Wastes from Industry,* The Encyclopedia of Environmental Science and Engineering, Gordon and Breach, Science Publishers, New York, Second Edition, 1983, pp. 560–582.
15. U.S. Environmental Protection Agency, "Handbook for Monitoring Industrial Wastewater", *Technology Transfer,* August 1973.
16. Task Force on Pretreatment, "Pretreatment of Industrial Wastes", POP FD-3, *Water Pollution Control Federation,* Washington, D.C., 1981.
17. Environmental Protection Agency, "Method for Chemical Analysis of Water and Wastes", *Technology Transfer,* EPA 16020-07/71, 1971.
18. Environmental Protection Agency, "Handbook for Analytical Quality Control in Water and Waste Water Laboratories", *Technology Transfer,* June 1972.
19. APHA, AWWA, and WPCF, "Standard Methods for Examination of Water and Wastewater", *American Public Health Association, Washington, D.C.,* 19th Edition, 1995.
20. Devore, J.L., Probability and statistics for engineering and the sciences, Duxbury Press, Belmont, 4th edition, 1995.
21. Neville, A.M. and J.B. Kennedy, *Basics Statistical Methods for Engineers and Scientists,* 4th Printing, International Textbook Co., Scranton, PA 1994.
22. Dowdy, S. and S. Wearden, "Statistics for Research", John Wiley & Sons Inc., 1991.
23. Snedeor, G.W. and W.G. Cochran, "Statistical Methods", Iowa State University, Ames, Iowa, 1989.
24. Graham, A.T., Statistics/Alan Graham, NTC Publishing Group, Chicago, IL, 1994.
25. De Renzo, D.J. (Editor), "Pollution Control Technology for Industrial Wastewater", Noyes Dala Corporation, Park Ridge, N.J., 1981.
26. Dyer, Jon C., A.S. Vernick and H.D. Feiler, "Handbook of Industrial Wastes Pretreatment", Water Management Series, Garland STPM Press, New York, 1981.
27. Theodore, L. and Y.C. McGuinn., Pollution Prevention, Van Nostrand Reinhold, N.Y., 1992.
28. Sell, N.J., Industrial Pollution Control: Issues and Technologies, 2nd Edition, Van Nostrand Reinhold, N.Y., 1992.
29. Conway, R.A. and R.D. Ross, "Handbook of Industrial Waste Disposal", Van Nostrand Reinhold Co., New York, 1980.
30. Eckenfeller, W.W., "Industrial Water Pollution Control" *McGraw-Hill Book Co.*, 1989.
31. Metcalf & Eddy, Inc., "Wastewater Engineering: Treatment Disposal and Reuse", McGraw-Hill Book Co., 3rd Edition, 1994.
32. Azad, H.S. (Editor), "Industrial Wastewater Management Handbook", McGraw-Hill Book Co., New York, 1976.
33. U.S. Department of Interior, "The Cost of Clean Water", Industrial Waste Profile Series, FWPCA, Publication No I.W.P. 1 to I.W.P. 10, 1967.
34. U.S. Environmental Protection Agency, "Development Documents for Effluent Limitations Guidelines and Standards for Point Source Category", Effluent Guidelines Division, U.S. EPA, Washington, D.C., 1979.
35. U.S. Environmental Protection Agency Technology Transfer Seminar Publications of Various Industries, U.S. EPA, Environmental Research Information Center, Technology Transfer, Cincinnati, Ohio.
36. Krenkel, P.A. and F.L. Parker, "Biological Aspects of Thermal Pollution", Vanderbilt University Press, 1969.
37. Wright, J.H. and S.J. Dea, *Water Pollution Control in the Power Generation Industry,* Chapter 10, "Industrial Wastewater Management Handbook", H.S. Azad (Editor), McGraw-Hill Book Co., New York, 1976.

CLINTON E. PARKER
SYED R. QASIM
University of Texas at Arlington

INSTRUMENTATION: WATER AND WASTEWATER ANALYSIS

INTRODUCTION

In the observation of our pollution problems there seems to be an attitude of separation on the part of the human observer from the polluted lake or stream. In reality water is so pervasive in our life; it is such a large part of our bodily mass and surrounds us in clouds, fog, rain, snow, lakes, rivers, and oceans. We seem to accept its presence without much thought. However, we all are part of the ecosystem and, therefore, pollution is an intimate condition of our lives—not something unconnected to us. Much of the human population appears to have been separated from their ecological heritage and membership. Perhaps this is the reason pollution is so endemic to our world; many people had seen pollution as something displaced from their intimate reality.

In the last thirty years the threat and cause of damage to ecological and human health from polluting surface and ground water and acid rain and snow, as well as air pollution, global warming, and the destruction of the ozone layer has increasingly occupied our consciousness and our everyday life. The society from young school children to adults reading newspapers and watching television are aware that we are heirs to serious environmental problems. Polls indicate the great extent of this concern. Recently the concerns of various national governments have led to international conferences dealing with the ozone problem and discussion of global warming. Perhaps the convergence of several environmental conditions that threaten to change planet earth's ecological system have awakened the irresponsible amongst the citizenry, government administrators, scientists and engineers, and the industrial establishment to finally realize that we are all part of the ecological system and have a vital interest in the control of pollution.

The Clean Water Acts of the U.S. Congress and environmental action of various States and similar actions in Canada have resulted in some improvement in natural water quality in North America. The role of the Green parties and the citizenry has had a similar effect in Western European nations. In Eastern Europe there is increasing concern about pollution problems. Much remains to be done in the areas of irrigation, non-point source pollution, acid rain and snow, the effect of air pollution on water pollution, protection of ground water from hazardous wastes, and the further reduction of pollution from industrial sources. Extensive human effort and resources have been dedicated to detect and measure water pollutants and understand their effect on human populations and on the ecological system, as well as on the collection and rectification of wastewater in treatment facilities. However, much more remains to be done.

A realistic primer may help us to visualize the overall effects of water pollution. Sitting by an ecologically healthy lake or stream, we observe a proliferation of life—plants and animals familiar and cherished by us. Comparing that to our experience of being next to a polluted water body, we would notice different plants, not attractive to us and the presence of foul offensive odors. (However for a lake acidified by acid rain, very clear waters, devoid of life, are observed.) The system has changed from being aerobic (presence of dissolved oxygen) to anaerobic (lack of dissolved oxygen). The water body has changed so that it is no longer attractive to us nor can it serve as a water resource. A lack of dissolved oxygen in the water has changed the living conditions so that anaerobic fauna and flora can reside there. Two conditions can cause this situation: i.e., an excess of nutrients (such as nitrates or phosphates) serving to facilitate the growth of plants and an excess of biodegradable organic matter serving as food for the microbial population. These pollutants originate from human biological waste and human activities such as agriculture and industry.

An excess of biodegradable organic matter leads to an accelerated growth of the microbial population. Since they are aerobic and require dissolved oxygen in the water for respiration, a large population could deplete the dissolved oxygen supply leading to the asphixiation of fish, other animals and insects and the death of plants. Then anaerobic fauna and flora will flourish producing reduced gaseous substances, such as ammonia and hydrogen sulfide. These gases are toxic and unpleasantly odiferous. Although water can be reaerated by the air above its surface to provide a supply of dissolved oxygen, the process is very slow allowing for the conditions of oxygen depletion to exist for long periods of time.

Another mechanism leading to the same result is caused by an excess of nutrients. The presence of excessive amounts of nitrates and phosphates spur algae growth in the water body. The upper layers of algae shield the lower layers from sun light. This situation causes death of the lower layers of algae adding large amounts of biodegradable organic matter to the water body and an explosion in microbiological growth. Thus, through the action described above the dissolved oxygen content is greatly reduced and anaerobic conditions develop. Another category of pollution are the toxic substances entering water bodies, such as some synthetic organic materials and toxic metals and non-metals: they cause the death of aquatic plants and animals disrupting the water ecosystem. Non-biodegradable substances may be toxic, cause problems due to their physical nature, or detract from the beauty of nature.

From a consideration of the foregoing descriptions of the mechanisms of pollution effects, a number of parameters for the determination and control of water pollution can be listed. For example degradable organic matter, non-biodegradable substances, dissolved oxygen, nitrates and phosphates, and toxic metals, non-metals and organic matter are classes of substances requiring methods of analysis.

However the previous and rather bare outline of the pollution scenario does not expose the complex problems in describing the ecological mechanisms affected by pollution and their attendant solutions. The definition of a problem is necessary if one is to prescribe a solution. The more complete the definition, the more precise and comprehensive the proposed interpretation can be. Unfortunately, we do not have the luxury of unlimited time to adequately define the various environmental problems; we must institute actions using the knowledge at hand and update and improve our interim solutions as we approach a more complete definition of each of the problems. Indeed, the answers to the problems of water pollution and abatement have been undertaken in this vein.

The large question, what do we measure, brings us to the complexity of the issue, since what we measure is connected to why we measure a particular property or component. The attempt to answer these questions cannot be undertaken in this relatively short article, however, a very limited response will be given to these questions.

This article will describe the operation and use of chemical instrumentation both in the laboratory and in monitoring instrumental systems, for data collection necessary for refining the definition of the environmental water problems, monitoring of processes to treat wastewaters and drinking water, and the ecological monitoring of natural waters.

WATER AND WASTEWATER ANALYSIS

In the last fifty years the advances in electronics have made possible the development of the sophisticated instrumentation and computer systems which serve very well the purposes described herein. Development of chemical sensors and their combination with instrumentation has resulted in the laboratory and monitoring chemical measurement instruments so commonly found in laboratories, environmental monitoring systems and manufacturing plants. In addition the interfacing of these instruments and computer systems results in effective and creative data handling, computation, and prediction.

A general consideration of the analysis of water and wastewater samples brings forth several factors to consider. What characteristics need to be monitored and for what reasons? How do we obtain a representative sample of the source to be analyzed and how do we preserve its integrity until an analysis is completed? What constitutes our present methodology and with what biological, chemical, physical and instrumental means do we carry out these measurements? However, a primary consideration in answering these questions relates to the nature of water and wastewater samples.

The Nature of Water-Related Samples and Sampling Considerations

Nature of Water-Related Samples The category of water and wastewater samples can include water samples, sludges, benthic muds, plant matter and so forth. Samples may be taken from a number of systems: for example, natural water bodies, process streams from wastewater treatment and manufacturing plants, benthic environments, marshes, etc. Different procedures for sampling can be required for each variety of sample based on their unique chemical, physical, and/or biological nature.

Water is alluded to as the universal solvent for good reasons; it is the best solvent humans experience. In addition to dissolved substances water can also transport insoluble, suspended, and/or colloidal matter. Thus, a sample can contain components in a number of physical states: i.e.; dissolved, in ionic and molecular form; insoluble, as are bubbles of gas, and suspended and colloidal chemical substances; and biological organisms in a variety of sizes. The determination of the identity and concentration of unique chemical and biological components is important. The presence of these components give to the water sample biological, chemical, and physical characteristics—such as physiological qualities, acidity, alkalinity, color, opacity and so forth.

Sludges and mud samples are heterogeneous mixtures containing water with dissolved matter and the sampling procedure must not change the composition of the mixture. Plant matter, having its own unique characteristics, requires the proper procedures for sampling.[1] Many samples display time-based changes once taken from the source for a host of reasons. Changes are evident over various time scales. For example suspended matter settles during a time period as determined by particle size giving a change in opacity and/or color, chemical reactions may occur amongst components, gases and volatile substances may diffuse to the surface of the sample and evaporate, gases or volatile substances in the air space above the sample may condense and dissolve at the sample interface, etc. Substances in benthic samples can experience air oxidation and plant matter can lose moisture and so forth. All of these changes present a deviation of the sample's composition and characteristics from the source. The seriousness of the changes depend on the purpose of the analysis and its use. There must be no perturbation of the sample composition or physical characteristics which will nullify or seriously distort the analysis results and be detrimental to the purposes of the analysis.

Preservation is a means of preventing decomposition and change of various biological or chemical components in the sample. Standard Methods provides a treatment for the preservation of specific components in a variety of samples where preservation is possible (see Table 1).[2]

Sampling Considerations Two types of changes must be considered in order to define and carry out a proper sampling program, i.e., time-based changes in the source to be sampled and in the sample taken from the source.

Time-based changes in the source may be short or long term. The short-term changes taking place are indicative of biological, chemical, and/or physical interactions in the time span assigned to repetitive sampling. Long-term changes in the sources are related to ecological trends. It is important to be aware of each of these types of changes so that sample storage changes are not mistaken for sample source changes.

In order to obtain a legitimate sample of water or wastewater for analysis one must understand the nature the sample and its time-based changes. Once a sample taken from the source it may begin to change for many reasons as discussed in the previous section, IIA,*1*. The sample must be representative of the sample source, that is it must have the same biological, chemical, and physical characteristics of the source at the time the sample is taken. Therefore the sampling procedure must not cause a change in the sample relative to the source if there is to be an accurate correspondence between the sample analysis and the nature of the sample source at the time the sample was taken.

To prevent confusing the two variables, ecological changes in the source and interactions in the sample that can occur in the same time span in a series of samples, preservation of the samples is undertaken to deter changes after sampling. However, preliminary sampling and testing may be necessary to indicate the type of time-based changes occurring. Then a reliable program can be established with some certainty.

Three kinds of sampling schemes are undertaken in consideration of the sample source characteristics—grab, composite and integrated samples.

A grab sample is one taken at a given time and place. If the source doesn't change greatly during a long passage of time or within a large distance in all directions from the sampling point, the grab sample is useful. The results from grab samples are said to represent the source for the given values of distance from sampling location and time. However a time and place series of grab samples is needed to establish the constancy of analytical values for different times and distances from the original grab sampling point. With that information the sampling frequencies and times for a sampling program can be established. The sampling of solids, such as, benthic muds and sludges, requires great care in order to obtain truly representative samples.

A series of samples collected and blended to give a time-averaged sample is known as a composite or time-composite sample. In another procedure the volume of each sample of the series collected is proportional to flow of the sample source, namely the water body or waste stream. The samples of various volumes are composited to provide the final time-composite sample. This type of sample gives an average value over a time period and saves analysis time and cost. Sampling frequencies and the total time span of the series depends on the source. Composite sampling may be used for process streams to determine the effect of unit processes or to monitor a plant outfall for daily or shift changes. However, biological, chemical, and physical parameters that changes on storage during the sampling time period can't be reliably determined in time-composited samples and another sampling protocol is needed (see Table 1).

At times, simultaneous samples are needed from various locations within a given source, such as a river or lake. Grab samples are then composited, usually based on volumes proportional to flow, and are called integrated samples. These samples are used to determine average composition or total loading of the source which varies in composition in its breadth and depth. The sampling program for such sources is complex and requires careful consideration for each unique source.

Water and Wastewater Parameters

A large number of water quality parameters are utilized in the characterization, management and processing of water and wastewater. Table 2 lists a number of these parameters separated into three categories—physical, chemical and biological. It is obvious that only some parameters are considered to be pollution factors because they indicate conditions of water during processing or in the natural state. The STORET system of the USEPA lists more than four hundred parameters separated into six major groups and is used for the analysis, collection, processing and reporting of data.[3] Table 3 gives a sampling of groups of parameters in this system. Not all of these parameters are used frequently, since many are rather unique to particular waste effluents. In actuality, a very small number are used in the analysis of a particular sample.

Water quality parameters may be divided into two groups, specific and non-specific water quality parameters. Specific parameters refer to chemical entities of all types, e.g., ions, elements, compounds, complexes, etc. For example, in Table 2 some specific parameters are ammonia, all metals listed, dissolved oxygen, nitrates, sulfates, and so forth. Non-specific parameters are included in three categories and some examples are as follows: chemical (hardness, alkalinity, acidity, BOD [biochemical oxygen demand], TOC [total organic carbon], COD [chemical oxygen demand], chlorine demand), physical (salinity, density, electrical conductance, filterable residue), and physiological (taste, odor, color, turbidity, suspended matter). Many of these non-specific parameters are

TABLE 1

Summary of special sampling or handling requirements[a,*]

Determination	Container[†]	Minimum Sample Size *mL*	Sample Type[‡]	Preservation[§]	Maximum Storage Recommended/ Regulatory
Acidity	P, G(B)	100	g	Refrigerate	24 h/14d
Alkalinity	P, G	200	g	Refrigerate	24 h/14 d
BOD	P, G	1000	g	Refrigerate	6 h/48 h
Boron	P	100	g, c	None required	28 d/6 months
Bromide	P, G	100	g, c	None required	28 d/28 d
Carbon, organic, Total	G	100	g, c	Analyze immediately: or refrigerate and add H_3PO_4 or H_2SO_4 to pH $<$ 2	7 d/28 d
Carbon dioxide	P, G	100	g	Analyze immediately	stat/N.S.
COD	P, G	100	g, c	Analyze as soon as possible, or add $H_2SO_4$to pH $<$ 2; refrigerate	7 d/28 d
Chloride	P, G	50	g, c	None required	28 d
Chlorine, residual	P, G	500	g	Analyze immediately	0.5 h/stat
Chlorine dioxide	P, G	500	g	Analyze immediately	0.5 h/N.S.
Chlorophyll	P, G	500	g, c	30 d in dark	30 d/N.S.
Color	P, G	500	g, c	Refrigerate	48 h/48 h
Conductivity	P, G	500	g, c	Refrigerate	28 d/28 d
Cyanide:					
Total	P, G	500	g, c	Add NaOH to pH $>$ 12, refrigerate in dark[#]	24 h/14 d; 24 h if sulfide present
Amenable to chlorination	P, G	500	g, c	Add 100 mg $Na_2S_2O_3$/L	stat/14d; 24 h if sulfide present
Fluoride	P	300	g, c	None required	28 d/28 d
Hardness	P, G	100	g, c	Add HNO_3 to pH $<$ 2	6 months/6 months
Iodine	P, G	500	g, c	Analyze immediately	0.5 h/N.S.
Metals, general	P(A), G(A)	500	g	For dissolved metals filter immediately, add HNO_3 to pH $<$ 2	6 months/6 months
Chromium VI	P(A), G(A)	300	g	Refrigerate	24 h/24 h
Copper by colorimetry*					
Mercury	P(A), G(A)	500	g, c	Add HNO_3 to pH $<$ 2, 4°C, refrigerate	28 d/28 d
Nitrogen:					
Ammonia	P, G	500	g, c	Analyze as soon as possible or add H_2SO_4 to pH $<$ 2, refrigerate	7 d/28 d
Nitrate	P, G	100	g, c	Analyze as soon as possible or refrigerate	48 h/48 h (28 d for chlorinated samples)
Nitrate + nitrite	P, G	200	g, c	Add H_2SO_4 to pH $<$ 2, refrigerate	none/28 d
Nitrite	P, G	100	g, c	Analyze as soon as possible or refrigerate	none/48 h
Organic,					
Kjeldahl*	P, G	500	g, c	Refrigerate; add H_2SO_4 to pH $<$ 2	7 d/28 d
Odor	G	500	g	Analyze as soon as possible; refrigerate	6 h/N.S.
Oil and grease	G, wide-mouth calibrated	1000	g, c	Add HCl to pH $<$ 2, refrigerate	28 d/28 d
Organic compounds:					
MBAS	P, G	250	g, c	Refrigerate	48 h
Pesticides*	G(S), TFE-lined cap	1000	g, c	Refrigerate; add 1000 mg ascorbic acid/L if residual chlorine present	7 d/7 d until extraction; 40 d after extraction
Phenols	P, G	500	g, c	Refrigerate, add H_2SO_4 to pH $<$ 2	*/28 d

(*continued*)

TABLE 1
Summary of special sampling or handling requirements[a,*] (*continued*)

Determination	Container[†]	Minimum Sample Size *mL*	Sample Type[‡]	Preservation[§]	Maximum Storage Recommended/ Regulatory[#]
Purgeables* by purge and trap	G, TFE-lined cap	2 × 40	g	Refrigerate; add HCl to pH < 2; add 1000 mg ascorbic acid/L if residual chlorine present	7 d/14 d
Oxygen, dissolved:	G, BOD bottle	300	g		
Electrode				Analyze immediately	0.5 h/stat
Winkler				Titration may be delayed after acdification	8 h/8 h
Ozone	G	1000	g	Analyze immediately	0.5 h/N.S.
PH	P, G	50	g	Analyze immediately	2 h/stat
Phosphate	G(A)	100	g	For dissolved phosphate filter immediately; refrigerate	48 h/N.S.
Salinity	G, wax seal	240	g	Analyze immediately or use wax seal	6 months/N.S.
Silica	P	200	g, c	Refrigerate, do not freeze	28 d/28 d
Sludge digester gas	G, gas bottle	—	g	—	N.S.
Solids	P, G	200	g, c	Refrigerate	7 d/2–7 d; see cited reference
Sulfate	P, G	100	g, c	Refrigerate	28 d/28 d
Sulfide	P, G	100	g, c	Refrigerate; add 4 drops 2*N* zinc acetate/ 100 mL; add NaOH to pH > 9	28 d/7 d
Taste	G	500	g	Analyze as soon as possible; refrigerate	24 h/N.S.
Temperature	P, G	—	g	Analyze immediately	stat/stat
Turbidity	P, G	100	g, c	Analyze same day; store in dark up to 24 h, refrigerate	24 h/48 h

* See text for additional details. For determinations not listed, use glass or plastic containers; preferably refrigerate during storage and analyze as soon as possible.

[†] P = plastic (polyethylene or equivalent); G = glass; G(A) or P(A) = rinsed with 1 + 1 HNO_3; G(B) = glass, borosilicate; G(S) = glass, rinsed with organic solvents or baked.

[‡] g = grab; c = composite.

[§] Refrigerate = storage at 4°C, in the dark.

[#] Environmental Protection Agency, Rules and Regulations. 40 CFR Parts 100–149, July 1, 1992. See this citation for possible differences regarding container and preservation requirements. N.S. = not stated in cited reference; stat = no storage allowed; analyze immediately.

[a] If sample is chlorinated, see text for pretreatment.

very important in water and wastewater characterization and instruments are available to measure specific and non-specific parameters.

Methodology

The large variety of tests carried out on water and wastewater samples and sources have been codified and are included in the laboratory reference in the United States entitled "Standard Methods for the Examination of Water and Wastewater" and is commonly referred to as Standard Methods. This compendia of methods is regularly updated. At the present time the 19th edition published in 1995 is in use[2] and a supplement was issued in 1996. Supplements are used to update methods on an ongoing basis in order not to unduly prolong the publication of the new edition. However not more than one supplement appears to have been published for each edition.

Three professional organizations jointly write and edit this manual—the American Public Health Association, the American Water Works Association and the Water Environment Federation (formerly the water pollution Control Federation). It is published by the American Public Health Association. Over five hundred professionals belonging to these organizations and others participate in Standard Methods. It was first published in 1905 and an interesting history of its genesis is given in the preface to the 19th edition.[2]

At one time methods were segregated between water and wastewater test methods, however, since the 14th edition in 1976, that division ceased. In the 19th edition, methods are classified in ten groups: Introduction, Physical Aggregate Properties, Metals, Inorganic Nonmetallic Constituents, Aggregate Organic Constituents, Individual Organic Constituents, Radioactivity, Toxicity, Microbiological Examination, and Biological Examination.

TABLE 2
Some water quality parameters[a]

Physical	Chemical	Biological
Color	Acids or alkali	Algae
Conductivity	Ammonia	Bacteria
Odor	Biochemical oxygen demand (BOD)	Pathogens
Radioactivity	Calcium	Protozoa
Solar radiation intensity	Chloride	Viruses
Suspended solids or sludges	Chlorophyll	
Temperature	Chemical oxygen demand (COD)	
Turbidity	Dissolved oxygen	
	Hardness	
	Heavy metals:	
	Chromium	
	Copper	
	Iron	
	Lead	
	Manganese	
	Mercury	
	Magnesium	
	Nitrate, nitrite	
	Organic compounds:	
	Detergents	
	Herbicides	
	Pesticides	
	Phenol	
	Oils and greases	
	Oxidation–reduction potential	
	pH	
	Phosphates	
	Potassium	
	Sodium	
	Sulfate	
	Total organic carbon (TOC)	

[a] Reprinted from Ref. (4), p. 1438 by courtesy of Marcel Dekker, Inc.

Twenty-five years ago the dearth of instrumentation was used in Standard Methods.[4] However, in the present edition the following instrumentation is employed in the methodologies: molecular spectroscopy (visible, uv, ir), atomic spectroscopy (absorption, flame, ICP), chromatography (gas, ion, liquid), mass spectrometry (GC/MS, gas chromatography/mass spectrometry), electro-analytical techniques (polarography, potentiometric and amperometric titrations, selective ion electrodes), radio-activity counters (gas filled and semiconductor detectors and scintillation counters), and automated continuous-flow methods.

Also included in Standard Methods are aspects such as safety, sampling, mathematical treatment of results, reagents, apparatus etc. It is fortunate, indeed, that such a comprehensive work is available and that it is regularly revised. Enlightened editorial leadership and the many members of the Standard Methods committees in the last twenty years can be credited for the steady increase in the inclusion of instrumentation in Standard Methods. In the increasing complexity of environmental and ecological problems and guidance of Standard Methods is a valuable and practical support in obtaining necessary analytical data.

The American Society for the Testing of Materials, ASTM, is an important compendium for the analysis of raw and finished material products. A large section is devoted to the analysis of water and wastewater in the context of processing and usage.[5]

The EPA has published instrumental methods for the analysis of priority pollutants and other substances controlled by Federal legislation.[6]

Types and Purposes of Instrumentation and Computer Systems

The development of a large variety of analytical instrumentation has been a boon to water and wastewater characterization, research, management, and process control. In addition to the requirements of the process industries, the needs of the water and wastewater area have spawned the development of some specialized laboratory, monitoring and process control instrumentation. Some examples are the total carbon and organic carbon analyzers, biological oxygen analyzers and the residual chlorine analyzer. Monitoring and data acquisition systems, in conjunction with this instrumentation, are increasingly used in wastewater management and plant process control. Certainly a number of physical parameters such as temperature, flow rate, pressure and liquid level have been measured instrumentally in the process industries, including wastewater treatment plants, predating the development of this wide variety of analytical instrumentation.[7]

Instruments utilized in the measurement of parameters important to wastewater analysis, treatment and management can be divided into two categories based on application. Monitoring of water bodies and waste treatment processes require monitoring instruments which are characterized by ruggedness and capability of unattended operation and data storage and/or transmission. A second category, laboratory instruments, in many instances, may be more sophisticated, sensitive to the surrounding environment and also have data storage and transmission capabilities. Each type has its specific utility in the scheme of analysis and data acquisition for wastewater characterization and processing. In many instances monitoring instruments are laboratory devices which were ruggedized and prepaared for field use. Thus the variables measured and the principles of operation are the same in many cases. Some examples of variables measured by laboratory and monitoring instrumentation are pH, conductivity, DO (dissolved oxygen), specific cations and

TABLE 3
Number of STORET listings for water analysis

Parameters by groups	Example	Number of parameters in group
General physical and chemical	Alkalinity, COD, iron turbidity, zirconium	149
Physical observations	Algae, foam, oil	12
Radionuclides	Gross alpha and beta, strontium-90	141
Microbiological	Coliform by MPH and MF, total plate count	18
Organic materials		
Carbon adsorption data	Chloroform and alcohol extractables	12
Natural organics	Chlorophyll, tannins	4
Synthetic organics	ABS, phenols	2
Halogenated hydrocarbons	Aldrin, heptachlor, toxaphene	62
Phosphorated hydrocarbons	Malthion, parathion	10
Miscellaneous pesticides	Silvex	8
Treatment-related observations	Available chlorine	6

anions, and a variety of specific and nonspecific parameters by automated analyzers. However, the instrumental appearance and some unique functions related to unattended operations may differ.

Monitoring and data acquisition systems are also considered in this article. In the control of wastewater treatment systems and plants, data may be obtained exclusively from monitoring instrumental systems or a combination including data from laboratory instruments via a laboratory data system and/or from data entered through terminals.

Analytical instrumentation can be classified according to principles based on various physical phenomena. These general categories are spectroscopy, electrochemistry, radiochemistry, chromatography, and automated chemical analysis. The instrumentation described in this article is organized according to these categories.

INSTRUMENTATION

Structure of Instruments

An instrument is a device that detects a physical property or chemical entity through the conversion of a physical or chemical analytical signal to an energy signal, usually electrical, with subsequent readout of the energy signal.

Three main parts comprise an instrument: that is a chemical or physical sensor, signal conditioning circuits, and readout devices. The sensor develops a signal, usually electrical, in response to a sample property and the signal conditioning circuit modifies the signal in order to allow convenient readout display of the signal. Finally, a readout device displays the signal, representative of the sample, in terms of a reading on an analogue or digital meter, a recorder chart, an oscilloscopic trace, etc. Figure 1 delineates the three major parts and functions of an instrument, sample properties (measurand) to be measured, and instrumental criteria.

Sensors A sensor, the primary contact of the instrument with the sample, is a device that converts the input energy derived from a sample property to an output signal, usually electrical in nature. The relationship between the input energy (measurand), Q_1, and the output energy, Q_0, is expressed in the form:

$$Q_0 = f(Q_1) \tag{1}$$

and is known as the transfer function. The sensitivity is given in the equation

$$S = dQ_0/dQ_1. \tag{2}$$

When the transfer function is linear, the sensitivity is constant throughout the sensor's range. However, the sensitivity (gain or attenuation factor) is dependent on the value of the differential fraction in equation 2. The sensor threshold is the smallest magnitude of input energy necessary to obtain a measurable change in the output.

Readout signals may be digital, *D* (discrete), or analog, *A* (continuous), in form and are a function of the nature of the input signal and the sensor and the design of the signal conditioning circuits. These signals are interconvertible using *A*/*D* or *D*/*A* devices. Fast reacting sensors and circuits, however, are utilized for producing digital signals, where, formerly, analog signals were obtained.

Two varieties of sensors, chemical and physical, are in use on various instruments. The physical sensor allows the conversion of physical energy from one to another. One example is a photocell that converts an impinging light beam

Measurand Sample property	$\xrightarrow[\text{Signal}]{\text{Input}}$	**Sensor** Energy transducer	$\xrightarrow[\text{Signal}]{\text{Transducer}}$	**Signal conditioning** Signal modification	$\xrightarrow[\text{Signal}]{\text{Output}}$	**Readout** Types
Light		Light		Amplification		Analog
Absorption		Photocell		Arithmetic operation		Meter
Emission		Photographic plate		Chopping		Oscilloscope
Thermal		Thermal		Comparison to reference		Recorder
Conductivity		Thermocouple		Digitization		Digital
Temperature		Katharometer		Rectification		Nixie display
Heat capacity		Thermister		Stabilization		Point plotter
		Bolometer				Printer
Electrical		Electrical				Tape, paper, or magnetic
Redox		Electrode pair				*Instrument*
Electrolytic conductivity		Electrodes, AC system				*Criteria*
Ionic activity		Membrane electrode				Band width
Mechanical		Mechanical				Noise figure
Mass		Balance force transducer				Sensitivity
Density		Force transducer				Signal-to-noise
Viscosity		Hydrometer				Time constant
Surface tension		Viscosity pipet				Error
Nuclear		Nuclear				Hysteresis
(X-ray, $\beta.\gamma$)		Ionization tubes				Nonlinearity
Emission		Scintillation counters				Scale
Absorption		Photographic plates				Zero displacement
		Cloud chamber				
		Semiconductor detectors				

FIGURE 1 Diagram of instrumental functions. Reprinted from Ref. (4), p. 1442 by courtesy of Marcel Dekker, Inc.

to an electrical signal and is used in spectrometers. A second example is a piezoelectric crystal-based sensor that converts a mechanical force to an electrical charge translatable to a potential. The piezoelectric effect is reversible; an electric charge will cause a mechanical dislocation in the crystal. Another example of a physical sensor is a platinum resistance thermometer where the resistance of a platinum wire is altered by a change in temperature.

Chemical sensors are devices that allow the analyte or target material through one of its specific chemical parameters to ultimately generate an energy signal, usually electrical, in a transducer through the agency of a selective chemical or physical chemical reaction. A transducer is a material structure inside of which or on whose surface the specific chemical or physical chemical reaction takes place leading to the generation of the energy signal. Thus, there are two parts to the chemical sensor, the interface zone or area where the selective reaction takes place and the usually non-specific transducer.[8] Figure 2 illustrates, functionally, the parts of a chemical sensor.

An example of a chemical sensor is a potentiometric electrode. Here the selective chemical reaction, the redox reaction of the analyte, is in equilibrium at the electrode surface imposing a potential that is proportional to the logarithm of the concentration of the analyte as described by the Nernst equation. For example, a copper electrode in a solution of copper ions will take on a potential in response to the concentration of copper ions. The logarithm of the copper ion concentration is proportional to the electrode potential. Another illustration of a chemical sensor is an amperometric electrode, where a current arises due to the redox reaction of the analyte when the electrode is at the appropriate potential. The concentration of the analyte is proportional to the magnitude of the current. A platinum electrode maintained at the redox potential for the silver/silver ion redox system will detect the concentration of silver ions. A membrane electrode is another type of chemical sensor. The fluoride electrode consists of a lanthanum fluoride (LaF_2), thin, crystal membrane. On the outside surface, the sample side of the membrane, the fluoride ions, F^-, from the sample are attracted electrostatically to the lanthanum ion, La^{3+}, at the surface of the membrane to form a complex. The complexed entities do not penetrate very deeply into the surface. The amount of F^- complexed is a direct function of its activity (see Section III,B,2,*a*) and represents a selective physical chemical reaction. A membrane potential arises because the opposite side of the membrane is exposed to a standard activity of F^- giving a net difference in potential between the two sides. The membrane potential is the non-specific electrical signal of the sensor.

Signal-Conditioning Circuits These circuits modify the signal produced by the sensor so as to provide an accurate representation of the sensor signal with optimal electrical characteristics to drive the readout device. In Figure 1 a number of signal conditioning modes are given and can be

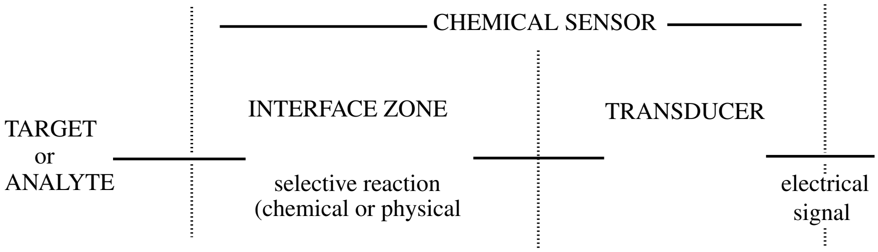


FIGURE 2 Chemical sensor.

placed in four categories—modification of sensor output, amplification, mathematical operation, and signal modification for readout.

The electrical components used in these circuits are of two types, active and passive elements. Active elements, such as solid state devices add energy to a circuit; whereas passive elements, such as resistors, capacitors, inductors, diodes add no energy. Both elements are combined to form active and passive circuits. Active circuits change signals in a complex way. Passive elements are used in active circuits to provide necessary conditions for the proper functioning of active circuits. Some active devices are ionization chambers, vacuum phototubes, operational amplifiers and gas discharge tubes.

Readout Devices The sensor signal modified by the conditioning circuits is ultimately converted into a visual form by the readout device or output transducer. The readout signal may be analog or digital requiring a compatible readout device. Analog readout devices comprise recorders, meters, oscilloscopes, photographic plates and integrators; printers, computers and digital meters with optical displays provide digital readouts. A digital computer may be interfaced to an instrument, in order to compute values from a digital output signal and produce a hard (printed) copy of the data using a printer. Analog output signals may be digitized in order to utilize a computer. The advantages of digital outputs are the statistical benefit derived from counting and analog outputs are advantageous in feedback control systems.

Analog Devices The automatic recording potentiometer or potentiometric recorder has been, over the years, the most frequently used readout device providing a continuous trace on a chart of an analog signal. Its operation is based on a low power servomechanism utilizing a feedback system. The instrumental signal to be measured is compared to a standard reference signal. The amplified, difference or error signal activates the pen-drive motor moving the pen on the chart to a position representing the magnitude of the analog signal. The control of the pen, based on the error signal, denotes the feedback system and the total system is referred to as a servomechanism.[9,10] Two types of recorders, the *Y*-time or *X–Y,* allow the recording of a signal, *Y,* as a function of time or of two signals representing the ordered (data) pair, *x, y,* respectively. In the *Y*-time device, a constant-speed motor moves the chart in the *x* direction while the servomechanism deals with the *y* signal. The *X–Y* recorder has two servo-systems, one for each signal, *x* and *y.* However, recorders may be limited by the rate that the data flows from the instrument. Some recorders can adequately respond to signals during fast scans. For example fast scans in cyclic voltammetry of about 1 volt/sec. can be transcribed using a recorder, however, at faster rates an oscilloscope is necessary.

Almost any instrument can utilize a potentiometric recorder. A *Y*-time analog recorder is commonly used to trace gas and liquid chromatograms; the abscissa, *X* axis, is for retention volume or time and the ordinate is for the detector response.

The oscilloscope is a measuring device with complicated circuitry that allows accurate display and measurement of non-sinusoidal or complex waveforms. The oscilloscope's basic part is the cathode ray tube, CRT. A CRT is a vacuum tube containing an electron gun pointing to a fluorescent screen at the tube's end. The electron gun provides a beam whose movement is controlled by two sets of deflector plates perpendicular to each other. The plates receive the signals representing the waveforms. These analog signals are displayed on a fluorescent screen as *Y*-time or *X-Y* curves. The display is photographed to provide a hard copy of the analog data. The oscilloscope can display data that is generated at high rates, since there are no mechanical movements used in manipulating the electron beam. Where very fast events must be recorded, an oscilloscope is an effective readout device.[11] (See the previous paragraph on the potentiometric recorder.) Oscilloscopes have facilities to store, compare, and manipulate signals.

Analog meters are based on the D'Arsonval meter movement. The electrical current signal passing through a moving coil, to which is fixed a pointer, induces a magnetic field in the coil. A static magnetic field from a permanent horseshoe magnet surrounds the coil. The interaction between the two fields causes the movement of the coil: the degree of movement is determined by the magnitude of the signal current. Analog meters require the analyst to interpret or read the output signal value by the position of the indicator needle or pointer using a calibrated scale mounted on the meter. A resistance placed in series with the meter movement allows the measurement of voltage. Resistance may also be measured with the meter. One weakness of this device is its low internal resistance causing loading errors by high impedance signals.[11]

A meter is used in the analysis of single samples or samples analyzed, serially, at a slow rate on a spectroscopic instrument at one frequency or wavelength. Meters also are employed to indicate proper adjustment of potentials, currents, temperatures, etc. for various instruments.

An electronic voltmeter, EVM, is more sensitive and accurate than the D'Arsonval-based meter previously described, particularly for signals with high impedance. The internal resistance is 10 Mohms (megaohms, 10^6 ohms) or more for d.c. (direct current) signals and 1 Mohm for a.c. (alternating current) signals. The circuits use solid state devices compared to the earlier device, a VTVM (vacuum tube voltmeter). Current and resistance is measurable with the EVM. Its application parallels those for the D'Arsonval-based meter.[11]

A photographic plate or film may be used to collect data in the time domain where all the data are displayed simultaneously, that is a spectrum in emission spectroscopy. The radiation in the dispersion pattern of the sample reflected or transmitted from the prism or grating impinges on the photographic plate.

Electronic integrators determine the area under a curve and are superior in precision to the ball and disk integrator and the several hand methods widely utilized. They may be based on operational amplifier or transistor circuitry. Some potentiometric recorders have a second pen controlled by an integrator and the density of the pen's excursions determine the area under the curve. This last type is not as convenient as the electronic integrators that can correct for baseline changes. Chromatographic peak areas for GC and HPLC (high performance liquid chromatography), anodic stripping analysis peaks, spectroscopic curves, etc. are integrated as a means of quantitation and analysis of an analyte.

Analog computers are available but are not used now to any great extent.

Digital Devices The digital computer or microprocessor interfaced to the instrument brings a broad capability to the display and processing of instrumental data. Data reception and storage is convenient when real time computation and display are not required. Mathematical calculations, including the areas under curves, graphic and tabular displays, correlation with previously collected data, and many other operations can be carried out at one's convenience. Real time processing can be accomplished on a time-sharing basis or with a dedicated computer. The visual display is at a video monitor and a printer provides a hard (printed) copy of the raw and calculated data, graphs, and other information. Computer devices include microprocessors and micro-, mini-, and mainframe computers. The instrument must be carefully interfaced to the computer and this task requires much electronic skill. Instruments providing spectral readouts, the need for number crunching and repetitive analyses can benefit greatly from a computer interface. Some instruments that utilize Fourier transform analysis require a computer capability and many instrumental techniques have been revolutionalized by computer use. The use of the computer[12] in the reduction of noise in instrumental signals by ensemble and boxcar averaging has greatly improved the quality of instrumental data.[12]

Digital meters measure analog signals and provide a digital readout. A/D conversion of the analog input is accomplished electronically. The digital data is displayed as numeric images using solid state devices such as LEDs, light emitting diodes, and LCDs, liquid crystal displays, and lamps such as, NIXIE, neon, and incandescent bulbs. The LED is the more convenient device because its seven segment readout display uses lower currents and voltages than the lamp displays. The LED's red image, due to the semiconductor gallium arsenide doped with phosphorus, may be increased in intensity by using more semiconductor in the LED. The image color of LEDs may be fabricated to be green or yellow, also.[11,13] LCDs operate by means of polarizing light. They use reflected light for viewing, a seven-segment and dot matrix readout display, an a.c. voltage, consume very little power and are more fragile than LEDs.[14] The LCDs and LEDs are the newest and most convenient display devices.

Digital meters can be used in place of the analog variety. The former are more accurate and easier to read.

Instrumental Parameters and Definitions Instrumental characteristics of operation and data treatment and statistics are defined by a number of parameters. A definition of each term is as follows:

- The range of frequencies (information) in the signal is called the bandwidth. During amplification, some amplifiers cannot respond to the range of frequencies in the signal producing an amplified signal with a narrower bandwidth.
- The baseline is the signal obtained when no sample is being examined and reflects the noise inherent in the instrument.
- Calibration is the process relating instrument response to quantity of analyte. In general a series of standard solutions or quantities of analyte are analyzed on the instrument taking reagent blanks into account and using a similar matrix as the sample under consideration. The quantity-response data are plotted to provide a calibration curve where error bars indicate the precision of the method.[15] Other calibration procedures such as the methods of standard additions[16] and of internal standards[17] have advantages in specific situations. The former is helpful in ameliorating interferences from the sample matrix and the latter in correcting for changes in instrument response particularly in GC, and ir (infrared) and emission spectroscopy.[18]
- The gain refers to the ability to amplify a signal and is the ratio of the output to input signal. The

gain may refer to voltage, current or power amplification and its input and output impedances.

- Noise refers to random signals, usually continuous, that restricts the lower detection limit and accuracy of the signal. Noise arises from electronic components and environmental sources and cannot, at times, be completely eliminated.
- The ratio of the amplitude of the signal to that of the noise is called the signal-to-noise, S/N, ratio. This ratio gives the ability to distinguish between signals and noise, that is a measurement of the quality of an instrument. One cannot usually distinguish the signal from the noise when the ratio is less than about 2 or 3.
- Resolution or resolving power is the capability of displaying two signals differing slightly in value. The resolving power, *R,* of a monochrometer concerns absorption of emission spectral signals,

$$R = \lambda / d\lambda \tag{3}$$

is the wavelength under consideration and $d\lambda$ is the difference of wavelength between the two signals. In mass spectrometry resolution refers to the separation of two mass peaks *Ms* and *Ms* + *dMs,* where *dMs* is the difference in masses so that

$$R = Ms/dMs. \tag{4}$$

For resolution for chromatographic methods see Part Two Section III,B,4,*a*.

- Response time refers to the time needed for a pen of a potentiometric recorder to travel the total vertical distance on the *Y* axis.
- Sensitivity, *S,* describes the ratio of the change in the response or output signal, dI_0 of the instrument to a small change in the concentration or amount of the analyte, *dC.* The ratio is given as follows:

$$S = dI_0/dC. \tag{5}$$

- Linear dynamic range, LDR, describes the mathematical relationship between amount or concentration of the analyte and the response of the instrument. An increase in the analyte quantity results in a linear increase in response. The size of the range of quantities accommodated by the instrument response is the key factor for this parameter. For example in voltammetry the LDR is 10^{-8} to 10^{-3} M (molar), five orders of magnitude, in (ultraviolet) uv–visible spectrophotometry, about 10 to 100, and for a GC with a FID (flame ionization detector) the LDR extends from 10^{-1} to 10^{7} ng (nanograms, 10^{-9} grams) or eight orders of magnitude. Obviously the sensitivity remains constant in contrast to a non-linear dynamic relationship.
- The reagent blank or blank in a spectroscopic determination is the signal obtained by the solution of the reagents without any analyte. The sample matrix is important to include, if known, in the blank. In many instances the effect of the matrix is determined indirectly.
- Accuracy defines, mathematically, the absolute error, e_a, inherent in the method when comparing the analytical result, x_i, with the true value, x_t, of the analyte content of the sample.

$$e_a = (x_i - x_t). \tag{6}$$

Preparing a standard sample containing an accurately known concentration of the analyte is required. This is not a simple task, because homogeneity of any mixture is difficult to obtain and ascertain.

- The precision of a method is concerned with the repeatability of the analytical results for a number of analyses on the same sample. There are several ways of expressing precision; standard deviation is a very effective and meaningful measure. The standard deviation, *sd,* for small sets of data is given as follows:

$$sd = \left[\sum_{i=1}^{N} (x_i - x_a)^2 / No - 1 \right]^{1,2}. \tag{7}$$

Here x_i is the experimental value, x_a, the average of the experimental values, and *No,* the number of values. The standard deviation is a measure of the average uncertainty of all the measurements in the data set, x_i, that is $x_1, \ldots, x_N$.[18,19]

Types of Instruments

Analytical instruments can be classified according to categories based on various physical phenomena. The general categories used in this article are spectroscopy, electrochemical analysis, radiochemical analysis, chromatography, and automated analysis. Table 4 illustrates these categories.

Spectroscopy

Introduction Spectroscopic instruments include optical and other types of instruments. The optical instruments analyze electromagnetic radiation, emr, while other spectroscopic instruments deal with sound, mixtures of ions, electrons, and other forms of energy. Other optical methods utilize instruments that make refractometric and polarimetric measurements. Refractometric measurements will be discussed in the section on liquid chromatography.

Spectroscopy, classically, is that area of science where the electromagnetic radiation, emr, emitted from or absorbed

TABLE 4
Bases for instrumental methods

Energy interaction	Process	Instrumental method
EMR, range	Absorption of emr	x-ray, uv/vis atomic & ir spectrophotometry
Emr/magnetic field	Absorption of emr in a magnetic field	NMR spectroscopy
e^-, ions, or electric field	Ion formation/seperation in electric or magnetic field	Mass spectroscopy
Electricity (arc, spark), heat (flame, plasma)	Emission of emr	x-ray, uv/vis, flame emission spectroscopy
Emr, x-ray	Emission of electrons	x-ray photoelectron spectroscopy (XPS or ESCA)
Emr, uv	Emission of electrons	UV photoelectron spectroscopy (UPS)
x-ray or e^-	Emission of electrons	Auger spectroscopy
Emr, uv/vis	Emission of acoustic energy	Photoacoustic spectroscopy
None	Emission via radioactive decay	Radiochemical methods
Emr, range	Fluorescence & phosphorescence of emr	x-ray uv/vis[1], & atomic fluorescence spectroscopy
Emr, vis	Scattering of emr by particles	Nephelometry, turbidimetry
Emr, vis	Scattering of emr by molecules	Raman spectroscopy
Emr, x-ray	Diffraction	x-ray diffraction
Emr, vis	Refraction (bending of light beam)	Refractometry
Emr, uv/vis	Rotation of plane-polarised light	Polarimetry
Emr, uv/vis	Rotation as a function of wavelength	Optical rotatory dispersion
Emr, uv/vis	Rotation using circularly polarized light	Circular dichroism
Electricity	current measurement	Amperometry, coulometry, polarography, voltammetry
Electricity	pass current/weigh-plated material	Electrogravimetry
Electricity	potential measurement	Chronopotentiometry, potentiometry
Electricity	resistance/conductance measurement	Conductometry
Heat	weight loss vs increasing temperature differential temperature vs increasing temperature heat flow to sample vs increasing temperature. Temperature vs volume of reagent	Thermogravimetric analysis. Differential thermal analysis. Differential scanning calorimetry. Enthalpimetric methods

by a substance is resolved into its component wavelengths indicating its intensity and presented as a spectrum. In this category absorption, emission, and photoluminescence (fluorescence and phosphorescence) spectroscopy using x-ray, ultraviolet–visible (uv–vis), and ir radiation, and the measurement of turbidity, or suspended matter by nephelometry and turbidimetry, are included. However, today in a broader sense, spectroscopy includes the following: resolution of electrons of many energies by uv and x-ray photoelectron, Auger etc. spectroscopy; sound waves by acoustic spectroscopy; ions by mass number by mass spectroscopy; and absorption of radiowaves by atoms and electrons exposed to a magnetic field in nuclear magnetic resonance and electron spin resonance spectroscopy. The phenomena of absorption, emission, photoluminescence (fluorescence and phosphorescence), and scattering are the bases of spectroscopic instruments.

b. Spectroscopic instruments

Spectroscopic instrumentation is differentiated with respect to the wavelength range of the instrument, that is x-ray, uv, visible, and ir and type of instrument, i.e. absorption, emission, photoluminescence (fluorescence and phosphorescence), and turbidity. The energy sources, sample cells, wavelength selection devices (gratings, prisms, filters, crystals) and sensors may differ for these various instruments. These parts are listed in Figure 3 for the wavelength regions of from 100 to 40,000 nm (nanometer, 10^{-9} meters). X-ray and non-optical spectroscopic instruments are not included.

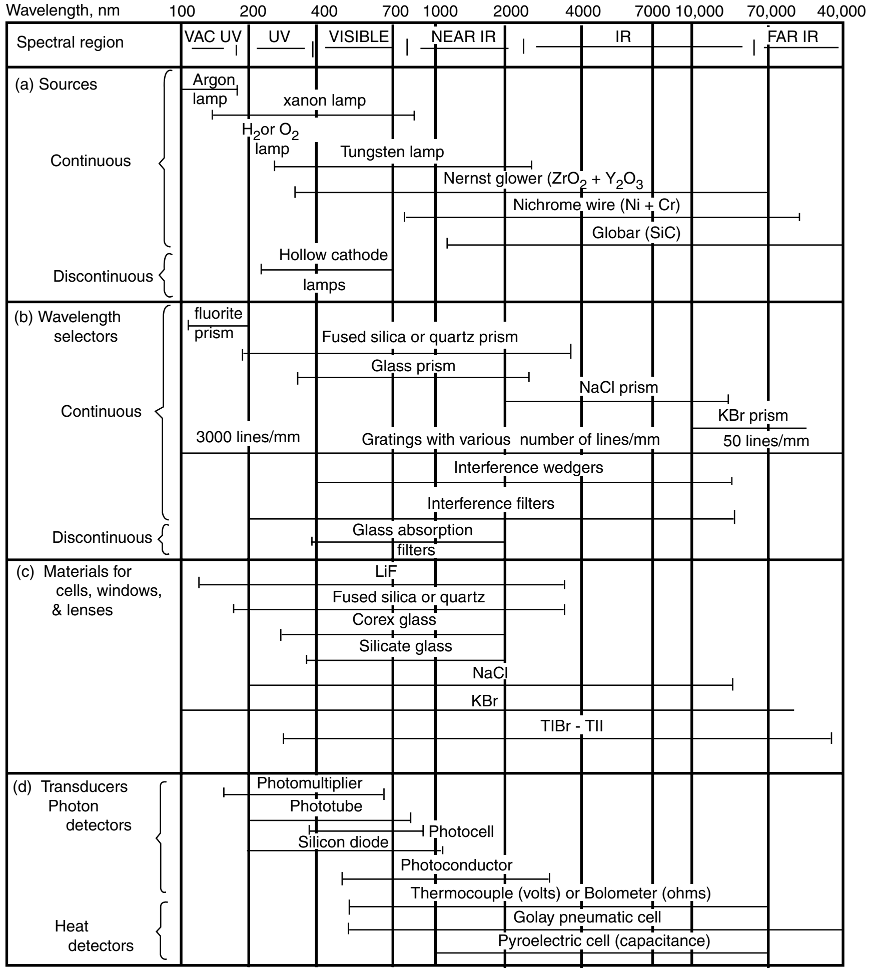


FIGURE 3 Components and materials for optical spectroscopic instruments. (Courtesy of Prof. A. R. Armstrong, College of William and Mary.)

The types of instruments can be characterized by some simple diagrams regardless of wavelength range as given in Figure 4. The basic difference between absorption and emission spectroscopy is the use of a transmitted emr energy source in the former, while in the latter, the sample is stimulated in a thermal or electrical energy source to emit radiation. Photoluminescence (fluorescence and phosphorescence), stimulated by emr, is observed usually perpendicular to the stimulating beam. In nephelometry instrumentation similar to photoluminescence is utilized; turbidimetry can employ absorption instrumentation.

A brief description of the basic parts of the instruments using these phenomena follows.

(1) Energy sources

As noted from the diagrams above, all but emission instrumentation use energy sources that irradiate the sample. In Figure 3 the sources are indicated as a function of the wavelength range of their radiant emissions. Various lamps, i.e., argon, xenon, H_2 (hydrogen) and D_2 (deuterium), and solid state radiators give continuous emissions, i.e., a range of contiguous wavelengths and are used in molecular spectroscopic instrumentation. Photoluminescence and nephelometric instruments use these sources.

For atomic absorption instruments hollow cathode lamps are utilized. They are line (discontinuous) sources providing unique radiation with a narrow bandwidth characteristic of particular element. An individual lamp is usually employed for each element. Some multielement lamps are available.

X-ray sources include x-ray tubes or radioactive sources. The x-ray tube consists of a tungsten cathode that emits electrons when heated. The electrons accelerated by a large potential strike the metal anode generating x-rays characteristic of

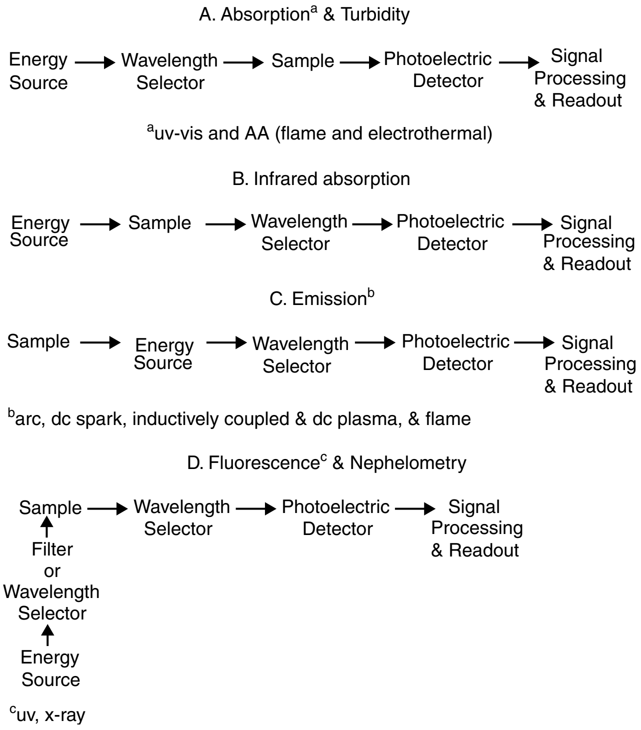


FIGURE 4 Outline of spectroscopic instrumentation.

the particular metal target. Metals such as chromium, copper, iron, molybdenum, rhodium, silver, tungsten and others comprise the anode target. A number of radioisotope sources emit useful x-rays, e.g., iron-55 yields manganese K radiation, cadmium-109 gives silver K radiation, and cobalt-57 provides iron K radiation.

(2) Sample interface

The sample is usually presented as a solution contained in a cell made of material transparent to source radiation for absorption and photoluminescence (see Figure 3). Solid samples are also used. Potassium bromide disks containing homogeneously distributed powdered analyte are used in ir absorption methods.

However in atomic absorption spectroscopy the sample is atomized in a flame, plasma or thermal heat source. In effect the sample container is that volume of flame, plasma or heat source.

Solutions, as well as solid samples in the form of pressed disks, pieces of solids, or solid solutions in borax, are conveniently analyzed in an x-ray fluorescence instrument. Solutions of sufficient thickness are the best sample preparations because of their homogeneity; they may be contained in mylar cells. Obviously the solvent must not contain heavy atoms that fluoresce. Sample surfaces are directly exposed to the x-ray beam (see Figure 5).

In emission instruments the solid sample is placed in an energy source environment, e.g., an electrical arc or spark, a flame, or plasma.

(3) Wavelength selectors

The wavelength selector allows isolation of a particular wavelength segment of the source or transmitted beam. A monochrometer is a selector comprising a grating or a prism which disperses or separates the radiation continuously over a considerable wavelength region. The effective bandwidth of the wavelength, isolated by slits placed before the sample, is quite narrow, 1 nm or less. The grating operates on the principle of interference and the prism by dispersion.

Other wavelength selectors are interference and absorption filters. Their effective bandwidths are about 20 to 50 nm, respectively; they are not continuous. An interference wedge is continuous over a region with an effective bandwidth of 20 nm.

The dispersing device, a single crystal mounted on a rotating table or goniometer (see Figure 5a), is the wavelength selector used in x-ray spectrometers. A specific wavelength and its second and third orders of reflection are diffracted at a given angle of the beam to a particular plane of the crystal. The angle of diffraction depends on the "d" or interplanar spacing of the crystal and the wavelength and is defined by Bragg's law. Some examples of diffracting crystals with their unique wavelength ranges are topaz—0.24 to 2.67 Å, sodium chloride—0.49 to 5.55 Å, and ammonium diphosphate—0.93 to 10.50 Å. (An angstrom, Å, is 10^{-8} cm.) Unlike a prism or a grating that disperses a total spectrum in the spectral regions of the source of radiation, the x-ray monochrometer diffracts a unique wavelength and its orders of reflection depending on the angle of the beam to the crystal plane.

(4) Detectors

In Figure 3 the variety of transducers are listed with their wavelength range of detection. Following is a description of the most commonly used detectors grouped according to their wavelength range.

(a) Uv/visible

(i) Photovoltaic (barrier layer) cells

This detector, which generates its own signal, is sensitive to radiant energy in the visible (350 to 750 nm) region. Light shining on a semiconductor coating, such as selenium or copper(I) oxide plated on an iron or copper electrode, generates a current at the metal–semiconductor interface. A second electrode, a transparent coating of gold or silver on the outer surface of the semiconductor, collects the electrons formed by the action of radiant energy on the semiconductor. The magnitude of the photocurrent is proportional to the number of photons/sec impinging on the semiconductor. This detector is insensitive to low light levels, slow in response, shows a tendency to suffer fatigue, and has a high temperature coefficient. However, photovoltaic cells are rugged, require no separate source of energy and are low in cost. They are used in inexpensive filter photometers.

(ii) Vacuum photoemissive tubes[20]

In a photoemissive detector two electrodes, a cathode with an electron emissive coating and an anode, are enclosed in an evacuated tube. When the saturation potential is applied

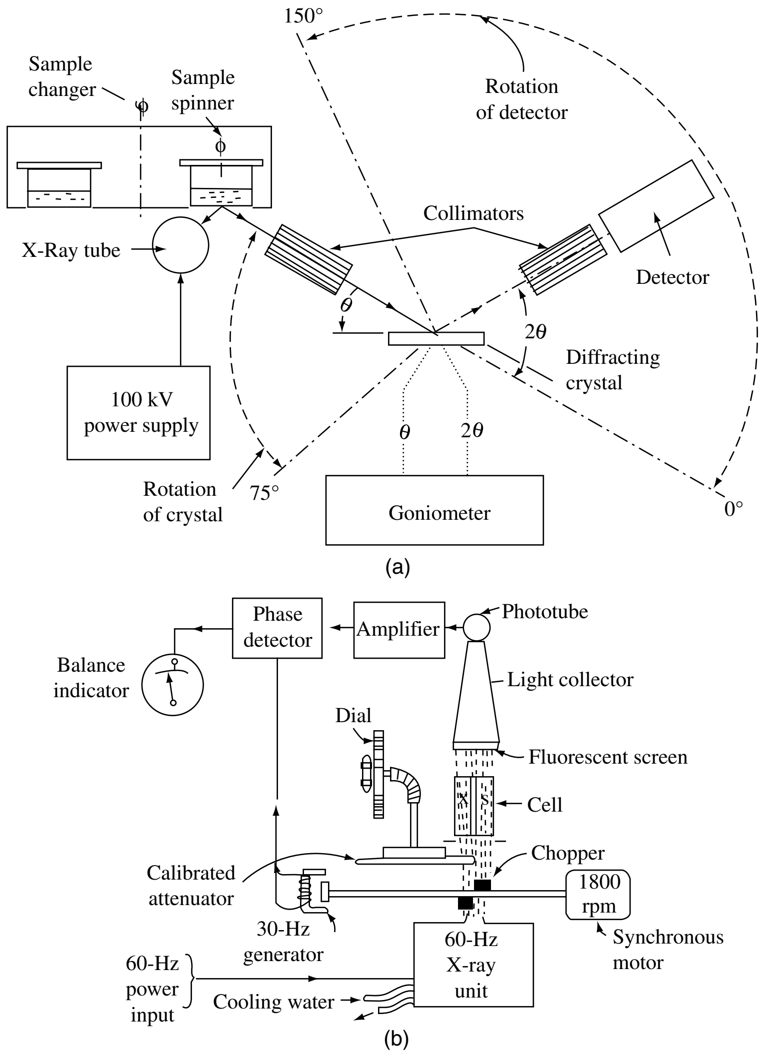


FIGURE 5 (a) Geometry of a plane-crystal x-ray fluorescence spectrometer. Note that the angle of the detector with respect to the beam, 2θ, is twice that of the detector to the crystal face, θ. (Courtesy of Philips Electronic Instruments.) (b) Nondispersive x-ray absorptiometer. (Courtesy of General Electric Co.)

across the electrodes, the radiant energy or photons cause emission of photoelectrons. The photoelectrons are collected at the anode giving rise to a photocurrent. The photocurrent is proportional to the power or radiant energy of the light beam and is independent of the applied potential (see Figure 6). The eleven chemical compositions of various photoemissive cathode coatings determine the wavelength range and sensitivity varying from the uv to the near ir spectral regions. The window in the tube must be transparent to wavelength of interest. The dark current is a small current flowing when no light falls on the cathode and is due to thermal energy and electron emission from potassium-40,[40] K, in the glass tube. It limits the sensitivity of the detector. Although this detector has about one tenth the sensitivity of the photovoltaic cell, its signal may be amplified because of its large internal electrical resistance compared to the photovoltaic detector. The photoemissive detector is used for higher intensity radiation and lower wavelength scanning rates than used with other detectors.

(iii) Photomultiplier tubes[20]

A photomultiplier tube contains a photoemissive cathode followed by a sequential, electron multiplying assemblage of about nine dynodes (electrodes) as illustrated in Figure 7. The voltage of each succeeded dynode increases by 75 to 100 volts. Photoelectrons from the photoemissive cathode are accelerated by the voltage increase of the first dynode causing the release of several electrons for each impinging photoelectron. This multiplier effect continues as the electrons

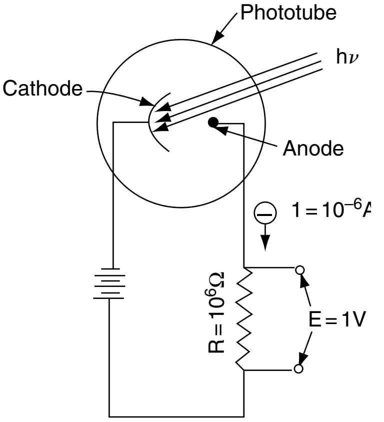


FIGURE 6 Simple phototube circuit. (Reprinted from Ref. (176), p. 441 by permission of Prentice Hall, Inc., Englewood Cliffs, New Jersey.)

contact the succeeding dynodes accelerated by ever higher voltages. A cascade of a large number of electrons is collected by the anode of the ninth dynode. The final photocurrent can be amplified, electronically, before readout. The gain, *G,* can be calculated as follows:

$$G = (fs)^n \tag{8}$$

where *fs,* the secondary emission factor for each stage, depends on the dynode emissive coating and *n* is the number of dynode stages. Using values for *fs* of 3 to 10 for older dynode emissive coatings and 50 for newer coatings and *n* equal to 9 results in gains of about 10^4, 10^9 and 10^{15}, respectively. The response times can vary from 0.5 to 2 nsec (nanosec, 10^{-9} sec). The dark current can be decreased considerably by cooling the photomultiplier detector. Since the dark current is a fairly constant value it may be subtracted or automatically nulled using a potentiometer. The

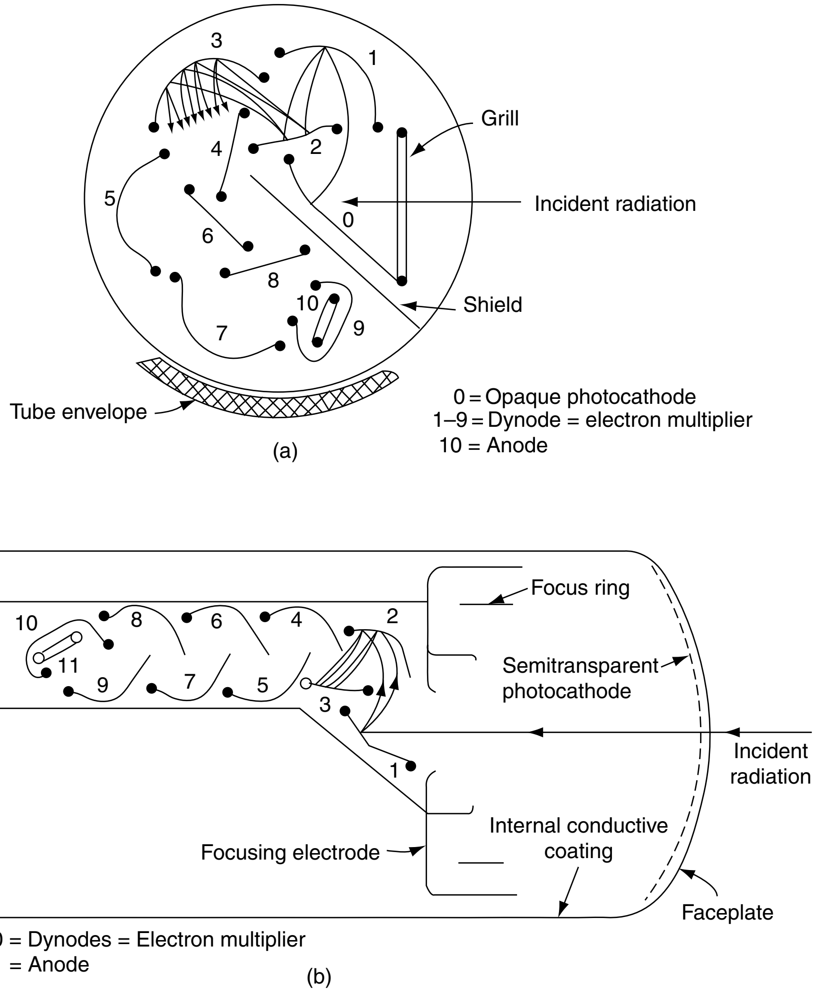


FIGURE 7 Photomultiplier Design. (a) The Circular-Cage Multiplier Structure in a Side-on Tube and (b) The Linear-Multiplier Structure in a Head-on Tube. (Courtesy of the General Electric Company.)

photomultiplier tube is the most widely utilized detector in optical spectroscopic instruments.

(iv) Photodiodes

The photodiode, PD, is a small wafer of silicon dioxide with a shallow layer of *p* and *n* material of the top and bottom surfaces, respectively, to which are attached electrodes. The device is reverse biased. When photons impinge on the optically-active *p* diffusion region, electrons promoted to the conduction band generate a photocurrent that is proportional to the intensity of the optical light beam. The PD detectors are about ten times more sensitive than the vacuum photoemissive tube and are mostly responsive in the visible and near ir regions. Some tubes are sensitive to the uv region at about 200 nm. A lens is optically coupled to each small PD wafer.

Linear arrays or vidicon tubes of these multichannel detectors allow nearly simultaneous detection of a spectrum of wavelengths in instruments operated in the spatial mode (see Section III,B,1,*b,5* Instrumental ensembles), where the detectors are swept electronically. Optical multichannel analyzers consisting of a monochromater, a multichannel detector and a computer are used in flame emission and uv/visible spectrophotometers.[21]

(b) Infrared detectors[22,23]

There are two categories of detectors used for the spectral region above 1.2 μm (micrometer, 10^{-6} m) namely, heat and semiconductor detectors.

(i) Thermocouples and thermopiles

A thermocouple is formed when two wires of a metal are separately joined to the opposite ends of a wire of a dissimilar metal. If the two dissimilar metal junctions are maintained at different temperatures, a thermoelectric current will flow in the circuit. Therefore, if one junction is maintained at a constant temperature, a thermoelectric current will be generated proportional to the temperature of the second junction. The changes in the intensity of incident ir radiation can be detected in ir spectrophotometers using this type of detector. The sensitivity is 6 to 8 microvolts per microwatt and a temperature difference of 10^{-6}°C is detectable (see Figure 8). A thermopile, consisting of a number of series-connected thermocouples, may be miniaturized through thin film techniques to provide an effective ir detector. It has an 80 msec (millisec, 10^{-3} sec) response time with a flat response below a frequency of 0.35 Hz (Hertz).

(ii) Golay cell

The Golay cell is a pneumatic device similar to a gas thermometer. The ir radiation shining on the blackened surface of a sealed cell containing xenon gas causes the gas to expand and distort a diaphragm, a part of the cell wall. The moving diaphragm may be coupled to one plate of a capacitor transducing an ir intensity to a capacitance. In another mechanism the beam of ir radiation is reflected from the mirrored diaphragm surface to impinge on a photocell. The area of coverage of the beam on the photocell changes as a function of the movement of the diaphragm. The intensity of the ir radiation affects the area of the beam that impinges on the photocell and ultimately the magnitude of the photocurrent. The ir beam must be optically focused on the detector. Its response time is 20 msec. Sensitivity is about equal to that of the thermocouple detector. In the far ir it is an excellent detector (see Figure 3).

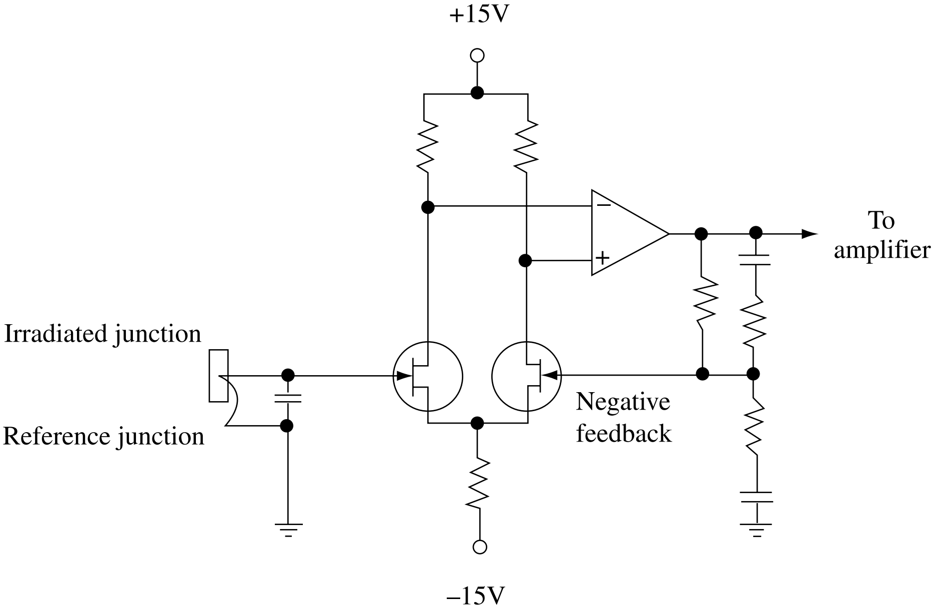


FIGURE 8 Thermocouple and preamplifier (Reprinted from Ref. (180). With permission from the *Journal of Chemical Education*.)

(iii) Pyroelectric detector

The pyroelectric detector is a thin wafer of material such as, $LiTaO_3$ and $LiNbO_3$, placed between two electrodes to form a capacitor. The detector material is a non-centrosymmetrical crystal whose internal electric field is changed as a function of temperature when it is below its Curie temperature. This detector is based on the change of the capacitance of a substance with temperature and is sensitive to the rate of change of the detector temperature. The changing radiation is modulated by chopping or pulsing because the detector ignores steady unchanging radiation. Therefore this detector has a much faster response time than those dependent on temperature directly. Depending on the circuit parameters, the response times may be 1 msec or 10 μsec and the responsivity (detector output/incident radiation) is 100 or 1, respectively. This detector has a large ir range (see Figure 3).

(iv) Photoconductive and photovoltaic detectors

Photoconductive detectors are crystalline semiconductor devices that experience an increase in conductivity upon interaction with a photon. The increase of conductivity is due to the freeing of bound electrons by energy absorbed from the radiation. A Wheatstone bridge is used to measure the change in conductance (see following Section III,B,*2,d.*). The semiconductor materials include the metallic selenides, stibnides or sulfides of cadmium, gallium, indium or lead.

Lead sulfide is commonly used as a detector in the near infrared, 800 to 2000 nm, where it exhibits a flat response. The cell consists of a thin layer of the compound on a thin sheet of quartz or glass kept under vacuum.

Photovoltaic detectors have been discussed for uv/vis spectroscopic instruments (see Section III,B,1,*b.(4),*(a), (*i*)). For ir applications the *p*-type indium antimonide detector, cooled by liquid nitrogen, is available with a sensitivity limit at 5.5 μm. However the two types of lead in telluride detectors extend the ir range. One detector, cooled with liquid nitrogen, has a range of 5 to 13 μm and a second one, cooled with liquid helium, has a range of 6.6 to 18 μm. A minimum response time of 20 nsec (nanosec. 10^{-9} sec) is achieved with these detectors.

(c) X-ray detectors

Gas-filled and semiconductor detectors and signal processors and readout used in the measurement of radioactivity (see Sections III,B,*3,*b,(*1*),(a) and (b) and (2)) are the applicable in x-ray spectroscopy.

(5) Instrument ensembles

The design of an instrument depends on its use and monetary considerations. The several main modes of design are designated temporal, spatial and multiplex. In turn each of these are of the dispersive or nondispersive type.[24]

In the temporal category the instrument scans sequentially, in time, the wavelength in order to determine the intensity. Dispersive systems employ monochrometers that are rotated so as to position the selected wavelength on an aperature or slit preceding the sample or detector. Nondispersive systems utilize a series of absorption or interference filters that are interchangeable.

Spatial systems display the total spectrum with simultaneous determination of the radiation intensities. For a dispersive instrument a monochrometer provides the dispersed radiation and a multichannel detector to detect their intensities. Multichannel detectors utilized are a detector array (silicon diode array or vidicon tube), a number of individual detectors properly positioned, or a photographic plate. In nondispersive systems the radiation beam is divided into a number of beams and each passes through a unique filter followed by a detector.

Multiplex systems employ a single data channel where all the components of the signal are observed simultaneously. A Fourier transform is usually employed to resolve the complex signal into its components requiring the use of a computer. There are distinct advantages to Fourier transform spectroscopy: namely, increased S/N ratio, increased energy throughput, large precision in wavelength measurement, and facility in its use. However, thus far the instrument is costly to acquire and maintain. No further comment will be made in this article about these instruments.[25,26]

Dispersion instruments give more spectral detail because the wavelength selected has a narrower bandwidth wavelength spread. However non-dispersive instruments are usually cheaper, more rugged and have a higher signal to noise ratio. Filter instruments are used frequently in monitoring equipment.

(6) Absorption instrumentation

In the absorption process, radiation passes through the sample and a specific pattern of absorption of different wavelengths occurs leading to a spectrum for that sample. For each wavelength the amount of light absorbed will differ and therefore the amount of transmitted light vary for each wavelength. The intensity of transmitted light is inversely proportional to the concentration of sample and is measured in absorbance, Ab, units. The spectrum is the qualitative factor of identification while the intensity of the transmitted radiation is the quantitative measure. Light is absorbed in the uv and visible region by electrons in the atoms or molecules of a sample. Some elements are identified by atomic absorption, AA, and some functional groups and species by uv/vis spectroscopy. Absorption in the ri is due to vibrational and rotational activity of atoms in molecular groupings, such as functional groups, double, triple, and conjugated bonds, etc.

The spectroscopic curve or an instrument reading provides an absorbance value for a chosen wavelength from which the concentration of the absorbing substance can be computed. The absorbance value represents the degree of attenuation of the radiation of specific wavelength by absorbing substances in the sample solution in the cell. A constant, the absorptivity, *a,* or molar absorptivity, *e,* can be calculated for a pure substance for a given wavelength and solvent. The mathematical relationship between the concentration of a substance, *C,* the path length of the sample cell, *b,* and its absorbance units, *Ab,* is expressed as follows:

$$\text{molar absorptivity, } e = Ab/(b,\text{cm})(C,\text{moles/L}) \qquad (9)$$

$$\text{absorptivity, } a = Ab/(b,\text{cm})(C,\text{g/L}). \qquad (10)$$

The absorptivity, *a,* can be expressed in a number of concentration terms (g/L, grams/Liter and mg/ml, milligrams/milliliter) and path length terms (cm, centimeters and mm, millimeters).

Equations 9 and 10 are a statement of Beer's Law indicating a linear relationship between the absorbance and the concentration for fixed conditions. In analytical determinations the concentration can be calculated using equations 9 or 10. Also used in analysis is a calibration curve, whose slope is absorptivity. It can be drawn using concentrations and corresponding absorbance values. Beer's Law prevails for many substances, but there are deviations for some substances due to chemical, instrumental, and physical phenomena.

(a) Uv/visible instrumentation[27]

A uv/visible instrument consists mainly of uv and visible energy sources (lamps), a wavelength selector (grating, prism or filter), reference and/or sample cell, a detector (photodetector or photomultiplier) and a readout device (recorder, analog or digital meter, etc.) (see Figure 4A). Figures 9 and 10 illustrate the arrangement of these parts for photometers and spectrophotometers, respectively. The distinction between the two types of instruments is that a photometer uses a filter and a spectro photometer a grating or prism as a wavelength selector. In Figure 9 the difference between a single and double beam instrument is shown and the arrangement also refers to a spectrophotometer. The various parts that transmit the light beam such as lens, cells, mirrors, transmission gratings and prisms must be transparent to uv light and be fabricated of fused silica or quartz. Flint glass can be used in the visible region.

(b) Infrared instrumentation[28]

Figure 11 is a schematic of a double beam ir spectrophotometer. The parts and functions are similar to a un/visible instrument, however the arrangement differs—the light beam passes through the sample and then the wavelength selector in contradistinction to the uv/visible instrument (see Figures 4A, 9, and 10). Materials transparent to ir are the alkali metal chloride, bromide, and iodide salts.

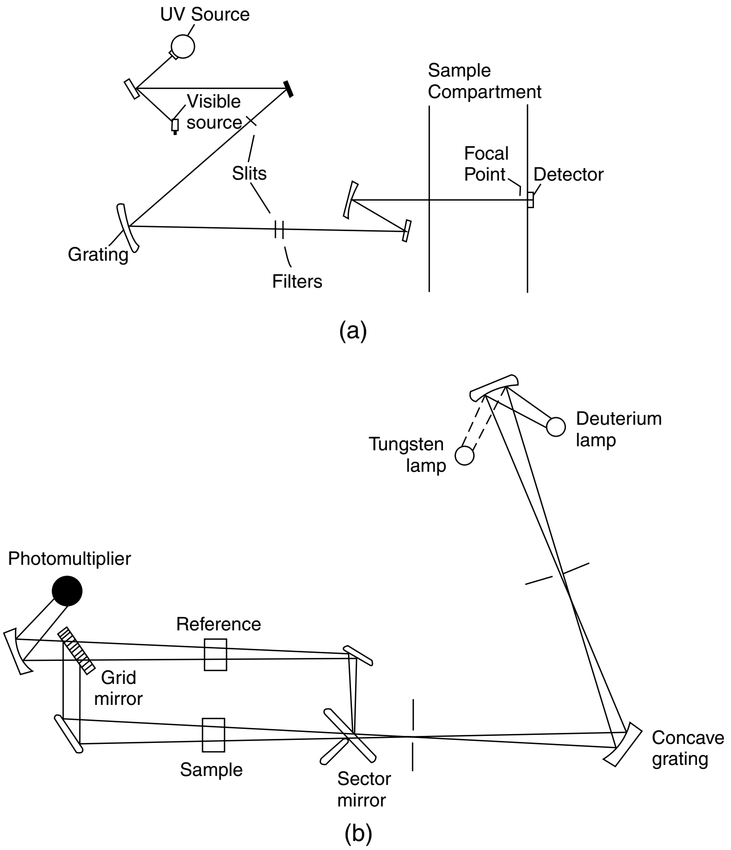


FIGURE 9 Single- and double-beam uv-visible spectrophotometers. (a) Beckman DU^{R} Series 60, single-beam. (Courtesy of Beckman Instruments, Inc., Fullerton, CA.) (b) Hitachi Model 100–60, double-beam. (Courtesy of Hitachi Instruments, Inc., Danbury, CT.)

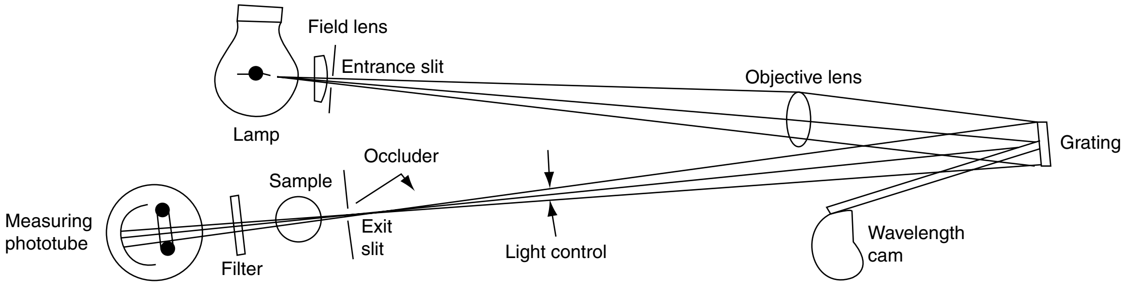


FIGURE 10 An example of a simple spectrophotometer. The SPECTRONIC^R 20. (Courtesy of the Milton Roy Co., Rochester, NY)

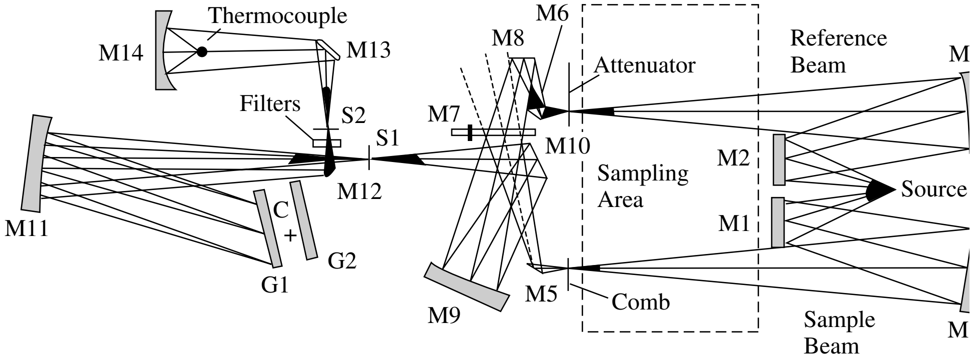


FIGURE 11 Schematic diagram of a typical double-beam infrared spectrophotometer. The symbol M1 through M14 indicate mirrors; S1 and S2 indicate slits; and G1 and G2 indicate gratings. (Courtesy of the Perkin-Elmer Corporation.)

(c) Atomic absorption instrumentation[29]

A schematic for a single beam atomic absorption spectrophotometer is given in Figure 12. Both flame and electrothermal atomizers may be utilized in this instrumentation (see Figure 4A). The hollow cathode lamp is the energy source generating uv or vis radiation that passes through the sample. The flame or thermal area of the electrothermal device acts as the sample cell where the sample solution that has been nebulized (formed into a fine aerosol) is atomized. (Atomization is the formation of free atoms through thermal energy in the flame or thermal area.) The flame is generated from various fuel mixtures: acetylene and the oxidants air, oxygen or nitrous oxide; hydrogen and the aforementioned oxidants; and natural gas and air or oxygen. Each of these mixtures as well as the fuel/oxidant ratio determines the flame temperature, a critical condition for atomization.

(7) Emission instrumentation[30,31]

Emission of emr by elements and some chemical entities, energized by flames, plasmas, and arc, is the basis of emission spectroscopy. The electrons are energized and move to higher energy levels on absorption of the energy. On relaxation, the electron returns to a lower energy level and the absorbed energy is emitted as radiation.

Emission methods give rise to atomic spectra by a series of atomization techniques: namely, flame, inductively coupled argon plasma (ICP), electric arc and spark, and direct current argon plasma, DCP (see Figure 4B). An emission spectrophotometer capable of using plasma and arc and spark sources is illustrated in Figure 13.

(8) Photoluminescence instrumentation

The occurrence of fluorescence and phosphorescence (photoluminescence) refer to substances which on excitation by radiation emit light on relaxation of the excited species. In resonance fluorescence the wavelengths of excitation and emission are the same. However, in many cases the wavelength of emitted radiation is longer than that of the exciting radiation. The difference in fluorescence and phosphorescence is the time delay between excitation and emission. The former is quite small ($<10^{-6}$ sec), while the latter has a time delay of several seconds or longer. Fluorescence can occur in a number of organic and metal-organic complex molecules, and gases on excitation with uv light.

(a) Uv/visible[32]

A general schematic for this instrumentation is given in Figure 4C. Fluorescence of molecular substances is measured

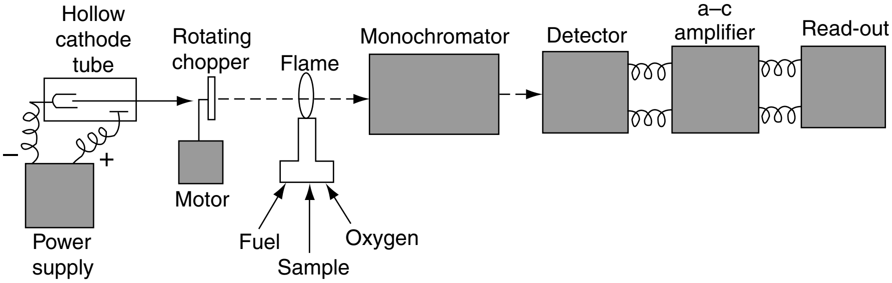


FIGURE 12 Components of an atomic absorption spectrophotometer. The flame may be replaced by a furnace. (Reprinted from Ref. (176), p. 464 by permission of Prentice Hall, Inc., Englewoods Cliffs, NJ.)

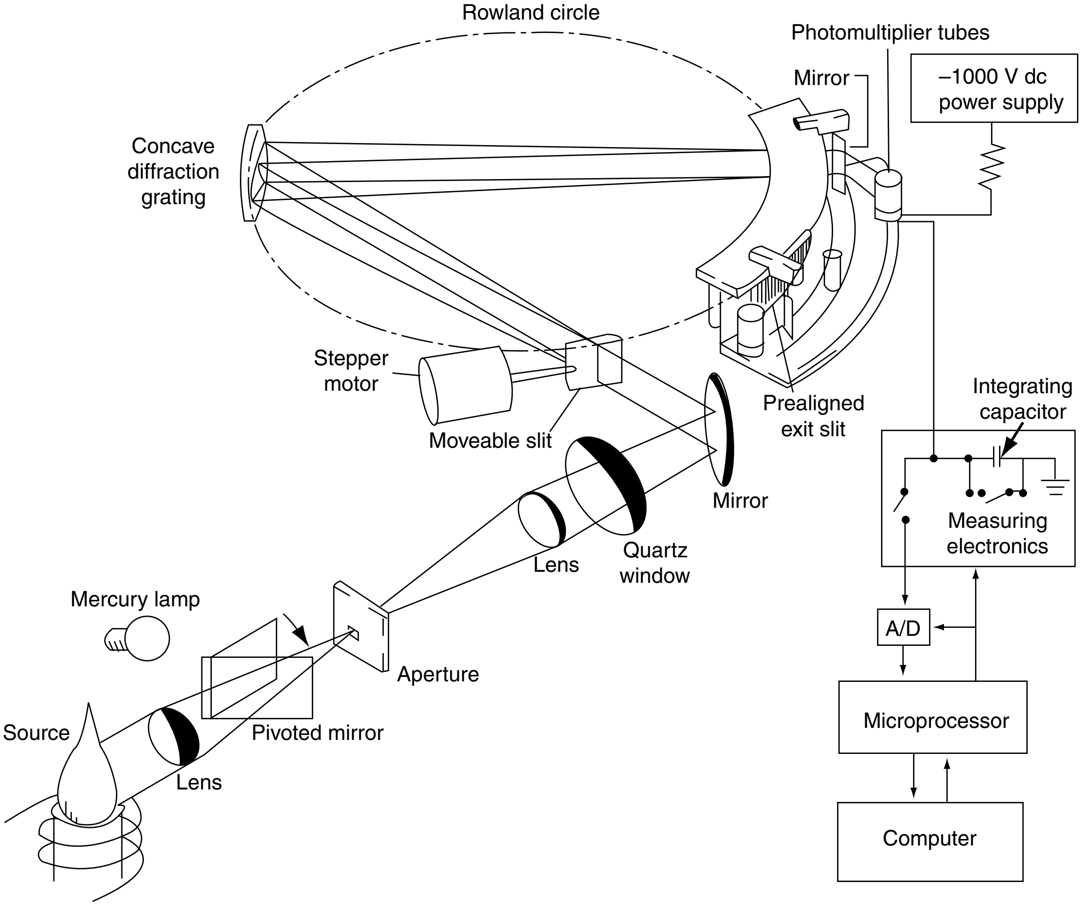


FIGURE 13 A plasma multichannel spectrometer based upon Rowland circle optics. (Courtesy of Baird Corp./IMC Bedford, MA.)

in a fluorometer (see Figure 14) or a spectrofluorometer (see Figure 15).

(b) Atomic[33,34]

Atomic fluorescence is a fairly new instrumental method that has been used for environmental samples. The instrumental arrangement is given in Figure 4C. Because of the dearth of commercial sources of the instrument and no large benefits compared to other atomic spectroscopic instruments there is a small number of literature citations for its use.

(c) X-ray[35]

X-ray fluorescence spectroscopic instrumentation utilizes an x-ray source, i.e., x-ray tube or radioactive source, to energize electrons of the inner orbitals of atoms which on

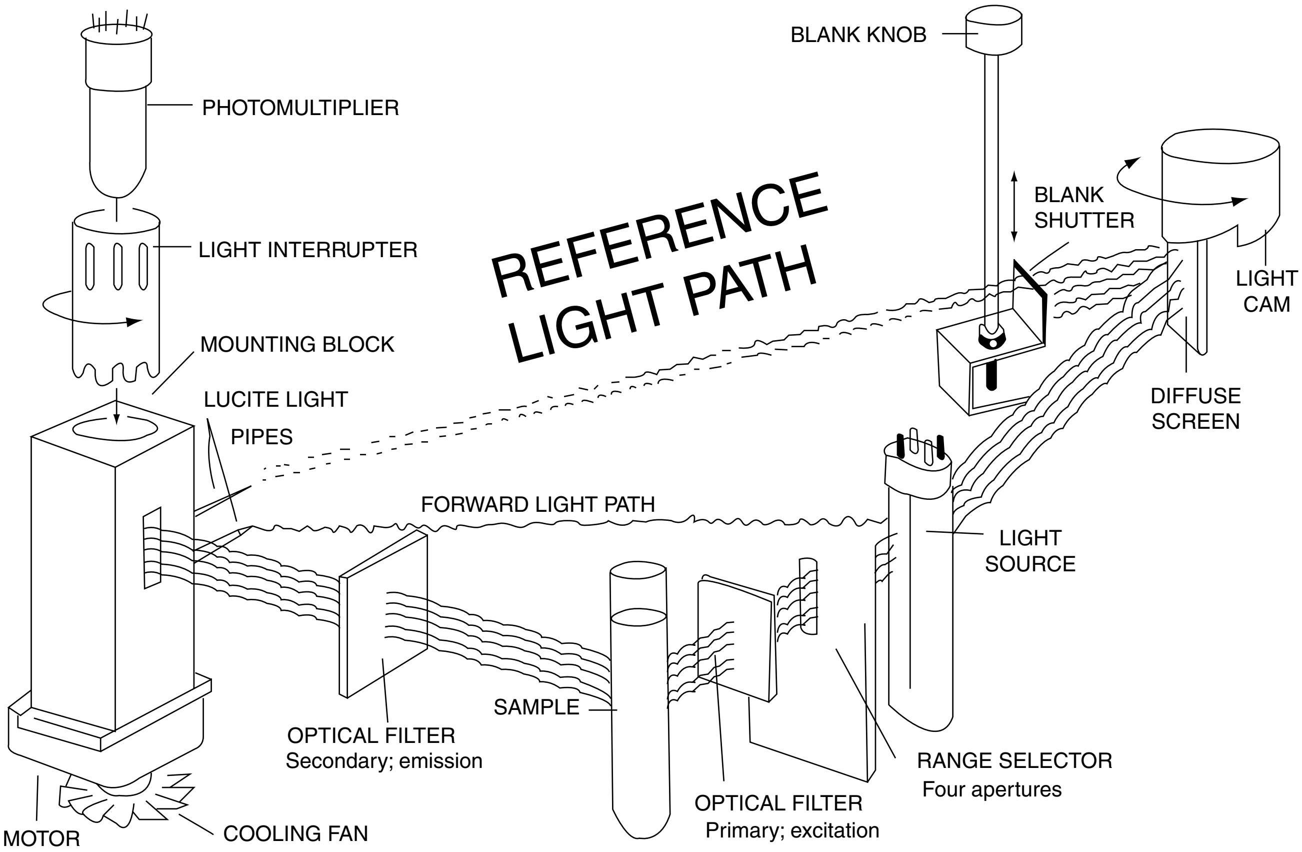


FIGURE 14 The Model 112 Turner Digital Filter Fluorometer. (Courtesy of the Turner Division of Unipath, Mountain View, CA.)

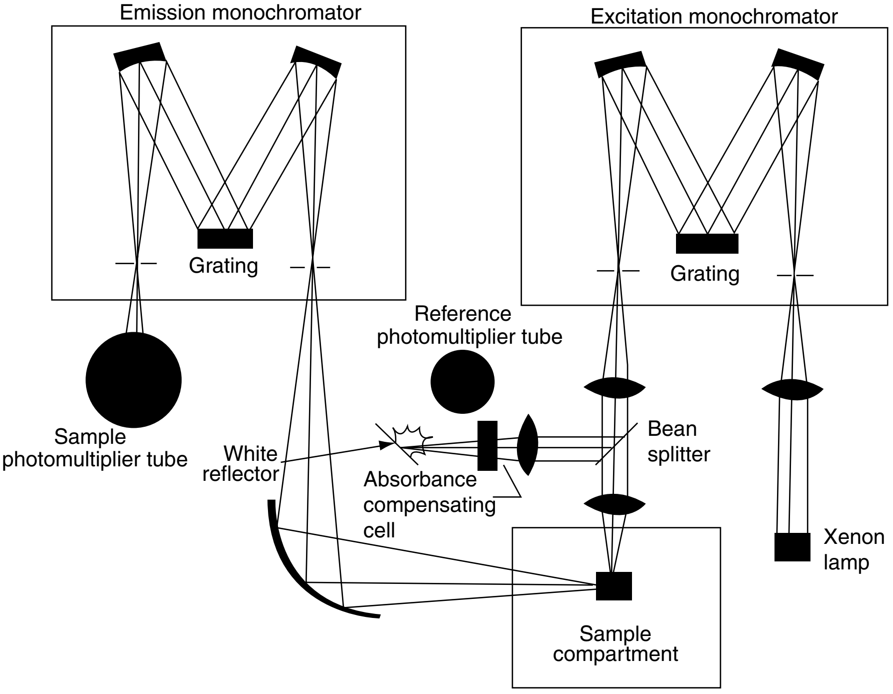


FIGURE 15 A spectrofluorometer. (Courtesy of SLM Aminco Instruments, Urbana, IL.)

relaxation emit fluorescence emission. The wavelengths in the fluorescence emission are unique for different elements. This information may be delineated by wavelength or energy dispersion instruments (see Figure 4C). Wavelength dispersion is carried out with a crystal as a diffraction grating with subsequent detection by a gasfilled detector (see Figure 5), while energy dispersion may be accomplished by a lithium-drifted silicon detector and energy-discriminative electronic circuits (see Figure 16). Non-dispersive instruments use filters.

Elements with atomic numbers greater than 7 (oxygen = 8) fluoresce on irradiation with x-rays. Useful irradiating wavelengths extend from 0.5 to 2.5 Å. Due to the large absorption of wavelengths greater than 2.5 Å by air and spectrometer windows, elements of atomic numbers below 22 (titanium) cannot be detected. Elements with lower atomic numbers can be detected with a change of atmosphere; down to aluminium (atomic no. = 13) in helium and to boron (atomic no. = 5) in a vacuum.

(9) Nephelometric & turbidimetric devices[36]

The presence of turbidity, suspended matter and colloids, in liquid results in the scattering of a beam of incident light passing through the liquid. The scattering process is elastic; the wavelength of incident and scattered light is the same. Particle shape and size distribution, size relative to the wavelength of the incident light, concentration of particles, and molecular absorption effect the angular distribution of scattered light intensity. Since the scattering phenomenon is so complicated analytical results are empirical depending on the use of standards. However, differences in the design of instruments leads to different values for the same standard.

Turbidimetry refers to the measure of the decrease in the intensity of a beam of light undergoing scatter by suspended or colloidal particles in a liquid. If the transmittance is less than 90%, this method is effective. A filter photometer illustrated in Figure 9 is a suitable instrument.

If the intensity of the scattered beam is measured at an angle to the transmitted beam, then the phenomenon is known as nephelometry. Right angle scatter is commonly used for a number of readings, although forward scatter is more sensitive to large particles. Stray light caused by scratches, dirt or condensation on cell walls leads to a positive error. Figure 14, a simple fluorometer, measuring scattered light at 90° can be employed for nephelometric measurements. A surface scatter instrument shown in Figure 17 is used to eliminate stray light. No cell is employed since the flowing water sample surface is directly illuminated with the light beam. Figure 18 illustrates a low range turbidimeter.

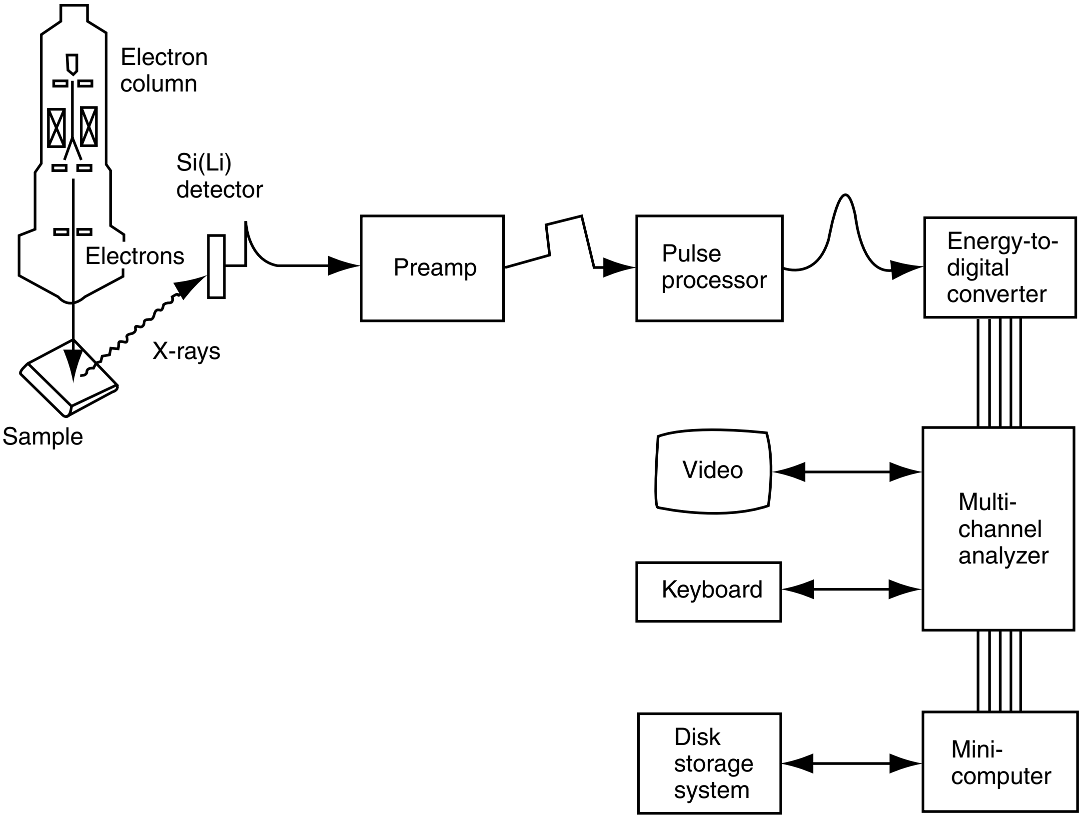


FIGURE 16 Components of a typical energy-dispersive microanalysis system. The Si(Li) detector is cooled in a liquid nitrogen cryostat. The charge pulse from the Si(Li) detector is converted in the preamp to a step on a voltage ramp. The pulse processor converts the signal to a well-shaped voltage pulse with an amplitude proportional to the energy of the x-ray. (Courtesy of the Kevex Instruments, Inc.)

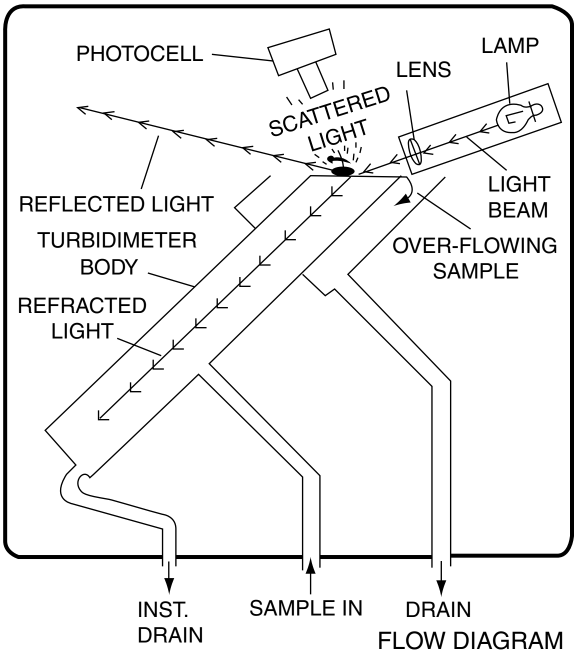


FIGURE 17 Surface scatter turbidimeter. (Used with permission of Hach Co., Loveland, CO.)

(The names turbidimeter and nephelometer appear to be used interchangeably to describe a device measuring turbidity from the intensity of scattered light.) The unit of measurement is the NTU, nephelometric turbidity unit.

Colored constituents in the sample can cause error by absorbing light. A correction can be made in a number of ways: namely, by using a wavelength of light not absorbed by the solution, by making an absorption reading of the clarified solution, or using an instrument that combines both readings.[37]

(10) Other spectroscopic instruments

Mass spectrometry is treated in the gas chromatography/mass spectrometry Part Two Section, *4,c,(2),(a)*. NMR (nuclear magnetic resonance) spectroscopy is not discussed in this article. The technique is, indeed, a most fruitful means of identifying chemical entities and their structures. However, the use in water analysis is not a primary activity. In the identification for known natural and anthropogenic materials its use would be invaluable. No doubt the next edition of this article or an expansion of this article will contain a section on NMR spectroscopy.

(c) Applications of spectroscopic instruments

Standard Methods[2] includes a number of spectroscopic methods using various instruments for the analysis of metals (see Table 5). Colorimetric methods using uv/vis absorption spectroscopy are available in Standard Methods for the following non-metals: bromide, fluoride, iodide, residual chlorine, cyanide, ammonia, nitrate, nitrite, phosphate, sulfide and sulfite. The determination of turbidity in water by nephelometry and the analysis of sulfate ion by a turbidimetric method appears in Standard Methods.[2] See Part Two Section C for more applications.

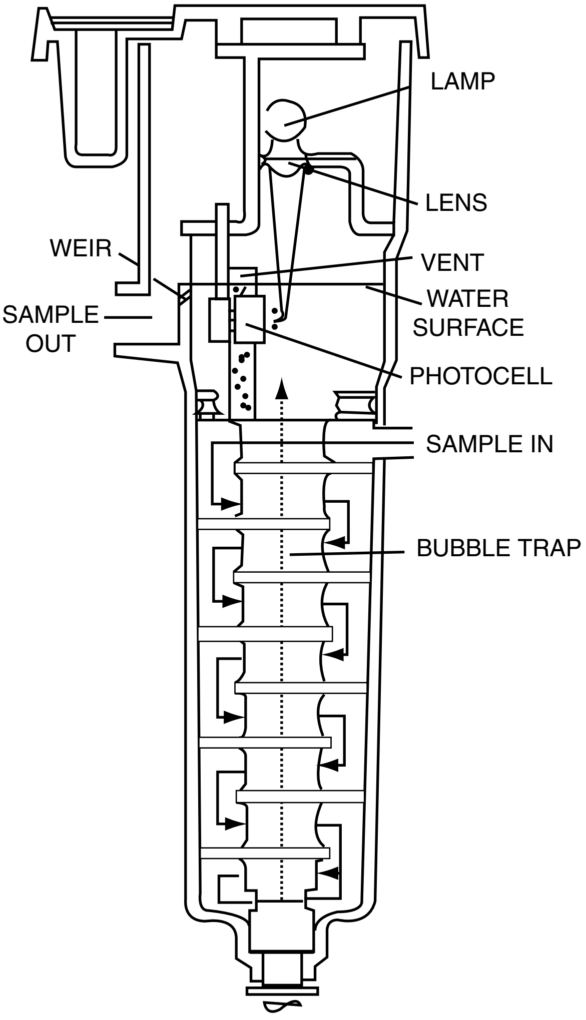


FIGURE 18 Low range turbidimeter. (Used with permission of Hach Co., Loveland, CO.)

2. Electroanalytical instrumentation

Electroanalytical chemistry encompasses a wide variety of analytical measurements and includes three different types of correlations. The first type concerns the relationship between potential, current, conductance (or resistance), charge (or capacitance) and the analyte. For the second is the determination, during the titration, of the analyte and, ultimately, the endpoint by electrochemical means. The conversion of the analyte by an electric current to a convenient gravimetric or volumetric form is the third type. In this section a number of methodologies and their corresponding instruments will be discussed. They include potentiometry, voltammetry, amperometry, coulometry, conductance measurements, and titrations using potentiometry and amperometry for endpoint detection.

TABLE 5

Metals Analysis by Spectroscopy[a]

Metals	Atomic Absorption	Method	Flame
Sb, Bi, Cd, Ca, Cs, Cr Co, Cu, Au, Ir, Fe, Pb Li, Mg, Mn, Ni, Pd, Pt K, Rh, Ru, As, Na, Sr Tl, Sn, Zn		direct aspiration	air-acetylene
Al, Ba, Be, Mo, Os Rh, Si, Th, Ti, V		direct aspiration	N_2O-acetylene
[b]Al, Sb, As, Ba, Be Cd, Cr, Co, Cu, Fe, Pb Mn, Mo, Ni, Se, Ag, Sn		electrothermal	none
[c]Cd, Cr, Co, Cu, Fe Pb, Mn, Ni, Ag, Zn		chelation[d]/ extraction[e]	air-acetylene
[c]Al, Be		chelation[f]/ extraction[e]	N_2O-acetylene
Hg		cold vapor	none
As, Se		hydride formation	argon-hydrogen or N_2-hydrogen
	Inductively Coupled Plasma (atomic emission)		
	Al, Sb, As, Ba, Be, B, Cd, Ca, Cr, Co, Cu, Fe, Pb, Li, Mg, Mn, Ni, K, Se, Si, Ag, Na, Sr, Tl, V, Zn		
	Flame Emission		
	Li, K, Na, Sr		
	Colorimetric/Spectrophotometric		
	Al, As, Be, B, Cd, Cr, Cu, Se, Si, Ag, V, Zn		

[a] From Ref. (2); [b] microquantities; [c] low conc.; [d] ammonium pyrrolidinedithiocarbamate; [e] methyl isobutyl ketone; [f] 8-hydroxyquinoline.

Potentiometric Instruments An instrument consisting of an electrochemical cell containing indicating and reference electrodes and electronic means for measuring cell potential to within 0.001 volt (depending on the accuracy desired) may be optimized to measure concentrations of various ions and molecules. The indicating electrode is the sensor which gives the specific electrochemical detection of the analyte in question, while, the function of the reference electrode is to provide a stable reference potential for the indicating electrode. The potential of the indicating electrode changes as a function of the concentration of the analyte according to the Nernst equation,

$$E = E^0 - (0.0591/n)\text{Log }\alpha_{red}/\alpha_{ox}. \quad (11)$$

The activities of the reduced and oxidized species are α_{red} and α_{ox}, respectively, n is the number of electrons in the redox reaction and E^0 is the standard reduction potential of the redox couple when the activity is one.

For species, Sp, its activity, α_{Sp}, is a variable related to its concentration in Moles/L. The symbol, [Sp], represents the concentration of Sp in Moles/L. The α_{Sp} is defined by the equation,

$$\alpha_{Sp} = f[Sp]. \quad (12)$$

The activity coefficient, f, varies inversely with the ionic strength of the solution. (Ionic strength is a function that can be calculated from the concentration of ions in solution and their ionic charges.) Therefore, at high concentrations of electrolytes, the ionic strengths are high, and the activity coefficients, small, less than one. Activity and concentration are approximately equal when the ionic strength is low and f is equal to values between 0.90 to 1.

We shall make a simplifying, practical, and slightly erroneous substitution replacing activity with concentration [Moles per liter] as follows:

$$E = E^0 - (0.0591/n)\text{Log}[\text{red}]/[\text{oxid}]. \quad (13)$$

A convenient expression for the log of the concentration of some species if the p function. By definition it is

$$pF = \text{Log } 1/[F]. \quad (14)$$

The most common p function is pH, defined by Log $1/[H^+]$ as measured by a glass electrode (a membrane electrode). However this function applies to other species, e.g., pOH, pCa, pNO_3 etc. The p functions can replace the log of the reciprocal of the concentration terms.

The specificity of the indicating electrode is determined by the materials of construction and the structure. There are several types of indicating electrodes, metallic electrodes of the first, second, and third kinds and membrane (selective) electrodes including glass, solid-state, liquid membrane, gas sensing, and enzyme and microbial electrodes. The latter two electrodes can be of the potentiometric and amperometric types. For organizational convenience and simplicity all these electrodes shall be discussed in this section entitled, potentiometric instrumentation.

(1) Metallic electrodes[12]

An electrode of the first kind of a metal wire, rod or plate in equilibrium with its metallic ion in solution. For example, a silver indicating electrode, a silver wire, in contact with silver ions, Ag^+, gives the potential mathematically described by the Nernst equation,

$$E = E^0 - (0.0591/n)\text{Log } 1/[Ag^+] \quad (15a)$$

$$e = E^0 - (0.0591/n)\text{ pAg} \quad (15b)$$

Metallic electrodes of inert metals such as platinum, gold, palladium, etc. and non-metals such as carbon and boron carbide may be used as indicating electrodes when the redox couples are soluble species as Fe^{2+}/Fe^{3+}. However the electrode response is not always reversible for some combinations of electrodes and redox couples leading to nonreproducible potentials.

An electrode of the second kind consists of a metal electrode whose surface is coated with a slightly soluble salt of that metal or a metal electrode in contact with a solution containing a low concentration of a complex of that metal. An example of the former is a mercury electrode coated with slightly soluble mercurous chloride making the electrode potential sensitive to chloride ion concentration in the solution. Most reference electrodes are of the former type. Two examples are the calomel (mercury/mercurous chloride) and the silver/silver chloride electrodes. Figure 19 is a schematic of a calomel reference electrode.[38] A solution of known concentration of chloride ion bathes the electrode allowing the potential to remain constant. Potassium chloride concentrations at several different levels, 0.1 or 1.0 M or saturated are often used providing several different reference electrodes. As predicted by the equilibrium and Nernst equations for the calomel reference electrode (see equations 16a and 16b), different chloride ion concentrations will result in different potentials.

$$Hg_2Cl_2(s) + 2e \rightleftharpoons 2Hg(1) + 2Cl- \quad (16a)$$

$$E = E^0 - (0.0591/n)\text{Log } [Cl^-]^2 \quad (16b)$$

Since $Hg_2Cl_2(s)$ and Hg(1) are a solid (s) and liquid (1), respectively, insoluble and not in ionic form in the electrode electrolyte (see Figure 19), they are not represented in the Nernst equation. (2e represents the number of electrons taking part in the reaction.) Similar equations can be written for the silver/silver chloride reference electrode.

A metal electrode made sensitive to a second metal is an electrode of the third kind. The metal electrode is immersed in the analyte solution containing a small concentration of the metal complex and of a similar complex of the second metal. The electrode potential is dependent on the concentration of the ion of the second metal.[39]

(a) Oxidation–reduction potential, ORP[40]

The measurement of the ORP of a system can provide valuable information about its oxidative state. It must be understood that the ORP value is the overall potential of the system being measured, not necessarily at equilibrium, and very seldom relates to one species. With caution in interpretation in mind, the ORP value can be used essentially in two area: biological and chemical inprocessing measurements and control.

The ORP measurement is made in a cell using a nonpolarizing indicating electrode, a reference electrode and a suitable potential readout device—an electronic voltmeter. The indicating electrode, an electrode of the third kind, is usually a noble metal such as gold or platinum in the form of a wire or button. The redox system equilibrates electronically with the indicating electrode and the reference electrode. For the following system,

$$M^{b+} + (b - a)e \rightleftharpoons M^{a+} \quad (17)$$

one can write the reduction potential, $E_{Mb\,+\,/Ma\,+}$, using the Nernst equation, where b and a are the number of charges and $b > a$, and E^0 is the standard reduction potential of the system.

$$E_{Mb+/Ma} + = E^0_{Mb+/Ma+} - 0.0591/(b - a)\log \alpha_{Ma+}/\alpha_{Mb+} \quad (18)$$

If one assembles a cell to measure the ORP of this system using an indicating electrode and a standard hydrogen electrode, the cell potential value measured is the potential of the

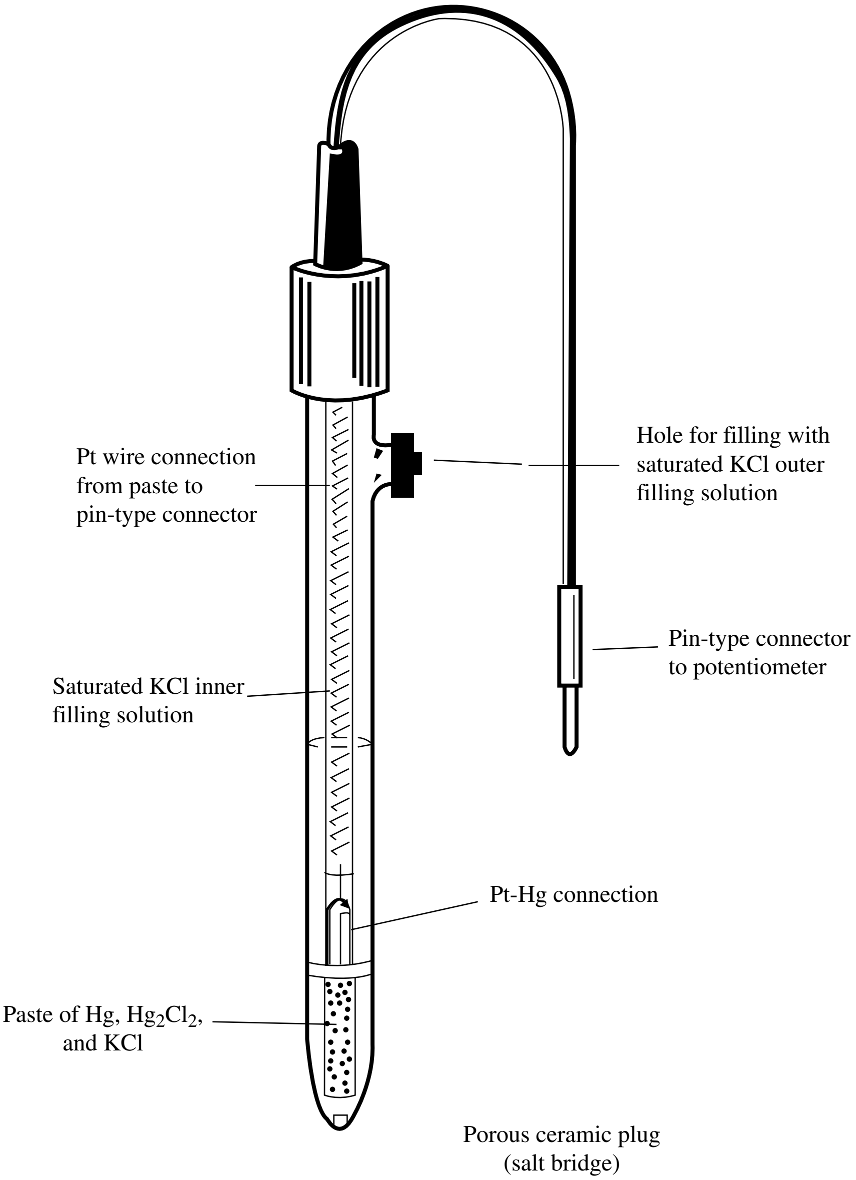


FIGURE 19 Commercial saturated calomel reference electrode. (Courtesy of Thomas Scientific.)

indicating electrode, $E_{Mb+/Ma+}$, since the potential of the standard hydrogen electrode is zero. This potential is referred to as, E_h, the "h" referring to the standard hydrogen electrode. The sign, however depends on the sign of the standard reduction potential, E^0, and the relative concentrations of M^{b+} and M^{a+}. In the cell the sign of E_h depends on the direction of the flow of the electron current; if the flow is from the hydrogen to the indicating electrode, the ORP is negative and flow in the opposite direction indicates a positive value.

However, since the standard hydrogen electrode is not practical to use, a calomel or silver-silver chloride reference electrode is conveniently employed. The cell potential is then designated, E_c, when a calomel reference electrode is used. The E_c value is converted to the E_h or ORP value using the standard potential for the calomel reference electrode, as follows:

$$E_h = E_c + E_c^o . \tag{19}$$

The ORP or E_h measuring cell can be standardized readily using a saturated solution of quinhydrone. However, gold electrodes show no concentration dependence, whereas platinum electrodes show a marked dependence in the quinhydrone solution.[41]

The ORP values can be expressed as pE. The value of pE measures redox power, at equilibrium. Large positive values indicate a strongly oxidizing environment and smaller or negative values a strongly reducing environment. Varying values for pE of −10 to 14 are found in water systems. E_h − pH limits for natural environments and E_h values for sanitary engineering processes have been published.[42,43,44]

(2) Membrane electrodes[45,46,47]

Membrane electrodes are those electrodes that contain a selective membrane that responds to the presence of specific ions or molecules. In these electrodes a potential develops across a selective membrane, ideally, in response to a specific ion or molecule. However, the mechanism of physical and/or electrochemical action differs for the different types of membrane electrodes. The membrane, the differentiating feature of these electrodes, can be crystalline, glass, or a porous polymer. All membrane electrodes contain, internally, a reference electrode and a known standard activity of the analyte ion (see Figure 20). Membrane electrodes are quite different from metal electrodes in their electrochemical and physical mechanisms and structural design. However, as in all potentiometric measurements the membrane (indicating) electrode forms a cell with the reference electrode. Therefore, potentiometric measurements using selective electrodes have two reference electrodes; the external reference in the sample solution and the internal reference that is part of the selective electrode. Those substances important in wastewater analysis and detected by these electrodes are listed in Table 6.

Two subdivisions of membrane electrodes can be envisioned: namely, ion- and molecular-selective electrodes. In this article the electrodes are categorized as follows: glass, solid-state, liquid membrane, gas sensing and enzyme and microbial electrodes. The latter three electrodes detect molecules, whereas the first three are sensitive to cations and anions. Enzyme and microbial electrodes are designed to operate in either the potentiometric or amperometric mode. For simplicity both types will be treated in this, the potentiometric instruments section. However the type will be clearly indicated.

(a) Glass electrodes[48]

The glass electrode commonly refers to the glass membrane electrode used to measure pH (or hydrogen ion activity or concentration). A glass electrode may also be sensitive to cations of sodium, potassium, lithium, and silver[49]. This electrode has a selective membrane, a thin glass bulb of specific chemical composition, that responds most sensitively to a specific cation. Inside of the bulb is an internal standard analyte solution of known activity and a reference electrode; the outside surface contacts the sample solution (see Figure 20).

The mechanism of the function of the glass membrane is quite involved and the following explanation is offered.[50] The glass membrane has an inner dry layer, about 50 μm thick, sandwiched between 5–100 nm thick layers of hydrated glass (silicic acid gel): that is, the hydrated layers are on the inside and outside surfaces of the glass bulb. The hydrated glass layer attracts specific cations via an ion exchange or adsorption mechanism to, most likely, an ionic fixed sites. Conductivity is due to transport through the dry layer, ideally by the specific ion, but actually by a cation such as sodium, Na^+; hydrogen ion, H^+, is not transported through the dry layer. The inside layer interfaces with a fixed standard analyte concentration and therefore a fixed quantity of cations on its sites, while the outside layer contacts the sample solution whose analyte activity may vary with each sample. A membrane potential develops due to the difference of charge between inside and outside hydrated layers.

The glass electrode must be standardized with solutions of known cationic activity, since the glass membrane does not respond ideally giving rise to errors due to asymmetry potentials, non-ideal transport effects, etc. These electrodes have very high resistances, therefore direct-reading, solid-state instruments with large internal resistances are required: that is, an electronic voltmeter. The instrument has an adjustment to corret the meter reading during the standardization.

(b) Solid-state electrodes[46,47]

Solid-state electrodes are of two kinds: The homogeneous type is fabricated from insoluble, crystalline, metallic salts, while the heterogeneous type is of insoluble, non-crystalline

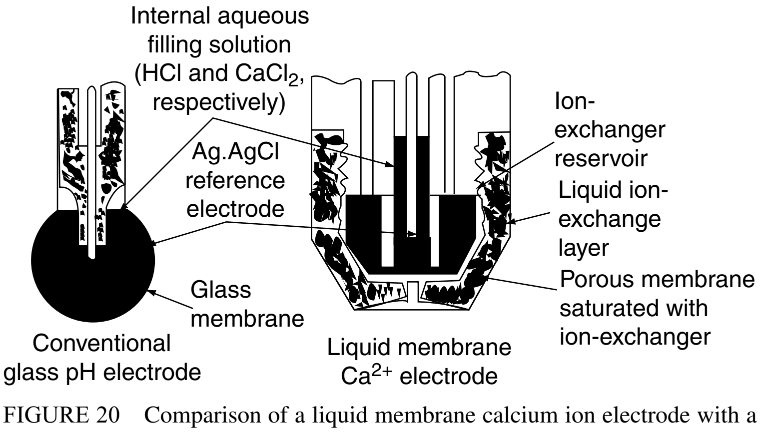


FIGURE 20 Comparison of a liquid membrane calcium ion electrode with a glass electrode. (Orion Research Inc.: Boston, MA with permission.)

precipitates of metallic salts. The first type is made of a slice of a single crystal, a pressed pellet, or a casting from molten material of the crystalline metallic salt, that comprises the selective electrode membrane. Insoluble metallic precipitates encapsulated in inert, polymer matrices constitute the second type.

The fluoride electrode, a homogeneous solid-state electrode, is composed of a 10 mm diameter disk of lanthanum fluoride, LaF_3, doped with europium fluoride, to increase conductivity. The crystalline slice of LaF_3 is 1–2 mm in thickness. Figure 21 illustrates a solid-state electrode. The internal solution in a fluoride electrode serves two purposes: It bathes the membrane with a known standard activity of fluoride ions and provides the internal silver/silver chloride reference electrode with a known activity of chloride ion.

The inside and outside layers of the LaF_3 membrane ionize as follows:

$$\underset{\text{surface}}{LaF_3} \rightleftharpoons \underset{\textit{surface}}{LaF_2^+} + \underset{\text{solution.}}{F^-} \tag{20}$$

The amount of charge on each face depends on the fluoride concentration in the sample (outside) or the standard (inside) solutions. The layer interfacing with the lower concentration of fluoride ions will have a positive charge with respect to the other face; the equilibrium will favor the ionization leaving a surplus of positively charged LaF_2^+ ions in the layer. This difference of charge mirrors the variation in fluoride concentration of the two solutions giving rise to the membrane potential. The differences in the chemical potentials for the two ion exchange equilibria for the two faces results in the membrane potential (see equation 20). There is some similarity to the mechanism encountered in the glass electrode.

The pF value is expressed in terms of the cell potential, Ev, a constant, 0.0591, and Le, a combination of the internal and external reference electrode potentials and junction potentials, and an expression for the activity of the internal standard.

$$PF = (Le - Ev)/0.0591 \tag{21}$$

The fluoride electrode, which may be used in a temperature range of 0 to 80°C, detects fluoride ion concentrations of saturated to 10^{-6}M (0.02 ppm, parts per million). Limitation of detection is controlled by the solubility of the lanthanum fluoride. A lower solubility leads to a lower detection limit, other factors being constant. Except for hydroxyl ion, OH^-, this electrode is uniquely selective for fluoride ion; measurements at pH values above 8 (OH^- activity $>10^{-4}$) result in serious error. For 0.1 M OH^- and 0.001 M F^- there is, approximately, a 10% error. At pH values below 5 (H^+ activity $>10^{-5}$) hydrogen ions associate with fluoride ions to form the hydrogen fluoride molecule. The fluoride ion activity is, therefore, decreased. Since the fluoride electrode does not respond to the HF molecule, the electrode can only indicate the free fluoride ions in solution. The error is due to this chemical equilibrium pointing to the specificity of this electrode. The fluoride electrode most closely fulfills the conditions of an ideal, ion specific electrode.

Other solid-state electrodes of interest are Br^-, Cd^{2+}, Cl^-, Cu^{2+}, CN^-, F^-, I^-, Pb^{2+}, Ag^+/S^{2+}, and SCN^-. (See Table 7 for a list of commercially available solid-state electrodes.) These electrodes provide measurement of their respective ions, overall, in the concentration range of 1 to 10^{-8} M. Each electrode has its own unique interferences.[51] Some mixed, insoluble, silver salt electrodes have multiple detection capabilities. A membrane containing an equivalent molar content of silver sulfide and halide (Cl^-, Br^- or I^-) ion will detect silver ion (similar to a metal electrode of the first kind) and sulfide, S^{2-}, ion (Similar to an electrode of the second kind). An electrode, similar to one of the third kind, is a mixture of silver sulfide with cadmium, lead or copper(II) sulfide yielding an electrode sensitive to cadmium, lead or copper(II), respectively.

TABLE 6
Selective electrodes for water parameters

Metal ions		Nonmetal ions	
Analyte	Electrode	Analyte	Electrode
Cd^{2+}	Cd^{2+}	Cl^-	Ag/AgCl
Cu^{2+}	Cu^{2+}	F^-	F^-
Pb^{2+}	Pb^{2+}		
Ca^{2+}	Ca^{2+}	CN^-	CN^-
Hardness	Divalent ion	Br^-	Br^-
		I^-	I^-
Hg^{2-}, RHg^+	I^-	S^{2-}	S^{2-}
(R = allyl & aryl)		Residual Chlorine	I^-, indirect

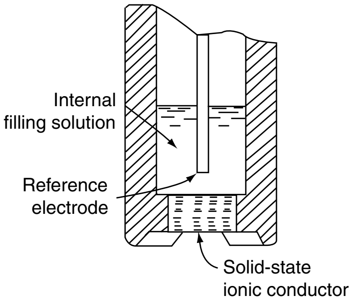


FIGURE 21 Cross-sectional view of a solid-state sensor. (Orion Research Inc.: Boston, MA with permission.)

(c) Liquid membrane electrodes[51]

Liquid membrane electrodes can detect a variety of mono and divalent anions and cations for example, BF_4^-, Ca^{2+}, $C10_4^-$, NO_3^-, K^+ and water hardness ($Ca^{2+} + Mg^{2+}$).[51] The membrane is a porous polymer disk which is filled with a water-immiscible, organic solvent containing a dissolved organic ion exchanger or chelating agent (neutral carrier). Figure 22

TABLE 7
Commercial solid-state electrodes[a]

Analyte Ion	Concentration Range, *M*	Interferences[a]
Br^-	10^0 to 5×10^{-6}	mr: 8×10^{-5} CN^-; 2×10^{-4} I^-; 2 NH_3; 400 Cl^-; 3×10^4 OH^-. mba: S^{2-}
Cd^{2+}	10^{-1} to 10^{-7}	$Fe^{2+} + Pb^{2+}$ may interfere.mba: Hg^{2+}, Ag^+, Cu^{2+}
Cl^-	10^0 to 5×10^{-5}	mr: 2×10^{-7} CN^-; 5×10^{-7} I^-; 3×10^{-3} Br^-; 10^{-2} $S_2O_3^{2-}$; 0.12 NH_3; 80 OH^-. mba: S^{2-}
Cu^{2+}	10^{-1} to 10^{-8}	high levels Fe^{2+}, Cd^{2+}, Br^-, Cl^-. mba: Hg^{2+}, Ag^+, Cu^+
CN^-	10^{-2} to 10^{-6}	mr: 10^{-1} I^-; 5×10^3 Br^-; 10^6 Cl^-, mba: S^{2-}
F^-	sat'd to 10^{-6}	0.1*M* OH^- gives <10% interferenc when $[F^-] = 10^{-3}$ *M*
I^-	10^0 to 5×10^{-8}	mr: 0.4 CN^-; 5×10^3 Br^-; 10^5 S_2O_3; 10^6 Cl^-
Pb^{2+}	10^{-1} to 10^{-6}	mba: Hg^{2+}, Ag^+, Cu^{2+}
Ag^+/S^{2-}	10^0 to 10^{-7} Ag^+ 10^0 to 10^{-7} S^{2-}	Hg^{2+} must be less than 10^{-7} *M*
SCN^-	10^0 to 5×10^{-6}	mr: 10^{-6} I^-; 3×10^{-3} Br^-; 7×10^{-3} CN^-; 0.13 $S_2O_3^{2-}$; 20 Cl^-; 100 OH^-. mba: S^{2-}

[a] From Ref. (51). (Orion Research Inc.: Boston, MA with permission.)

[b] mr: maximum ratio $\left(\frac{M\text{ interference}}{M\text{ analyte}}\right)$ for no interference. mba: must be absent.

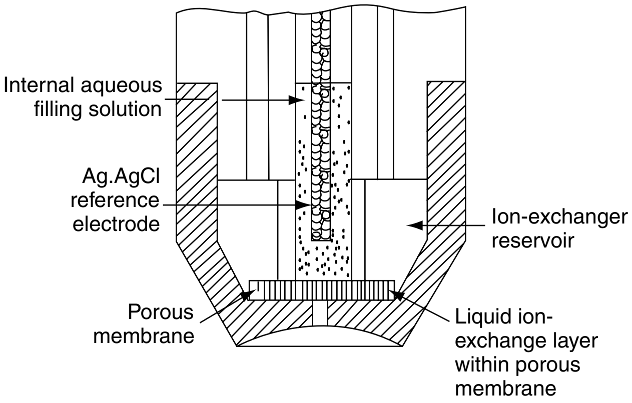


FIGURE 22 Construction of liquid ion-exchange electrode. (Orion Research Inc.: Boston, MA with permission.)

depicts the general structure of a liquid membrane electrode. Another rendition of the membrane is a hydrophobic polymer (e.g. polyvinyl chloride) film formed with the organic ion exchanger or chelating agent as an integral part of the membrane ensemble disk.

As seen in Figure 22, the membrane mediates between the standard analyte solution and the sample solution. The ion exchanger or chelating agent dissolved in the immiscible organic solvent layer (a straight chain alcohol) permits cations or anions to diffuse between the two aqueous phases, sample and standard. After some period of time, equilibrium will be established to provide a stable, membrane potential, meter reading. A calcium ion, Ca^+, electrode utilizes the ion exchanger, bis(2-ethylhexyl) phosphoric acid, $[(RO)_2POOH]$, where R = 2-ethylhexyl.

At each interface the following equilibrium is established.

$$[(RO)_2POO]_2Ca \rightleftharpoons 2(RO)_2POO^- + Ca^{2+} \qquad (22)$$

The pCa value is expressed in terms of the cell potential, E_v, and two constants, 0.0591 and Le (see equation 21) as follows:

$$pCa = 2(Ev - Le)/0.0591. \qquad (23)$$

The potassium electrode utilizes a chelating agent, valinomycin, an antibiotic and a microcyclic ether, that selectively complexes potassium ions. Thus the potassium electrode is composed of a 10^{-4} to 10^{-7} M solution of valinomycin dissolved in decane in the porous polymer disk (see Figure 22). The internal solution contains a known activity of potassium chloride; the chloride ion is for the silver/silver chloride reference electrode and the potassium ion, K^+, for the reference potassium activity. The electrode selectivity for K^+ over Na^+ is 3,800 times, an excellent specificity factor.

(d) Gas-sensing electrodes[12]

A gas-sensing electrode assembly featured in Figure 23 is effective for the analysis of gases. This electrode assembly is actually a cell with an indicating glass electrode, a silver-silver chloride reference electrode and a cell containing internal solution. The gas diffuses through the membrane into the thin layer of internal solution where a chemical reaction takes place generating some soluble reaction products. A specific product of this reaction that is proportional to the concentration of the analyte in the sample is detected by a selective electrode sensitive to the specific reaction product.

Two types of gas-permeable membranes are used in this assembly, homogeneous and microporous. The former is a thin (0.01 to 0.03 mm) silicon rubber membrane which allows the gas to "dissolve" in the membrane and leave by dissolving in the internal solution. A microporous membrane, about 0.1 mm thick, is prepared from hydrophobic polymers (polypropylene or polytetrafluoroethylene) and contains pores about 1 μm in diameter. Water, electrolytes and other polar substances are excluded from these membranes, but gases easily penetrate the membrane effusing through pores into the internal solution. In this manner the gaseous analyte in a solution or in a gas sample is separated from the sample matrix.

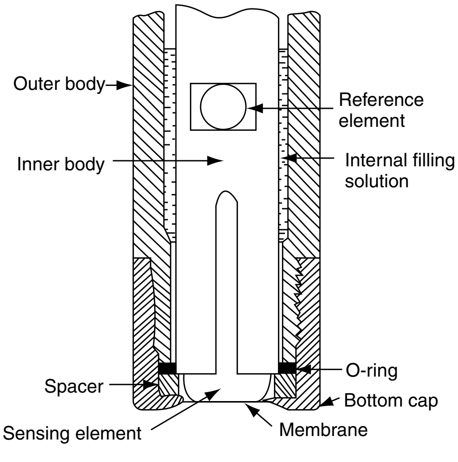


FIGURE 23 Construction of the gas-sensing electrode. For example, in the CO_2 gas-sensing electrode, the filling solution containing bicarbonate ion, HCO_3^-, forms a thin film between the sensing element, a glass electrode, and the membrane. A suitable reference electrode, e.g., SCE or Ag/AgCl can be the reference element. The CO_2 gas diffusing into the thin film reacts chemically with the HCO_3^- forming hydrogen ion, H^+, that is detected by the glass electrode. See the text and equations 24–26 for the chemical mechanism of the CO_2 gas-sensing electrode. (Orion Research Inc.: Boston, MA with permission.)

An example of a gas sensing probe is the ensemble illustrated in Figure 23 used in physiological investigations to measure carbon dioxide, CO_2, content of gas or water samples. The CO_2 in the sample travels through the membrane dissolving in the internal solution. The internal solution contains a relatively high concentration of bicarbonate ion HCO_3^-. The reaction which takes place between the dissolved CO_2 and water in the internal solution is

$$CO_2\,(aq) + H_2O \rightleftharpoons HCO_3^- + H^+. \tag{24}$$

The equilibrium (K and K′ are equilibrium constants) taking place in equation 24 is written

$$K = \left[HCO_3^-\right]\left[H^+\right]/\left[CO_2\right]. \tag{25}$$

Since the $[HCO_3^-]$ is relatively high in the internal solution, then the equation can be written

$$[H^+] = K'[CO_2]_{\text{external.}} \tag{26}$$

Measurement by the glass electrode for the hydrogen ion activity will yield the CO_2 analyte concentration. An important feature of this gas sensing assembly is that the analyte never comes in contact with the selective ion electrode.

Nitrogen dioxide, NO_2, can be detected in the presence of the gases, namely, CO_2, SO_2 and NH_3, using a nitrate electrode because of the following equilibrium

$$2NO_2\,(aq) + H_2O \rightleftharpoons NO_2^- + NO_3^- + 2H^+. \tag{27}$$

In Table 8 gas-sensing ensembles available commercially are listed.[12]

TABLE 8
Gas-sensing electrodes

Gas	Detected ion	Sensing Electrode[b]
CO_2	H^+	glass
NH_3	H^+	glass
Diethyl amine	H^+	glass
HCN	CN^-	silver (Ag_2S)
HF	F^-	fluoride (LaF_3)
H_2S	S_2^-	silver (Ag_2S)
NO_2	H^+ or NO_3^-	glass or nitrate
SO_2	H^+	glass

The substances in the parentheses indicate the composition of the solid-state electrode.

(e) Enzyme and microbial electrodes[45,54–56]

Enzyme and microbial electrodes can be an amperometric, membrane, or a conductometric electrode convered by a film or membrane containing the immobilized enzyme or microbial organisms. A permeable membrane intervenes between the enzyme or microbial film and the electrode (see Figure 23). In the course of measurement the analyte diffuses through the immobilized enzyme undergoing an enzymatic reaction. One of the participants of the reaction, product or reactant, is sensed by the electrode. These devices provide a unique combination of two selection processes—specificity of enzyme or microbial reactions and of the selective electrode.

Measurements for two electrochemical modes, amperometric and potentiometric, are made as follows: In the potentiometric mode the log of the concentration of the detected species is a linear function of the measured electrode potential (Nernst equation). In the amperometric mode the concentration of the detected species is proportional to the electrode current. The potential of the amperometric electrode chosen corresponds to the redox potential of the detected species. Through the stoichiometry of the enzyme or bacterial reaction, the concentration of the analyte is easily calculated from the concentration of the detected species.

However, the compatibility between conditions for enzymatic and microbial action and the optimal electrode sensing environmenta in the form of pH, solvent, ionic strength, temperature, etc. must be taken into account. At times the compromises made in the conditions do not serve well the measuring objective of the enzyme or microbial electrode ensemble. This point will be illustrated.

(i) Enzyme electrodes[45,54,56]

Enzymes are immobilized in several ways, e.g., absorption on a porous inorganic structure or polymer gel membrane or copolymerization with a compatible monomer to form a thin membrane. The immobilized enzyme film is viable for from about two to four weeks. The modular construction of the electrode permits easy replacement of the immobilized enzyme membrane.

(x) Potentiometric type

A potentiometric enzyme electrode may be formed from a gas-sensing electrode assembly containing an ion- or molecular-sensitive electrode. The immobilized enzyme membrane is inserted in the space allotted for the thin film of internal solution (see Figure 23). An example of this electrode is the measurement of urea where reaction with the enzyme urease is illustrated as follows:

$$\text{urea} + H_2O + \text{urease} \rightarrow CO_2 + 2NH_4^+. \qquad (28)$$

At an alkaline pH the forms NH_3 which is detected by an ammonia, gas-sensing electrode. Alternatively, a pH/glass electrode can be used to determine the change in pH due to the hydrolysis of the NH^4_+ ion. Potentiometric enzyme electrodes are also available for the analysis of L-lysine and L-tryptophane.

Incompatibility of enzyme or bacteria with the selective electrode can occur because of the differences in pH for optimal enzyme, microbial, and selective electrode activity. For example in the determination of urea, the enzyme, urease, converts urea to ammonium ion, NH^4_+, and bicarbonate ion, HCO_3^-. This reaction optimally occurs at a pH of about 7. An ammonia gas-sensing electrode is the electrode used and the NH^4_+ ion, enzymatically formed, requires a pH of 8 or more to change NH^4_+ to ammonia, NH_3, which is sensed by the electrode. Thus none of the components are operable, optimally, at each other's best pH. An alternative selective electrode in this enzyme electrode is a glass electrode which detects NH^4_+. However, Na^+ and K^+, are also sensed thereby obviating some of its selectivity.

An automatic analyzer is available for the analysis of urea that addresses the problem of optimal pH conditions (see Part Two Section 5,*b,(3),*(*c*),(*iii*)).

(y) Amperometric type

The glucose electrode consists of an amperometric probe covered with immobilized glucose oxidase. The enzymatic reaction is as follows:

$$\beta\text{-D-glucose} + O_2 + H_2O + \text{glucose oxidase} \rightarrow H_2O_2 + \text{D-glucono-}\Delta\text{-lactone}. \qquad (29)$$

Either oxygen or peroxide may be detected as a measure of the glucose content. The decrease in oxygen or the increase in peroxide is proportional to the glucose concentration. Therefore, an oxygen or peroxide amperometric electrode can be used. Each electrode has a specific advantage; the oxygen-based electrode doesn't respond to other reaction substituents but the peroxide does respond to other oxidizable components of low molecular weight. However, the oxygen electrode has a lower sensitivity than the peroxide probe, since it has a high background signal.

The oxygen electrode (see Section III,B,*2,*b,*(2),*(a)) is used in a number of enzyme electrodes because of its selectivity.[56] Amperometric enzyme electrodes are also available for the analysis of alcohols, L-amino acids, ascorbate, L-glutamate, lactate, lactose/galactose, oxalate, salicylate, sucrose, and uric acid.[56]

Enzyme electrodes have been incorporated into an automatic or semi-automatic analyzer (see Part Two Section 5,*b,(3),*(*c*),(*iii*)).

(ii) Microbial electrodes[52,53]

Microbial electrodes are similar in construction and operation to enzyme electrodes. In the microbial electrodes a polymeric membrane containing immobilized organisms is placed between the electrode and the analyte solution. The nature of the analyte dictates the species of organism. Several examples of analytes, immobilized organisms and electrodes that comprise a number of microbial electrodes are listed in Table 9.

(x) Potentiometric type

Glutamic acid and glutamate can be detected by a microbial electrode utilizing a CO_2, gas-sensing electrode.[57,58] The

TABLE 9

Microbial electrodes[a]

Analyte Ref.	Organism	Electrode[b]	
Glucose	Pseudomonas fluorescens	oxygen, Amp.	(59)
Alcohol	Trichosporon brassicae	oxygen. Amp.	(60)
Acetic acid	Trichosporon brassicae	oxygen, Amp.	(61)
Formic acid	Clostridium butyricum	hydrogen, FC	(62)
Glutamic acid	Escherichia coli	CO_2, GS	(57) (58)
Ammonia	Nitrosomonas sp. and Nitrobacter sp.	oxygen, Amp.	(63) (64, 65)
Methane	Methylomonas flagellata	oxygen, Amp.	(66) (67)
Nitrite	Nitrobacter sp.	oxygen, Amp.	(68) (69)
BOD	Clostridium butyricum or Trichosporon cutaneum	bio, FC	(70) (71)
Mutagen	B. subtilis, recombination deficient	oxygen, Amp.	(72) (73) (74)

[a] From Ref. (52).

[b] Amp.—amperometric; GS—gas sensing; FC—fuel cell.

immobilized Escherichia coli, having glutamate decarboxylase enzymic activity, participate in the following reaction:

$$\text{Glutamic acid}+\text{enzyme}\rightleftharpoons\text{CO2}+\text{g-aminobutyric acid.} \quad (30)$$

The sample solution is at pH 4.4, low enough to assure the presence of CO_2 (pK = 6.34) instead of HCO_3^- (see equation 24). This electrode responded very slightly to some other amino acids.

A linear dynamic range is obtained in the following terms: a linear relationship exits between the log of the concentration range, 100–800 mg/L, vs. the potential of the CO_2 electrode. The standard deviation, sd, is 1.2 mg/L. Measurement time is 5 minutes. The sensor functioned consistently for more than three weeks and 1,500 determinations.

(y) Amperometric type

An example of a microbial electrode of use in wastewater analysis is the ammonia electrode illustrated in Figure 24. The nitrifying reactions are as follows:

$$NH_3+1.5O_2+\text{Nitrosomonas sp}\rightarrow NO_2^-+H_2O+H^+ \quad (31)$$

$$NO_2^- + 0.5O_2 + \text{Nitrobacter sp} \rightarrow NO_3^-. \quad (32)$$

Oxygen concentration is detected, amperometrically, by the Pt electrode. The current decrease is linearly related to the NH_3 concentration when the NH_3 concentration is below 42 mg/L; the minimum detectable NH_3 concentration is 0.1 mg/L. This sensor has about a 4 min. response time.

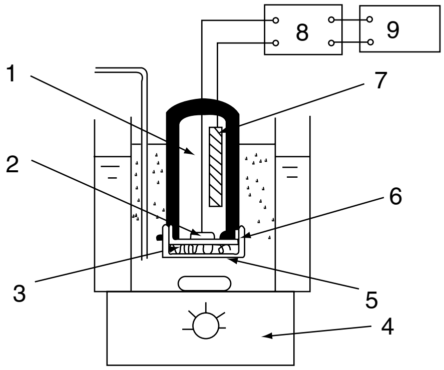


FIGURE 24 The microbial sensor system for ammonia. 1. Electrolyte (NaOH); 2. Cathode (Pt), 3. Immobilized cells of nitrifying bacteria; 4. Magnetic stirrer; 5. Gas permeable Teflon membrane; 6. Teflon membrane; 7. Anode (Pb); 8. amplifier; 9. recorder. (Reprinted with permission from reference 52. Copyright 1986 American Chemical Society.)

Karube uses a bacterial electrode employing Trichosporon cutaneum to measure biochemical oxygen demand in wastewater treatement. The difference between the steady state and initial electrode current was correlated to 5-day BOD, BOD_5.[71] A BOD_5 measuring system[52] utilizing a different microbial electrode is described in Section 5,*b,(3)*,(*c*),(*ii*).

b. Voltammetric and amperometric procedures

Voltammetry and amperometry are used to determine the concentration of an electroactive analyte through current measurement under various potential regimes. In voltammetric procedures the current is measured as a function of the potential scanned between two potential values. Amperometric methods measure current at a fixed potential.

Amperometric methods apply to electrochemical sensors which are used directly or in a variety of titrimetric methods. Voltammetry can be divided into two categories:

In polarography the dropping mercury electrode, DME, is utilized and other voltammetric methods employ a variety of solid and paste electrodes. Polarography includes linear-scan (classic), current-sampled, and normal- and differential-pulse, and alternating-current techniques. Other voltammetric methods comprise fast linear-sweep, cyclic, and stripping voltammetry.

As described in Section III,B,*2*,a, the working electrode is the sensor in these electrochemical methods. The cell may consist of two or three electrodes. There are two basic types of cells, electrolytic and galvanic. In an electrolytic cell the redox reaction occurs due to the imposition of a potential from an outside source. The galvanic cell is based on a spontaneous reaction occurring through the proper choice of anode and cathode materials and the electroactive analyte. Electrolytic and galvanic cells are used in amperometry, while electrolytic cells are employed in voltammetry.

Voltammetric and amperometric currents are measured using working microelectrodes (surface areas of a few square millimeters) under conditions leading to polarization of the working electrode. The condition of polarization of an electrode exists when a change of potential of the electrode will not result in a change in the current generated.

Polarization occurs when the analyte is impeded in its electrochemical reaction with the electrode surface. The analyte–electrode reaction rate is dependent on the rates of a number of rate processes giving rise to several types of polarization. These are as follows: Concentration polarization is the limited access of the analyte to the electrode surface due to the rate of mass transfer of analyte by diffusion. Charge or electron transfer polarization arises through the slow, heterogeneous transfer of electrons between the analyte and the electrode surface. Physical or surface polarization occurs when adsorption, desorption, or crystallization (electrodeposition) processes occur. Reaction polarization takes place when the analyte or intermediates undergo chemical reactions before or after electron transfer.

(1) Voltammetry[50,75]

(a) Polarography[76,77,78]

Polarography is a voltammetric method where the dropping mercury electrode, DME, is the indicating electrode. The DME is a length of capillary tube connected by tubing to a mercury reservoir situated above the capillary. The polarographic cell contains the DME and a reference electrode for a two electrode system and a third electrode, usually a platinum wire, in a three electrode system (see Figure 25). Drops

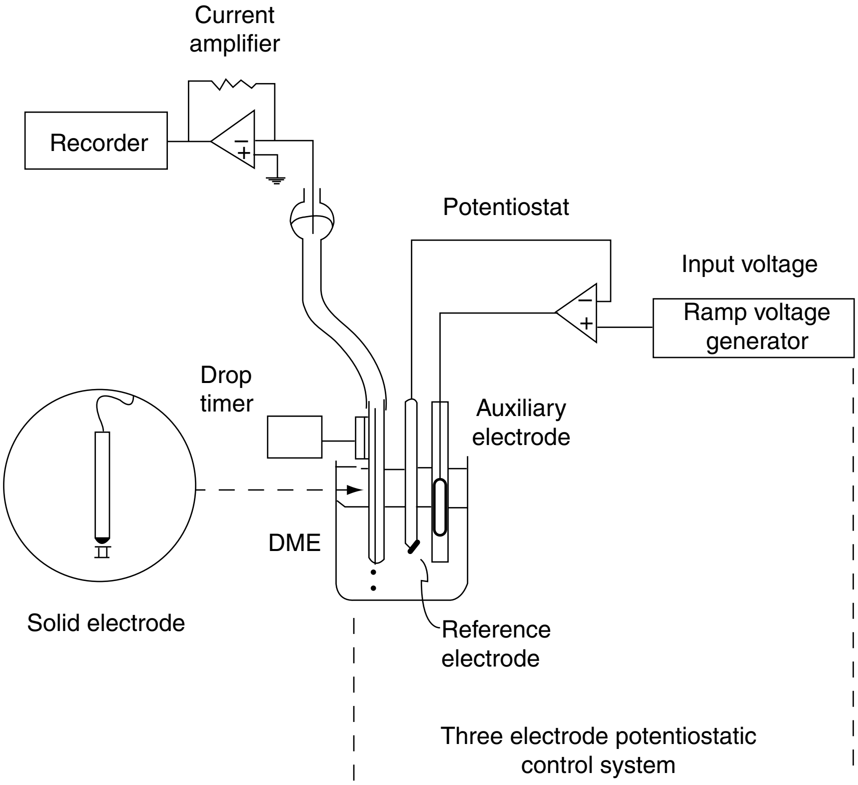


FIGURE 25 Three-electrode polarographic and voltammetric analyzer. A DME equipped with a drop timer or solid electrodes (gold, platinum, glassy carbon, wax impregnated carbon, etc.) are employed for polarographic or voltammetric analysis, respectively.

issue constantly from the DME. The rate of mercury flow, *m,* is 1 to 3 mg/second and is dependent on the lumen and length of the capillary and the height of the mercury column (the applied pressure) in classical (linear scan) polarography. Drop diameters are between 0.1 and 1 mm. The large excursions seen in Figure 26A on the plateau of the curve are due to the growth of the mercury drop. In current-sampled, and normal- and differential-pulse polarography drop time, t, is controlled by a mechanical or electrical device causing dislodgement of the drop at chosen drop times of 1 to 5 sec. The excursions on the curve are diminished because drops times are decreased as compared to classical polarography (see Figure 26A). The DME provides a renewable surface electrode that prevents the anomalous surface effects experienced, at times, with solid electrodes.

The polarographic cell solution containing the analyte and an indifferent electrolyte are purged with pure nitrogen to remove all dissolved oxygen. (A reduction wave for oxygen occurs and interferes with the analyte wave.) The indifferent electrolyte, e.g., potassium chloride, removes the possible condition of electrostatic attraction or repulsion of a charged analyte resulting in a migration current, i_m. This then allows the analyte to approach the electrode by the process of diffusion leading to a diffusion current, i_d, solely. The voltage is scanned with a linearly increasing d.c. ramp of 1 to 5 mv/sec (mv, millivolts). In the voltage scanning process, before reduction or oxidation of the analyte takes place, impurities in the solution contribute to a small current known as the residual current, i_r, resulting in a baseline of slowly increasing current as a function of potential (see Figure 27). The electrochemical redox reaction of the analyte begins at the decomposition potential in a heterogeneous transfer of electrons between the analyte in solution and the electrode surface. The current increases, sharply, until the limiting current, i_L, is reached where it remains constant due to concentration polarization. Concentration polarization is a necessary condition in some types of voltammetry. Figure 27 shows a polarogram and the measurement of the diffusion current of a polarographic wave for a substance i_d.

A number of scanning and measuring modes, i.e., linear scan (classic), current-sampled (tast), normal and differential pulse, and fast-linear sweep, are available. Figure 26A and B illustrates the various polarographic curves for each of these scanning modes. The classical "S" shaped curve is obtained with linear (classic), current-sampled (tast), normal pulse techniques. The fast linear scan gives a distorted "S," while the differential pulse produces a differential curve. The height of the wave, the diffusion current,

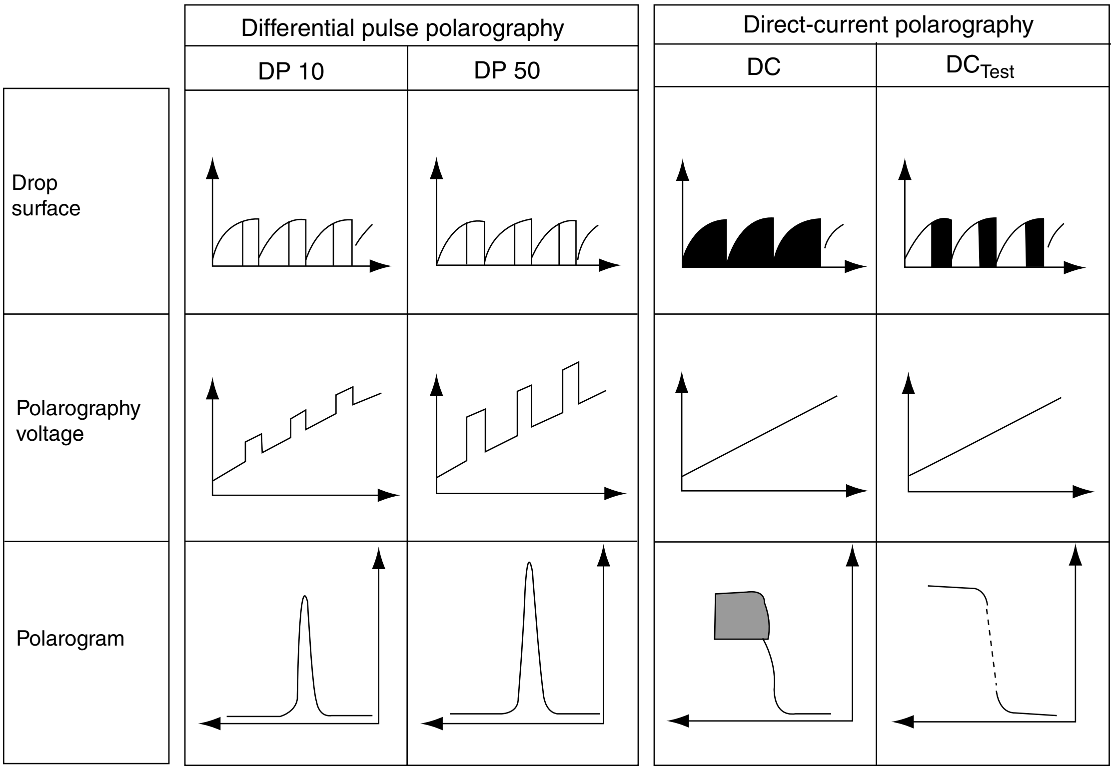


FIGURE 26A Direct-current, DC, and differential pulse, DP, polarographic methods. (Last and current-sampled polarography are synonymous.) The following are detection limits and $E_{1/2}$ resolution values i.e., 2×10^{-6} M and 0.2 V, 1×10^{-6} M and 0.2 V, and 1×10^{-7} M and 0.05 V for DC, DC tast, and DP polarography, respectively. (Courtesy of Brinkmann Instruments, Inc.)

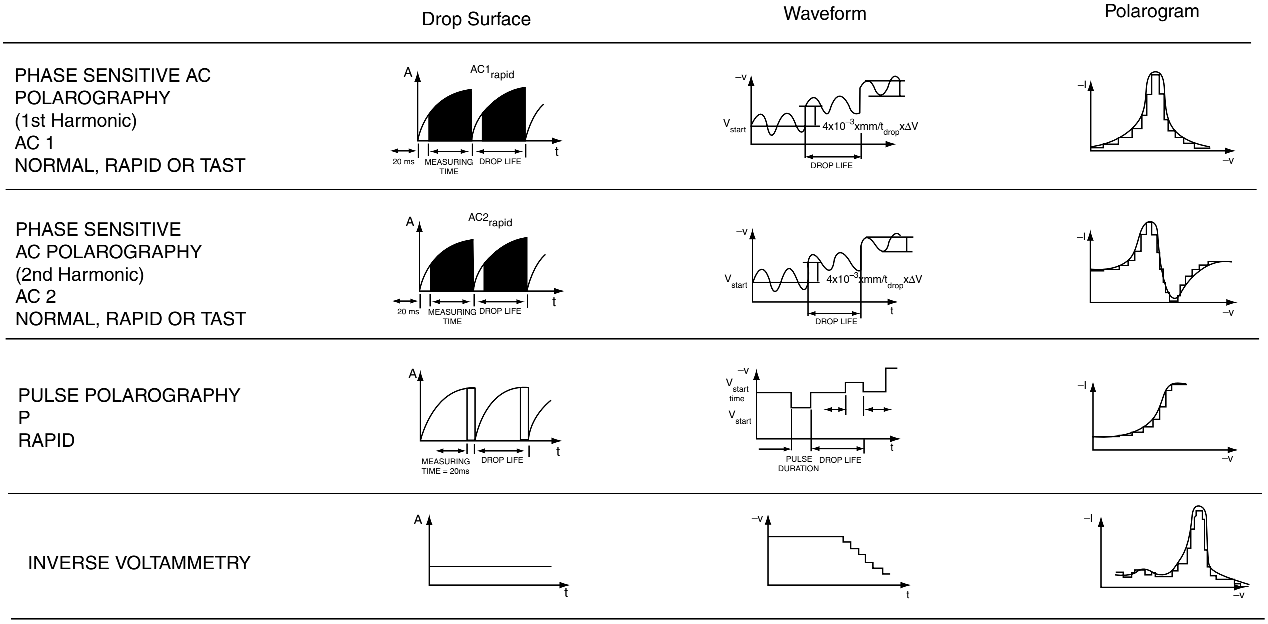


FIGURE 26B Other polarographic and voltammetric methods. (Pulse polarography refers to normal pulse as differentiated from differential pulse polarography. Inverse and stripping voltammetry are synonymous.) The following are detection limits and $E_{1/2}$ resolution values i.e., 5×10^{-7} M and 0.05 V, 5×10^{-7} M and 0.2 V, and 10^{-9} M and 0.1 V for AC and normal pulse polarography, and stripping voltammetry, respectively. (Courtesy of Brinkmann Instruments, Inc.)

i_d, is proportional to the concentration of the electroactive analyte, while the half wave potential, $E_{1/2}$ in volts, is the qualitative factor identifying the analyte (see Figure 27). For the differential pulse curve, the height of the peak, i_p, is proportional to the concentration of the analyte and the peak potential, E_p, is almost equal to the $E_{1/2}$. The standard reduction potential, E^0, and the $E_{1/2}$ value are approximately equal (see equation 11). The Ilkovic equation relates the i_d, in amperes, with the diffusion coefficient of species S, D_s in cm^2/seconds; the concentration of the analyte, C in moles/L; the number of electrons transferred in the redox reaction, n; the flow rate of mercury, m in mg/seconds; and the drop time, t in seconds as follows:

$$i_d = 708nD_s^{1/2}\,Cm^{2/3}\,t^{1/6}. \qquad (33)$$

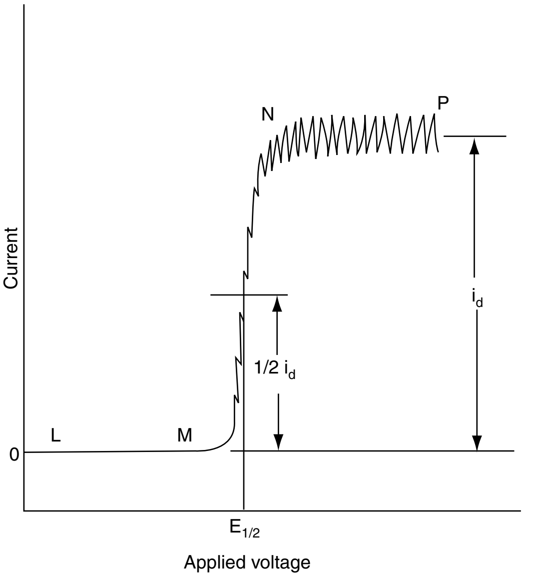


FIGURE 27 Typical polarogram. The line LM represents the residual current, i_r and i_d the diffusion current. (Reprinted from Ref. (176), p. 394 by permission of Prentice-Hall, Inc., Englewood Cliffs, NJ.)

In linear scan or classic polarography an impediment to low level detection is the changing non-faradaic charging current that electrostatically charges the mercury drop. (Consider the drop as a condenser.) As the area of the mercury drop increases with time, the rate of area increase decreases with time. Thus, the non-faradaic charging current decreases with time and is quite low at the end of the drop time. The total polarographic current is a sum of the diffusion current, i_d, the charging current and the residual current, i_r. (The i_r, due to electroactive contaminants, is rather small and invariant with time.) Since the charging current is an unknown quantity the signal can not be amplified to increase the sensitivity of the method without introducing a considerable but variable error. To obviate the charging current problem, normal-and differential-pulse[79] and current-sampled modes were developed (see Figure 26). In each of these modes the current is measured during the last 5 to 20 msec of each drop before the drop is dislodged. At that point the charging current is at a very low level and doesn't cause much error on amplification to increase sensitivity.

In the normal pulse technique, an increasing d.c. pulse is imposed on the linearly increasing d.c. ramp on each successive drop. A 20-to-100 mv pulse is imposed on each mercury drop in the differential pulse mode. The voltage pulse leads to an increased surge in current by the electroacive species undergoing reaction at the electrode surface due to the change in voltage. This enhanced current occurs in both normal- and differential-pulse. There is also a decrease in non-faradaic charging current because the measurement is made at the end of the drop life-time about 40 msec after the pulse is applied to allow for the exponential decay of the capacitive current. In the normal pulse technique the current is measured as in the current-sampled method, in the last 5–20 msec; whereas in the differential pulse mode the current is the difference between the diffusion current before the pulse and the current at the end of the pulse on each drop. The enhancement of the polarographic current in normal and differential pulse techniques is due to two effects, enhancement of the faradaic current and a decrease in the nonfaradaic charging current. The differential method is less sensitive than the normal pulse method, however its resolution is higher. The pulse techniques are more sensitive than classical polarography (see Figures 26A and B).

Fast linear sweep, LS, polarography carries out a fast sweep of the voltage range on one drop of mercury.[12] A differential curve is obtained whose summit or peak current is composed of the surge in current previously mentioned in pulse techniques and the i_d. The peak current, i_p, is given by the equation

$$i_p = (2.69)\left(105\mathrm{n}^{3/2}\ \mathrm{ArD}_s^{1/2}\ \mathrm{Cv}^{1/2}\right). \tag{34}$$

Here Ar is the area of the electrode, v is the voltage scan rate in volts/sec, and the remaining terms are those given in equation 33. E_p for a reversible reaction is written

$$E_p = E_{1/2} - 0.065/\mathrm{n}. \tag{35}$$

This procedure has the advantage of good resolution as in the pulse methods, speed, and increased sensitivity (see the discussion in the following paragraph on LSV and Figures 26A and B).

Alternating current polarography employs 1 to 35 millivolts of a.c. of 10 to 60 Hz superimposed on a d.c. ramp of 5 mv/sec used in the previous methods. The a.c. faradaic current responsible for the redox reactions of the analyte is recorded as a function of the d.c. ramp voltage and provides a maximum peak voltage, E_p, similar to the $E_{1/2}$ of the analyte. The concentration of the analyte is proportional to the a.c. faradaic current. Species undergoing reversible redox reactions are sensitive to the a.c. technique. The detection limit is quite low and the a.c. peak accords high resolution when several analytes are present. Therefore, a low concentration of a reducible species can be detected in the presence of a high concentration of a less easily reducible species. This method responds to a kinetic factor. Very slow reactions do not give peaks but intermediate ones give a.c. currents limited by the reaction rate. A.c. polarography finds advantageous use in analysis and the study of electrode kinetics.

(b) Other votammetric methods[50,75,80,81]

A number of procedures use electrodes other than the DME employing stripping, and linear sweep, LSV, and cyclic, CV, voltammetry. These electrodes include stationary hanging mercury drop, HMDE, and mercury film electrodes, solid electrodes of noble metals such as platinum, palladium, silver, and gold in the form of cylindrical (wires) or flat surfaces, and various types of carbon electrodes, such as treated spectroscopic amorphous carbon, glassy carbon, and carbon paste.[81] These electrodes can be used in the quiescent solution or as a rapidly rotating wire or disk electrode. The wide variety of electrodes offer unique advantages for dealing with different analytes, concentration ranges, and interferences present in sample solutions.

Linear sweep and cyclic voltammetry are related. For LSV a rapid scan at rates of several mv/sec to several hundred mv/sec is used. However, in cyclic voltammetry a triangular (cyclic) voltage scan is impressed on the electrode amounting to two linear sweeps, forward and reverse. The process allows the selective reduction of the analyte followed by the oxidation of the reduced analyte, if the redox process is reversible. However, the order of the scan can be reversed, oxidation followed by reduction. The cyclic voltammagram permits one to interpret redox reaction mechanisms for complicated electrode reactions. The HMDE and solid electrodes, referred to previously, are used in quiescent solutions in this techniques.

Stripping analysis[82] is a method that concentrates the analyte on or near the electrode (HMDE or solid) surface in a deposition (redox electrolytic) step. The voltammagram is obtained by the oxidation or reduction of the electrolytic product by a fast linear sweep scan in the stripping step. Stripping analysis is of two kinds: Anodic (ASV) and cathodic (CSV) stripping voltammetry where cation and anion analysis occurs, respectively.

For ASV in the deposition step a stirred dilute cupric ion solution ($>10^{-9}$ M) is subjected to electrolysis at an electrode for a predetermined time while stirring the solution. (The lower the concentration of the analyte, the longer the deposition time.) The fixed plating voltage is more positive than the reduction potential of the ion. For example, the cupric ion is reduced to copper metal and the metal forms an amalgam on the surface of the HMDE or a film on a solid electrode. At the end of the electrolysis step the stirring is ended. In the stripping step a fast linear voltage sweep in the negative direction causes the copper to be oxidized. The recorded voltammagram of this oxidation yields a peak whose height is proportional to the concentration of analyte cation (see Figure 26B). In CSV halides are deposited on the mercury anode as the mercury(I) halides. The remainder of the procedure is similar to ASV.

There is an on-line instrument for laboratory and field use to carry out ASV.[83] In the cell a glassy carbon electrode is automatically coated with a thin film of mercury. (When the replating of the electrode is necessary it is carried out automatically.) A pump and a set of five solenoid stream-switching valves direct sample and reagents to the cell. By means of a microprocessor and keyboard all steps and analytical parameters can be facilitated. The detection limits are 1 ppb for bismuth, cadmium, copper, lead, thallium and zinc that plate on mercury and 50 ppb for gold and silver that do not plate on mercury and mercury itself.

A.c. voltammetry employs a variety of electrodes with similar operations and response described in the polarography section above.

Figures 26A and B give information relating to the detection limits in terms of the molarity of the analyte and the resolution between peaks or waves expressed as volts. Therefore, comparisons can be made among the techniques illustrated.

(2) Amperometry

(a) Dissolved oxygen electrode

Dissolved oxygen, DO, in water is determined by DO probes using galvanic or electrolytic type cells. These DO probes have their electrodes protected by an oxygen permeable membrane and are referred to as membrane electrode systems. These systems contain a cathode working electrode and a Ag/AgCl reference electrode and an electrolyte of 1M KCl.

In an electrolytic probe the working electrode may be a platinum (Clark electrode[84]), gold, carbon, or silver microelectrode (see Figure 28). Gold seems to give the best service. A membrane electrode system using a gold electrode, an Ag/AgCl reference electrode with a KCl gel electrolyte, and a polarizing voltage of −0.80 is commercially available.[85] The cell reactions are as follows:

Gold cathode

$$O_2 + H_2O + 4e \rightarrow 4OH^- \qquad (36)$$

silver anode (Cl^-,a = 1)

$$4Ag + 4Cl^- \rightarrow 4AgCl + 4e. \qquad (37)$$

The design, based on the Clark electrode, is commonly used by a number of manufacturers.[85,86,87]

A DO probe operating on the galvanic principle is one developed by Mancy *et al.*[88] In this cell a spontaneous electrochemical reactions occurs as follows:

Pb anode

$$2Pb + 60H^- \rightarrow 2PbOOH^- + 2H_2O + 4e \qquad (38)$$

Ag cathode

$$O_2 + 2H_2O + 4e \rightarrow 4OH^-. \qquad (39)$$

The electrolyte, 1 M KOH, is used since hydroxide, OH^-, ion is formed during the reduction of the oxygen. For higher sensitivities a saturated potassium bicarbonate electrolyte solution is employed because of the insolubility of the lead salt formed. A 1 mil, polyethylene membrane separates the cell from the analyte sample. Several galvanic electrode systems are available.[89,90,91]

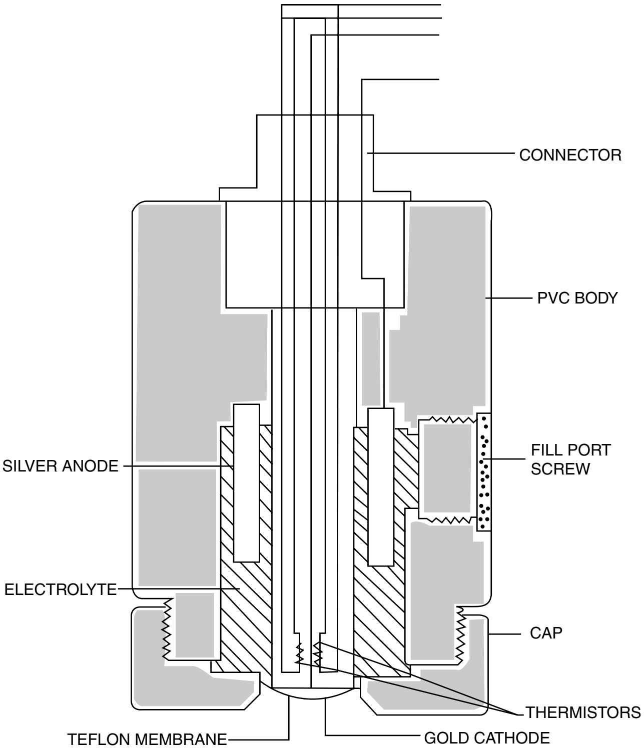


FIGURE 28 Clark DO membrane probe. (Courtesy of Rosemount Analytical, Inc., La Habra, CA.)

The DO probe output current, from both types of cells, is sensitive to temperature changes. To compensate for this effect, thermisters are placed in series with the load circuit[92] or in a bridge circuit.[93]

(c) Potentiometric, amperometric and conductometric titrations

Electrochemical means are used in potentiometric, amperometric and conductometric titrations to determine the endpoint of a titration. These titrations are so named to indicate the mechanism of endpoint determination. The electrodes used in the potentiometric and amperometric methods are sensitive to analyte and/or titrant redox species; whereas in conductometric titrations total ion content is detected. In amperometric titrations the magnitude of current flow is determined as a function of the volume of titrant. The electrode potentials are measured as a function of the titrant volume in potentiometric titrations. Conductance measurements are used in conjunction with titrant volumes in this third method.

(1) Potentiometric titrations[94]

The electrodes discussed in the previous section on potentiometric instruments may be used in potentiomeric titrations. A reference and the appropriate indicating electrode allows the measurement of the potential for each volume of titrant added. The potential changes in response to the changes in the ratio of concentrations of oxidized to reduced species. The Nernst equation can be used to calculate the magnitude of these changes. A common classical example of a potentiometric titration is an acid-base titration using pH/reference electrode pair. The pH electrode indicates the hydrogen ion concentration in terms of pH as a function of volume. The titration apparatus and the resulting "S" shaped curve is shown in Figure 29. Another example concerns the complexometric titration of calcium ion with the titrant, EDTA (the sodium salt of ethylene diamine tetraacetic acid). A calcium ion selective and reference electrode are employed. The titration curve, is similar in shape to the curve in Figure 29. The pCa value is measured by the calcium electrode as a function of volume of EDTA titrant.

Automatic titrators allow the analyst to carry out a variety of potentiometric titrations without the consumption of a great deal of time and effort. However the reaction must not be too slow and the concentrations of the analyte between 0.1 and 0.001N. Two types of automatic titrators are available—recording and non-recording. Many commercial titrators have the following features: A delivery system consisting of a buret with a solenoid operated delivery valve or a calibrated syringe whose plunger is motor-driven by a micrometer screw. The potential from the indicator-reference electrode pair is compared to a set end point potential by an anticipator, null-sensing amplifier circuit yielding an amplified error signal. The error signal controls the buret delivery. Anticipation of the endpoint prevents over titrating the endpoint (see Figure 29).

Recording titrators plot the titration curve and are completely automatic accepting a number of samples for titration in a serial manner. Some titrators contain microprocessors providing computational ability and processing of a number of samples. The error signal is used to drive the recorder pen and control buret delivery. Advantageously, unknown systems may be titrated yielding a curve that can be interpreted by the analyst.

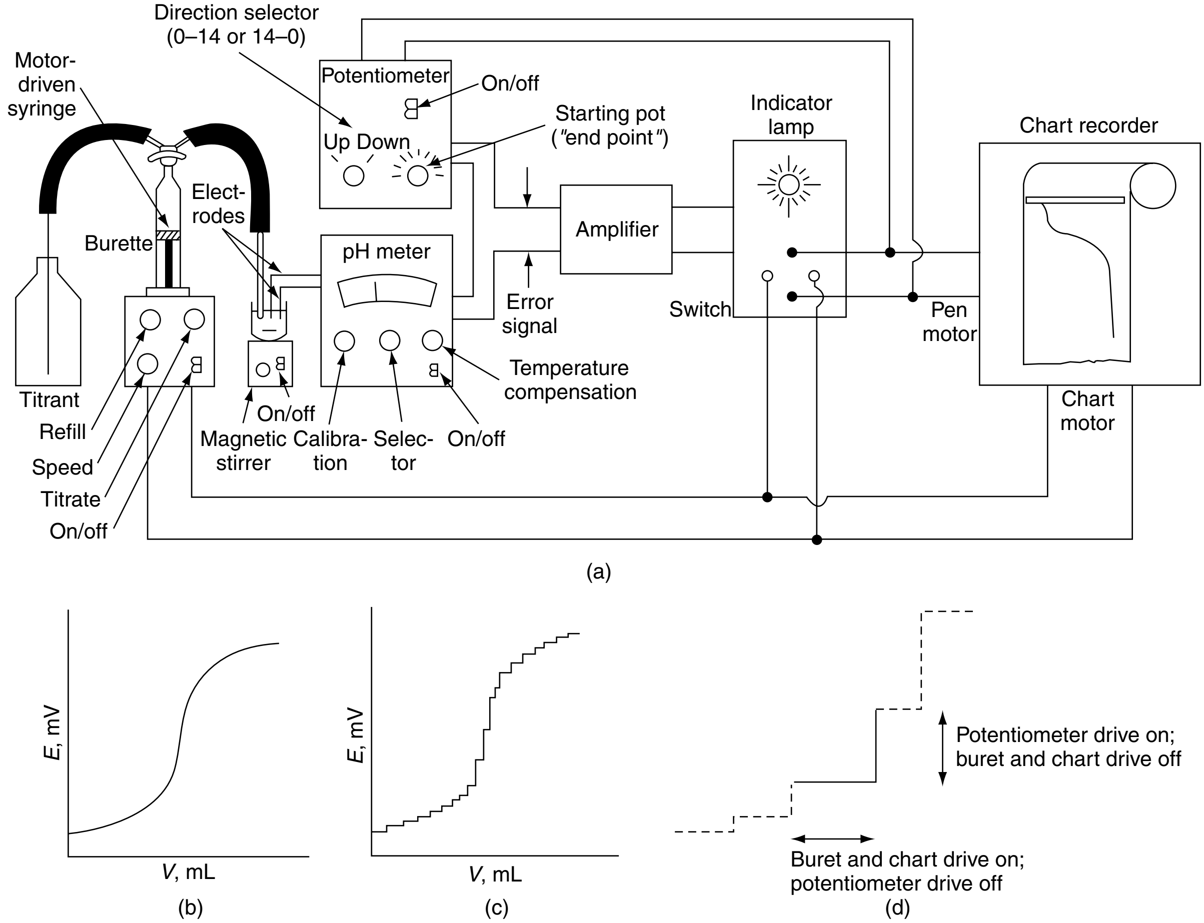


FIGURE 29 (a) An automated, curve-recording titrator. (b) Theoretical titration curve. (c) Recorded curve with endpoint anticipation. (d) Enlarge portion of (c) with explanation. (From G. Svehla, Automatic Potentiomeric Titrations, p. 176, Pergamon Press. Copyright © 1978 by Prof. G. Svehla. Reprinted with permission.)

The accuracy of automatic titrators depends on the concentration of the analyte and the sensitivity of their electrode sensing systems. Accuracies are 0.1% for a 10^{-2} N solution and 1% for the limiting concentration of 10^{-3} N.

Titrimetric methods are given in Standard Methods[2] for the following parameters—carbon dioxide, cyanide, COD, sulfite, and ammonia.

(2) Amperometric titrations[50]

Amperometric titrations may utilize one or two polarizable micro electrodes configured differently, electrically. These several electrode systems will lead to differently shaped titration curves. Titration curves show the current flow as a function of the volume of titrant.

A single polarized electrode may be a DME, a solid-state electrode of carbon, platinum or other noble metal or a rotating noble metal electrode. The potential imposed on the indicating electrode is such that the limiting current, i_L, is obtained. As in voltammetry, the current obtained is proportional to the concentration of the electroactive species. Therefore as the titration proceeds, the diffusion current, i_d, changes. The shape of the titration curve depends on the electroactivity of the titrant and the analyte as seen in Figure 30. In general two straight lines are obtained and their interpolated intersection indicates the endpoint.

Two polarized microelectrodes result in titration curves whose shapes depend on the reversibility of the electrode reactions of the titrant and the analyte. When there is a reversible reaction for the analyte, (I_3^-/I^-), and an irreversible one for the titrant, ($S_2O_0^-/S_4O_6^{—}$), the result is a "deadstop" endpoint where the current ceases to flow. When analyte and titrant show reversible electrode behavior, e.g., Fe^{2+}/Fe^{3+} and Ce^{3+}/Ce^{4+}, respectively, the titration curve shows no current flow at the endpoint. There is current flow on either side of the endpoint.

(3) Conductometric titrations[95]

The instrumental apparatus and conductance cells for conductometric titrations are given in Figures 31 and 32. In Figure 33 a precipitation and several acid–base titration curves are illustrated. Since ions are detected by the electrodes, any change in ionic concentration in the analyte solution during the titration is the basis for a conductometric titration. In a neutralization reaction two ions, hydronium, H_3O^+ and hydroxide, OH^-, are removed to form unionized water, H_2O. Two ions, silver, Ag^+ and chloride, Cl^-, form a precipitate of silver chloride, thereby decreasing the conductivity of the solution in the course of the titration. Analogously, the formation of an unionized complex can be the basis for a conductometric titration.

(d) Conductometric methods[96]

Conductometric methods detect ionic species in solution. The conductivitivity of a solution indicates the presence of ionic species and can, under certain conditions, be used to estimate the concentration of the dissolved electrolyte. Conductance is also used in conductometric titrations (see the previous section III,B,*2*,c,*(3)*). Solutions containing electrolytes conduct electricity and obey Ohm's law. Electrical resistance or conductance is measured by placing the solution between two electrodes and using a Wheatstone bridge to carry out the measurement.

Conductivity, C_n, is reciprocally related to resistance, Rs, so that $C_n = 1/Rs$. Units for Rs and C_n are ohms and mhos (ohms^{-1}), respectively. (The IS unit for C_n is siemans, S.) Resistance and conductivity of a solution are sensitive to the dimensions of the volume of solution included between the plate electrodes (see Figure 32). Conductivity is proportional to the area of the electrodes and reciprocally related to the distance between them.

By normalizing conductivity to a given dimension, a cube that is one centimeter on its edge, and designating a new parameter, specific conductance, C_{sp} in mhos/cm, the following calculation can be made,

$$C_{sp} = (1/R_m)(d/ar) = (1/R_m)(k_c) \text{ mhos/cm} \qquad (40)$$

where R_m is the measured solution resistance in the conductivity cell, d and ar are the distance between and the area of the

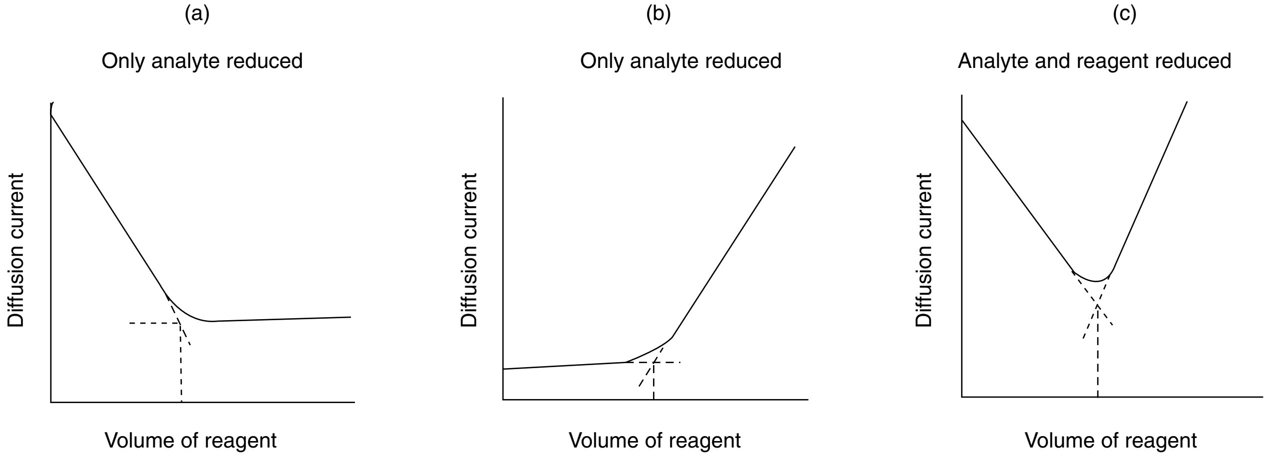


FIGURE 30 Amperometric titration curve.

electrodes, respectively. The cell constant is $k_c = d/ar\ cm^{-1}$. In analytical determinations of solution conductivities, C_{sp} is usually calculated as micromhos per centimeter (μmhos/cm) or in SI units of millisemens per meter (mS/m).

Temperature control during measurements is necessary, since the temperature coefficient for conductance measurements is about 0.5 to 3% per °C. Conductance data is usually obtained at 25°C. Temperature corrections can be made, however the further the data temperature is from 25°C, the greater the uncertainty.

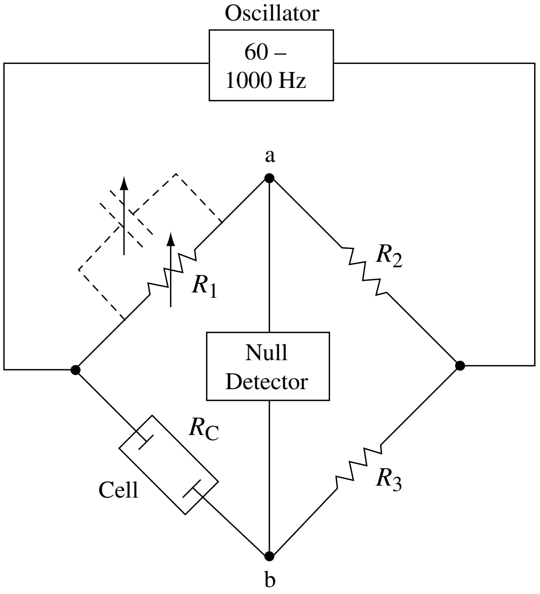


FIGURE 31 Basic Wheatstone ac bridge circuit for measuring conductance. (From Gary D. Christian and James E. O'Reilly, Instrumental Analysis, 2nd edition. Copyright © 1986 by Allyn and Bacon. Reprinted with permission.)

The measurement of solution resistance requires a conductance cell to contain the solution and a Wheatstone bridge for measurement. The Wheatstone bridge, shown in Figure 31, has an alternating current source of 6 to 10 volts with a frequency of 60 to 1000 Hz. R_2 and R_3 are fixed resistors of known values, R_1 is a variable resistor with up to 4 decades of resistance, and R_c is the resistance of the analyte solution. The null detector may be a headphone used with 1000 Hz, a cathode-ray tube or a micrometer. A variable capacitor is connected across R_1 and adjusted to balance out any phase shift in the a.c. signal caused by the capacitance of the electrode surfaces. This adjustment is made to provide the sharpest minimum in the null signal.

A conductivity cell essentially consists of two square plates of platinum of the same area, ar and platinized with platinum black to prevent polarization (see Figure 32). The plates are arranged parallel to each other at a fixed distance, d. (Other durable metals such as stainless steel and nickel are used in field and continuous monitoring operations.) The cell constant is not easily nor accurately obtained by measurement of the area and distance. A standard procedure is to fill the cell with a potassium chloride solution of known molarity and specific conductivity, C_{sp}. The measured resistance, R_m, along with C_{sp}, when substituted in equation 40, will give the cell constant, k_c. Standard Methods gives an excellently, detailed description of the measurement of conductivity.[2]

The measurement of specific conductances has a number of uses in water analysis: i) Ascertaining the mineral content of water in order to determine the effect of total ionic content on corrosion rates, physiological effects on animals and plants and the effects on chemical equilibria. ii) Appraising daily and seasonal variations in the mineral content in raw waste and natural waters. iii) Estimating the mineral content of high purity (distilled and deionized) water. iv) Checking the results of chemical analysis and estimating sample size. v) Determining the endpoint in conductometric titrations. vi) Estimating the total dissolved solids content by multiplying the specific

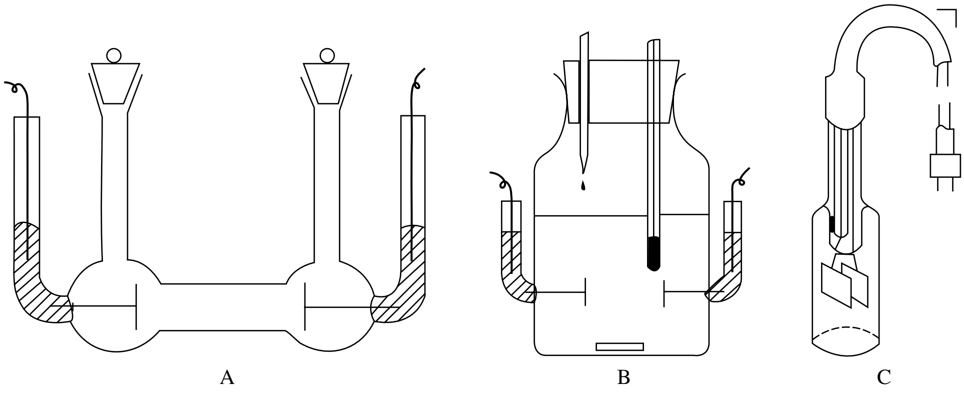


FIGURE 32 Three types of conductance cells, A: Precision conductance cell. B: Conductometric titration cell. C: Concentration dip cell. (From Gray D. Christian and James E. O'Reilly, Instrumental Analysis, 2nd edition. Copyright © 1986 by Allyn and Bacon. Reprinted with permission.)

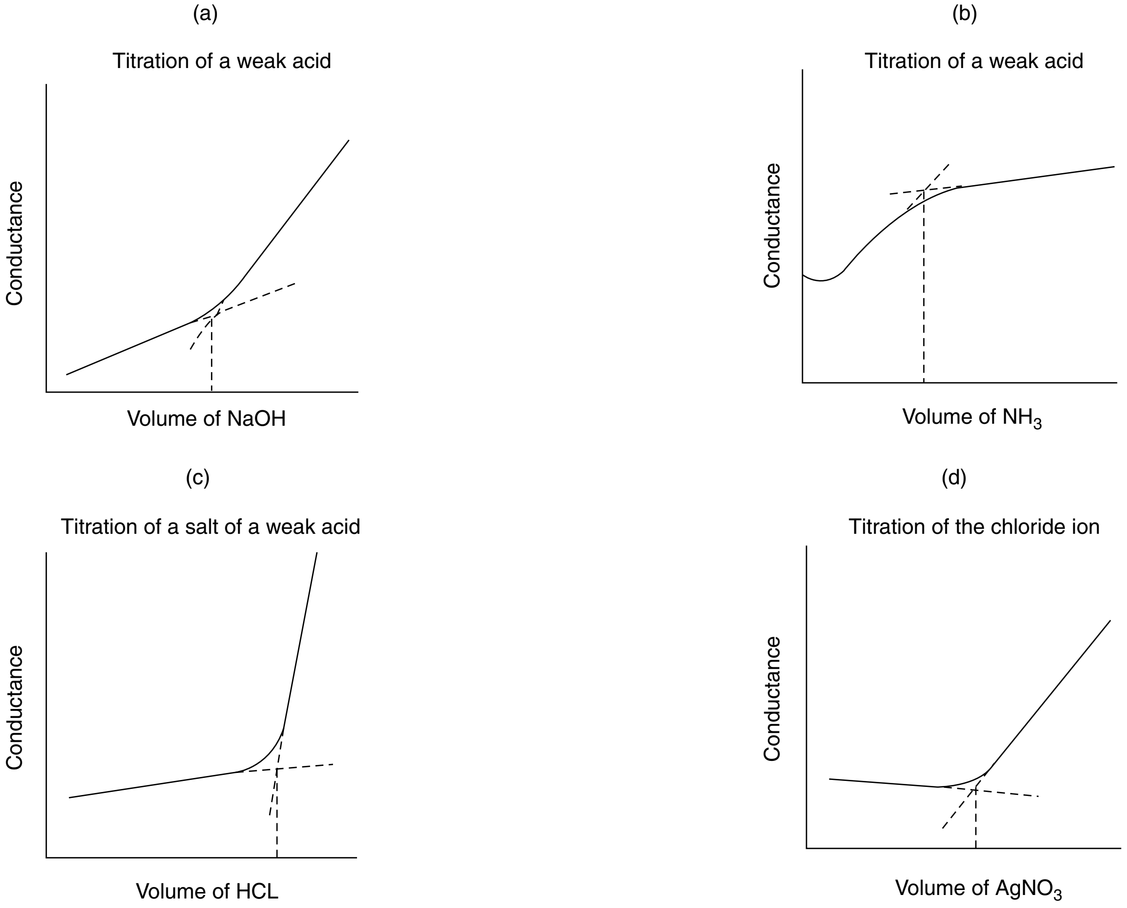


FIGURE 33 Examples of conductometric titration curves. The first three curves *a, b,* and *c* represent acid/base titrations and curve *d* a precipitation titration.

conductance by developed empirical factors.[2] Conductivity is also used to control plant processes, such as fluoride additions to drinking water and coagulant dosage control.[97]

(e) Coulometric methods[98]

Coulometry embraces three methods coulometric titrations of constant-current (amperostatic) coulometry, constant-potential (potentiostatic) coulometry, and electrogravimetry. In these methods the oxidation or reduction of an analyte to a new species or the reaction of an analyte with an electrolytically generated substance takes place. In the two coulometric methods the number of coulombs, necessary to directly or indirectly affect and electrochemical change of the analyte or a species that reacts with the analyte, is determined. In electrogravimetry the analyte undergoes an electrolytic reaction and the new species is deposited on an electrode and measured gravimetrically. The analyte concentration is calculated from a knowledge of the chemical composition of the deposit.

(1) Electrogravimetry

The analyte, metallic cations, halide ions, etc., are completely removed from the solution by plating on an inert electrode, usually platinum, in electrogravimetry. The increase in electrode weight is due to a species containing the reduced or oxidized form of the analyte. Electrolytic deposits can be pure metals on the cathode or compounds such as lead peroxide, thallium oxide and manganese dioxide on an anode, and silver halide salts on a silver anode. Complete removal of the analyte by electrolytic deposition from solution requires an unpolarized process where large electrodes and vigorous stirring are used. Two conditions are necessary concerning the electrolytic deposit: The deposit must adhere well to the electrode so that an accurate weight of the plated substance can be obtained, and a predictable, constant chemical composition of the deposit is required. The composition of the electrolyte solution does have, at times, a dramatic effect on the nature of the deposit. Substances are added to the electrolysis solution to improve the adherence of the plated substance to the electrode. Electrolysis in electrogravimetry can be achieved under two conditions—constant current through the cell or at a controlled working electrode potential.

Constant-current electrolysis has the advantage of a constant rate of reaction and is faster than the controlled-potential method. The limitation is an uncontrolled, working electrode potential. As the reaction progresses, the voltage must be increased to provide a constant current because of polarization effects. This can lead to the deposition of other species. However, this method is advantageously applicable when the species is alone in the solution, more readily reduced than hydrogen ion, and therefore, not requiring a controlled working electrode potential. Some of the cations amenable to this technique are cadmium(II), cobalt(II), copper(II), iron(III), lead(II), nickel(II), silver(I), tin(II), and zinc(II). If interfering ions are present they may be complexed in order to prevent electrodeposition or removed by prior chemical precipitation. Another application of this technique is the removal of a species which would interfere in a subsequent analytical operation. A mercury electrode is frequently used in this last application.

Controlled, working electrode potential electrolysis has the greatest applicability to mixtures of analytes. The control of the working electrode permits selective deposition of metals with standard potentials that differ by several tenths of a volt. A schematic diagram in Figure 34 embodies the elements of the apparatus. One example is the analysis of a sample containing copper, bismuth, lead, cadmium, zinc, and tin. The first three metallic ions are deposited at selective potentials; tin is held in solution as the tartrate complex. Cadmium and zinc are selectively deposited from ammoniacal solutions. Finally, the solution is acidified to decompose the tin-tartrate complex and tin is deposited.[99] Table 10 lists analyses amenable to this technique.[100] Controlled, working electrode potential electrolysis is a slower process than the constant current method, since the current flow tends to decrease in the controlled-potential technique due to polarization and other effects.

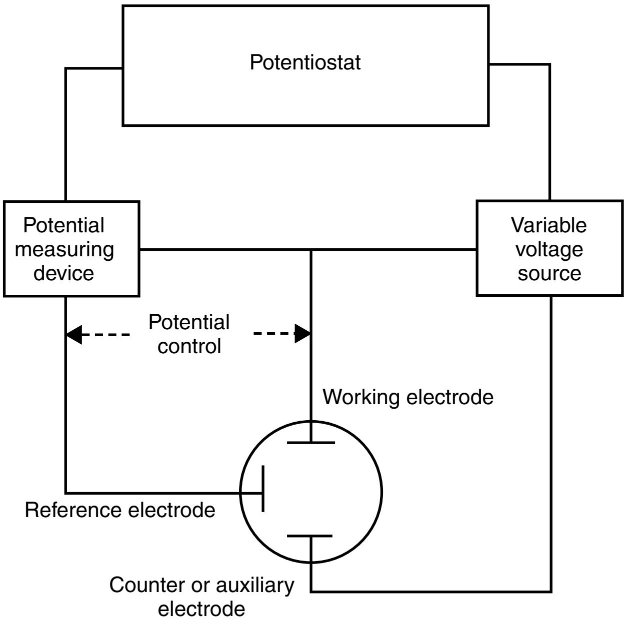


FIGURE 34 Schematic for a controlled-potential electrolysis apparatus.

(2) Coulometry

In coulometry one measures the quantity of electricity (number of coulombs) that is required to carry out a redox reaction of the analyte or generate a reagent that reacts with the analyte. Two general methods are used in coulometry, namely controlled-potential (potentiostatic) and controlled-current (amperostatic), commonly known as coulometric titrations. All the current must be used, solely, either directly or indirectly, for the reaction concerned with the analyte, that is, the current efficiency must be 100% for a quantitatively accurate, analytical result. In these methods standards are not needed; the proportionality constant (Faraday's constant) relating the number of coulombs and the amount of the analyte is derivable from known physical constants. In the cell the effects of polarization are decreased by using electrodes with large surface areas and by vigorous stirring of the solution during electrolysis.

Faraday's law relates the amount of electricity in coulombs to the number of equivalents of reactant in a redox reaction. Faradays's constant, 96,487 coulombs/equivalent, is the stoichiometric factor relating electrical charge passing in an electrolysis and the equivalence of substance reduced or oxidized. Since one ampere is the rate of flow of one coulomb per second, the time integrated flow of the current during a redox reaction will yield the number of coulombs, Q. When the current varies over the reaction time period an integration is necessary. With constant current the number of coulombs can be calculated by the equation,

$$it = Q = 96{,}487\ W/ew. \tag{41}$$

The quantities are i, amperes; t, seconds; W, sample weight in grams, and ew, equivalent weight in grams per equivalent. The equivalent weight is the atomic weight, aw, or formula weight, fw, divided by the number of electrons in the redox or half cell reaction. For example in the reactions,

$$Fe^{3+} + e = Fe^{2+} \tag{42}$$

TABLE 10
Controlled-potential electrolysis for metal separation and determination

Metal Separated	Metal Matrix
Sb	Pb, Sn
Bi	Sb, Cd, Cu, Pb, Sn, Zn
Cd	Zn
Cu	Sb, Bi, Cd, Pb, Ni, Sn, Zn
Pb	Al, Cd, Fe, Mn, Ni, Sn, Zn
Ni	Al, Fe, Zn
Ag	Cu and more active metals

and

$$MnO_4^- + 8H^+ + 5e = Mn^{2+} + 4H_2O \qquad (43)$$

the ew of Fe^{3+} is the aw/1 and ew of MnO_4^- or $KMnO_4$ is the fw/5. The number of equivalence is represented by the fraction. W/ew, in equation 41.

(a) Potentiostatic coulometry

The working electrode is maintained at a constant, present potential in potentiostatic coulometry. In this process the analyte changes oxidation state reacting quantitatively with the current: i.e., at 100% current efficiency. Redox reactions with less reactive species in the analyte solution are precluded. Maintenance of 100% current efficiency allows the amount of analyte present to be determined unambiguously by the quantity of coulombs consumed application of Faraday's law (see equation 41). As the redox reaction proceeds, the current and, therefore, the reaction rate decreases because of increases in polarization effects, changes in solution resistance, etc. Nevertheless the working electrode potential is maintained restricting the rate of the redox reaction of the analyte. The amount of current consumed usually decreases with time and is measured by an integrating device: namely, an electronic, chemical or electromechanical coulometer. An electronic controlled-potential coulometric titrator developed by Kelley *et al.*[101] includeds a potentiostat, which provides and maintains a constant potential, a d.c. current source, an electrolytic cell, and an integrating coulometer providing the readout in coulombs. This device has a range of 10 microamperes to 10 milliamperes, an accuracy of 0.01%, and a response time of 10 milliseconds.

An application of this method is the analysis of a mixture of antimony(III) and antimony(IV) accomplishable in two steps.[102] From the voltammetric data, the centered plateau reduction voltages for the following reactions are as follows:

$$Sb^{5+} + 2e = Sb^{3+}, -0.21 \text{ V vs SCE} \qquad (44)$$

and

$$Sb^{3+} + 3e = Sb^0, -0.35 \text{ V vs SCE.} \qquad (45)$$

The supporting electrolyte (6M hydrochloric acid plus 4M tartaric acid) is reduced at −0.35 V, before addition of the sample to remove impurities. Upon addition of the sample of antimony species, the solution is deaerated with nitrogen. The potentiostat described previously[101] is set at −0.21 V and the experimental voltage rises slowly to that value during the reduction of Sb^{5+}. The current starts to decrease when the set voltage is reached because of the decreasing concentration of Sb^{5+}, other polarization effects, etc. When all the Sb^{5+} has been reduced to Sb^{3+}, the current will decrease to a negligible value signaling the end to the first analytical step. The second reduction is carried out by setting the potentiostat at −0.35 V and repeating the process. In calculating the Sb^{3+} content of the sample a correctin for the Sb^{3+} generated in the first reduction must be made. When interfering substances are present, e.g., the analysis of plutonium in the presence of iron, an indirect approach can be used.[103]

Potentiostatic coulometry has the same advantages of controlled-potential electrogravimetry. In employing this method rather than electrogravimetry, the attendant problems of poorly adhering electrolytic deposits are eliminated. In addition analytes that do not yield electrolytic deposits, but are amenable to coulometric analysis can be analyzed. This technique requires a longer analysis time compared to amperostatic coulometry because of the decrease in current flow as the reaction proceeds.

(b) Coulometric titrations

Coulometric titrations (amperostatic or controlled current coulometry) are carried out at constant current (see Figure 35). At the working electrode a reagent is generated that reacts with the analyte in one of several types of reactions namely, oxidation/reduction, acid/base, complexation or precipitation (see Figure 36). When all the analyte has reacted with the generated reagent, the endpoint or completion of the reaction may be detected by potentiometry, indicator color changes, amperometry, or conductance. In some coulometric titrations, however, part of the current arises from direct reduction or oxidation of the analyte at the electrode and the remainder through the generation of the reagent. Subsequent reaction of the reagent and the remaining unreacted analyte ends the titration and 100% current efficiency is maintained. The measured quantity for a coulometric titration is the number of coulombs necessary to generate the reagent; it is comparable to the volume of titrant in a classical titration. (A current efficiency of 100% is, therefore, required.) Classical and coulometric titrations are comparable in a number of ways; these two methods have similar endpoint detection methods and stoichiometric reactions between titrant and analyte must be rapid, complete, and free of side reactions. Since constant current is used in this technique, an accurate timer is used in the coulometric titrator in Figure 35. The product of time and current (see equation 41) will give the number of coulombs. An integrator is not needed, as in potentiostatic coulometry, where the current can change with time.

The generation of reagents can be internal, in the cell containing the analyte, or external to the cell. Outside generation of reagents is often convenient for several reasons: Electrolytic interference of substances in the sample solution and incompatibility in conditions fostering efficient generation of the reagent can occur in internal generation.

(x) Internal generation

Two modes of internal generation are utilized and are referred to as primary and secondary coulometric titrations. In the primary mode the analyte reacts directly with a species generated from the electrode material. Therefore, no other species should be present which will react electrolytically with the working electrode within about 0.5 V of the

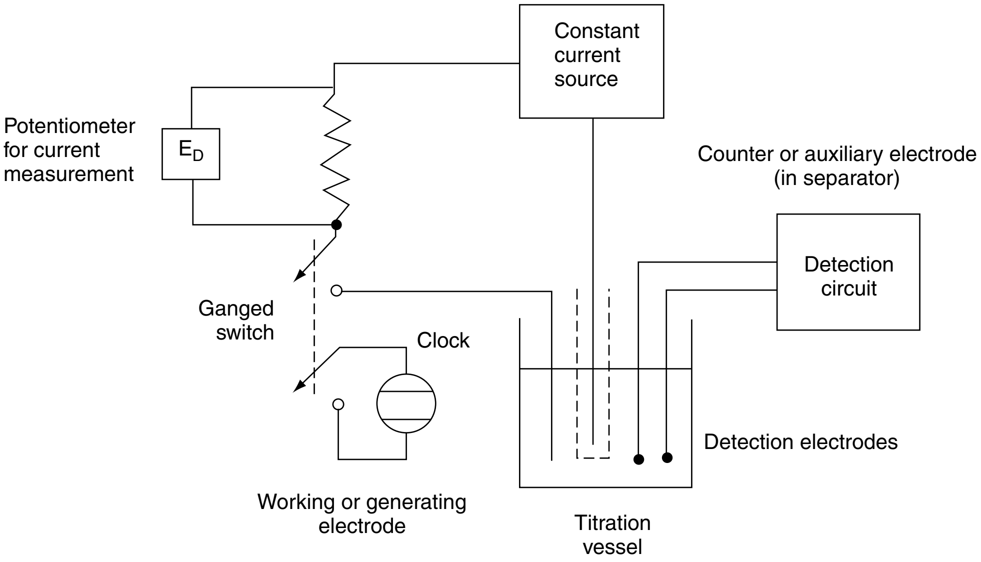


FIGURE 35 Coulometric titration apparatus.

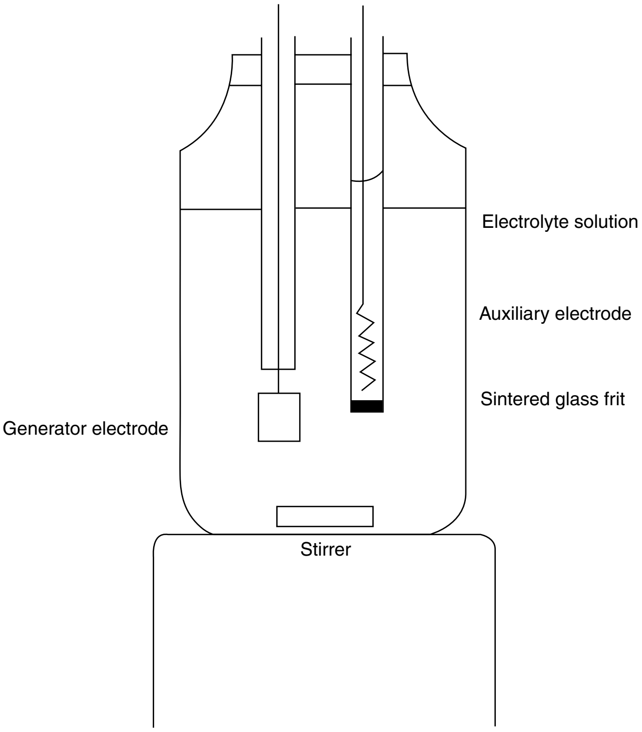


FIGURE 36 Coulometric titration cell.

potential of the analyte reactivity. An example of this mode is the reaction of ionic halide ions, mercaptans, or sulfhydryl groups with electrolytically generated silver ion from a silver working electrode. The endpoint may be determined amperometrically.

In the secondary mode an active intermediate is generated which reacts directly with the analyte. The intermediate is generated at a potential between that of the potential of the analyte redox reaction and the potential of an interfering, unwanted reaction. The interfering reaction can be due to electrolysis of the supporting electrolyte or another electroactive substance in the sample. An illustration of a secondary titration is as follows: The coulometric titration of ferrous ion uses the Ce^{3+}/Ce^{4+} system as the intermediate reactant. At the onset of the titration the ferrous ion is directly oxidized at the platinum working electrode. Since the current is constant, the potential increases as the concentration of ferrous decreases. At the oxidation potential of cerous ion and ceric ion is generated and in turn oxidizes the ferrous ion. The summation of number of coulombs consumed by each electrolytic reaction, direct and ceric oxidation, is that needed to titrate the ferrous ion present. The endpoint chosen is the presence of excess ceric ion indicating the complete oxidation of ferrous ion. The ceric ion can be detected photometrically or amperometrically. Since an excess of cerous ion is present, the working electrode is maintained at the unique potential of the cerous/ceric couple. The potential of the working electrode is prevented from increasing to a value where the interfering substances would be electrolyzed.

In secondary coulometric titrations reagents can be generated that are inconvenient to use in classical titrimetry. Some examples are the generation of hydroxide ion free of contamination by carbonate and the reactive titrants bromine, chlorine and titanous. More on this subject is given in the section on external generation of titrants.

(y) External generation

External generation of reagents can be necessary because of the presence of interfering substances in the sample. In Figure 37 a double-arm apparatus for the outside generation of reagents is shown.[104] Specifically the equations for the generation of acid (hydrogen ion) or base (hydroxide ion)

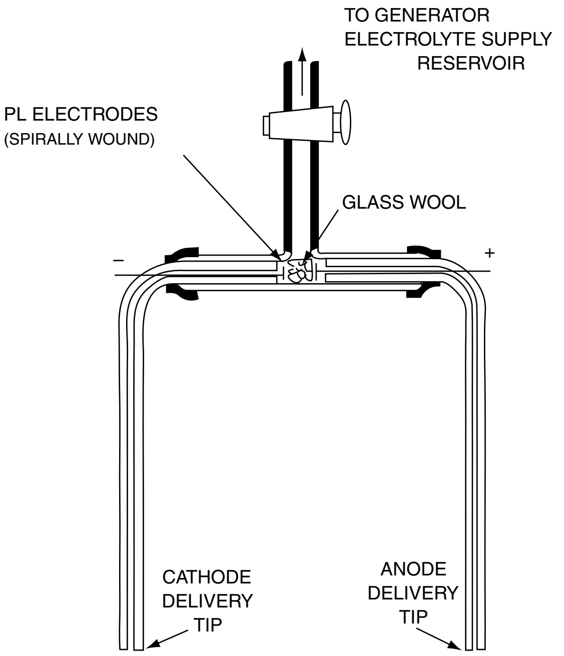


FIGURE 37 Generator cell for external production of titrants for coulometric titrantions. (Adapted with permission for reference 104. Copyright 1931 American Chemical Society.) For example in acid/base titrimetry: (a) generation of OH^- at the cathode: $2H_2O + 2e \rightarrow H_2 + 2OH^-$ (b) generation of H^+ at the anode: $H_2O \rightarrow 1/2O_2 + 2H^- + 2e$.

are illustrated. A single arm electrolytic cell is also available.[105] Unstable reagents, namely, bromine, dipositive silver, titanium(III), chlorine, etc., can be conveniently generated and utilized in this technique. In Table 11 a number of applications are listed. The ability to utilize these unstable reagents expands the horizon of tritrimetry and analytical methodology.

(3) Measurement of radioactivity[106–8]

a. Introduction

Radioactive species in water and wastewater samples arise from natural and anthropogenic sources. The radiation emitted from radioisotopes is deleterious to all life forms and has led to the development of a number of radiochemical methods. Standard Methods[2] includes procedures for the detection of tritium, radium, and radioactive species of cesium, iodine, strontium, and uranium, as well as, the general measurement of the amount of ionizing radiation.

Several kinds of ionizing radiation emanate from radioisotopes. They include gamma and x-rays, beta rays (electrons and positrons), and alpha rays (helium nuclei). Ionizing radiation, when penetrating matter, causes the formation of ions that are dangerous to life forms leading to cancer in people and animals. Gamma and x-rays are at the high frequency ($>10^6$ Hz) end of the electromagnetic spectrum. The former arise from decay events occurring in the nucleus and are more energetic higher frequency than x-rays that arise from rearrangement of electrons surrounding the decaying nucleus. Both kinds of radiation are highly penetrating due to their high energy, and therefore are quite dangerous. Gamma and x-rays are emitted as continuous spectra.

Beta particles are of two types; negatively charged electrons or negatrons and positively charged positrons. They arise from decay processes in nuclei. A positron has a very short life time and is annihilated on collision with an electron to form two gamma photons. Beta particles have a greater penetrating ability than alpha particles and are poorer at ionizing matter. Since beta particles are scattered in air, their range is difficult to assess. Beta particles, emitted during nuclear decay, yield a continuous spectrum of particle energies.

Alpha particles are the nuclei of helium atoms and have a positive charge of two. They are the product of the decay of natural isotopes, such as uranium, radium, and others. Alpha particles have high energies and, therefore, high ionizing power. Their penetrability, however, is low, about 5–7 cm in air. Alpha particles emitted from radioisotopes do not yield continuous spectra. The particles are monoenergetic or have a group of several discrete energies.

b. Instrumentation

Ionizing radiation interacting or colliding with matter causes the production of charged particles or energetic species; measurement of these entities is the basis of radiation detectors. Two phenomena are utilized in detecting ionizing radiation: namely, ionization of gases and solids to yield an ion current pulse and the excitation of crystals to provide a luminescence pulse. The ionization of gases is used in ionization chambers, proportional counters, and GM (Geiger Muller) tubes. Semiconductor radiation devices produce electron–hole pairs on the impingement of ionizing radiation. Crystal and liquid scintillation detectors are employed for insoluble and soluble samples, respectively. An instrument for radiation measurement consists of several modules: namely, the detector responding to the radiation, circuits that count pulses, and pulse height analysis circuits that discriminate between the energies of the pulses.

The nature of the pulse must be understood. In a number of detectors the pulse height of the radiation event is proportional to the energy of the particle or radiation and independent of the applied potential, whereas in other detectors the pulse height is independent of energy. The pulse of photoelectrons or luminescence has a pulse width due to the random nature of the collision process. This fact leads to small statistical differences in the pulse heights for a series of radiation events with the same energy. The smaller the pulse width the greater the resolution of the detector.

Radiation detectors operate under two modes, pulse and mean level. Of interest here is the pulse type that detects the interaction of radiation with the detector as a unique event. These pulses are counted and measured in counts per unit time, usually, counts per minute, cpm. The type of radiation (beta, gamma, etc.), the frequency or wavelength and the intensity of the radiation can also be determined.

TABLE 11

Typical electrogenerated titrants and substances determined by coulometric titration*

Electrogenerated Titrant	Generating Electrode and Solution	Typical Substances Determined
Bromine	Pt/NaBr	As(III), U(IV), NH_3, olefins, phenols, SO_2, H_2S, Fe(II)
Iodine	Pt/KI	H_2S, SO_2, As(III), water (Karl Fischer), Sb(III)
Chlorine	Pt/NaCl	As(III), Fe(II), various organics
Cerium(IV)	$Pt/Ce_2(SO_4)_3$	U(IV), Fe(II), Ti(III), I^-
Manganese(III)	$Pt/MnSO_4$	Fe(II), H_2O_2, Sb(III)
Silver(II)	$Pt/AgNO_3$	Ce(III), V(IV), $H_2C_2O_4$
Iron(II)	$Pt/Fe_2(SO_4)_3$	Mn(III), Cr(VI), V(V), Ce(IV), U(VI), Mo(VI)
Titanium(III)	$Pt/TiCl_4$	Fe(III), V(V, VI), U(VI), Re(VIII), Ru(IV), Mo(VI)
Tin(II)	$Au/SnBr_4(NaBr)$	I_2, Br_2, Pt(IV), Se(IV)
Copper(I)	Pt/Cu(II)(HCl)	Fe(III), Ir(IV), Au(III), Cr(VI), IO_3^-
Uranium(V), (IV)	Pt/UO_2SO_4	Cr(VI), Fe(III)
Chromium(II)	$Hg/CrCl_3(CaCl_2)$	O_2, Cu(II)
Silver(I)	$Ag/HclO_4$	Halide ions, S^{2-}, mercaptans
Mercury(I)	$Hg/NaClO_4$	Halide ions, xanthate
EDTA	$Hg/HgNH_3Y^{2-}$[a]	Metal ions
Cynaide	$Pt/Ag(CN)_2^-$	Ni(II), Au(III, I), Ag(I)
Hydroxide ion	$Pt(-)/Na_2SO_4$	Acids, CO_2
Hydrogen ion	$Pt(+)/Na_2So_4$	Bases, CO_3^{2-}, NH_3

[a] Y^{4-} is ethylenediamine-tetra-acetate anion.

* From A.J. Bard and L.R. Faulkner, Electrochemical Methods, p. 390. (Copyright 1980 by John Wiley & Sons, Inc., with permission.)

(1) Detectors

Radiation detectors operate on two principles; ionization of a gas or solid to generate a small ion current and excitation of a substance to cause a short term luminescence in a crystal or a solution. Ionization detectors are of two varieties; those that use gas as the ionization medium, i.e., gas-filled detectors and those that use a crystal, i.e., semiconductor detectors. Measurements of luminescence due to radiation are made in crystal and liquid scintillation counters.

(a) Ionization and detectors

(i) Gas-filled detectors

Gas-filled detectors respond to ionizing radiation by the formation of ion pairs, a photoelectron and a positive ion (cation) by the gas molecules. The gas in the detector tube may be a mixture: e.g., argon and a low concentration of an organic substance or methane. In Figure 38 the gas-filled detector is shown with a central electrode (anode) to which is applied a voltage. The pulse of photoelectrons migrate to the electrode due to the electric field, are collected, and produce a small ion current. The current represents the radiation event or particles deposited in the detector. Several types of gas-filled detectors are available and are operated in a number of voltage ranges leading to different kinds of detector responses; the operation potentials of these detectors are illustrated in Figure 39.

In the saturation (ionization) chamber region a radiation event of given energy gives rise to a number of photoelectrons independent of the applied potential. They are collected on the electrode yielding an ion current or pulse height proportional to the energy of the radiation event. The pulse of photoelectrons, however, has a pulse width due to the random nature of the collision process. This situation leads to a statistical variation in pulse heights for the same energy source defining the resolution of the detector. An ionization chamber in combination with a pulse height analyzer can operate as an alpha particle spectrometer.

When the applied voltage is increased, producing a field greater than 200 V/cm, the electrons formed in ion pair formation are accelerated. These accelerated electrons collide with gas molecules causing increased ionization leading to an increase in the collected electrons. This effect is called gas amplification and is present in the proportional counter and GM tube regions.

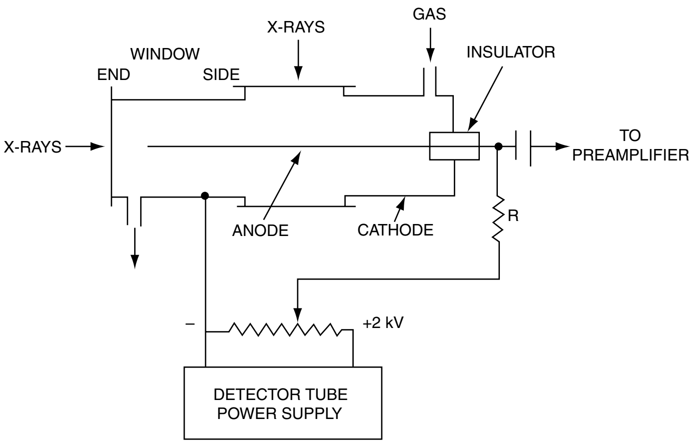


FIGURE 38 Structure of the gas-filled x-ray detector (proportional and Geiger counters, and ionization chamber). The detector is shown to have both end and side windows for illustration only.

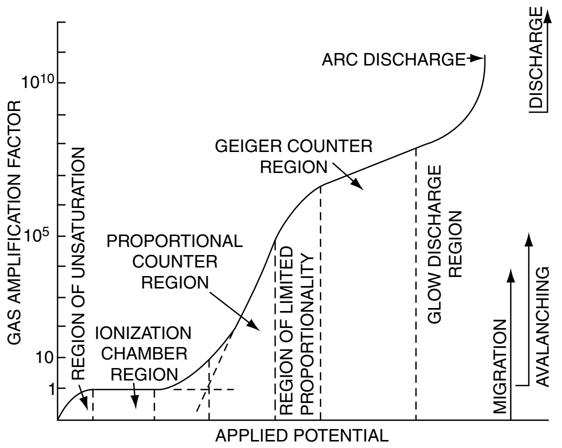


FIGURE 39 Gas-amplification factor as a function of applied potential for the gas-filled detector.

The proportional counter region yields a pulse height that is proportional to the energy of the radiation or particle. The amplification factor is 500 to 10,000 and the detector has a dead or non-conducting time of 0.5 μsec. Alpha and beta particles can be measured separately by using two different voltages and a pure alpha emitter in a proportional counter. The limited proportional region is not useable because the presence of secondary charges hinders the gas amplification process.

In the GM tube region the amplification factor is 10^9 and the pulse height is independent of the energy and type of radiation. Therefore pulse height analysis can not be performed using the GM tube but it can be used for general counting. The plateau is 300 V in length and the counting rate increases less than 3% for a 100 V increase in applied voltage. The positive space charge formed in the detector causes a non-conducting or dead time of 250 microseconds leading to a loss in radiation events. Serious rate errors are experienced when counting rates are larger than 10^4 cpm. A correction to the rate may be calculated using the resolving time. This detector is useful for the analysis of separated radionuclides.

(ii) Semiconductor detectors

There are several types of semiconductor detectors, namely the surface barrier and p-n junction detectors and the lithium drifted silicon or germanium detectors. They all function, generally, according to the following model.

Semiconductor detectors are analogous to gas-filled detectors in their principle of operation. In the semi-conductor detector a radiation event leads to an ion pair formation of an electron–hole pair, whereas in the gas-filled detector electron–cation pair are formed. A model of a semiconductor detector, a lithium drifted detector, is given in Figure 40. A central zone of ultrapure intrinsic semiconductor is flanked by thin layers of p and n type semiconductor material. A bias voltage is imposed across the ensemble to form a high field. The radiation event causes the formation in the central zone of a highly energetic photoelectron which gives rise to a large number of electron–hole pairs. The large number of highly mobile electrons are raised to the conduction band due to the transfer of kinetic energy from the photoelectron. The electrons and holes "move" to the p and layers, respectively, and are collected under the effect of the high field giving a current pulse. The size of the current pulse is proportional to the energy of the radiation event as in the proportional counter and has a pulse width due to the randomness of the process as discussed previously.

In the lithium-drifted detectors lithium, an n-type substance, is used to form the ultrapure intrinsic semiconductor from p-type silicon or germanium. Ultimately lithium becomes the doppant for the n-type semiconductor after several involved processes. These detectors function more efficiently if the detector and preamplifier are kept at liquid

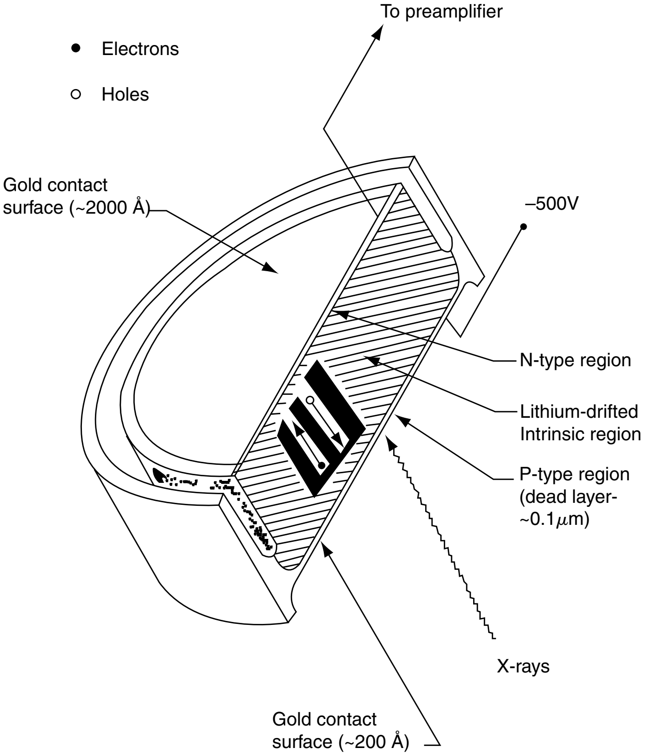


FIGURE 40 Cross-section of a typical lithium-drifted silicon, Si(Li), detector, x-rays create electron-hole pairs in the intrinsic region of the semiconductor; these charge carriers then migrate to the electrodes under the influence of an applied bias voltage. (Courtesy of the Kevex Instruments, Inc.)

nitrogen temperature, -196°C. At this low temperature thermal electronic noise is reduced and the lithium is prevented from diffusing into the central zone of intrinsic semiconductor material affording increased resolution capabilities. The energy required for an electron-hole pair generation for several radiation detectors are as follows: 2.9 eV (eV, electron volt) for germanium, 30 eV for a gas ionization chamber and 500 eV to produce a photoelectron in a NaI(T1) scintillation detector. Thus the germanium detector produces 170 times the number of ion pairs than the scintillation detector per unit of eV and the resolution is better by 13 times. Different pulse widths from each type of detector caused by the randomness of the process lead to the differences in energy resolution. For electrons, x-rays, and photons the energy resolution in the germanium detector is 3.8, 0.6 and 20 keV, respectively. The line widths for the germanium and scintillation detectors are 3.3 and 46 keV (keV, 1000 ev), respectively. The germanium detector in conjunction with a pulse height analyzer can be used to measure x- and gamma rays in energy dispersive spectrometers. In these systems radiation with a spectrum of energies can be resolved, since the germanium detector produces a current pulse proportional to the energy of the radiation. One disadvantage of this system is that radiation with a wave-length above one Angstrom is better resolved in a crystal spectrometer. Energy dispersive systems have excellent resolution for wavelengths below one Angstrom but demonstrate poor resolution for wavelengths above 1 Å.

(b) Scintillation counters

Some inorganic crystals and organic crystals or molecules emit a pulse of light on interaction with ionizing radiation. The energy of the radiation causes ionization or activation of the scintillating substance; it relaxes emitting a fluorescent or phosphorescent light pulse with life times of about 10^{-8} and 10^{-4} seconds, respectively, in the visible or near uv region. The number of photons emitted in each light pulse is proportional to the energy of the radiation event. Some crystals, namely sodium iodide, are doped with an activator to shift the light pulse to a longer wavelength.

(i) Crystal scintillation counters

Scintillation counters utilize a crystal optically coupled to a photomultiplier tube that converts the light pulse to a current pulse (see Section III,B,1,b,*(4)*,(*a*),(iii)). Further coupling to a pulse height analyzer results in an energy dispersive spectrometer. Inorganic scintillation crystals of alkali halides doped with thallium, lead or europium as activators, are commonly used. Potassium halides containing the natural radioisotope, ^{40}K, which radiates electrons, positrons, and γ rays, are not used in scintillation detectors.

A sodium iodide crystal doped with 0.1 to 1% of thallium (I) iodide, NaI(T1), is most widely used. The NaI has a high density that absorbs gamma radiation and the iodide ion provides an efficient conversion of radiation to light. The radiation event first activates the iodide ion that emits a light pulse in the uv region. The uv pulse excites the thallous ion that on relaxation emits fluorescent light at about 410 nm. It is compatible to the photomultiplier tube. Since NaI is hygroscopic, the crystal must be sealed well. For most efficient counting a "well type" detector is frequently used. Isotopes emitting x-, gamma, beta, and alpha radiation are detected by NaI(T1), and in particular x- and gamma radiation are most beneficially measured. Energy dispersive spectrometers using this detector are employed in x- and gamma-ray spectrometers.

(ii) Liquid scientillation counters

Liquid scintillation counting is a convenient and efficient means of detecting low levels of radiation from small amounts of samples. The liquid scientillation solution consists of a primary, or a primary and secondary scintillator or phosphor dissolved in an organic solvent. The radioactive sample is dissolved in the scintillation solution. The radiation excites the organic solvent molecules, such as toluene, xylene, terphenyl, etc. The excited molecules transfer their energy, mainly non-radiatively, to the primary scintillator causing emission of a fluorescent light pulse detected by the photomultiplier tube.

The newer models of liquid scintillation counters contain photomultiplier tubes that respond to the emissions of a primary scintillator. PPO (2,5-diphenyloxazole), a primary scintillator, emits light pulses in the range 330 to 400 nm. Blue sensitive photomultiplier tubes employed in older model counters necessitated the use of a secondary phosphor which absorbed the light pulse of the primary phosphor and emitted a light in the blue wavelength region. POPOP, the 2^0 scintillator, 2,2′p-phenylenebis(4-methyl-5-phenyl-oxazole), absorbs at about 360 nm and emits fluorescence light at 410 to 420 nm. Therefore, PPO and POPOP are a good pair to provide an effective liquid scintillator for the older blue sensitive photomultiplier tubes. A diagram of a scintillation counter is shown in Figure 41.

(2) Signal processors and readout

Radiation detection instruments are able to measure two characteristics; the number of particles or radiation events occurring in a unit time and the energy of each particle or event.

The counting of events is carried out in electronic countering devices. Counting circuits convert voltage pulses to square waves, scale or reduce the number of input pulses, and count the frequencies of pulses using binary digital circuits. A scaling circuit decreases the frequency of pulses arriving from the detector by a known fraction in order to accommodate the counting devices detection rate. Binary digital circuits containing JK flip-flops accomplish this reduction. A binary counter with four JK flip-flops produces one output pulse for every sixteen input pulses. Arrangement of the binary circuits to provide a decade counting unit provides a decimal readout.

In ionization chambers, proportional counters, semiconductor detectors, and scintillation counters the size or height of the pulse is proportional to the energy of the radiation event. Pulse height analyzers are coupled to the output of these detectors in order to form an energy dispersive instrument. Pulse height analyzers employed to separate radiation of differing energies are equivalent to monochrometers used to disperse uv, vis, and ir radiation. Instruments are also available that carry out wavelength dispersion. The radiation is dispersed using a crystal for x- and gamma rays.

A description of the operation of a pulse height analyzer is as follows: the radiation event is converted to a signal by the detector and amplified resulting in pulses as large as 10 volts. A pulse height selector provides a narrow voltage window by rejecting voltage pulses between minimum and maximum values using electronic discriminator circuits. This window can have a voltage range of 0.1 to 0.5 volts. A pulse height analyzer can consist of one or several pulse height selectors. One pulse height selector comprises a single channel pulse height analyzer. The voltage range of 10 volts can be scanned, automatically or by hand, using a window of 0.1 volt yielding an energy dispersion spectrum. Two to hundreds of channels comprise a multichannel analyzer. Each channel with its own counting circuit is set for a specific voltage window. Therefore, the total spectrum can be simultaneously counted and recorded.

Readout of counting rates can be displayed on solid state devices, printers, etc. Spectra display on a potentiometric recorder or storage in a computer with subsequent printout is available.

(3) Applications

Liquid scintillation counting is used for a variety of tasks: namely; measurement of the low energy beta emitters ^{3}H, ^{14}C, ^{32}P and ^{35}S; gamma and x-ray energy dispersive spectrometry; and beta-gamma coincidence scintillation counting for ^{131}I.

Standard Methods[2] has procedures for the analysis of cesium-134 and 137, and iodine-129 through 135, total radioactive strontium (^{89}Sr and ^{90}Sr) and ^{90}Sr alone, tritium, radon-222, uranium-234, 235 and 238, and total radium, radium-226, radium-228, and total alpha and beta content of water.

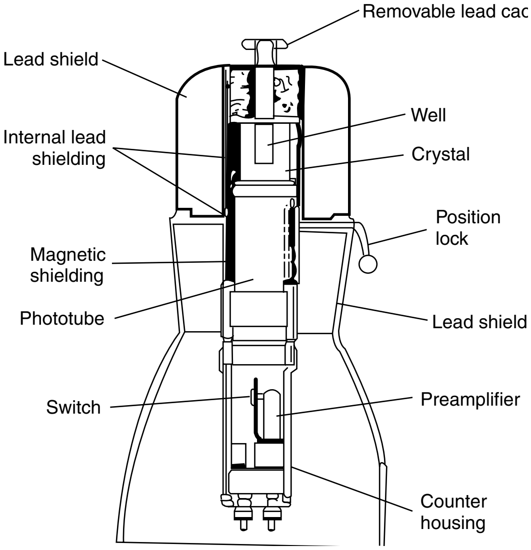


FIGURE 41 A well-type scintillation counter. (Courtesy of TN Technologies, Inc.)

REFERENCES

1. Kieth, L.H., ed. (1988), Principles of Environmental Sampling, American Chemical Society, Washington, D.C.
2. Amer. Pub. Health Assoc., (1995), Standard Methods for the Examination of Water and Wastewater, 19th Ed., Washington, D.C.
3. Mentink, A.F. (1968), Specifications for an Integrated Water Quality Data Acquisition System, U.S. Depart. of Interior, FWQO, Washington, D.C.; STORET System Handbook (1973), U.S. Environmental Protection Agency, Washington, D.C.
4. Ciaccio, L.L., Raul R. Cardenas, Jr. and J.S. Jeris (1973), Automated and Instrumental Methods in Water Analysis in Water and Water Pollution Handbook, Vol. 4, L.L. Ciaccio, ed., Chap. 27, Marcel Dekker Inc., New York.
5. American Society for Testing Materials (1996), Book of ASTM Standards, Vols. 11.01 and 11.02, American Society for Testing Materials, Philadelphia, PA.
6. EPA methods: Publications on the use of instruments in the analysis water by the U.S. Environmental Protection Agency are too numerous to list here. The Agency can supply a list of these publications.
7. Technical Practice Committee (1984), Process Instrumentation & Control System, Manual of Practice No. OM-6, Water Pollution Control Federation, Washington, D.C.
8. Edmonds, T.E., editor (1988), Chemical Sensors, Blackie & Son, London (Chapman and Hall, New York).

9. Malmstadt, H.V., C.G. Enke and E.C. Toren (1963), Electronics for Scientists, W.A. Benjamin Inc., New York.
10. Lewin, S.Z. (1961), *Anal. Chem.,* **33(3),** 23A–43A.
11. Diefenderfer, A.J. (1979), Principles of Electronic Instrumentation, 2nd Ed., W.B. Saunders Company, New York.
12. Skoog, D.A. (1985), Principles of Instrumental Analysis, 3rd ed. Saunders College Publishing, New York.
13. Horowitz, P. and W. Hill (1989), The Art of Electronics, 2nd ed., Cambridge Univ. Press.
14. Watson, J. (1983), Mastering Electronics, 2nd ed., McGraw-Hill.
15. Willard, H.H., L.L. Merritt, Jr., J.A. Dean and F.A. Settle, Jr. (1981), Instrumental Methods of Analysis, 6th ed., Wadsworth Publishing Company, Belmont, Calif.
16. Bader, M. (1980), *J. Chem. Educ.,* **57,** 703.
17. Ettre, L.S. (1967), The Interpretation of Analytical Results: Qualitative and Quantitative Analysis, Chap. 7, in The Practice of Gas Chromatography, Ettre, L.S. and A. Zlatkis, Eds., Wiley-Interscience, New York.
18. Walpole, R.E. and R.H. Myers (1972), Probability and Statistics for Engineers and Scientists, The Macmillan Co., New York.
19. Taylor, J.R. (1982), An Introduction to Error Analysis, University Science Books, Mill Valley, California.
20. Lytle, F.E., (1974), *Anal. Chem.,* **46,** 545A.
21. Talmi, Y. (1982), Appl. Spect., 36, 1; indem, (1979, 1982), Multichannel Image Detectors, Vol. 1 & 2, American Chemical Society, Washington, D.C.
22. Ewing, E.W. (1971), *J. Chem. Educ.,* **48(9),** A521.
23. Levinstein, H. (1969) *Anal. Chem.,* **41(14),** 81A.
24. Winefordner, J.D., J.J. Fitzgerald and N. Omenetti (1975), *Appl. Spectrosc.,* **29,** 369.
25. Marshall, A.G. (1982) Fourier, Hadamard and Hilbert Transforms in Chemistry, Plenium Press, New York.
26. Griffiths, P.R. (1983), *Science,* **222,** 297.
27. Meehan, E.J. (1981) in Treatise on Analytical Chemistry, 2nd ed., P.J. Elving, E.J. Meehan, and I.M. Kolthoff, Eds., Part I, vol. 7, Chap. 2, Wiley, New York.
28. Smith, A.L. (1981), in Treatise on Analytical Chemistry, 2nd ed., P.J. Elving, E.J. Meehan, and I.M. Kolthoff, Eds., Part I, vol. 7, Chap 5, Wiley, New York.
29. Thompson, K.C. and R.J. Reynolds (1978), Atomic Absorption, Fluorescence, and Flame Emission Spectroscopy, 2nd ed., Wiley, New York.
30. Alkemade, Th.J. and R. Herrmann (1979), Fundamentals of Analytical Flame Spectroscopy, Wiley, New York.
31. Sacks, R.D. (1981), in Treatise on Analytical Chemistry, 2nd ed., P.J. Elving, E.J. Meehan and I.M. Kolthoff, Eds., Part I, vol. 7, Chap 6, Wiley, New York.
32. Seitz, W.R. (1981), in Treatise on Analytical Chemistry, 2nd ed., P.J. Elving, E.J. Meehan, and I.M. Kolthoff, Eds., Part 1, vol. 7, Chap. 4, Wiley, New York.
33. Van Loon, J.C. (1986), *Anal. Chem.,* **53,** 332A.
34. Cocoran, Jr., F.L. (1974), *Amer. Lab.,* **6(3),** 69.
35. Tertian, R. and F. Claisse (1982), Principles of Quantitative X-Ray Fluorescence Analysis, Heyden, London.
36. Hochgesang, F.P. (1964), in Treatise on Analytical Chemistry, I.M. Kolthoff and P.J. Elving, Eds., Part I, vol. 5, Chap. 63, Interscience, New York.
37. Hameyer, K., H. Madlin and L.L. Ciaccio (1967), *Food Technol.* **21,** 95.
38. Ives, D.J.G. and G.J. Janz (1961), Reference Electrodes, Academic Press, New York.
39. Reilley, C.N. and R.W. Schmid (1958), *Anal. Chem.,* **30,** 947; Reilley, C.N., R.W. Schmid, and D.W. Lamson, (1958), *ibid.*, 30, 953.
40. Jones, R.H. (1966), *J. Instr. Soc. Amer.,* **13,** 40; Pohland, F. and K. Mancy, (1968), Proc. 3rd Intern. Fermentation Symp., Rutger's Univ., Paper G3-9.
41. Moloff, A.H. (1960), Ph.D. Thesis, Univ. of Michigan, Ann Arbor, Michigan.
42. Wood, E.J. Ferguson (1965) Marine Microbial Ecology, p. 131, Van Nostrand Reinhold, New York.
43. Bostrum, K. (1967) in Equilibrium Concepts in Natural Water Systems, (W. Stumm, ed.), Advances in Chemistry Series, No. 67, American Chemical Society, Washington, D.C.
44. Dirasian, H.A. (1968), *Water Sewage Works,* **115,** 453.
45. Rechnitz, G.A. (1982), *Anal. Chem.,* **54,** 1194A–1200A; idem, (1981), *Science,* **214,** 287–291.
46. Freiser, H., Ed. (1978) lon Selective Electrodes in Analytical Chemistry, Plenum Press, New York.
47. Durst, R.A., Ed. (1969) lon Selective Electrodes, Nat. Bureau Stand. Special Publ. 314, U.S. Government Printing Office, Washington, D.C.
48. Eisenman, Ed, G. (1967) Glass Electrodes for Hydrogen and Other Cations, Marcel Dekker, New York.
49. Rechnitz, G.A. (1967) *Chem. Eng. News,* June 12, 149.
50. Bard, A.J. and L.R. Faulkner (1980), Electrochemical Methods, John Wiley & Sons, New York.
51. Orion Research (1982), Handbook of Electrode Technology, Cambridge, Mass.
52. Karube, I. (1986), Microbial Sensors for Process and Environmental Control in Fundamentals and Applications of Chemical Sensors, D. Schuetzle & R. Hammerle, Eds., Chap. 20, ACS Symposium Series 309, American Chemical Society, Washington, D.C.
53. Arnold, M.A. (1986), *Ion-Sel. Electrode Rev.,* **8(1),** 85–113.
54. Sayers, W.T. (1971), *Environ. Sci. Technol.,* **5,** 114.
55. McNelis P.J. (1973), Monitoring of Water Systems in Water and Water Pollution Handbook, Vol. 4, L.L. Ciaccio, Ed., Chap. 26, Marcel Dekker Inc., New York.
56. Universal Sensors, Inc., Metairie, Louisiana.
57. Hikuma, M.T. Yasuda, I. Karube, and S. Suzuki (1981), *Ann. New York Acad. Sci.,* **369,** 307.
58. Hilcuma, M., H. Obana, T. Yasuda, I. Karube and S. Suzuki (1981), *Anal. Chim. Acta,* **116,** 61.
59. Karube, I., S. Mitsuda, and S. Suzuki (1979), *Europ. J. Appl. Microbiol. Biotechnol.,* **7,** 343.
60. Hikuma, M., T. Kubo, T. Yasuda, I. Karube and S. Suzuki (1979), *Biotechnol. Bioeng.,* **21,** 1845.
61. Hikuma, M., T. Kubo, T. Yasuda, I. Karube and S. Suzuki (1979), *Anal. Chim. Acta,* **109,** 39.
62. Matsunanga T., I. Karube, and S. Suzuki (1980), *Europ. J. Appl. Microbiol. Biotechnol.,* **10,** 235.
63. Karube I., T. Okada, and S. Suzuki (1981), *Anal. Chem.,* **53,** 1852.
64. Pkada T., I. Krube, and S. Suzuki (1982), *Anal. Chim. Acta,* 135, 159.
65. Hikuma, M., T. Kubo, T. Yasuda, I. Karube, and S. Suzuki (1980), *Anal. Chem.,* **52,** 1020.
66. Okada, T., I. Karube, and S. Suzuki (1981), *Europ. J. Appl. Microbiol. Biotechnol.,* **12,** 102.
67. Karube, I., T. Okada and S. Suzuki (1982), *Anal. Chim. Acta,* **135,** 61.
68. Karube, I., T. Okada and S. Suzuki (1982), *Europ. J. Appl. Microbiol. Biotechnol.,* **15,** 127.
69. Okada, T., I. Karube and S. Suzuki (1983), *Biotechnol. Bioengineering,* **25**(6), 1641–51.
70. Karube, I., T. Matsunanga and S. Suzuki (1977), *J. Solid Phase Biochem.,* **2,** 97.
71. Hikuma, M., H. Suzuki, T. Yasuda, I. Karube, and S. Suzuki (1979), *Europ. J. Appl. Microbiol. Biotechnol.,* **8,** 289.
72. Karube, I., T. Matsunanga, T. Nakahara, and S. Suzuki (1981), *Anal. Chem.,* **53,** 1024.
73. Karube, I., T. Nakahara, T. Matsunanga, and S. Suzuki (1982) *Anal. Chem.,* **54,** 1725.
74. Karube, I., (1984), *Trends in Analytical Chemistry,* **3,** 40.
75. Delahay, P. (1954), New Instrumental Methods in Electrochemistry, Interscience, New York.
76. Kolthoff, I.M. and Lingane, J.J. (1952), Polarography, vol. I & II, Interscience, New York.
77. Meites, L. (1958), Polarographic Techniques, 2nd. Ed., Wiley-Interscience, New York.
78. Bond, A. (1980) Modern Polarographic Methods in Analytical Chemistry, Marcel Dekker, New York.
79. Burge, D.E. (1970), *J. Chem. Educ.,* **47,** A81.
80. Southampton Electrochemistry Group (1985), Instrumental Methods in Electrochemistry, John Wiley & Sons, New York.
81. Adams, R.N. (1969), Electrochemistry at Solid Electrodes, Marcel Dekker, New York.
82. Wang, J. (1985), Stripping Analysis, VCH Publishers, Inc., Deerfield Beech, Florida.

83. Trace metals analyzers, Lachat Instruments, Mequon, Wi.
84. Clark, L.C. (1953), *J. Applied Physiol.,* **6,** 189; Clark, L.C., R. Wold, D. Granger, and Z. Taylor (1953), ibid., **6,** 186.
85. Beckman Instruments Inc., Model 956 Process Oxygen Analyzer, Bull. 0-4044, Fullerton, California.
86. Analytical Chemistry, LabGuide, 1989–90, American Chemical Society, Washington, D.C.
87. Beckman Instruments Inc., Model 778 Process Oxygen Analyzer, Bull. 0-4041, Fullerton, California.
88. Mancy, K.H., D.A., Okun, and C.N. Reilley (1962), *J. Electroanal. Chem.,* **4,** 65.
89. Precision Scientific, The Galvanic Cell Oxygen Analyzer, Bull. 638B, Chicago, Il.
90. Neville, J. (1962), *Rev. Sci. Instr.,* **33,** 51.
91. Johnson, L.F., J.R. Neville, R.W. Bancroft, and T.H. Allen (1963), NASA Doc. N63- 22411.
92. Carritt, D. and J. Kanwisher (1959), *Anal. Chem.,* **31,** 5.
93. Briggs, R. and M. Viney (1964), *J. Sci. Instr.,* **41,** 78.
94. Sergeant, E.P. (1984), Potentiometry and Potentiometric Titrations, Wiley, New York.
95. Ross, J.W. (1966), Conductometric Titrations, in Standard Methods of Chemical Analysis, 6th ed, F.J. Welcher, Ed., Vol. 3, Part A, Van Nostrand Reinhold, New York.
96. Loveland, J.W. (1963), in Treatise on Analytical Chemistry, 2nd ed., I.M. Kolthoff and P.J. Elving, Eds., Part I, vol. 4, Chap. 51, Wiley, New York.
97. Babcock, R.H. and K.F. Knowlton (1956), *J. Am. Water Works Assoc.,* **48,** 622.
98. Bishop, E. (1975), in Comprehensive Analytical Chemistry, C.L. Wilson and D.W. Wilson, Eds., vol. 11D, Elsevier, New York; J.T. Stock (1984), *Anal. Chem.,* 56, 1R; Lingane, J.J. (1958), Electroanalytical Chemistry, 2nd ed., Chap. 19–21, Wiley, New York.
99. Lingane, J.J. and S.L. Jones (1951), *Anal. Chem.,* **23,** 1798.
100. Skoog, D.A. and D.M. West (1982), Fundamentals of Analytical Chemistry, 4th edition, Saunders College Publishing, New York.
101. Kelley, M.T., H.C. Jones and D.J. Fisher (1959), *Anal. Chem,* **31,** 489.
102. Dunlap, L.B. and W.D. Shults (1962), *Anal. Chem.,* **34,** 499.
103. Shults, W.D. (1961), *Anal. Chem.,* **15,** 33.
104. DeFord, D.D., J.N. Pitts and C.J. Johns (1951), *Anal. Chem.,* **23,** 938.
105. Pitts, J.N., D.D. DeFord, T.W. Martin and E.A. Schmall (1954), *Anal. Chem.,* **26,** 628.
106. Friedlander, G., J.W. Kennedy, E.S. Macis and J.M. Miller (1981) Nuclear and Radiochemistry, 3rd edition, John Wiley & Sons, New York.
107. Birks, J.B. (1964), The Theory and Practice of Scintillation Counting, Pergamon Press, New York.
108. Kruger, P. (1971) Principles of Activatino Analysis, John Wiley & Sons, New York.

LEONARD L. CIACCIO
Ramapo College

L

LEGAL ASPECTS OF THE ENVIRONMENT

The environment has become a concern of the law and Environmental Law is now a recognized and established legal discipline.

The body of Environmental Law is growing in response to the felt needs of society which is beginning to acknowledge that the technological advances of this generation are destroying the legacy of the past and the capital assets of the future.

Historic forces forged common law doctrines suited to a world of endless frontiers, where ruthless exploitation of natural resources was a manifestation of the desperate need to subdue all nature as the means of survival in the wilderness of a new world. Such survival techniques are no longer appropriate on an earth whose life support systems are as fragile as those in a space capsule.

As natural resources are being utilized at rates unprecedented in the history of civilization, and the quality of environment deteriorates, there is increasing public demand for legal aid to the environment. Attorneys are being called upon to act as public defenders of the environment and the law is being asked to restore the quality of life.

Environment and the law interface in three significant areas: Legislation, Administration and Executive action, and Judicial determination.

There already exists a substantial body of statutory law at the federal, state and local levels dealing with matters of environmental concern, and there has been a perceptible evolution of ecologically sophisticated, environmentally responsible, socially relevant and politically feasible legislation at all operative levels of government.

Administrative agencies characterized by a combination of delegated legislative, executive and judicial powers, functions, and responsibilities affect the environment in the furtherance of their statutory mission.

The administrative agencies are legislative creations. In theory, they exist to effect policy established by the elected legislative representatives of the people. To accomplish this the legislature ceded rule-making power from its legislative mandate under the Constitution, the executive ceded a certain amount of administrative power, and the judiciary ceded certain judicial functions, in particular fact-finding and preliminary hearing. As a result of this tripartite grant of power, administrative agencies represent not a fourth branch of government as some seem to think, but the foundation of all practical government operations: Administrative agencies provide the substantial bulk of bureaucracy.

The judiciary of the United States and that of the several states furnish the forum for environmental litigation. Such litigation involves the interpretation of statutes and the adjudication of liabilities for damages resulting from nuisance, negligence, trespass and other traditional common law torts; as well as general actions seeking declaration of the rights of the people (in the form of class actions for declaratory judgment) and equitable relief (injunction, reparations) based on such ancient common law equitable principles as the Trust Doctrine and the maxim *sic utere to alienam non laedas*—so use your own property as not to injure that of another. Following the trail blazed by the American Labor Movement and retracing many of the judicial steps of the Civil Rights struggle, while seeking judicial vindication of a fundamental human right, to a salubrious environment, as one of those rights "so basic and important to our society that it would be inconceivable that it is not protected from unwarranted interference" and is a right retained by the people of the United States under the Ninth Amendment of the Constitution of the United States and protected from disparagement by the actions of the federal government by operation of the *due process* and *equal protection* clauses of the Fifth Amendment of the Constitution and protected from disparagement by the actions of any of the several states by operation of the *privileges* or *immunities, due process* and *equal protection* clauses of the Fourteenth Amendment of the Constitution.

Environmental Law and Environmental Litigation became recognized elements of the Anglo-American legal system in the Spring of 1966 when a suburban housewife brought an action on behalf of all the citizens of Suffolk County, New York seeking equitable relief from a toxic insult to the community ecosystem; challenging not merely the local mosquito control commission still routinely using DDT in an increasingly futile attempt to control a mosquito population that had long since become resistant to that pesticide, but the broad spectrum, persistent chemical biocide, 1,1,1-trichloro-2, 2-bis parachlorophenyl) ethane: DDT itself.

The New York State Supreme Court issued a temporary injunction restraining the County of Suffolk from using DDT for mosquito control on August 15, 1966 and continued this "temporary" injunction until December 6, 1967, finally holding that:

> DDT has, by its inherent chemical stability, become a continuing factor in some ecological life cycles so as to profoundly alter them and the environmental equilibrium. Thus, it is reasonably apparent that DDT is capable of and actually has to some extent caused extraordinary damage to the resources of this country. If in no other way, the chemical by its very stability has introduced an element of instability in the general ecosystem. For instance, by reducing a food source of some of the larger wildlife and so reducing the overall large wildlife population, lesser elements multiply more quickly. These lower forms are presumably more of a nuisance, assuming they in turn survive. Furthermore, DDT affects wildlife directly. Its ingestion, from whatever source has the capability, it seems, to disrupt reproductive processes or even more simply act as a poison. It is fairly apparent then that the application of DDT in Suffolk County has and is continuing to have a demonstrable effect on local wildlife, reducing it slowly but surely, either directly across the board or indirectly from the top down, but reducing it nevertheless.
>
> We have a situation where plaintiff has at least minimally sustained a massive effort to validiate the allegation that DDT does in fact do biological harm [*Yannacone v. Dennison et al.* (1967) 55 Misc 2d 468, 471–472, 285 NYS 2d 476].

ENVIRONMENTAL LEGISLATION

National Environmental Policy Act of 1969

The principal national legislative statement on the environment is the National Environmental Policy Act of 1969 (National Environment Policy Act) the purpose of which is:

> To declare a national harmony which will encourage productive and enjoyable harmony between man and his environment; to promote efforts which will prevent or eliminate damage to the environment and biosphere and stimulate the health and welfare of man; to enrich the understanding of the ecological systems and natural resources important to the Nation; and to establish a Council on Environmental Quality.

The National Environmental Policy Act (NEPA) is a synthesis by a Conference Committee of bills introduced in the Senate by Senator Jackson and in the House by Congressman Dingell.

The declaration of a national environmental policy contained in section 101 of the act speaks for itself.

> Sec. 101. (a) The Congress, recognizing the profound impact of man's activity on the interrelations of all components of the natural environment, particularly the profound influences of population growth, high-density urbanization, industrial expansion, resource exploitation, and new and expanding technological advances and recognizing further the critical importance of restoring and maintaining environmental quality to the overall welfare and development of man, declares that it is the continuing policy of the Federal Government, in cooperation with State and local governments, and other concerned public and private organizations to use all practicable means and measures, including financial and technical assistance, in a manner calculated to foster and promote the general welfare, to create and maintain conditions under which man and nature can exist in productive harmony, and fulfill the social, economic, and other requirements of present and future generations of Americans.
>
> (b) In order to carry out the policy set forth in this Act, it is the continuing responsibility of the Federal Government to use all practicable means, consistent with other essential considerations of national policy, to improve and coordinate Federal plans, functions, programs, and resources to the end that the Nation may
>
> 1) fulfill the responsibilities of each generation as trustee of the environment for succeeding generations:
> 2) assure for all Americans safe, healthful, productive, and esthetically and culturally pleasing surroundings:
> 3) attain the widest range of beneficial uses of the environment without degradation, risk to health or safety, or other undesirable and unintended consequences:
> 4) preserve important historic, cultural, and natural aspects of our natural heritage, and maintain, wherever possible, an environment which supports diversity and variety of individual choice;
> 5) achieve a balance between population and resource use which will permit high standards of living and a wide sharing of life's amenities; and
> 6) enhance the quality of renewable resources and approach the maximum attainable recycling of depletable resources.
>
> (c) The Congress recognizes that each person should enjoy a healthful environment and that each person has a responsibility to contribute to the preservation and enhancement of the environment.

Sec. 101 (b) (95) is recognition by Congress that uncontrolled magnitude and distribution of population underlies many of this nation's environmental and resource problems. To insure that a high standard of living is made available to all citizens and that all citizens have the opportunity for aesthetic enjoyment from a quality environment, the intent of NEPA is that the Federal Government must strive to maintain the magnitude and distribution capacity to provide such benefits. By section 101 (b) (5) Congress has thus implied that the rights of underprivileged citizens to a high standard of living are not to be sacrificed as a result of the national environmental policy to protect and environment.

The provision of NEPA most debated by the members of Congress was section 102.

Sec. 102. The Congress authorizes and directs that, to the fullest extent possible; (1) the policies, regulations, and public laws of the United States shall be interpreted and administered in accordance with the policies set forth in this Act, and (2) all agencies of the Federal Government shall—

A) utilize a systematic, interdisciplinary approach which will insure the integrated use of the natural and social sciences and the environmental design arts in planning and in decision making which may have an impact on man's environment;

B) identify and develop methods and procedures, in consultation with the Council on Environmental Quality established by title II of this Act, which will insure that presently unquantified environmental amenities and values may be given appropriate consideration in decision making along with economic and technical considerations;

C) include in every recommendation or report on proposals for legislation and other major Federal actions significantly affecting the quality of the human environment, a detailed statement by the responsible official on—

    i) the environmental impact of the proposed action,

    ii) any adverse environmental effects which cannot be avoided should the proposal be implemented,

    iii) alternatives to the proposed action,

    iv) the relationship between local short-term uses of man's environment and the maintenance and enhancement of long-term productivity, and

    v) any reversible and irretrievable commitments of resources which would be involved in the proposed action should it be implemented.

    Prior to making any detailed statement, the responsible Federal official shall consult with and obtain the comments of any Federal agency which has jurisdiction by law or special expertise with respect to any environmental impact involved. Copies of such statement and the comments and views of the appropriate Federal, State, and local agencies, which are authorized to develop and enforce environmental standards, shall be made available to the President, the Council on Environmental Quality and to the public as provided by section 552 of title 5, United States Code, and shall accompany the proposal through the existing agency review processes;

D) study, develop, and describe appropriate alternatives to recommend courses of action in any proposal which involves unresolved conflicts concerning alternative uses of available resources;

E) recognize the worldwide and long-range character of environmental problems and, where consistent with the foreign policy of the United States, lend appropriate support to initiatives, resolutions, and programs, designed to maximize international cooperation in anticipating and preventing a decline in the quality of mankind's world environment;

F) make available to States, counties, municipalities, institutions, and individuals, advice and information useful in restoring, maintaining, and enhancing the quality of the environment;

G) initiate and utilize ecological information in the planning and development of resource-oriented projects; and

H) assist the Council on Environmental Quality established by title II of this Act.

NEPA strengthened the right of public access under the Freedom of Information to records and information of Federal agencies Act. In addition, Congress has required all Federal agencies, to the fullest extent possible, to "make available to States, Counties, municipalities, institutions and individuals, advice and information useful in rest and maintaining, and enhancing the quality of the environment . . ."

These provisions enable citizens, under the Freedom of Information Act, to gain access to those records of Federal agencies that are relevant to environmental protection in addition to the environmental impact statements required to be filed under Section 102.

NEPA requires all Federal agencies and officials to consider environmental values in reaching decisions or in planning agency action.

Though Congress has enacted numerous laws over the past years that constitute Congressional mandates on various aspects on environmental policy, areas of Federal policy and action exist which have no environmental goals or policies and in which the conflicting operational necessities of different agencies complicate and often frustrate attainment of environmental quality objectives which are in the interest of the entire country. Many of the older operating agencies of the Federal Government, for example, do not, at present, have a mandate within the body of their enabling laws permitting them to give adequate attention to environmental values. In other agencies, especially when the expenditure of funds are involved, the latitude of a public official to deviate from the most economical alternative to reach an environmental goal may be strictly circumscribed by Congressional authorizations which have overlooked existing or potential environmental problems. There is also reason for serious concern over the activities of those agencies which do not feel they have sufficient authority to undertake needed research and action to enhance, preserve, and maintain the quality of the environment in connection with development activities.

The National Environmental Policy Act reemphasizes the importance of existing statutory programs relating to the environment. Prior to the passage of the National Environmental Policy Act of 1969, existing legislation involving certain areas of Federal activity did not provide "clear authority for the consideration of environmental factors which conflict with other objectives," and many Federal agencies had not given substantial and consistent consideration to environmental factors during decision making in certain areas of their responsibility.

Section 102 (1) remedies these shortcomings in the statutory foundations of existing agency programs by incorporating the policy and goals set forth in section 101 into the actions and programs of all Federal agencies. Section 102 (2) establishes procedures which will help to insure that the policies enunciated in sections 101 are implemented.

Prior to the passage of the National Environmental Policy Act of 1969, planning and decision-making that might have an effect on the quality of the environment was too often the exclusive province of the engineer and cost analyst. The National Environmental Policy Act requires Federal agencies to consider all relevant points of view and draw upon the broadest possible range of social and natural scientific knowledge and design arts in planning and decision making. In the past, Federal agencies have all too frequently ignored environmental factors in planning and decision making or omitted them from consideration during the early stages of planning because of the difficulty in evaluating such factors as compared with the ease of evaluating economic and technological factors. Under NEPA, Federal agencies and officials are now required to develop the methodology and techniques necessary to determine the total environmental impact and full course of actions by the Federal government.

One of the most frequent sources of environment litigation involves the construction and application of the National Environmental Policy Act of 1969 and the sufficiency of environmental impact statements prepared pursuant to section 102 (2) (C). Although the initial determination of environmental impact is made by the agency itself, should the agency find that the proposed activity will have a significant effect on the environment, the agency report of recommendation supporting the proposal must make findings with respect to the environmental factors set forth in section 102 (2) (C) (i)–(v). If adverse environmental effects cannot be avoided by reasonable alternative actions the agency must demonstrate that such adverse environmental effects are justified by other considerations of national policy, which must be stated in detail. The use of local, short-term resources must be found to be consistent with the maintenance and enhancement of the long-term productivity of the environment. Proposals involving significant commitments of resources that will be irreversible and irretrievable under conditions of known technology and reasonable economics must be found necessary.

The National Environmental Policy Act supplements the statutory authority of most Federal agencies and requires them to follow the policies, goals and procedures of the Act, unless the existing law governing such agency's operations expressly prohibits full compliance, however section 103 provides that no Federal agency is to interpret its existing statutory authority in such a way as to avoid compliance with the directives of NEPA. If the agency finds a clear conflict between its existing statutory mandate and the Act, the agency is required to propose to the President of the United States such measures as will be necessary to bring the statutory authority and mandate of the agency into conformity with the provisions of the National Environmental Policy Act.

The National Environmental Policy Act is a relevant statute within the meaning of the Administrative Procedure Act, thus conferring standing upon representatives of the general public seeking judicial review of actions and decisions by federal agencies which may have failed to follow the policies, goals and procedures of the Act, where such action or decision of the agency threatens to have adverse effects upon the ecological system an individual may reside in or use for recreation. Section 101 (C) of the National Environmental Policy Act recognizes that "each person should enjoy a healthful environment and that each person has a responsibility to contribute to the preservation and the enhancement of the environment," indicating that Congress intends to recognize the interests of individual citizens in the protection of the ecosystem they may reside in or use for recreation.

Such an individual citizen is within the class of persons which the National Environmental Policy Act was meant to protect, and is thus a person adversely affected or aggrieved within the meaning of the Administrative Procedure Act and as such a person is entitled to judicial review of the agency action.

The national goal and policies for the protections, establishment and enhancement of this Nation's environment established by the National Environmental Policy Act are in the words of its Senate sponsor, Senator Jackson.

More than the statement of what we believe as a people and as a nation, it establishes priorities and gives expressions to our National goals and aspirations. It provides a statutory foundation to which administrators may refer...for guidance in making decisions which find environmental values in conflict with other values.

What is involved is a congressional declaration, that we do not intend, as a government or as a people, to initiate actions which endanger the continued existence or the health of mankind: that we will not intentionally initiate actions which will do irreparable damage to the air, land, and water which support life on earth.

An environmental policy is a policy for people. Its primary concern is with man and his future. The basic principle of the policy is that we must strive in all that we do to achieve a standard of excellence in man's relationship to his physical surroundings. If there are to be departures from this standard of excellence, they should be exceptions to the rule and policy.

Administrative Procedure Act

The Administrative Procedure Act provides both an independent source of federal jurisdiction and a grant of standing to conservation organizations and aggrieved persons seeking to challenge decisions of, or actions by, Federal agencies which may result in environmental degradation.

Senator McCarren, the author of the bill which ultimately became the Administrative Procedure Act, explained that it conferred no administrative powers, but provided definitions of, and limitations upon, administrative action, to be interpreted and applied by the agencies in the first instance, but to be reviewed by the courts in the final analysis.

The scope of judicial review of agency action is governed by the language of the Administrative Procedure Act:

To the extent necessary to decision, and when presented the reviewing court shall decide all relevant questions of law, interpret constitutional and statutory provisions, and determine the meaning or applicability of the terms of an agency action. The reviewing court shall . . .

2) hold unlawful and set aside agency action, findings and conclusions found to be—
    A) Arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with the law;
    B) Contrary to constitutional right, power, privilege or immunity;
    C) In excess of statutory jurisdiction, authority, or limitations, or short of statutory rights; . . .
    F) Unwarranted by the facts to the extent that the facts are subject to trial *de novo* by the reviewing court . . . (5 United States Code § 706).

Under the Administrative Procedure Act any person "suffering legal wrong because of agency action or adversely affected or aggrieved by agency action within the meaning of a relevant statute" may seek judicial review.

Agency action made reviewable by statute and final agency action for which there is no other adequate remedy in a court are subject to judicial review. A preliminary, procedural, or intermediate agency action of ruling not directly reviewable is subject to review on the review of the final agency action. Except as otherwise expressly required by statute, agency action otherwise final is final for the purposes of [the Administrative Procedure Act] whether or not there has been presented or determined an application for a declaratory order, for any form of reconsideration, or, unless the agency otherwise requires by rule and provides that the action meanwhile is inoperative, for an appeal to a superior agency or authority. (5 United States Code § 104).

The "legal wrong" contemplated by the Administrative Procedure Act is something more than the mere personal damage normally contemplated in the law of torts. The damage necessary to constitute legal wrong under the provisions of the Administrative Procedure Act must produce some illegal effect and be demonstrably contrary to law in either substance or procedure. However, the law considered relevant upon judicial review of agency action under the Administrative Procedure Act is not only the statutory law directly involved with the agency action, but general constitutional law as well. The legal wrong necessary to confer jurisdiction on the court and standing to the party complaining can be the invasion of any legally protected right.

Jurisdiction can also be obtained over agents and officers of the United States Government such as contractors of the Atomic Energy Commission, the Secretary of Agriculture, the Secretary of Interior, and divisions and departments of Federal agencies such as the Division of Wildlife Services, the Bureau of Sport Fisheries and Wildlife, and the Bureau of Land Management of the United States Department of the Interior, the Corps of Engineers of the United States Army of the Department of Defense, and the Federal Aviation Administration and the Bureau of Public Roads of the Department of Transportation.

There is a basic presumption in the law favoring judicial review of agency or administrative action even where the statute prescribing agency action does not on its face provide for judicial review. For the right of judicial review to be denied, the agency must produce clear and convincing evidence of legislative intent to foreclose judicial review. The Supreme Court of the United States [*Association of Data Processing Service Organizations, Inc. v. Camp* (1970) 397 US 150, 25 h Edzd 184, 90 SCt 827] has reemphasized the presumption of judicial review of agency action in recent litigation which construed the Administrative Procedure Act not grudgingly but as serving a broadly remedial purpose holding that "the mere failure to provide specially by statute tor judicial review is certainly no evidence of intent to withhold review."

No presumption in favor of "administrative absolutism" and against judicial review exists. A congressional purpose to preclude judicial review of agency action must be "fairly discernable in the statutory scheme," since the right to judicial review is ordinarily inferred where congressional intent to protect the interests of the class of which the complainant is a member can be found.

Administrative action is not immune from judicial review because it may be committed to agency discretion or simply because the statute authorizes a public official to "prescribe such regulations as he may deem proper to carry out the provisions of [the legislation]."

The right to judicial review of agency action is all that stands between the citizen and government by administrative fiat not subject to correction by the elective process. Although a court usually will not substitute its judgement for that of an administrative agency, it can compel the proper exercise of statutory authority by the agency, since it is not for the administrative agency to finally determine the limits of its own power. That is a judicial function.

Where a party aggrieved has no other adequate remedy following administrative action there is certainly the right, upon proper application, to judicial review of agency action. Some confusion existed in the Federal Courts for a number of years after the passage of the Administrative Procedure Act because the provisions for judicial review apply, "According to the provisions thereof, except to the extent that agency action is committed to agency discretion by law" (5 United States Code § 701).

Some Federal Courts held that agency action involving the exercise of discretion could not be reviewed for any reason under the Administrative Procedure Act, however, the majority of Federal Courts have adopted the rule that any agency action, even that involving the exercise of discretion and dependent upon the special technical and scientific expertise of the agency, is subject to judicial review to the extent provided by the Administrative Procedure Act. The mere fact that the acts of some administrative official, even a cabinet-level Secretary, require the exercise of discretion and judgment does not preclude judicial review of official action. Any contrary decision would lead to establishment of a shadow government by administrative officials not subject to popular election or judicial review and therefore totally outside the systems of checks and balances so carefully developed during the Constitutional Convention. The rights and the public must receive active and affirmative protection at the hands of the administrative agencies.

Much environmental litigation has involved the standing of parties to bring suit against administrative agencies, their officials, employees and agents. Recent decisions have established the right of public benefit organizations whose purposes include protection of the environment, or protection the lives, health, and property of human beings or animals, as well as individuals seeking to proceed in the public interest as "private attorneys general," to bring actions to persons "adversely affected" or "parties aggrieved" without the meaning of the Administrative Procedure Act.

Neither economic injury nor a specific individual legal right are necessary adjuncts to standing. A complainant need only demonstrate that it is an appropriate person (corporate or human) to question the alleged failure of an agency to protect those values recognized by law as in the public interest.

Environmental Protection Agency

By virtue of the executive authority to reorganize Federal agencies, President Nixon created an Environmental Protection Agency to which was transferred, effective December 2, 1970, significant environmental responsibilities formerly scattered throughout the several executive and administrative branches of government.

> Sec. 2 *Transfers to Environmental Protection Agency.* (a) There are hereby transferred to the Administrator [of the Environmental Protection Agency].
>
> 1) All functions vested by law in the Secretary of the Interior and the Department of the Interior which are administered through the Federal Water Quality Administration, all functions which were transferred to the Secretary of the Interior by Reorganization Plan No. 2 of 1966 (80 Stat. 1608), and all functions vested in the Secretary of the Interior or the Department of the Interior by the Federal Water Pollution Control Act or by provisions of law amendatory or supplementary thereof.
> 2) (i) The functions vested in the Secretary of the Interior by . . . 16 USC 742d-1 (being an Act relating to studies on the effects of insecticides herbicides, fungicides, and pesticides upon the fish and wildlife resources of the United States), and (ii) the functions vested by law in the Secretary of the Interior and the Department of the Interior which are administered by the Gulf Breeze Biological Laboratory of the Bureau of Commercial Fisheries at Gulf Breeze, Florida.
> 3) The functions vested by law in the Secretary of Health, Education, and Welfare or in the Department of Health, Education, and Welfare which are administered through the Environmental Health Service, including the functions exercised by the following components thereof:
>   i) The National Air Pollution Control Administration,
>   ii) The Environmental Control Administration;
>   A) Bureau of Solid Waste Management,
>   B) Bureau of Water Hygiene,
>   C) Bureau of Radiological Health,
>
> except that functions carried out by the following components of the Environmental Control Administration of the Environmental Health Service are not transferred: (i) Bureau of Community Environmental Management, (ii) Bureau of Occupational Safety and Health, and (iii) Bureau of Radiological Health, insofar as the functions carried out by the latter Bureau pertain to (A) regulation of radiation from consumer products, including electronic product radiation, (B) radiation as used in the healing arts, (C) occupational exposures to radiation, and (D) research, technical assistance, and training related to clauses (A), (B), and (C).
>
> 4) The functions vested in the Secretary of Health, Education, and Welfare of establishing tolerances for pesticide chemicals under the Federal Food, Drug and Cosmetic Art, as amended, 21 USC 346, 346a, and 348, together with authority, in connection with the functions transferred, (i) to monitor compliance with the tolerances and the effectiveness of surveillance and enforcement, and (ii) to provide technical assistance to the States and conduct research under the Federal Food, Drug, and Cosmetic Act, as amended and the Public Health Service Act, as amended.

5) So much of the functions of the Council on Environmental Quality under section 204(5) of the National Environmental Policy Act of 1969 . . . as pertains to ecological systems.

6) The functions of the Atomic Energy Commission under the Atomic Energy Act of 1954, as amended, administered through its Division of Radiation Protection Standards, to the extent that such functions of the Commission consist of establishing generally applicable environmental standards for the protection of the general environment from radioactive exposures or levels, or concentrations of quantities of radioactive material, in the general environment outside the boundaries of locations under the control of persons possessing or using radioactive material.

7) All functions of the Federal Radiation Council (42 USC 2021 (h)).

8) (i) The functions of the Secretary of Agriculture and the Department of Agriculture under the Federal Insecticide, Fungicide, and Rodenticide Act, as amended (7 USC 135–135k), (ii) the functions of the Secretary of Agriculture and the Department of Agriculture under section 408(1) of the Federal Food, Drug and Cosmetic Act, as amended (21 USC 346a(1)), and (iii) the functions vested by law in the Secretary of Agriculture and the Department of Agriculture which are administered through the Environmental Quality Branch of the Plant Protection Division of the Agricultural Research Service.

9) So much of the functions of the transfer of officers and agencies referred to in or affected by the foregoing provisions of this section as in incidental to for necessary of the performance by or under the Administrator of the functions transferred by those provisions or relates primarily to those functions. The transfers to the Administrator made by this section shall be deemed to include the transfer of (1) authority, provided by law, to prescribe regulations relating primarily to the transferred functions, and (2) the functions vested in the Secretary of the Interior and the Secretary of Health, Education, and Welfare by section 169(d)(1)(B) and (3) of the Internal Revenue Code of 1954 (as enacted by section 704 of the Tax Reform Act of 1969, . . .); but shall be deemed to exclude the transfer of the functions of the Bureau of Reclamation under section 3(b)(1) of the Water Pollution Control Act (33 USC 446a(b)(1)).

b) There are hereby transferred to the Agency:

1) From the Department of the Interior, (i) the Water Pollution Control Advisory Board (33 USC 466t), together with its functions, and (ii) the hearing boards provided for in sections 10(c)(4) and 10(f) of the Federal Water Pollution Control Act, as amended (33 USC 446g(c)(4); 446g(f)). The functions of the Secretary of the Interior with respect to being or designating the Chairman of the Water Pollution Control Advisory Board are hereby transferred to the Administrator.

2) From the Department of Health, Education, and Welfare, the Air Quality Advisory Board (42 USC 1857c), together with its functions. The functions of the Secretary of Health, Education and Welfare with respect to being a member of the Chairman of that Board are hereby transferred to the Administrator.

Sec 6. *Abolitions* (a) Subject to the provisions of this reorganization plan, the following, exclusive of any functions, are hereby abolished:

1) The Federal Water Quality Administration in the Department of the Interior (33 USC 446–1).

2) The Federal Radiation Council (. . ., 42 USC 2021(h)).

Among the procedural landmark established in the Project Rulison litigation was the right of the plaintiff COSCC to take the depositions of experts prior to the determination of the defendants motion for summary judgment to dismiss the complaint. COSCC argued that the motion for summary judgment could not be decided without considering the relevant testimony of certain experts under the control of the defendants and the Court directed the defendants to produce those experts for pre-trial oral examination. The information obtained in those pre-trial examinations established the need for a full hearing on the merits as a condition precedent to determination of any AEC motion for summary judgment dismissing the complaint.

Sovereign Immunity and Popular Sovereignty

Throughout the relatively short history of environmental litigation involving agencies of government, sovereign immunity has been raised as a defense by the government agency asserting that the action is a suit against the sovereign United States of America brought without the consent of congress and that the relief sought would invade the powers of officers of the executive branch of the Federal Government to whom the actions complained of had been delegated. This defense is usually buttressed by an appeal for judicial restraint under the doctrine of Separation of Powers.

The general doctrine of the immunity of the United States from suit without consent of Congress is a rule propounded in a diction by United States Supreme Court Chief Justice John Marshall in the case of *Cohens v. Virginia* [19 US (6 Wheat)

264, 411–412 (1821)]. However, in the earliest discussion of sovereignty by the United States Supreme Court [*Chisholm v. Georgia* (1973) 2 US 419], the court stated,

> . . . the term sovereign has for its correlative, subject. In this sense, the term can receive no application; for it has no object in the Constitution of the United States. Under the Constitution, there are citizens but no subjects.
>
> . . . the people of the United States . . . have reserved the supreme power in their own hands; and on that supreme power have made these [governments] dependent, instead of being sovereign . . .
>
> Even in almost every nation which has been denominated free the state has assumed a supercilious preeminence above the people who have formed it; hence the haughty notions of State independence, State sovereignty, and State supremacy. In despotic governments, the government has usurped in a similar manner both upon the state and the people: hence all the arbitrary doctrines and pretensions concerning the supreme, absolute and incontrollable, power of government. In each, man is degraded from the prime rank which he ought to hold in human affairs . . .

Another instance, equally strong, but still more astonishing is drawn from the British Government as described by Sir William Blackstone and his followers. As described by him and them, the British is a despotic Government. It is a Government without a people. In that Government, as so described, the sovereignty is possessed by the Parliament: In the Parliament, therefore, the supreme and absolute authority is vested: In the Parliament resides that incontrollable and despotic power, which, in all Governments, must reside somewhere. The constituent parts of the Parliament are the King's Majesty, the Lord's Spiritual, the Lord's Temporal, and the Commons. The King and these three Estates together form the great corporation of body politic of the Kingdom. All these sentiments are found; the last expressions are found verbatim, in the *Commentaries Upon the Laws of England.* The Parliament forms the great body politic of England. What then, or where, are the people? Nothing! No where! They are not so much as even the "baseless fabric of a vision!" From legal contemplation they totally disappear! Am I not warranted in saying, that, if this is a just description; a Government, so and justly so described, is a despotic Government?

Chief Justice Jay also recognized the people of the United States, not the Federal Government, as the sovereign:

> From the crown of Great Britain, the sovereignty of their country passed to the people of it . . . [T]he people, in their collective and national capacity, established the present Constitution. It is remarkable that in establishing it, the people exercised their own visits, and their own proper sovereignty, and conscious of the plenitude of it, they declared with becoming dignity. "We the people of the United States, do ordain and establish this Constitution." Here we see the people acting as sovereigns of the whole country . . . [T]he sovereignty of the nation as [in] the people of the nation . . .
>
> Chief Justice Jay noted that in England the doctrine of sovereignty was based on feudal principles that considered the prince as sovereign and the people as his subjects. These feudal principals contemplated the sovereign.
>
> As being the fountain of honor and authority; and from his grace and grant derives all franchises, immunities and privileges; it is easy to perceive that such a sovereign could not be amenable to a Court of Justice, or subjected to judicial control and actual constraint. It was of necessity, therefore, that suability, became incompatible with such sovereignty. Besides, the Prince having all the Executive powers, the judgment of the Courts would, in fact, be only monitory, not mandatory to him, and a capacity to be advised, is a distinct thing from a capacity to be sued. The same feudal ideas run through all their jurisprudence, and constantly remind us of the distinction between the Prince and the subject. No such ideas obtain here. At the Revolution, the sovereignty devolved on the people and they are truly the sovereigns of the country, but they are sovereigns without subjects . . . and have none to govern but themselves, the citizens of America are equal as fellow citizens, and as joint tenants in the sovereignty.
>
> . . . Sovereignty is the right to government; a nation or State-sovereign is the person or persons in whom that resides. In Europe the sovereignty is generally ascribed to the Prince; here it resides with the people . . . Their Princes have personal powers, dignities, and pre-eminences, our rulers have none but official; nor do they partake in the sovereignty otherwise, or in any other capacity, than as private citizens.

A majority of the Supreme Court of the United States, just five years after the adoption of the Constitution of the United States, rejected the idea of the United States as sovereign—and necessarily the rule of sovereign immunity from suit because "suability became incompatible with such sovereignty." The people of the United States are the sovereign, not the United States government.

In 1882, the Supreme Court of the United States further said,

> Under our system the *people,* who are there [in England] called *subjects,* are the sovereign. Their rights, whether collective or individual, are not bound to give way to a sentiment of loyalty to the person of a monarch. The citizen here knows no person, however near to those in power, or however powerful himself, to whom he need yield the rights which the law secures to him when it is well administered. When he, in one of the courts of competent jurisdiction, has established his right to property, there is no reason why deference to any person, natural or artificial, not even the United States, should prevent him from using the means which the law gives him for the protection and enforcement of that right [*United States v. Lee* (1882), 106 US 196, 205 to 209].

Land Use and Resource Management

Throughout history, land and natural resources have been considered the property of the sovereign to be used, abused, given or taken at the sovereign whim. The history of civilization, to a certain extent, can be considered a chronicle of wars and revolutions fought over the ownership and utilization of land and natural resources.

The Constitution of the United States provides that the rights not explicitly given by the people of the United States to the federal government are retained by the people of the United States, individually, and as collectively assembled in the several states.

The enumeration of the Constitution, of certain rights, shall not be constructed to deny or disparage others retained by the people (*United States Constitution,* Ninth Amendment).

The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people. (*United States Constitution,* Tenth Amendment).

As the individual state constitutions were formulated, the rights of individual property owners were strengthened, but at no time did the sovereign people of the United States relinquish the ultimate right to determine the highest and best use of the land and natural resources subject to the jurisdiction of the United States. The justification for any legal restriction on the private use of land or natural resources is rooted in this concept of sovereignty in the people.

According to the report commissioned by the Council on Environmental Quality as a preparation for consideration of a National Land Use Policy Act,

> This country is in the midst of a revolution in the way we regulate the use of our land. It is a peaceful revolution . . . a quiet revolution . . .
>
> The *ancien regime* being overthrown is the feudal system under which the entire pattern of land development has been controlled by thousands of individual local governments, each seeking to maximize its tax base and minimize its social problems, and caring less what happens to all the others.

And if we are to live in harmony with that which has been given to us from preceding generations, and from the earth before mankind appeared, we must make certain assumptions with respect to every accessible parcel of real property and every natural resource capable of exploitation.

1) The area of resource is vulnerable.
2) Development, exploitation or utilization of some kind is inevitable.
3) The highest and best use of the land or resource as an element of human ecology must be accommodated.
4) Development must be consistent with natural ecological constraints.
5) Planned achievement of the highest and best use of land and resources is more profitable to mankind than unplanned development.
6) The police power of the state, the ultimate sovereignty of the people, and the maintenance of private property concepts are compatible and can result in the harmonious, beneficial development of land and resources.

Recognizing the limited availability of land itself and the place of land as the basic capital asset of civilization, land use historically has been limited by executive, legislative and judicial process.

An early example of land use legislation—zoning—has been upheld by the courts on the grounds that it represents action by an individual community to assure the highest and best use of limited resources for the greatest number of people without undue infringement upon the rights of individuals.

Zoning is not just an expansion of the common law of nuisance. It seeks to achieve much more than the removal of obnoxious gases and unsightly uses. Underlying the entire concept of zoning is the assumption that zoning can be a vital tool for maintaining a civilized form of existence only if we employ the insights and the learning of the philosopher, the city planner, the economist, the sociologist, the public health expert and also the other professions concerned with urban problems . . .

This fundamental conception of zoning has been present from its inception. The almost universal, statutory requirement that zoning conform to a "well-considered plan" or "comprehensive plan" is a reflection of that view. [See Standard State Zoning Enabling Act, US Dept of Commerce (1926).] The thought behind the requirement is that consideration must be given to needs of the community as a whole. In exercising their zoning powers, the local authorities must act for the benefit of the community as a whole following a common, deliberate consideration of the alternatives, and not because of the whims of either an articulate minority or even majority in the community . . . Rather, the comprehensive plan is the essence of zoning. Without it, there can be no rational allocation of land use. It is the insurance that the public welfare is being served and that zoning does not become nothing more than just a Gallup poll [*Udell v. Hass* (1968), 21 NY2d 463, NE2d 897, 288 NYS2d 888].

The key to successful land use and resource utilization legislation is the community determination of the highest and best use of land and resources in terms of intrinsic suitability and naturally imposed constraints. Of necessity this must be done by a team of individuals trained in the several disciplines necessary to define the environmental parameters of a Regional Ecological System. The community itself, particularly its people, constitute elements of that regional ecological system just as surely as topography, hydrology, and climate. One of the most significant contributions of environmental litigation to the development of Anglo-American jurisprudence has been the broadening of the concept of "expert witness" to include an individual member of a multidiscipline scientific team testifying about the work of the entire team even though the individual scientist may not have performed the work testified about personally. Environmental litigation, particularly that involving such environmental toxicants as DDT and radionuclides has led to legal acceptance of the "environmental scientist" who is permitted to testify generally, upon a showing, to the satisfaction of the Court, of demonstrated competence in some recognized scientific discipline together with regular collaboration with other scientists of diverse disciplines upon general problems of environmental concern.

The adequate determination of the highest and best use of the limited natural resources of any regional ecological system mandates a systems approach supported by modern computer technology in order to determine the relevant boundary conditions and elemental optimizations of the complex, non-linear dynamic relationships that describe a natural system as it actually exists. Courts increasingly are accepting evidence in environmental litigation, and the techniques of general systems analysis have become key factors in the resolution of environmental controversy such as that concerning the continued use of DDT, the distribution of radionuclides from peaceful applications of nuclear energy, and the Cross-Florida Barge Canal.

The acceptance of systems methods in ecology by the courts now means that any land use or resource utilization law which does not fully reflect the system characteristics of the Regional Ecological System in which it will operate is fatally defective in the legal sense. It cannot be sustained in the face of sophisticated legal challenge; while any such law which does, in fact, reflect the system characteristics of the region can be sustained in the public interest, even where such a law appears to limit the rights incident to private property ownership.

The current litigation involving the adequacy of statements prepared under the mandate of section 102 of the National Environmental Policy Act (environmental impact statements) demonstrates that any failure to consider the ecological integrity of a region and fully determine the interrelationships among each element of the land, landscape and natural processes diminishes the legal value of the effort; while a full evaluation of any environmentally significant action with respect to its effect upon the overall ecological integrity of the region and the interrelationships among each element of land, landscape, and natural processes of the region can form the basis for legal restraints upon land use and resource exploitation, even where such restraints appear to infringe upon private property rights or government agency prerogatives.

ENVIRONMENTAL LITIGATION

Environmental litigation includes almost all the actions brought before the courts of the United States, the several states, and the various administrative agencies exercising quasi-judicial powers which involve a determination which may affect natural processes, resources or the environment.

The Scenic Hudson Preservation Case

One of the landmark cases in environmental law is *Scenic Hudson Preservation Conference* v. *Federal Power Commission* [354 F2d. 608 (1965, CA2), *certiorari denied* 384 US 941, 16 LEd 2d 540, 86 SCt 1462], which established the right of conservation organizations and other public interest groups to challenge the extent of consideration given to historic, scenic, aesthetic, conservation and recreational aspects of power development by the Federal Power Commission. The holding of that case has been extended to permit similar organizations to challenge the consideration of such values by the Secretary of Transportation (in approving federal-aid highways) and the Secretary of the Army and Chief of Corps of Engineers (in licensing and granting permits for projects on navigable waterways).

The *Scenic Hudson* saga began on January 29, 1963 with Consolidated Edison of New York (the major electric power utility servicing the City of New York) filing an application with the Federal Power Commission (FPC) for a license to construct a pumped-storage hydroelectric facility on the west wide of the Hudson River at Storm King Mountain. As information concerning the project spread, the Federal Power Commission directed that a public hearing on the project be held February 25, 1964. The Scenic Hudson Preservation Conference was hastily formed and filed a petition on February 6, 1964 seeking to intervene in the proceeding. The petition was granted on February 14, 1964; intervention was permitted; hearings were held in Washington, DC; and the hearing examiner made an initial determination approving the license application on July 31, 1964 subject to review by the entire Federal Power Commission which eventually licensed the project on March 9, 1965.

The pumped-storage plant proposed by Consolidated Edison would generate electric power for use during peak load periods in New York City using hydroelectric units driven by water from a reservoir. The project consisted of three major components—a storage reservoir, a powerhouse, and transmission lines. The reservoir would be located over the power house which was to be located on the shores of the Hudson River at the foot of Storm King Mountain, a mountain of striking aspect rising over 1,360 feet above the river. The reservoir would be connected to the powerhouse by a tunnel 40 feet in diameter. During slack periods of demand for electrical power in New York City, Consolidated Edison's steam generating plants in New York City would provide electric power for the pumps at Storm King to force water through the tunnel and up the mountain into the reservoir. The water would be stored in the reservoir until periods of peak demand for electrical power in New York City at which time the water would be released from the reservoir to rush down the mountain through the tunnel to power the generators. The project was to have a capacity of two gigawatts (billion watts, or million kilowatts), but would be capable of expansion to three gigawatts. At that time, the Storm King Project was to be the world's largest pumped storage electric generating facility and was estimated to cost an estimated $162,000,000.

Subsequently the United States Court of Appeals for the Second Circuit reviewed the determination of the Federal Power Commission holding, that

> if the Commission is properly to discharge its duty in this regard [approve those projects best adapted to a comprehensive plan for improving or developing a waterway], the record on which it bases its determination must be complete. The petitioners and the public at large have a right to

> demand this completeness. It is our view, and we find, that the Commission has failed to compile a record which is sufficient to support its decision. The commission has ignored certain relevant factors and failed to make a thorough study of possible alternatives to the Storm King Project.

The court found that the Commission must take into consideration the unique beauty and important historical significant of the Hudson River noting that the Federal Power Act sought to "promote the comprehensive development of the Nation's water resources."

The court directed the Federal Power Commission to reexamine all the issues that the court found had not been properly considered.

The commission's renewed proceedings must include as a basic concern the preservation of natural beauty and of national and historic shrines, keeping in mind, that, in our affluent society, the cost of a project is only one of several factors to be considered. The record as it comes to us fails markedly to make out a case for the Storm King Project on, among other matters, cost, public convenience and necessity, and absence of reasonable alternatives. Of course the Commission should make every effort to expedite the new proceedings.

Following the hearings on remand, the hearing examiner made an initial determination on August 6, 1968 recommending that the license be issued. New York City then petitioned to intervene prior to a decision by the Commission in order to introduce evidence of the effects project construction activities might have upon the safety of the Catskill Aqueduct which supplies a substantial part of the water for New York City. The proceedings were reopened and on December 23, 1969 a supplemental decision was issued by the hearing examiner recommending that the licence be granted with some modifications.

On August 17, 1970, the Federal Power Commission issued the license for construction of a pumped-storage plant by Consolidated Edison at Storm King Mountain. In a lengthy opinion, the Commission paid close attention to the decision of the Second Circuit Court of Appeals, giving detailed consideration to areas in which the court had found that the previous decision had been deficient, including alternative sources for the needed power, alternative sties for the project, fish protective devices, underground transmission lines, and the scenic, aesthetic, and historical values that would be affected by the project. The Commission found the project to be compatible with the natural environment, and not to have an adverse effect upon the ecology of the area, the water quality of the Hudson River, or the fishery resources of the Hudson River. The area would be replanted after construction; the power house would be totally underground and the Commission found that the project, as modified and as licensed, represented no real impairment to the environment and scenic aspects of the highlands and the natural beauty of the area. The Commission recognized man's already awesome impact on this section of the Hudson River shoreline, by stating that,

> . . . it should not be inferred that previous and existing environmental detriment permits additional detriment which would otherwise not be permitted but for existing conditions, pointing out that Consolidated Edison would rehabilitate the waterfront area in the vicinity of the project site, an area which the Commission euphemistically described as "blighted."

Although the hearings had been conducted prior to the passage of the National Environmental Policy Act, the Commission found that full and careful consideration had been given to all the concerns which the Act embodies and that the licensing of the facility complied with the requirements of that Act.

All parties appealed to the Second Circuit Court of Appeals which affirmed the Commission's issuance of the license. The majority of the Court found that the decision of the Federal Power Commission was supported by substantial evidence and had been based on a consideration of all relevant factors, and had in fact complied with the requirements of the Federal Power Act and the National Environmental Policy Act.

After 8 years of litigation, at enormous cost to the Scenic Hudson Preservation Conference, its supporters, and Consolidated Edison, not to mention the inconvenience and cost to the electric power-hungry people of the City of New York, the project would be constructed, little modified from the initial proposal.

The first decision of the Court of Appeals in the Scenic Hudson Preservation case represented a crucial selection event in the evolution of Environmental Law. If the conservationists were willing to submit the controversy to the Federal Power Commission, then, according to the Court, the Federal Power Commission should hear and consider evidence on natural values and environmental impacts in addition to evidence on the economics of electric power generation and distribution. The alternative choice open to the conservationists was to remain before the court and challenge the inherent suitability of the Federal Power Commission as a body capable of resolving environmental conflict. The Scenic Hudson Preservation Conference, a coalition of preservationists and aesthetically concerned conservationists, chose to yield to the Federal Power Commission the authority to make an ecological judgment binding upon generations yet unborn, cloaking the Federal Power Commission with a mantle of ecological competence. Since the Scenic Hudson Preservation Conference chose to take its "day in court" before an administrative agency rather than a court of equity, the findings of fact made by the Commission were binding upon the conservationists in subsequent review by the Courts.

Fortunately, the principles set forth by the Court of Appeals in its initial review of the *Scenic Hudson* case mandating consideration of all relevant issues including environmental issues, and requiring affirmative action by the Agency to protect the public interest are not limited in application to the Federal Power Commission. Other agencies, including the Corps of Engineers of the United States Army, the Atomic Energy Commission, the Department of Transportation, the Department of the Interior and the

Department of Agriculture, must also consider all relevant issues, including environmental issues and take affirmative action to protect the public interest.

The Defense of Florissant

The Florissant fossils, located a short distance west of Colorado Springs, Colorado, are found in more than 6,000 acres of an ancient lake bed where seeds, leaves, plants and insects from the Oligocene period (34 million years ago) are remarkably preserved in paper-thin layers of volcanic shale which disintegrate when left exposed to weather unless properly protected. A number of bills had been introduced over the years in Congress to protect the Florissant fossil bed but they did not receive extensive consideration until the United States National Park Service promulgated a master plan detailing the paleontological and palynogical values of Florissant.

At the time the Florissant Fossil Beds National Monument Bill passed the Senate, Park Land Company, a Colorado Springs real estate group had already contracted to purchase 1,800 acres of the ancient lake bed. While the House of Representatives was deliberating its version of the National Monument bill, Park Land Company announced it would bulldoze a road through a portion of the proposed national monument to open the area for development and immediate sale to anyone interested in recreational housing. A group of Colorado conservationists met with the principles of the Park Land Company in an attempt to persuade them to withhold excavation in the area to be included within the Florissant Fossil Beds National Monument at least until the House of Representatives acted on the bill. This request was refused as was a similar request to confine development activities to the area lying outside the ancient lake bed. The only alternative offered the conservationists was the opportunity to purchase the land—for cash immediately—at a price considerably in excess of any appraised value based on recent land sales in the area.

Faced with the irreparable loss of a substantial portion of these unique and irreplaceable fossil beds, a small group of concerned citizens formed a non-profit, public benefit corporation called the Defenders of Florissant and commenced an action for declaratory judgment and an injunctive against the Park Land Company and all the other land owners and contract vendees in the area to be included within the proposed National Monument.

The United States District Court for the District of Colorado heard the Defenders of Florissant application for a temporary restraining order on July 9, 1969, and although the plaintiffs' proof that the proposed excavations for roads and culverts would result in the destruction of some of the most valuable fossil areas in the proposed national monument was uncontradicted and unchallenged the District Court held that there was nothing in the United States Constitution preventing the owners from using their property in any way not prohibited by law. The District Court denied the application for a temporary restraining order and a subsequent application for a stay pending appeal, but did, however, note the importance of preserving the fossil beds.

Following the District Court decision, representatives of the plaintiffs' held an informal conference in the Courtroom with two of the partners in the Park Land Company who agreed to postpone excavation until Monday, July 14, if the plaintiffs gave some assurance of raising the purchase price by that day. Refusing to accept an offer they felt was a form of community blackmail, the Defenders of Florissant appealed to the Tenth Circuit Court of Appeals the following morning, July 10. At the hearing before three judges of that Court in the afternoon, the Court questioned whether it had authority to issue a restraining order in the absence of any statute protecting the fossils. . . .

Admitting that Congress ". . . in its infinite wisdom, has not seen fit to pass legislation protecting fossil beds in general," plaintiffs argued: ". . . if someone had found the original Constitution of the United States buried on his land and then wanted to use it to mop a stain on the floor, is there any doubt . . . they could be restrained?"

Legally, plaintiffs claimed that the right to preservation of the unique and irreplaceable Florissant fossils, a national, natural, resource treasure, was one of the unenumerated rights retained by the people of the United States under the Ninth Amendment of the Constitution and protected by the *due process* and *equal protection* clauses of the Fifth Amendment, and the *privileges or immunities, due process* and *equal protection* clauses of the Fourteenth Amendment. Plaintiffs also asserted that the Florissant fossil beds were subject to judicial protection under the Trust Doctrine and while the defendants could profit from their nominal title to the land and make reasonable use of the area, they were under a duty to maintain that portion of the property vested with the public interest—the 34-million-year-old fossil shales. Procedurally, the Defenders invoked the federal equity jurisdiction relying on the fundamental equitable maxim, "there shall be no wrong without a remedy."

> In summation, counsel for the Defenders of Florissant picked up a fossil palm leaf that had been uncovered at Florissant, and holding it up to the Court, pleaded:
>
> The Florissant fossils are to geology, paleontology, paleobotony, palynology and evoluton what the Rosetta Stone was to Egyptology. To sacrifice this 34 million year old record, a record you might say written by the mighty hand of God, for 30 year mortgages and the basements of the A-frame ghettos of the seventies is like wrapping fish with the Dead Sea Scrolls.
>
> After a short recess, the Court issued an order restraining the defendants from disturbing the soil, subsoil or geological formation of the Florissant fossil beds by any physical or mechanical means. . . .

After a trial on July 29, 1969, the District Court denied the Defenders application for a preliminary injunction for the same reasons it had previously denied the application of a temporary restraining order, and the Park Land Company announced that its bulldozer would begin excavation that afternoon. Several hours later, the Plaintiffs filed a motion for an emergency stay with the Tenth Circuit Court of Appeals, citing defendants threat, and the Court of Appeals for the Tenth Circuit dramatically issued an order extending the restraining order of July 10 indefinitely until further order.

On July 31, 1969, the House Interior and Insular Affairs Committee, through its Subcommittee on Parks and Recreation favorably reported an amended version of the Florissant Fossil Beds National Monument Bill, and floor action by the House of Representatives was scheduled for August 4.

During the argument of the appeal before the Tenth Circuit Court of Appeals, the plaintiffs amplified their legal position, asserting that the Federal Courts had a duty to cooperate with Congress, and that by issuing a preliminary injunction, pending the final deliberation of the Congress of the United States the court was thereby aiding the orderly operations of the Legislative and Executive branches of government. Plaintiffs pursued their original theory that the Trust Doctrine protected the fossil beds, arguing that the land had acquired a public character due to the actions of Congress on the bills pending to dedicate the land as a national monument. The Court reserved decision at the close of the arguments and continued the temporary restraining order. That afternoon the House of Representatives passed its version of the bill as a number of concerned Congressmen from all over the country turned out to suspend the rules and consider the bill out of the regular order because of the pending threat to the fossils. The Senate agreed to the House version of the bill on August 7, and the President signed the bill on August 14, 1969. The preliminary restraining order issued by the Tenth Circuit Court of Appeals remaining in effect while the United States of America instituted suit to acquire the Park Land Company land by condemnation.

The court order prohibiting excavation of the fossil beds may have deprived the landowners of the most profitable use of their land, but did not prohibit all uses of the land consistent with the protection of the fossil beds. The landowners were free to develop the land for tourism, scientific research, or other uses compatible with maintenance of the paleontological integriryt of the area. Such uses, while perhaps not the most profitable use of the land, would still return a reasonable yield on the defendants' speculative investment.

The mere fact that the landowners might not wish to use the land for this purpose does not make the restraint on the land development an unreasonable taking where the public interest in the land was so great.

Certainly where a natural resource is as unique as the Florissant fossil beds were, the value to the public of protecting such a resource is so substantial as to justify the resultant burden upon the private property interests involved, even if it could have been shown that there was no reasonable expectation of profitable use of the property from tourism or other ancillary commercial development.

The message of the Florissant litigation is that judicial protection of a unique, national, natural resource treasure such as the 34-million-year-old Florissant fossil beds warrants restraint upon the rights of private property ownership, particularly during the period of due deliberation by Congress or other legislative body representative of the people.

The mere fact that Congress could not move as fast as the developers' bulldozer does not prevent a federal court of equity from acting to protect a national, natural resource treasures threatened with irreparable damage.

Project Rulison

Project Rulison provided the first direct confrontation among the several theories currently urged in support of citizen action to protect the environment from federal agency operations. Three separate suits were filed concerned with the AEC-Austral Oil Co. experiment seeking to stimulate production of natural gas by underground nuclear explosion. The first action was supported by the American Civil Liberties Union and relied on conventional theories to establish standing: Individual plaintiffs alleged direct, personal, private injury and special damage, seeking injunctive relief to protect their own property rights. The ACLU. application sought to restrain detonation of the underground nuclear device as its principle request for relief, and had already been denied when the second action was filed by the Colorado Open Space Coordinating Council. The title of that action in itself indicates the contrasting theories.

Colorado Open Space Coordinating Council, on behalf of all those entitled to the protection of their health and safety and of the health and safety of those generations yet unborn, from the hazards of ionizing radiation resulting from the distribution of radioactive materials through the permanent biogeochemical cycles of the Biosphere as a result of the defendants conduct of *Project Rulison,* and on behalf of all those entitled to the full benefit, use and enjoyment of the national, natural resource treasures of the State of Colorado without degradation resulting from contamination with radioactive material released as a result of the defendants conduct of *Project Rulison,* and all others similarly situated,

Plaintiffs

—against—
AUSTRAL OIL COMPANY, INCORPORATED
and
CER GEONUCLEAR CORPORATION,
Defendants
US ATOMIC ENERGY COMMISSION,
BUREAU OF MINES, US DEPARTMENT
OF INTERIOR, and
LOS ALAMOS SCIENTIFIC LABORATORY,
as their several interests may appear.

The shift in emphasis in the COSCC action from an emotional outcry against the underground nuclear blast itself to a reasoned demand for care in the release of radionuclides to the environment led to a Court order restraining the "flaring" of the radioactive natural gas following the blast until the hearing and determination of the action brought by COSCC.

By amending their complaint, ACLU, on behalf of the other individuals concerned personally with the blast, remained in the action. Subsequently, the District Attorney of the Ninth Judicial District of the state of Colorado attempted to bring an action in the state court on public nuisance theory, but that action was summarily transferred to the US district Court the consolidated with the COSCC and ACLU actions at the request of the Atomic Energy Commission.

EQUITY

The truly unique element of the Anglo-American is the concept of equity. Equity jurisprudence as a system of remedial law evolved from a number of common sources. It can be found in the Talmud and the earliest writings of the Roman law. It can be found today, though it is somewhat less than obvious, in the current system of civil jurisprudence derived from the Code Napoleon and used throughout most of Europe.

In its most elementary form, the fundamental principle of equity jurisprudence is the command: to use your own property as not to injure that of another. And the law, in order to give effect to this right provides for appeal to the ultimate power of society, be it king, parliament, state, or people, with a corollary maxim: equity permits no wrong to be without a remedy! The effective assertion of equitable rights by an individual or group of individuals is limited only by the rule that a party seeking equitable relief must come forward with "clean hands"—the party must be morally right as well as legally justified. There is an additional rule of restraint self-imposed by courts of equity: the relief granted must be commensurate with the injury suffered by the party seeking relief and tempered by the needs of society.

The Origin of Jurisdiction in Equity

Aristotle asserted that all law is universal and thereby cannot admit of exceptions, yet laws promulgated to cover a broad range of human action frequently cause injustice to some innocent individual because a particular case does not appear to be covered by the application of the universal law. Aristotle reasoned that when lawmakers make a law, they make it for the good of the community, therefore some legal remedy must be available to the individual treated unjustly by the particularly application of a general law.

In order to determine whether the individual was indeed being treated unfairly by the law, one must look to the intent of the lawmaker and the operation of the law. The remedy for the individual unjustly treated in a particular case by application of a general law was termed by Aristotle *Epicheia,* which translates loosely as "There should be an exception." It was left up to the judge deciding each particular case to determine whether a general or universal law was applicable, and if not, to make the appropriate exception in order to preserve the intention of the lawmaker, which was to ensure the good of the community.

The early Roman lawgivers accepted this concept, named it equity, and used it as a cornerstone for the development of what we now call the Roman Law. The intention of the Roman lawgivers was that one person should not benefit by a law while another person was injured by that law unintentionally and unnecessarily.

During the Middle Ages the heir apparent to the Roman legal system, the Christian church, developed the concept of equity even further, establishing the principle that "for every injury there must be some legal remedy," on the philosophical grounds that if the lawmaker did not provide some remedy for injury to an innocent individual, the law would allow certain injustices to go unpunished while others were punished, and this would be unjust. Since it was then an accepted principle that lawmakers were just, they could not have intended an injustice to follow from their laws, therefore, they intended to provide the legal remedy and the court would simply serve as the means to provide that which they intended to provide all along. Although such circumlocution might have found favor with medieval philosophers, there was, nevertheless, a singular practical reason for development of he principle. For if there was no procedure for individual relief from the unintended application of a general law, and sufficient individuals were oppressed by the unintended application of that law, then the individuals would tend to look with disfavor on the lawmakers and, in spite of the repression inherent in the feudal system, there might be civil unrest, a condition not conductive to the maintenance of the tenuous existence of the feudal estates which represented civilization in western Europe during the Middle Ages.

Equity jurisprudence developed throughout the ecclesiastical courts following the decline and fall of the Roman Empire, but it was to see its most dramatic development in England following the Norman conquest.

The common law, as distinguished from the customary law of the popular courts, originated in the establishment by Henry II of a national court administering a law for the entire nation, and by the end of the reign of Henry II, we find established a *Curia Regis,* a court of the King, which was a true court of law in the modern sense, administering a national law, common to the entire country, and which had largely displaced the customary laws of the different parts of the country. This continued during the thirteenth century, so that by its close, the common law was definitely established as the law of the nation, displacing the customary law and the local courts which were limited to local petty maters. The law of the *Curia Regis,* which had been the law of the very great, extended and adapted to the needs of the people so as to become the common law of a nation.

So long as the common law remained a flexible system with its field undefined, its power of inclusion unlimited, and its organs undifferentiated, there was no reason for distinguishing between the common law and equity. But soon the common law became so fixed as inflexible and its practitioners so absorbed in nice questions of form and pleading, there was no longer room for equity. By the early fourteenth century, the common law, which had supplanted the ancient customary law, had now, in its turn, become the regular system of remedial justice, but with gaps and defects where sufficient remedies were not provided, and resort was necessary to the ancient power of the king as the fountain of all justice.

Equity did not directly contest the existence of settled legal right. Rather, after recognizing those rights, equity went on to insist that the holder of such legal rights if they were acquired or retained unconscionably, or if they were being used in an unconscionable attempt to interfere with the fundamental rights of others, should be subject to the jurisdiction the "Chancellor" or the "Keeper of the Conscience of the King" as the source of all legal rights.

Equitable Relief

It is a fundamental principle essential to the very existence of organized society, and civilization as we know it, that every person, in exercising personal rights and in the use of personal property shall respect the rights and properties of others. Every person must so conduct themselves in the enjoyment of the rights and privileges which they may enjoy as individual members of society in such a way that they shall prejudice no one in their possession and enjoyment of their personal rights or the rights they hold in common as members of society. When there is an invasion of the rights or privileges of the public or the rights and privileges of any individual held in common by reason of the existence of civilized society, the absence of exact precedent and the fact that commentators on the law do not discuss the subject is of no material importance in awarding equitable relief.

That the exercise of the preventive powers of a court of equity is demanded in a novel case is not a fatal objection. In social evolution, with the march of the arts and sciences, and in the resultant effects upon organized society, it is quite intelligible that new conditions must arise in personal relations, which the rules of the common law, cast in the rigid mold of an earlier status, were not designed to meet. It would be a reproach to equitable jurisprudence, if equity were powerless to extend the application of the principles of common law, or of natural justice, in remedying a wrong, which in the progress of civilization, has been made possible as the result of new social or commercial conditions.

Equity is the agency by which law is brought into harmony with society. It is one of the factors which operate in judicial evolution. It succeeds legal fictions—those judicial assumptions through which a rule of law is modified in its operation—and it precedes legislation.

Equity has neither fixed boundaries, nor logical subdivisions, and its origin, both in Rome and in England, was that there was a wrong for which there was no remedy at law. As Lord Chancellor Cottenham observed:

It is the duty of this court, [equity], to adopt its practice and course of proceeding to the existing state of society and not, by a strict adherence to forms and rule, under different circumstances, to decline to administer justice and enforce rights for which there is no other remedy . . . If it were necessary to go much further than it is, in order to open the doors of this court to those who could not obtain [justice] elsewhere, I should not shirk form the responsibility of doing so.[†]

A distinguishing feature of equity jurisdiction is that it will apply settled rules to unusual conditions, and mold its decrees so as to do justice between the parties.

Peculiar and extraordinary cases will arise in the complex and diversified affairs of men, which perhaps, cannot be classed under any of he distinct heads of equity jurisdiction, but which must be acknowledged, nevertheless, to come within the legitimate powers of a court of equity because complete justice cannot otherwise be done between the parties. Therefore, when no remedy exists at law, courts of equity, to prevent injustice and in many cases on principles of general policy, will go far in granting relief.

Such was the law of equity at the start of the eighteenth century. If that rule of law had been developed with vigor much of the human and environmental degradation of he Industrial Revolution might have been avoided.

Since that time, two anomalous rules have evolved, interfering with the principle that for every wrong there is a remedy. The first of these anomalies concerns the artificial distinction between public and private nuisance and the second is the doctrine of "sovereign immunity." Both of these doctrines can be attributed to the meddling of Sir William Blackstone (1723–1780).

Public and Private Nuisance

Until Blackstone there was no distinction made between public and private nuisance. The rule had been well established that any individual could apply to a court of equity to abate a nuisance. But during the later part of the eighteenth century, Blackstone created a new rule of law that was to represent a classic manifestation of the obfuscation of simple legal principles by "self-proclaimed" legal scholars. This new rule of law proclaimed by Blackstone was to contribute substantially to the environmental and social crises of today and represents the kind of antisocial perversion of the law that made the common law of England during the seventeenth, eighteenth, and nineteenth centuries such an inviting target for the diverse talents of William Shakespeare, Jonathan Swift, Charles Dickens, and W.S. Gilbert.

Blackstone created a difference between public and private nuisance, and the significance of that difference is to be fond in the criteria for abatement of the public nuisance: (1) only a public nuisance may be made the basis for a criminal prosecution, and (2) only the public, through he proper officer, may sue to enjoin or abate a public nuisance, in the absence of special damage to a particular private individual—damage which is substantially greater than that suffered by other individual members of society—a public nuisance is subject to correction only at the hands of public authority. The mischief done and the disastrous consequence of Blackstone's whim are still evident.

In 1965, an action was brought on behalf of the people of the Town of Brookhaven, in the County of Suffolk,

---

[†] *Wallworth* v. *Holt* 4 Myl & C619, 41 Eng. Rep. 238 (1984).

New York, by a group of citizens, to restrain the Long Island duck industry form discharging raw, untreated sewage equivalent to that of a city of one million people into the waters of Great South Bay, which was once one of the finest shellfishing and marine recreation areas on the east coast of the United States. The duck industry defended the action by admitting that the eight million ducklings grown along the estuaries of he Great South Bay did in fact deliver the raw sewage effluent equivalent of a city of one million people into the Bay, but that the Court should dismiss the action on the grounds that such an affront to the public waters was so great that it was a "public nuisance" not a "private nuisance" and as a "public nuisance" it could only be attacked by the Attorney General of the State of New York, not by any private citizens, unless of course those private citizens could establish "special damages" different from the damages sustained by the public at large.

A New York Supreme Court, the same court that less than a year later in a dramatic reversal of precedent would issue that first injunction against the use of DDT ever granted by a Court, dismissed the case against the duck industry, accepting without question the argument that a public nuisance could not be abated by a private citizen or group of citizens.

How could this strange concept have crept into the law equity? What is the justification for this strange anomaly in that body of law that holds no wrong may exist without a remedy? In 1858, the Court of Appeals of the State of New York identified the source and expounded the justification for this onerous rule.

A contrary rule would be productive of very great inconveniences . . . No private person or number of persons can assume to be the champions of the community and in its behalf, challenge the public officers to meet them in the courts of justice to defend their official acts.

The court continued and discussed the theory of the decision,

> The general rule is that for wrongs against the public, whether actually committed or only apprehended, the remedy, whether civil or criminal, is by a prosecution instituted by the state in its political character, or by some officer authorized by law to act in its behalf. . . .

The principle is further exemplified in questions respecting nuisances. Common or public nuisances, which are such as are inconvenient or injurious to the whole community in general, are, as all are aware, indictable [the People of State can take action in a criminal proceedings] only, and the not actionable [any citizen can sue]; for as Blackstone [*Blackstone's Commentaries,* Book 4, p. 167] says, "it would be unreasonable to multiply suits by giving every man a separate right of action for what damnifies him in common only with the rest of his fellow-citizens.

Just who was Sir William Blackstone that he should exert such a restraint on the general application of equitable principles? Referring to the eleventh edition of the *Encyclopaedia Britannica,* published in 1911 and drawing on continual revisions from the first English edition in 1771, the following information may be elicited under the entry, "Blackstone, Sir William [1723–1780]:

> . . . In 1746 he was called to the bar. Though but little known or distinguished as a pleader, he was actively employed during his occasional residences at the university (Oxford), in taking part in the internal management of his college. In May 1749, as a small reward for his services, and to give him further opportunity of advancing the interests of the college, Blackstone was appointed steward of its manors. In the same year, on the resignation of his uncle, he was elected recorder for the Borough of Wallingford. . . . He accepted a seat on the bench, and on the death of Sir Joseph (Yates) succeeded him (in the court of common pleas). He died on the 14th of February, 1780. . . .

Blackstone was by no means what would now be called a scientific jurist. He had only the vaguest possible grasp of the elementary conceptions of law. Austin, who accused him of following slavishly the method of Hale's Analysis of the Law, declares that he "blindly adopts the mistakes of his rude and compendious model; missing invariably, with a nice and surprising infelicity, the pregnant but obscure suggestions which it proffered to his attention, and which would have guided a discerning and inventive writer to an arrangement comparatively just."

From the small place which equity jurisprudence occupies in his arrangement, he would scarcely seem to have realized its true position in the law of England.

Bentham accuses him of being the enemy of all reform, and the unscrupulous champion of every form of professional chicanery. Austin says that he truckled to the sinister influences and mischievous prejudices of power, and that he flattered the overweening conceit of the English in their own institutions. "He displays much ingenuity in giving a plausible form to common prejudices and fallacies. . . ."

For more than a century the opinion of that one man has stood in the way of a proper disciplined application of a fundamental principle of equity jurisprudence, "equity will not suffer a wrong without a remedy."

Just what happened at the time of Blackstone? It was obvious that Blackstone set himself to the task of codifying the laws of England, but in the process of attempting to build a logical and consistent body of legal principles he lost the basic insight of the Anglo-American system of jurisprudence, the common law.

The common law grew with civilization and the practices and customs of society. As the oppressed peasants obtained certain remedies during the Middle Ages, these remedies became a part of the common law. That is why the Anglo-American legal system depends on case law or precedent rather than elaborate codes promulgated by a legislature or king. A system of administering justice based on precedent rather than stature is inherently more flexible and capable of meeting the needs of society because the court is free to interpret the law with reference to, and by analogy with, part cases of similar import, though not precisely identical.

There is no doubt that Blackstone did a reasonable job of tidying-up the law, but in his attempt to make of law an axiomatic science little different from geometry, he lost the essential elements of equity jurisprudence, particularly the rule that for every injury there must be a legal remedy.

Blackstone's justification for striking the principle that for every injury there must be a legal remedy from the law of England arose from his belief that the King could do no wrong. Reasoning that failure to provide an adequate legal remedy for an injury was wrong, and that the King could do no wrong, it was obvious that there was no need for a rule of law providing a remedy for every wrong, because the King had obviously provided the remedy, since the King could do no wrong and failure to provide a remedy for an injury would be wrong. This perfectly circular argument has been exposed by many legal scholars, yet it still appears in decisions of courts that should know better.

Consider for a moment the ideological basis for establishing American independence. The founding fathers of this country were asserting the fundamental equitable principle that no wrong should exist without an adequate legal remedy. They did not deny the right of George III to tax his American colonies, they asserted that certain taxes were the wrong taxes, at the wrong time and for the wrong purposes. Taxation without representation was wrong. In other words, the King could, in fact, under certain circumstances, do wrong, and the equitable rule of law that no wrong should exist without a legal remedy was a fundamental human right.

Environmental Law is a mixture of the new and the old. Affirming the timeless principles of equity jurisprudence and asserting the unenumerated rights reserved to the people by the Ninth Amendment of the Constitution and protected by the *equal protection* and *due process* clauses of the Fifth Amendment and the *privileges or immunities, due process* and *equal protection* clauses of the Fourteenth Amendment of the Constitution, the law is newly applying these established principles and traditional legal procedures directly to the environmental crises threatening the Biosphere.

The basic element of environmental jurisprudence is recognition by our courts that the people have an absolute right to a salubrious environment as one of the "fundamental and inherent rights with which all humans are endowed even though no specific mention is made of them in either the national or state constitutions. . . ." The inherent human freedoms with which mankind is endowed are "antecedent to all earthly governments' rights that cannot be repealed or restrained by human laws, rights derived from the Great Legislator of the Universe."

[*Colorado Anti-Discrimination Com. v. Case*, 151 Colo. 235, 380 P. 2d 34 (1962).]

The words of Lord Atkin in an English case are compelling: "When those ghosts of the past stand in the path of justice, clanking their medieval chains, the proper course for the judge is to pass through them undeterred. . . . The common law does not go on the theory that a case of first impression presents a problem of legislative as opposed to judicial power."

[*United Australia, Ltd. v. Barclay Bank, Ltd.* (1940) 4 All En. 20, (1941) A C 1, 29.]

"The law will protect a flower or a vine as well as an oak . . . ; [Such] damages are irreparable too, because the trees and vines cannot be replaced."

[*Campbell v. Seaman*, 63 NY 568 (1978).]

The above historical treatment presented an overview of the fundamental arguments of Environmental Law. Some examples of recent legislative activity are discussed below by the editors.

The Clean Air Act Amendments of 1990 were the latest of a series which also include those of 1970 and 1977. They deal with the broad reduction of emissions from acid rain precursors (i.e., reductions of 10 million tons/yr and 2 million tons/yr in SO_x and NO_x emissions, respectively within 10 years) and air toxic emissions from industrial sources including over 170 species not previously listed. The American Electric Power Company said of the CAA amendments, "probably the most complex piece of legislation we've seen in the last 40 years."

Pursuant to the 1990 CAA requirements, Maximum Achievable Control Technology (MACT) regulations for many sources emitting hazardous air pollutants (HAP's) have been promulgated and are being enforced. Also new NAAQS limits have been established for particulate matter less than 2.5 micron diameter. These regulations and their enforcement are discussed in the article by C.V Mathai and Elliott, E.D., **EM**, pp. 25–34, May 2002. The U.S. Supreme court may be called upon to settle some lawsuits brought forth by a coalition of states, local governments and environmental groups in the enforcement of EPA regulations, such as New Source Review (NSR) provisions—Smith, D.C.," US Climate Change Legislation," Refocus, pp. 16–19, Apr/May (2004) and Barcott, Bruce, "Changing all the Rules", *New York Times Magazine*, pp. 38–78, Apr. 4, 2004. Also in the global arena, an administration's failure to confront climate change by its rejection of the Kyoto agreement, may be the subject of a lawsuit before an international tribunal (Strauss, A., The legal Option: Suing the United States in International Forums for Global Warming Emissions," Environmental Law Reporter, **33**, 108185 (2002). David Grossman argues that there may be good reason to shift the costs of damages caused by fossil fuel companies that have received economic benefits of activities that negatively impact the environment (Grosssman, David A., "Warming up to a not so radical idea: Tort-based climate change litigation," *Columbia Journal of Environmental Law* **28**, 1 (2003).

REFERENCES

Detailed treatment of the material contained in this article, including generally:

Environmental Protection and the Law, The Trust Doctrine, The Ninth Amendment; Nuisance and Other Common Law Remedies; Environmental Legislation; Environmental Toxicants; Radiation; Water

Pollution; Air Pollution; Noise; Visual Pollution; Solid Waste; Land Use and Natural Resource Management; Environmental Litigation and Administrative Agencies and the Environment, will be found in the treatise *Environment Rights and Remedies* by Victor John Yannacone, Jr., Bernard S. Cohen, Steven Gebauer Davison (1971). Lawyers Cooperative Publishing Company, Rochester, New York/Bancroft Whitney Co., San Francisco, California.

Current material in the nature of news accounts, federal, state and significant local legislation as promulgated, and texts of the decisions of courts and selected administrative agencies, is published weekly by the Bureau of National Affairs, Washington, D.C., in their service *Environment Reporter.*

VICTOR J. YANNACONE, JR.
Yannacone and Yannacone

LAKES: see AQUATIC PRIMARY PRODUCTION, WATER—FRESH

LEGAL ASPECTS: see also ENVIRONMENTAL LAW

LIMNOLOGY

INTRODUCTION

Limnology is the scientific study of the physical, chemical, and biological factors that affect aquatic productivity and water quality in lakes. Lakes are important resources—much more than places for groundwater, surface water, and precipitation to collect. They control flooding, provide water for domestic and agricultural uses, and provide recreational opportunities such as swimming, fishing, boating, and waterskiing. Lakes also provide habitat for insects, fish, and wildlife such as frogs, turtles, waterfowl, and shorebirds. Lakes' commercial value in food supply, tourism, and transportation is worth many billions of dollars each year. Lakes also offer opportunities for relaxation and appreciation of natural beauty. According to the North American Lake Management Society, this quality is not a minor asset; over 60 percent of Wisconsin lake property owners who were asked what they valued in lakes rated aesthetics as especially important. (U.S. EPA, 1990).

However, a lake cannot be all things to all people. Desirable uses, even obtainable ones, can conflict. For example, swimmers may want no plants but some plants are needed in order to provide fish habitat. Lakeside property owners and lake associations often want their lake to do everything; they want aesthetic pleasure, great fishing, clean water, sandy shorelines and bottoms, and a healthy wildlife population—without insects or weeds. No lake can meet all of these demands.

This article will provide an overview of the physical, chemical, and biological components of lake ecosystems. An ecosystem is a system of interrelated organisms and their physical–chemical environment. It is impossible to alter one characteristic of a lake ecosystem without affecting some other characteristic of the ecosystem. The article will also explain how lake ecosystems get out of balance, and what can be also done to restore the balance.

LAKE WATERSHEDS AND ZONES

Lakes are receiving bodies—constantly receiving water, dissolved materials, and particulates from their watersheds and from the atmosphere, and energy from the sun and wind. A watershed is the area which drains to a lake. Watersheds come in all sizes. For example, the watershed that drains to Beaver Lake in western Washington near Seattle is less than two square miles in area, whereas the Lake Washington watershed is 350 square miles. Lakes are sensitive to existing conditions in the surrounding watershed and atmosphere.

Each lake has a unique watershed, size, and shape. The size and shape are often determined by the origin of the lake basin and, in turn, influence the lake's productivity, water quality, habitat, and lifespan.

The most common origin of lake basins in North America has been glacial activity such as the erosion of bedrock and deepening of valleys by expansion and recession of glaciers. Glacial lakes of Canada and the upper midwestern United States were formed about 8,000 to 12,000 years ago. For example, the Finger Lakes of upper New York State were formed when deep depressions left by receding glaciers filled with meltwater (U.S. EPA, 1990).

The depressions left by melting ice blocks form kettle or "pothole" lakes. This type of lake is common throughout the upper midwestern United States, the eastern portion of the state of Washington, and large portions of Canada. Kettle lakes and their watersheds are popular home sites and recreational areas. The size and shape of kettle lake basins reflect the size of the original ice block and how deeply it was buried in the glacial debris (U.S. EPA, 1990).

Some lakes are formed by volcanic activity; i.e., a volcano erupts creating a huge depression or caldera which then fills with water. Crater Lake in Oregon is an example of a volcanic lake. Movements of large segments of the earth's crust created Reelfoot Lake in Tennessee, Lake Tahoe in California and Nevada, and many other lakes (U.S. EPA, 1990).

Solution lakes are formed where groundwater has dissolved limestone; this is the case for many Florida lakes. Other lakes originate from shifting of river channels. For example, oxbow lakes are stranded segments of meandering rivers. The persistence of dam-building beavers can also create lakes (U.S. EPA, 1990).

A lake has four zones, each with different plants and animals (Figure 1). The littoral zone is the portion of the lake that extends from the shoreline lakeward to the greatest depth occupied by rooted plants. By contrast, the pelagic zone is the open area of a lake from the edge of the littoral zone to the center of the lake. The benthic or profundal zone refers to the deep waters at the bottom of a lake where photosynthesis does not occur because light does not penetrate. The marginal zone refers to the margins of the lake on the lake shoreline (U.S. EPA, 1990).

Shallow lakes tend to be more biologically productive than deep lakes because of the large area of bottom sediments

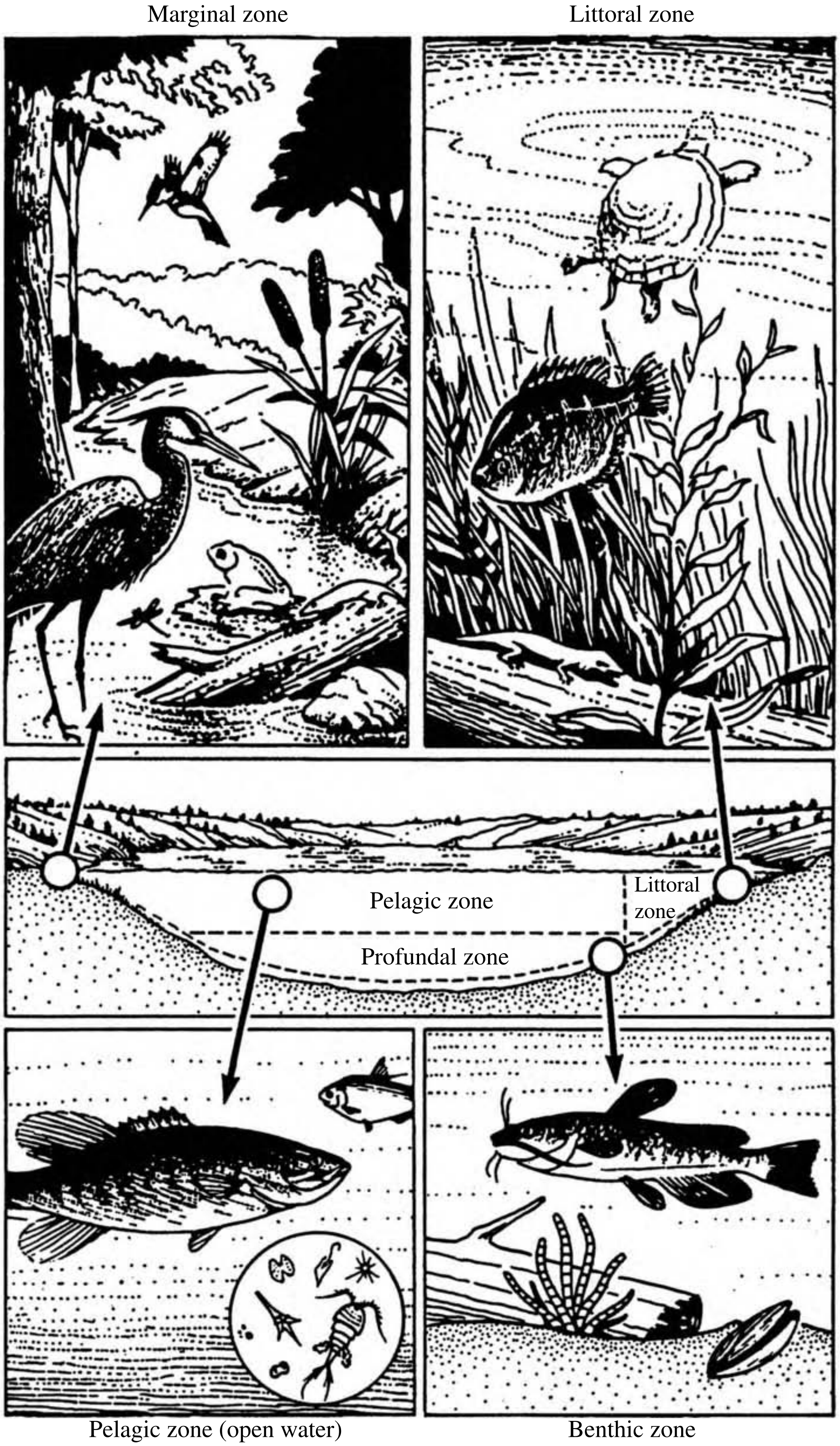


FIGURE 1 The location and nature of typical lake communities, habitats, and organisms. In addition to the lake's watershed, all of these components are part of the lake ecosystem. (U.S. EPA, 1990)

relative to the volume of water, more complete wind mixing of the lake water, and the large littoral zone along the lake perimeter that can be colonized by plants. Shallow lakes often have most of their plants in littoral areas and have little pelagic habitat. On the other hand, deep lakes have fewer areas that receive enough light for rooted aquatic plants to grow, and therefore have a high proportion of pelagic habitat and less littoral habitat.

HYDROLOGIC CYCLE AND WATER BUDGET

Since precipitation and surface water runoff have direct influences on lake ecosystems, understanding the hydrologic cycle and water budget are key concepts in limnology. The hydrologic (water) cycle refers to the circulation of water between the Earth's surface and the atmosphere. This is powered by the sun. Water falls to Earth as precipitation. About 75 percent of the precipitation is returned to the atmosphere as vapor through direct evaporation and transpiration from both terrestrial and aquatic plants during photosynthesis. The remaining 25 percent of the precipitation is stored in ice caps, drains directly off the land into lakes, streams, wetlands, rivers, and oceans, or infiltrates the soil and underlying rock layers and enters the groundwater system. Groundwater enters lakes and streams through underwater seeps, springs, or surface channels (Cooke *et al.*, 1986; U.S. EPA, 1990; Wetzel, 1983).

Drainage lakes are formed primarily by inflowing rivers and streams. Therefore, their water levels vary with the surface water runoff from their watersheds. On the other hand, seepage lakes form where groundwater intersects with the land surface. Since seepage lakes are maintained primarily by groundwater inflow, their water levels fluctuate with seasonal variations in the local water table. For both drainage and seepage lakes, the balance between water inputs and outputs influences the supply of plant nutrients (nitrogen and phosphorus) to the lake and the lake's hydraulic (water) residence time, thereby influencing the lake's water quality and biological productivity (U.S. EPA, 1990).

The hydraulic (water) residence time is the amount of time that water entering a lake will remain in it or the average amount of time required to completely renew a lake's water volume. The amount of water entering a lake from its watershed controls the volume of the lake. The hydraulic residence time is calculated by dividing the water volume of a lake by its flow rate, and varies greatly among lakes. For example, if a lake has a volume of 500 acre-feet and the outflow rate is 10 acre-feet per day, then the hydraulic residence time would be 50 days. If the hydraulic residence time of a lake is 100 days to several years, this means that plant nutrients and pollutants remain in the water column long enough to degrade water quality and to allow plants to accumulate (U.S. EPA, 1990; Wetzel, 1983).

Each lake has a water balance, in which water input = water output + the change in the amount of water stored in the lake. If inputs are greater than outputs, lake levels rise as water is stored in the lake. When outputs are greater than inputs, lake levels fall. This happens during summer droughts.

A related concept is the lake water budget, which is a measure of the sources of water entering and flowing out of a lake over the course of a year. A lake's water budget is affected by the hydrologic cycle, and the quantity and timing of water entering and leaving the lake. Types of data used in calculating a water budget include precipitation, stream flow into and out of the lake, and lake surface elevation (water level). Sources of water input or inflow include the lake inlet(s), precipitation, surface water runoff, point source discharges, and groundwater. Sources of water output or outflow include the lake outlet(s), evaporation, transpiration from lake plants, groundwater seepage, and water withdrawals for domestic, agricultural, and industrial purposes. The change in storage accounts for changes in surface elevation over the year. This change is positive if lake volume increases over the year, negative if lake volume decreases (U.S. EPA, 1990; Wetzel, 1983).

Land use and geology of the surrounding watershed affect the water budget. For example, lakes in areas with permeable soils receive inflowing groundwater throughout the year. Lakes in areas with impervious surfaces can receive large volumes of stormwater runoff.

PHOSPHORUS BUDGET AND LOADING

Another important characteristic of lakes is the phosphorus budget, a measure of the sources of phosphorus entering and leaving the lake over the course of a year. Phosphorus is a nutrient that is essential in plant growth. The amount of phosphorus in a lake directly influences biological productivity. The phosphorus budget will indicate if the phosphorus in the lake is coming from within the lake, from sources in the watershed, or from both internal and external sources.

For a given lake, phosphorus inputs (inflow loading) –phosphorus outputs (outflow loading) + net sedimentation + change in storage. This means that phosphorus inputs to the lake equal phosphorus losses from the lake plus or minus the change in the total amount of phosphorus stored in the lake. Change in phosphorus storage within the lake equals the amount of phosphorus entering the lake minus the amount of phosphorus leaving the lake minus the net loss of phosphorus to the lake sediments. Sources of phosphorus inputs to a lake are the lake inlet(s), point sources discharging directly to the lake, precipitation, surface water runoff, leachate from malfunctioning shoreline septic tanks, other groundwater inputs, and migrant waterfowl wastes. Sources of phosphorus outputs from a lake are the lake outlet(s), groundwater seepage, and water withdrawals for domestic, agricultural, and industrial purposes (Cooke *et al.*, 1993b; U.S. EPA, 1990; Wetzel, 1983). Figure 2 illustrates the phosphorus budget for Cottage Lake, near Seattle, Washington. As indicated in the figure, most of the phosphorus in the lake comes from the Daniels Creek inlet and from the lake sediments. Most of the phosphorus that leaves the lake does so via the lake outlet (KCM, 1994; Solomon *et al.*, 1996).

Net sedimentation refers to the amount of phosphorous accumulated in lake bottom sediments, i.e., the difference in the amount of phosphorus that binds to the sediments and

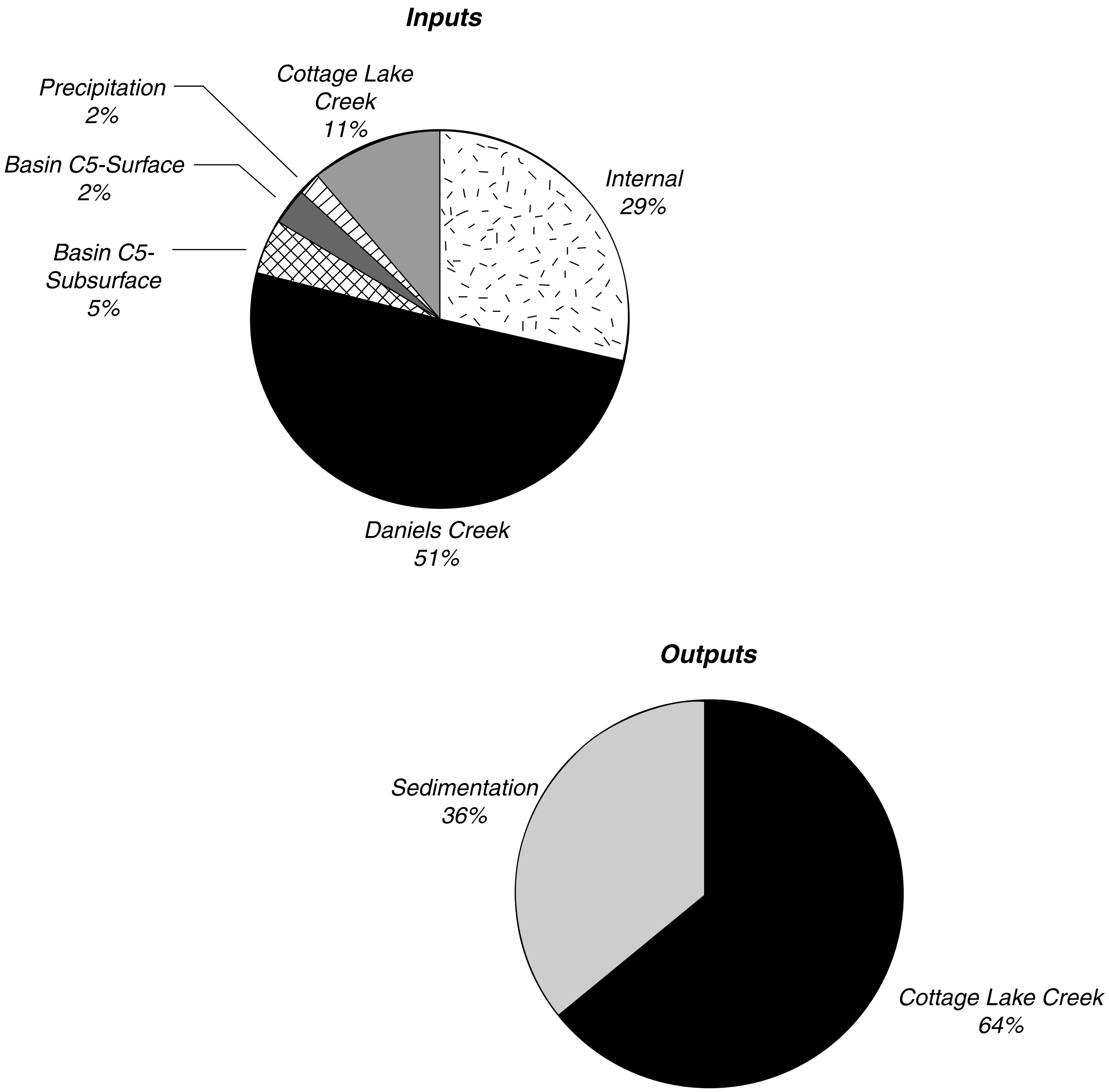


FIGURE 2 Cottage Lake total phosphorus inputs and outputs.

the amount of phosphorus released from the sediments to the water column. In general, lake water quality will improve as the magnitude of sedimentation increases because higher sedimentation means there is less phosphorus in the water column to stimulate overgrowth of aquatic plants (for more details, see section on eutrophication). The change in storage accounts for changes in the total amount of phosphorus in the lake water column between the beginning and end of the year.

Phosphorus concentration differs from phosphorus loading. Phosphorus loading to a lake is calculated on the basis of the water budget for the lake and measured phosphorus concentrations in the lake, its inlets, its outlets, precipitation, surface water runoff, and groundwater. Loadings are based on concentrations and flow rates and most accurately express the relative impacts of various watershed sources on lake water quality. For example, a stream that is an inlet to a lake may have a high concentration of phosphorus. This does not necessarily mean that the stream is a major contributor to the lake phosphorus budget. If the stream has a low flow, it will contribute a relatively low annual phosphorus loading.

The concept of phosphorus loading can be illustrated with an analogy to a grocery bill as shown in Table 1. For each grocery item, the cost is determined by multiplying the unit cost and the number of items purchased. The total grocery bill is the sum of the costs of all items. Likewise, for each source of phosphorus input to a lake, the phosphorus loading is determined by multiplying the flow rate for the source (lake inlet, groundwater, etc.) and its phosphorus concentration over annual and seasonal periods. The total "phosphorus bill" (total phosphorus loading from all sources) is the sum of the loadings from each source.

TABLE 1
Phosphorus Loading Concept (U.S. EPA, 1990)

Grocery Bill	Phosphorus Loading
Item	Source
Quantity	Flow
Unit Cost	Concentration
Cost of Item	Loading from Source
Total Cost of All Items	*Total Loading from All Sources*

Phosphorus loadings change in response to season, storm events, upstream point sources, and land use changes. For example, converting an acre of forest into residential or commercial land typically increases the phosphorus loading to a lake in that watershed fivefold to twentyfold. This is because there will be increases in both water flow (runoff from the newly created impervious surfaces) and phosphorus concentration (deposition of phosphorus on impervious surfaces). An evaluation of phosphorus loadings provides a basis for predicting lake responses to changes in land use.

STRATIFICATION

Many swimmers notice that when they dive into a lake during the summer, the deeper waters of the lake are much colder than the surface waters. This is due to stratification, an interesting temperature-related characteristic of most temperate climate lakes.

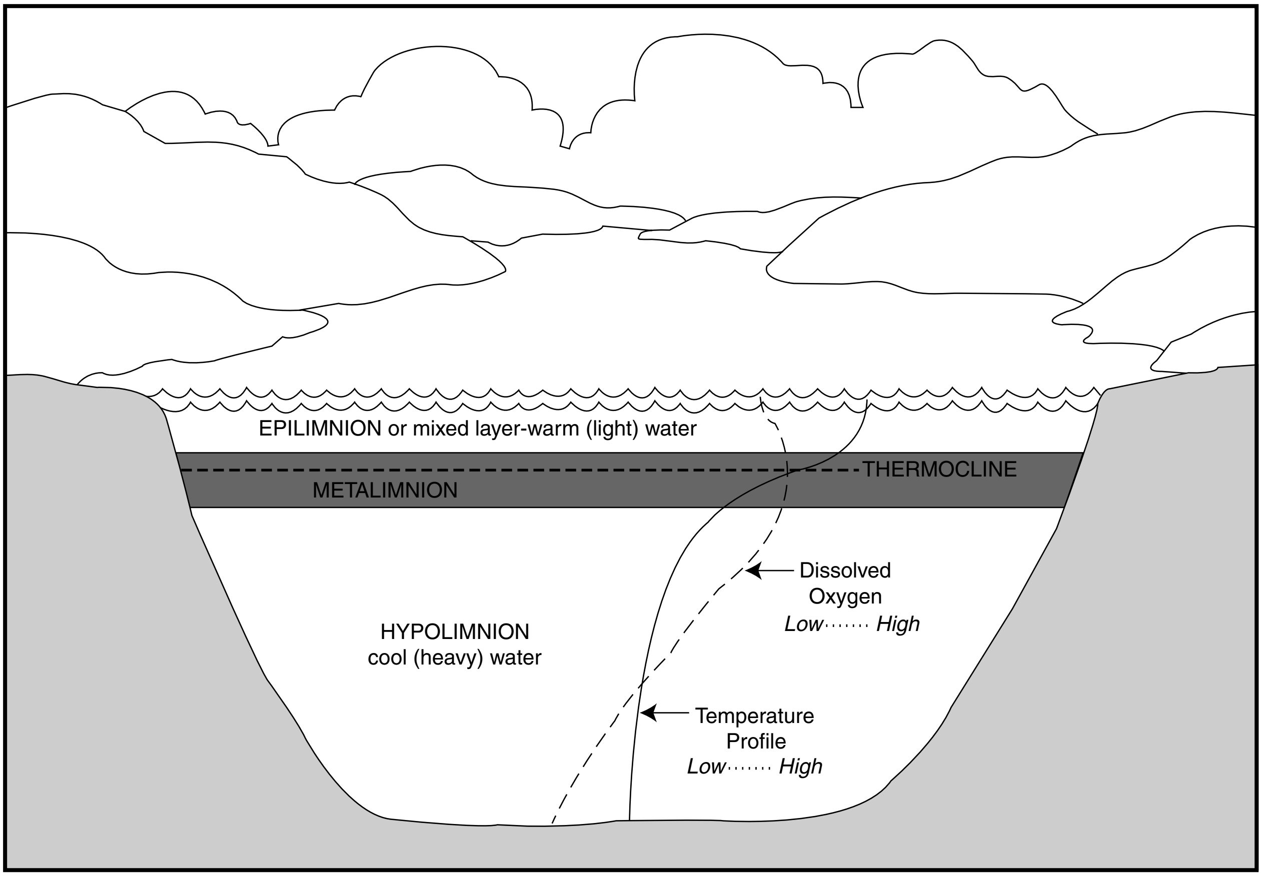


FIGURE 3 Thermal stratification.

Temperature-related characteristics of water have a large effect on the water quality and ecology of lakes. Water is at its densest at 4°C or 39°F, then expands (becomes less dense) until it freezes at 0°C (32°F). This anomalous expansion of water allows ice to float and form at the surface of lakes at 0°C (32°F) or less, and thermal stratification to occur during the warmer, summer weather. During spring and early summer, energy from the sun heats the upper water layer. The warmer, less-dense surface waters float on top of the cooler, denser bottom waters. This results in the upper layer, or epilimnion, becoming isolated from the lower layer, or hypolimnion (Figure 3). The two layers are separated by the middle layer, or metalimnion, where large temperature changes occur with changes in depth. The thermocline, which is located within the metalimnion, is a horizontal plane of water across the lake through the point of the greatest temperature change. The metalimnion presents a physical barrier to the mixing of the epilimnion and hypolimnion. Since there is little or no exchange of water between the epilimnion and the hypolimnion, water quality can be quite different in each layer.

In the temperate regions of the world where there are not strong contrasts in seasonal conditions (e.g., mild winters and summers), this type of thermal stratification is common during the summer and early fall. After the summer, the epilimnion tends to cool, and by late fall or early winter the temperature difference between the two water layers is small enough that the winds will mix the water throughout the lake, which will then remain fully mixed until the onset of stratification in late spring. Lakes that undergo this type of seasonal pattern (i.e., they stratify once and re-mix or turn over once each year) are called monomictic lakes. These include lakes in mountainous regions of the temperate zones, warm regions of the temperate zones, many coastal regions of North America and Europe, and mountainous areas of subtropical latitudes (U.S. EPA, 1990; Wetzel, 1983).

By contrast, in the temperate regions of the world with strong contrasts in seasonal conditions (e.g., very cold winters and very hot summers), lakes undergo complete turnover in the spring and fall separated by thermal stratification in the summer (i.e., warmer surface waters float on top of cooler bottom waters) and inverse thermal stratification in the winter. Ice cover forms and floats on the surface of such lakes under clam, cold conditions. Inverse stratification of water temperatures occurs under the ice, in which colder, less-dense water overlies warmer, more-dense water near the temperature of maximum density at 4°C. Some gradual heating of the water occurs during the winter under ice cover. When the ice cover melts in the spring, the water column is nearly uniform in temperature. If the lake receives sufficient wind energy, as is usually the case, then the lake circulates completely and undergoes spring turnover. Stratification occurs during the warmer days of summer, with another complete turnover in fall. Lakes that undergo this type of seasonal pattern (i.e., they stratify twice and re-mix twice each year) are called dimictic lakes and include most lakes of the cool temperate regions of the world (U.S. EPA, 1990; Wetzel, 1983).

LAKE BIOTA

The types of organisms found in a lake may include phytoplankton (algae), zooplankton, benthic infauna, fish, amphibia (such as tadpoles, frogs, and salamanders), reptiles (such as turtles and water snakes), and birds (waterfowl and shorebirds). Lake plants and animals are interrelated via a food chain.

Algae are microscopic plants found in the lake water column. Algal species may occur in many different forms including filamentous, colonial, and single-celled. Algae are easily carried by wind-generated currents and will often accumulate in windward areas of the lake, forming surface scums. When algae populations increase rapidly, the algae can become a nuisance by forming high concentrations in the water column, or even surface accumulations, called algal blooms.

Several different algal species can usually be found in a lake at any time of the year. A variety of environmental factors including light, temperature, and nutrient levels, affect phytoplankton production and the occurrence of algal blooms. Diatoms are algae that are golden in color and contain silica. They predominate in the spring and autumn due to their ability to reproduce and grow in cooler temperatures and less light. During the summer, increased water temperatures and available light create conditions that favor green algae or blue-green algae. Blue-green algae can form nuisance blooms; they are particularly problematic because they will float to the surface, forming scums that affect the recreational uses and aesthetic qualities of a lake. In some lakes with high biological productivity, blue-greens dominate in spring, summer, and fall.

In addition to algae, large vascular plants (plants with roots, stems, and leaves) or macrophytes are found in lakes. Macrophytes are classified as emergent, floating, or submersed. Emergent plants grow on the shoreline and include cattails, irises, and purple loosestrife. Floating plants are plants that float on the surface of the lake. They can be rooted in the lake bottom such as water lilies or watershield or free-floating such as duckweed. Submersed plants are rooted plants that live below the lake surface and include pondweed and water weed (*Elodea*). Figure 4 illustrates the community types and common examples of plants associated with each type (Washington State Department of Ecology, 1994).

Some macrophytes are native to the particular lake and geographic region; others, called exotics, have been imported or are transported to the lake from other lakes. For example, native macrophytes in lakes of the Pacific Northwest region of the United States include cattails, yellow water lilies, and pondweed. Exotic or non-native plants in Pacific Northwest lakes include purple loosestrife, white water lilies, and Eurasian watermilfoil. Some non-native plants are invasive, crowding out native plants and not providing useful habitat for fish and wildlife.

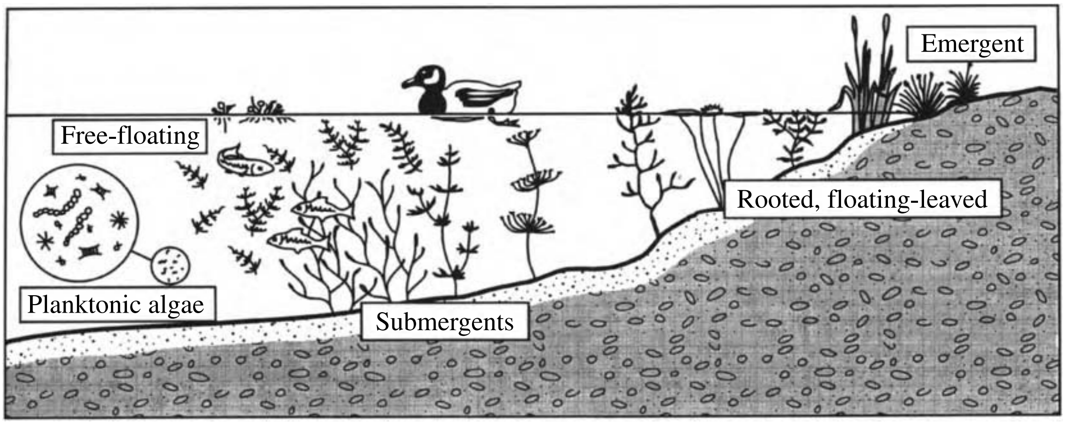


FIGURE 4 Macrophyte community types.

Aquatic plants provide many benefits, including sediment and shoreline stabilization; food source and habitat for benthic invertebrates, fish, and wildlife; oxygenation of the water column; and aesthetics. Most rooted macrophytes obtain their nutrients from lake sediments rather than the water column, and take up phosphorus that would otherwise have been available for algal growth, thus preventing the overgrowth of algae.

However, when there are too many aquatic plants, particularly non-native plants, the advantages turn into disadvantages. When a lake is shallow and nutrient-enriched, then there can be too many macrophytes. Too many aquatic plants can decrease the quality of fish and wildlife habitat, interfere with beneficial uses of a lake such as swimming and boating, and even create safety problems, i.e., swimmers can become entangled in milfoil and other plants. When the plants decay, they deplete the lake waters of oxygen and release nutrients into the water column which can promote algal growth. Too many macrophytes can make a lake look unsightly. The advantages and disadvantages of aquatic plants are opposite sides of the same coin; it's a matter of degree and balance.

Many types of animals are found in lakes. Zooplankton are microscopic animals found in the lake water column. Examples are rotifers and water fleas, e.g., *Daphnia.* They are visible to the naked eye on close inspection of a glass of lake water. Zooplankton are important in the food web of a lake because they eat algae and, in turn, are eaten by planktivorous fish. The types and number of zooplankton present are also indicative of lake water quality. Generally, large grazing species improve water quality by eating algae. On the other hand, a general decrease in the size of zooplankton species, with their reduced capacity to graze the phytoplankton, is a response to the greater availability of bacterial detritus resulting from the relatively ungrazed algae (Welch, 1992). Therefore, the presence of larger zooplankton in a lake usually indicates good water quality, while the presence of smaller zooplankton generally indicates more nutrient-rich waters.

Benthic infauna are small invertebrate animals such as molluscs, worms, and midges that live in the bottom sediments of lakes. They feed on detritus in the sediments and recycle nutrients to the water. The species of benthic animals found in a given area are usually indicative of the surrounding water quality. Some invertebrates, such as mayflies, are intolerant of low dissolved oxygen conditions; their presence in large numbers in lake ecosystems indicates good water quality. Other invertebrates, such as oligochaetes and chironomids, are more tolerant of low dissolved oxygen conditions; their presence in large numbers in a lake may indicate the presence of pollutants or degraded water quality.

The greatest density and diversity of benthic invertebrates is usually found in the littoral zone of a lake, where ample vegetation and oxygen are present. The benthic communities, in turn, provide food for larger invertebrates, fish, amphibians, and birds.

The types of fish found in a lake are influenced by water temperature and dissolved oxygen levels. Fish such as perch, bass, and smelt are warmwater fish and thrive in lakes where the summer water temperature exceeds 65°F. The dissolved oxygen level needs to be at least five parts per million (ppm) in order for the fish to remain healthy. Coldwater fish such as salmon and trout are found in lakes where the summer water temperature is less than 65°F. The dissolved oxygen level needs to be at least seven ppm for these fish. If the summer water temperature is too high, the dissolved oxygen level is often too low to support healthy fish populations.

Each organism in a lake is dependent on other organisms for its food. Each lake has a natural food chain. Algae are eaten by zooplankton. In some lakes, the efficient grazing of zooplankton by algae can help to maintain water clarity. Zooplankton are eaten by planktivorous fish such as longfin smelt and perch. Planktivorous fish are eaten by larger, piscivorous fish such as northern squawfish and largemouth bass. The larger fish are eaten by birds and by mammals, including humans. Figure 5 illustrates the aquatic food chain found in many lakes.

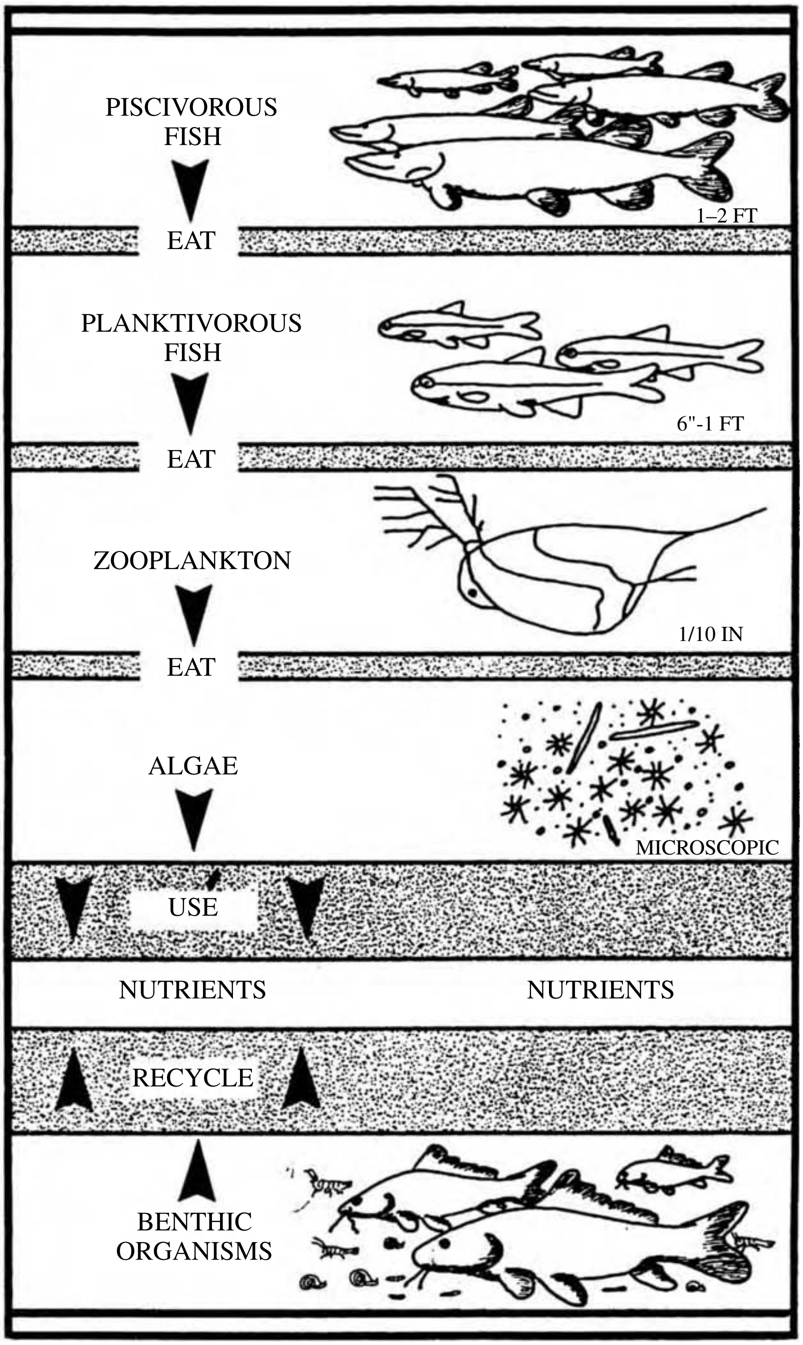


FIGURE 5 Aquatic food chain (U.S. EPA, 1990).

The food chain concept involves the flow of energy among the lake organisms and the recycling of nutrients. Each trophic level (food chain level) transfers only 10 to 20 percent of the energy received up the chain to the next trophic level (Kozlovsky, 1968; Gulland, 1970). This means that a few large piscivorous fish depend on a large supply of smaller planktivorous fish which depend on a very large supply of zooplankton which depend on a successively much larger base of photosynthetic production by phytoplankton and other aquatic plants. By constantly producing wastes and eventually dying, all of these organisms provide nourishment to detritus-eating organisms in the sediments, which obtain their energy by decomposing organic matter. Organic matter decomposition results in the recycling of nutrients that are required for further plant production (U.S. EPA, 1990).

When one level of the food chain of a lake is altered, it affects all other levels, sometimes positively and sometimes adversely. For example, in Lake Washington, *Daphnia* populations increased in the 1970s.Why? The longfin smelt population increased in the 1960s when flood control activities in the main inlet stopped and spawning beds were no longer damaged. Longfin smelt feed on a large crustacean called *Neomysis* which feeds on *Daphnia.* Predation on *Daphnia* was thereby reduced. This had a positive effect on Lake Washington because *Daphnia* grazed on the algae, resulting in improved water clarity in the lake.

Another example of food chain manipulation is stocking a lake with piscivorous fish. When the fish are removed by anglers, there will be more planktivorous fish which will result in a decreased zooplankton population. Fewer zooplankton will mean more algae in the lake, which could have adverse effects on water clarity in the lake. In sum, altering one part of a lake's ecosystem has repercussions throughout the ecosystem.

TROPHIC STATUS/EUTROPHICATION

Lakes are characterized according to their level of biological productivity, or trophic status. The trophic status of a lake depends on the concentration of chlorophyll *a* (the pigment found in green plants that traps energy from the sun to enable the plants to produce their own food by the process of photosynthesis), frequency of algal blooms, the concentrations of nutrients, particularly phosphorus, and water clarity (transparency). The phosphorus concentration determines how many algae and other plants will grow in the lake. The clarity of the water is influenced by a variety of factors including algae, turbidity from sediments or other suspended particles, and the natural color of the water in the lake. Water clarity is measured with the use of a Secchi disk, a 20-centimeter plastic or metal disk that is divided into alternating black and white quadrants. The disk is lowered into the water until the observer can no longer see it. The distance between the lake surface and the point at which the disk disappears from view is called the "Secchi transparency" or "Secchi depth" of the lake.

Three trophic classifications are commonly used for lakes. An oligotrophic lake is one in which there is clear water, low levels of chlorophyll *a* and nutrients, and hence little aquatic life. Oligotrophic lakes tend to be found in alpine and other wilderness areas. The lakes are beautiful to look at and are fine swimming and boating lakes, but are not good fishing lakes unless they are stocked with fish. There are few naturally occurring fish in oligotrophic lakes because there are few plants or insects for fish to eat.

At the other end of the scale are eutrophic lakes. A eutrophic lake has murky water, high levels of chlorophyll *a* and nutrients, and is full of aquatic life. Many lakes in urban and suburban areas are eutrophic, as evidenced by algal blooms. A mesotrophic lake is in between, i.e., is moderately transparent, with moderate levels of chlorophyll *a* and nutrients, and some aquatic life.

Transparency, chlorophyll *a* and total phosphorus (both organic and inorganic forms of phosphorus) are most frequently used to assign trophic status to lakes. The general relationship between these lake water quality parameters and trophic status index (TSI) is summarized in Table 2.

A lake's natural level of productivity is determined by a combination of factors, including the geology and size of the watershed, depth of the lake, climate, and water sources entering and leaving the lake. Some lakes are naturally eutrophic based on their inherent physical attributes and watershed characteristics.

Increases in a lake's natural productivity over time, a process called eutrophication, occurs naturally in some lakes, and may be accelerated in others by human activities. For many small lakes, natural eutrophication typically occurs over hundreds or thousands of years, and is hence not observable in a single lifetime. What is observable in a single lifetime is the human-induced, or cultural eutrophication of lakes. Our land-based activities, including home-building, agriculture, forestry, resource extraction, landscaping, gardening, and animal husbandry, all contribute nutrients and sediments to surface waters, which in turn contribute to increasing a lake's biological productivity. Land erosion and forest clearcutting contribute sediments to lakes. Surface water runoff from impervious surfaces such as construction sites, parking lots, and pavement contributes nutrients and pollutants to lakes. Agricultural practices such as horses grazing near lakes, cows wandering in streams, and extensive pesticide use contribute nutrients and toxic pollutants to lakes. If oil or other toxic chemicals are poured down storm drains, these end up in the nearby lake, stream, or bay. Gardening chemicals such as fertilizers and household toxic chemicals can end up in storm

TABLE 2

Trophic status and associated values (Carlson, 1977; Cooke *et al.*, 1993b; Porecella *et al.*, 1980)

Trophic Status	Transparency (meters)	Chl. a (μg/L)	Total Phosphorus (μg/L)	TSI (average)
Oligotrophic	>4	<3	<4	<40
Mesotrophic	2–4	3–9	14–25	40–50
Eutrophic	<2	>9	>25	>50

μg/L = micrograms per liter (parts per billion).

drains and thus in lakes. Failing septic systems can discharge nutrients from raw sewage to lakes.

The result of all these inputs of phosphorus, nitrogen, sediment, and organic matter in large algal blooms which are unsightly and can severely restrict lake beneficial uses including swimming, fishing, boating, and aesthetic appreciation. Beneficial uses of a lake may also be degraded by other water quality problems related to eutrophic conditions, including low dissolved oxygen levels, fish kills, algal toxicity, and excessive aquatic macrophyte growth.

The level of dissolved oxygen in lakes is one determinant of the habitat available to aquatic organisms. Oxygen is added to a lake from exposure to the air, and by the contribution or aquatic plants through photosynthesis. Oxygen is removed from a lake by the respiration of aquatic organisms and plants, and the bacterial decomposition of organic matter in the water and sediments. Eutrophic lakes with large algal blooms are characterized by high phosphorus concentrations and low dissolved oxygen concentrations in the lake hypolimnion in the summer. This happens because decaying algae and other plants fall to the bottom of the lake where they contribute phosphorus and remove oxygen. Photosynthesis does not take place in the hypolimnion of a eutrophic lake because light does not penetrate to that depth; hence, the oxygen that is being depleted is not replaced. Anoxic (lack of oxygen) conditions at the water—sediment interface on the lake bottom usually increase the potential for nutrient release by converting iron phosphate in the sediments from a water-insoluble to a water-soluble form.

The very low dissolved oxygen levels in the hypolimnion of eutrophic lakes during the summer months may be too low to support coldwater fish such as salmon and trout. The salmon and trout would then move to the lake epilimnion, but the water temperatures may be too high for them in the surface waters. Eutrophic conditions in lakes often lead to decreased quantity and quality of fish habitat and stressed fish populations.

RESTORING BALANCE TO LAKE ECOSYSTEMS

Management of Eutrophic Lakes

When a lake is eutrophic with unsightly algal blooms, water quality problems, and impaired beneficial uses, its ecosystem is out of balance. Lake restoration involves reducing the impact of human activities on lake water quality, with the goal of decreasing biological productivity and improving water quality and associated beneficial uses of the lake. Several methods are available to accomplish this goal. Each method has its advantages and drawbacks.

In order to determine the most effective method(s) to use in a given lake, it is first necessary to be knowledgeable about the physical, chemical, and biological components of the lake's ecosystem. This can be accomplished through one or two years of monitoring parameters such as transparency, lake temperature, acidity, alkalinity, dissolved oxygen, lake level, amount of precipitation, nutrient levels, chlorophyll *a,* fecal coliform bacteria (a group of bacteria associated with human, other mammal, and bird wastes), algae, zooplankton, benthic infauna, an fish. Once monitoring data are obtained, they need to be summarized and pollution sources prioritized for control.

Most lake water quality problems are associated with an overabundance of nutrients, which results in excessive plant growth. In managing such water quality problems, it is important to assess what nutrient limits plant growth. In eutrophic lakes, phosphorus is often the limiting nutrient; this means that the amount of phosphorus in the lake will determine the amount of plant growth. Therefore, most lake management strategies focus on reducing phosphorus loading.

If the lake's phosphorus budget shows that most of the phosphorus is coming from within the lake, in-lake restoration techniques should help to reduce phosphorus levels and make the lake less eutrophic. On the other hand, if most of the phosphorus is coming from the watershed (this is often true of small lakes with very large watersheds), then the focus should be on watershed best management practices (BMPs) to control sources of nutrients. In some lakes, phosphorus comes from within the lake and the watershed, so both types of actions are needed.

Watershed Best Management Practices Implementation of watershed best management practices (BMPs) improves water quality by reducing the quantity of pollutants entering the lake. Most pollutants within a watershed result from human activities. Pollutants originating on each parcel of land within a watershed can collectively become a serious threat to the receiving water quality. BMPs are structural and nonstructural methods, including common sense "housekeeping measures," used to prevent or reduce pollution by controlling erosion, surface water runoff, sources of nutrients, and sources of toxic chemicals. Watershed BMPs can be basin-wide or can target management of developed property. These measures can include native plant revegetation of lake shorelines, retention/detention ponds and biofiltration swales for stormwater treatment, and homeowner/business owner BMPs to enhance water quality through better landscaping methods, alternative household and gardening practices, better animal-keeping practices, drainage controls, and septic system maintenance and repairs. Local and state agencies can work in partnership with lake associations and other citizen groups in a watershed to educate residents and business owners about BMPs that are inexpensive, easy to implement, and make a difference in protecting lake water quality and aquatic biota. Ideally, this environmental education should include hands-on water quality activities (e.g., storm drain stenciling, lakeshore revegetation) and habitat monitoring activities for community volunteers including schoolchildren because these activities impart a sense of lake stewardship to people who live on or upstream of a lake. Following is a discussion of each type of BMP.

Many lakes have no native plants growing on the shoreline; houses may have manicured lawns leading to the water's edge. Where shoreline vegetation is absent, surface water runoff enters the lake directly, degrading lake water quality.

Ducks and geese may also graze on the shoreline and affect lake water quality via the nutrients and fecal coliform bacteria in their wastes. Native plantings along lake shorelines and streambanks of creeks that are tributary to lakes serve multiple functions: improving wildlife habitat; acting as a physical barrier to intrusion by ducks and geese; increasing shoreline soil stability thereby preventing erosion; and moderating impacts of surface water runoff by filtering out suspended solids, nutrients, and toxic chemicals.

Alternatives to standard lawn maintenance and landscaping practices include minimal use of fertilizers, reduction in lawn size, regular thatching and aeration, incorporation of native plants in new landscaping, soil enhancement through mulching and composting rather than chemical fertilizers, and integrated pest management techniques rather than chemical pesticides.

Household hazardous wastes should be properly disposed of at collection sites and never in storm drains. Homeowners should be educated about non-toxic alternatives to common household cleaning products.

When cars are washed near storm drains, wash water carries oils, greases, nutrients, heavy metals, suspended solids, and soaps to local water bodies including lakes and their tributaries. Residents of a lake watershed should be encouraged to wash their cars at commercial car wash facilities, which discharge wash water to the sanitary sewer system. People who do wash their cars at home should be informed about draining wash water to vegetated areas such as lawns, using a high pressure nozzle with trigger to minimize water usage, and using commercial products that clean vehicles without water.

The feeding of waterfowl by lakeside residents should be discouraged. Pet and domestic animal waste should be properly disposed of away from a lake and surface water pathways that reach a lake.

Business owners should be educated about BMPs, such as proper storage of toxic chemicals and proper maintenance and repair of oil-water separators in order to prevent the discharge of petroleum hydrocarbons, metals, and other toxics to lakes and their tributaries. Many local and state agencies provide technical assistance to businesses in preventing or reducing the discharge of pollutants to lakes and streams.

State and local agencies such as conservation districts should also educate agricultural landowners about agricultural BMPs to improve pastures, maintain healthy livestock, dispose of or recycle livestock water, restrict livestock access to lakes and their tributaries (e.g., building fences around streams which are inlets to lakes), and prevent discharge of pollutants from livestock waste and farm operations to lakes and tributaries. Cost-sharing incentives and technical assistance increase the success rate of these measures.

Maintaining on-site wastewater treatment systems (septic systems) in good working order is another way to reduce phosphorus loading to a lake. Lakeside and watershed residents and business owners should know about septic system operation and maintenance practices, such as using no- or low-phosphate detergents, composting organic wastes rather than using garbage disposals, selecting and maintaining optimal vegetative cover over drainfields, and inspecting and cleaning the system on a regular basis to ensure proper system functioning. This can be accomplished through articles in lake association newsletters, through brochures, and through workshops conducted by local and state agencies.

In-Lake Restoration Techniques In-lake restoration techniques that can be used to control internal phosphorus loading are phosphorus inactivation and precipitation (e.g., aluminum sulfate treatment), removing sediments from the lake bottom (dredging), hypolimnetic aeration, hypolimnetic withdrawal, dilution, and artificial circulation. Following is a discussion of the principles, advantages and disadvantages of each technique.

Adding aluminum sulfate (alum) to a lake reduces the lake's phosphorus content by precipitating phosphorus and retarding its release from the sediments (Cooke *et al.*, 1993a). When alum is added to the water column, a polymer forms that binds phosphorus and organic matter. The aluminum phosphate-hydroxide compound (commonly called alum floc) is insoluble and settles to the lake bottom. Dramatic increases in water clarity typically occur immediately following an alum treatment, as suspended and colloidal particles are removed from the water column by the floc.

Alum has been used extensively in the United States, with general success in controlling phosphorus release from lake sediments for several years (Cooke *et al.*, 1993a; Garrison and Knauer, 1984). If external sources of phosphorus are not controlled, the effectiveness of alum will decrease with time as the alum layer on the sediments becomes covered by nutrient-rich silt and organic material. The lake may therefore need to be treated again. Regular long-term monitoring is required in an alum-treated lake to evaluate the effectiveness of the treatment.

The alum dose should be based on the pH and alkalinity of the lake, and the potential toxicity of aluminum to the lake (Cooke *et al.*, 1986, 1993a; Kennedy and Cooke, 1982). As alum is added to a lake, pH and alkalinity decrease and dissolved aluminum concentrations increase; alkaline lakes can tolerate higher alum doses than can softwater lakes. Adding alum to a lake with low to moderate alkalinity requires careful planning to ensure that pH and alkalinity are not lowered to levels that would stress aquatic biota. The use of sodium aluminate as a buffer permits a greater alum dose to be used. Such buffering agents have been applied with alum in several northeastern lakes and high success in maintaining normal lake pH and alkalinity levels (Cobbossee Watershed District, 1988; Dominie, 1978). The use of sodium carbonate in the alum treatment of Long Lake in western Washington was also highly successful in maintaining safe pH and alkalinity levels, as well as in improving lake water quality (KCM, 1994).

Alum is a promising technique for reducing algae through physical settling and removal during the application and through the long-term control of internal nutrient loading. The treatment does not kill the algae instantaneously in the water column, but settles them on the lake bottom, where they die over a period of up to two weeks. This longer time period and the location at the lake bottom greatly reduce the hazard from toxins that might be released by the dying algal cells. Alum can provide long-term reduction in the occurrence of algal toxicity if internal phosphorus loading is reduced. Alum has also been found to reduce the sediment-to-water migration of blue-green algae in Green Lake in Seattle (KCM, 1994).

The use of alum salts may cause toxic conditions. Alum causes zooplankton to flocculate and settle out of the water column, along with sediment and phytoplankton, which can stress the food chain of a lake. To date, alum treatments have not resulted in adverse effects on fish and have not damaged invertebrate populations in well-buffered lakes (Cooke *et al.*, 1993a; Narf, 1990). Invertebrate populations may, however, be more sensitive to alum application in softwater lakes. For example, the alum/sodium aluminate treatment of Lake Morey in Vermont (alkalinity = 30 to 50 milligrams of calcium carbonate per liter) resulted in a short-term decrease in density and species diversity of benthic invertebrates (Smelzer, 1990).

Although most case studies of alum treatments demonstrate multiple-year success, failures have also occurred. These have been attributed to insufficient dose, lake mixing, inadequate reduction in external nutrient inputs, and a high coverage of macrophytes.

Other nutrient inactivation techniques have been used with less success than alum. Calcium hydroxide (lime) has recently been used in hardwater Alberta, Canada lakes to control nutrient supply and algal growth (Murphy *et al.*, 1990; Kenefick *et al.*, 1992). However, lime would not offer the same phosphorus-binding benefit in softwater lakes (Cooke *et al.*, 1993a).

The release of nutrients from lake sediments can also be controlled by removing the layer of the most highly enriched materials. This may result in significantly lower in-lake nutrient concentrations and less algal production. Several types of dredging equipment can be used to remove sediments from lakes; a hydraulic dredge equipped with a cutterhead is the most common choice. The cutter loosens sediments that are then transported as a slurry of 80 to 90 percent water through a pipeline that traverses the lake from the dredging site to a remote disposal area. In the United States, a permit from the U.S. Army Corps of Engineers is normally required before sediments can be dredged from a lake (Cooke *et al.*, 1993b; U.S. EPA, 1990).

Sediment removal to retard nutrient release can be highly effective. For example, in Lake Trummen (Sweden), the upper, nutrient-rich layer of sediments was removed, increasing the lake depth from 3.6 feet to 5.8 feet. The sediment was disposed of in diked-off bays and upland ponds. Return flow from the ponds was treated with alum to remove phosphorus. The total phosphorus concentration in the lake dropped sharply (U.S. EPA, 1990).

However, sediment removal has high potential for serious negative impacts on the treated lake and its surrounding watershed. The disposal area must be sufficiently large to handle the high volume of turbid, nutrient-rich water that accompanies the sediments. Unless the sediment-water slurry can be retained long enough for settling to occur. The turbid, nutrient-rich runoff water will enter the lake outlet and end up in a tributary stream or another lake downstream of the treated lake. Turbidity, algal blooms, and dissolved oxygen depletion may result in the receiving waters (Cooke *et al.*, 1993b; U.S. EPA, 1990).

Prior to dredging, the lake sediments must be analyzed for heavy metals (especially copper and arsenic, which have been extensively used as herbicides), chlorinated hydrocarbons (which have been used in pesticides), and other potentially toxic chemicals. Special precautions will be required if these substances are present in high concentrations (Cooke *et al.*, 1993b; U.S. EPA, 1990).

Another technique for preventing the release of phosphorus from lake sediments to the water column is hypolimnetic aeration. This technique involves oxygenating the bottom waters of a lake without causing destratification. Air is used to raise cold hyplimnetic water in a tube to the surface of deep lakes, where the water is aerated through contact with the atmosphere, loses gases such as carbon dioxide and methane, and is then returned to the hypolimnion. Phosphorus release from the sediments is limited by hypolimnetic aeration if there is sufficient iron in solution to bind phosphorus in the re-oxygenated waters. Aeration oxidizes the soluble ferrous phosphate to insoluble ferric phosphate, which would then precipitate out into the sediments and remain there. In addition, hypolimnetic aeration increases habitat and food supply for coldwater fish species. The technique has been used with varying levels of success. Unsuccessful treatments have been attributed to inadequate oxygen supplies, disruption of lake stratification, or lack of sufficient iron (Cooke *et al.*, 1993b; KCM, 1994; U.S. EPA, 1990).

It is important that hypolimnetic aeration not destratify the water column. Premature destratification (e.g., before fall turnover of the lake) can be stressful and become toxic to aquatic life when bottom waters with little dissolved oxygen, low pH, and high concentrations of toxic gases mix with surface waters. Destratification can also stimulate algal growth by supplying hypolimnetic nutrients to surface waters and mixing algae throughout the water column. In shallow lakes, destratification could occur due to wind mixing (KCM, 1994).

An alternative to aerating the hypolimnion is to remove this nutrient-rich, anoxic water layer either through a deep outlet in a dam or by a siphon, thereby accelerating a lake's phosphorus loss and perhaps producing a decrease in phosphorus concentration in surface waters. There are few documented case histories of hypolimnetic withdrawal (Cooke *et al.*, 1993b; Nurnberg, 1987; U.S. EPA, 1990).

There are major disadvantages to hypolimnetic withdrawal. The hypolimnion water that is discharged may be of poor quality and therefore may require aeration or other treatment. Federal, state or local regulatory agencies may require a permit to discharge this water. Hypolimnetic withdrawal could destratify the water column, thereby introducing nutrient-rich, oxygen-poor water to the surface of the lake and triggering on algal bloom (Cooke *et al.*, 1993b; Nurnberg, 1987; U.S. EPA, 1990).

Another approach to reducing the phosphorus concentration in eutrophic lakes is to dilute the lake water with sufficient quantities of another water source that is low in phosphorus; algal cells will be flushed out of the lake at the same time. When water low in phosphorus is added to the inflow, the actual phosphorus loading will increase, but the mean phosphorus concentration will decrease, depending upon initial flushing rate and inflow concentration. Concentration will also be affected by the degree to which loss of phosphorus to sediments decreases and counters the dilution. Lakes with low initial flushing rates are poor candidates for this technique because in-lake concentration could increase unless the dilution water is essentially devoid of phosphorus. Internal phosphorus release could further complicate the effort (Cooke *et al.*, 1993b; U.S. EPA, 1990).

Flushing can control algal biomass by cell washout; however, the flushing rate must be near the cell growth rate to be effective. Flushing rates of 10 to 15 percent of the lake volume per day are believed to be sufficient (U.S. EPA, 1990).

There are very few documented case histories of dilution of flushing because additional water is seldom available, especially water that is low in nutrients. One successful example is Moses Lake in eastern Washington. Low-nutrient Columbia River water was diverted through the lake. Daily water exchange rates of 10 to 20 percent were achieved in this eutrophic lake. Lake transparency dramatically increased and algal blooms dramatically decreased (Welch and Patmont, 1980).

For dilution and flushing to be successful, lake outlet structures must be capable of handling the added discharge. The increased volume of water released downstream could have negative effects. Water used for dilution or flushing must be tested to ensure that no toxics are present before the water is introduced into the eutorphic lake.

Another in-lake restoration technique is artificial circulation. This eliminates or prevents thermal stratification, through the injection of compressed air into lake water from a pipe or ceramic diffuser at the lake's bottom. The artificial circulation structure must be designed properly to ensure an air flow of about 1/3 cubic foot per minute per acre of lake surface; this is required to maintain oxygen within the lake. Algal blooms may be controlled through one or more of the following processes. First mixing of algae to the lake's bottom will decrease their time in full light, leading to reduced net photosynthesis. Introduction of dissolved oxygen to the bottom of a lake may inhibit phosphorus release from the sediments (i.e., have the same impact as hypolimnetic aeration), hence curtailing internal phosphorus loading. A third possible process is that rapid circulation and contact of lake water with the air, as well as the introduction of carbon dioxide-rich bottom water during the initial period of mixing, can increase the water's carbon dioxide content and lower the pH, leading to a shift from blue-green algae to less noxious green algae. Finally, when zooplankton are mixed to the lake's bottom, they are less vulnerable to planktivorous fish. If more of the zooplankton survive, then they may eat more algae (Cooke *et al.*, 1993b; U.S. EPA, 1990).

Results of artificial circulation have been highly variable. In about half the lakes where this technique has been attempted and where temperature differences are small between surface and bottom waters during the summer, algal blooms have been reduced. In other cases, phosphorus and turbidity have increased and water transparency has decreased (U.S. EPA, 1990).

Management of Aquatic Macrophytes

The watershed BMPs outlined above for reducing the quantity of algae in a lake are also effective in reducing the quantity of aquatic macrophytes in a lake. Watershed BMPs involve voluntary changes in behavior and are easy and inexpensive to implement. Any reduction in nutrient loading to a lake as a result of BMPs can maintain or extend the effectiveness of in-lake methods for managing aquatic macrophytes. The disadvantage of watershed BMPs is that they will not result in immediate, substantial reduction in nuisance aquatic plant growth because habitat has already been created in the lake that supports aquatic plant growth. Therefore, watershed BMPs usually must be combined with other methods to reduce aquatic plant growth.

There are physical, chemical, and biological methods for reducing aquatic plant growth in lakes and restoring balance to the lake ecosystem so that the aquatic plants are beneficial rather than harmful. Following is a description of examples, principles, advantages, and disadvantages of each method, summarized from *Aquatic Plant Control* (Washington State Department of Ecology, 1994), *A Citizens' Manual for Developing Integrated Aquatic Vegetation Management Plans* (Gibbons, Gibbons, and Systma, 1994), and *Crary WeedRoller Pilot Project Report* (Cooke, 1996).

Physical Methods Physical methods of reducing the amount of aquatic plants in a lake include hand-pulling, hand and mechanical cutting, mechanical harvesting, bottom barriers (sediment covers), water level drawdown, water column dyes, rotovating, diver-operated suction dredging, and weed rolling.

Hand-pulling aquatic plants is similar to pulling weeds out of a garden. This method involves digging out the entire plant with a spade or long knife and disposing of the residue onshore. In waters deeper than three feet, hand removal can best be accomplished by snorkelers or scuba divers carrying collection bags for plant disposal. The technique results in immediate clearing of the water column of nuisance plants and is most appropriate for small-area, low-plant density treatment, e.g., clearing pondweed from areas around docks and beaches.

Hand and mechanical cutting differ from hand-pulling in that plants are cut below the water surface (roots are usually not removed) with scythes, rakes or other specialized devices that can be pulled through the weed beds by boat or people. Rakes can be equipped with floats to allow easier plant and fragment collection. Mechanical cutters can be battery-operated and hand-held, portable and mounted on boats, or specialized underwater cutters using a sickle to cut weeds in water as shallow as ten inches and as deep as five feet. Cutting results in immediate removal of nuisance submerged plants.

Hand-pulling, hand-cutting, and mechanical cutting are inexpensive, easy to implement around docks and swimming areas, environmentally safe, and allow removal of undesirable aquatic plants while leaving desirable plants. On the other hand, these methods are labor-intensive, time-consuming, and may need to be repeated several times each summer. It may be difficult for the laborer to see and dig out all plant roots. Some plants, such as water lilies, are difficult to cut or pull. Visibility may become obscured by turbidity generated by cutting activities. All plant fragments must be removed from the lake to prevent them from rerooting or drifting onshore. This is particularly important in the case of invasive, non-native plants such as Eurasian watermilfoil. Environmental impacts of hand-pulling and cutting include short-term, localized increases in water turbidity and some disruption of benthic infauna.

Mechanical harvesting is a short-term technique to temporarily remove plants that interfere with recreational uses or aesthetic enjoyment of a lake. Mechanical harvesters are large machines which cut plants below the water surface and collect the plants and plant fragments for disposal. Harvested plants are removed from the water by a conveyor belt system and stored on the harvester until disposal. A barge stationed near the harvesting site for temporary plant storage in an efficient disposal method. Alternatively, the harvester can carry cut weeds to shore for disposal in landfills or for use as compost. Harvesting is usually performed in late spring, summer, and early fall when submersed and floating-leafed plants have reached or are close to the surface of the lake. Harvesters can cut and collect several acres per day depending on plant type, plant density, and the storage capacity of the equipment. Depending on the equipment used, the plants are cut from 5 to 10 feet below the water's surface in a swatch that is 6 to 20 feet wide. Harvesting is most appropriately used for large, open areas with few surface obstructions.

Like mechanical cutting, harvesting results in immediate open areas of water and can be targeted to specific locations. The lake can continue to be used for recreational purposes while harvesting is underway. Another advantage of mechanical harvesting is that removing plants from the water also eliminates a possible source of nutrients often released during fill dieback and decay. Furthermore, harvesting can reduce sediment accumulation by removing organic matter that normally decays and adds to the bottom sediments. Harvested vegetation can often be easily composted and used to enrich soil.

Mechanical harvesters share the previously mentioned disadvantages of creating plant fragments and requiring repeated application. Several disadvantages are unique to mechanical harvesting. Off-loading sites and disposal areas for cut plants must be available. On heavily developed shorelines, suitable off-loading sites may be few and require long trips by the harvester. Some large harvesters are not easily maneuverable in shallow water or around docks or other obstructions. Furthermore, harvesting can be detrimental to non-target plants, insects, and small fish; these are often removed from the lake along with target plants.

The use of bottom barriers (sediment covers) is an effective physical method of plant control. Barrier material is placed like a blanket over the lake bottom to prevent plants from growing. Applications of bottom barriers can be made up to any depth, with divers often utilized for deeper water placement. Readily available materials such as burlap, plastics, synthetic rubber, polypropylene, perforated black mylar, fiberglass screens, woven polyester, and nylon film can all be used as bottom barriers. There are also commercial bottom barriers which are specifically designed for aquatic plant control. These include Texel, a heavy, felt-like polyester material, and Aquascreen, a polyvinylchloride-coated fiberglass mesh which resembles a window screen.

The ideal bottom barrier should be durable, reduce or block light, prevent plants from growing into and under the fabric, be easy to install and maintain, and should readily allow gases produced by rotting weeds to escape without "ballooning" the fabric upwards. Even the most porous materials, such as window screen, will billow due to gas buildup. Therefore, it is very important to anchor the bottom barrier securely to the bottom of the lake. Unsecured barriers can create safety hazards for swimmers and boaters.

The duration of plant control depends on the type of material used, application techniques, sediment composition (bottom barriers are difficult to place on deep muck sediments), the rate that plants can grow through or on top of the bottom barrier, and the rate that new sediment is deposited on the barrier. Installation of bottom barriers is easiest in winter or early spring when plants have died back. In summer, plants should be hand-pulled or cut first in order to facilitate bottom barrier installation. Bottom barriers may also be attached to frames rather than placed directly onto the sediment. The frames may then be moved for control of a larger area.

Bottom barriers can provide immediate removal of nuisance plant conditions upon placement. It is easy to install bottom barriers in small, confined areas such as around docks, moorages or beaches. Other advantages of bottom barriers are that they are hidden from view, do not interfere with shoreline use, and do not result in significant production of plant fragments.

Disadvantages of bottom barriers include high cost of some materials, suitability only for localized plant control, possible regrowth of plants from above or below the barrier, and the need for regular inspection and maintenance to remove accumulations of sediment and any rooting plant fragments. Bottom barriers can also cause localized decreases in the populations of benthic infauna such as aquatic insects.

A fifth type of physical plant control is water level drawdown. This involves exposing plants and root systems to prolonged freezing and drying or hot, dry conditions to kill the plants. Drawdown is usually performed during the winter and is more common in management of aquatic macrophytes in reservoirs and ponds than in natural lakes. Accurate identification of target plant species is important because aquatic plants vary greatly in terms of susceptibility to drawdown.

In addition to controlling aquatic plant biomass, drawing down the water level makes it possible to use several other lake restoration or improvement procedures. For example, water level drawdown can be used for fish management to repair structures such as docks or dams, to facilitate localized dredging or bottom barrier placement, or to remove stumps or debris. This technique can result in compaction of certain types of sediments, such as mucky substrates and thus improve shoreline use. Drawdown can reduce nearshore vegetation, thereby reducing potential inputs of nutrients to the water from seasonal dieback of aquatic plants. Drawdown can also be used to attract waterfowl by enhancing growth of emergent plants such as cattails and bulrushes.

A disadvantage of water level drawdown is that it is not species-selective; hence beneficial plants may be removed along with nuisance plants. Wetlands adjacent to the lake can be exposed with adverse impacts on their plant and associated animal communities. Prolonged freezing and drying can kill benthic infauna that are important food sources for planktivorous fish. Lowering the water level in a lake may result in decreased levels of dissolved oxygen, with resulting negative impacts on fish and other aquatic biota. During the period of drawdown, recreational use of the lake may be limited or unavailable.

A sixth physical plant-control method is the application of dark-colored dyes to reduce the amount of light reaching the submersed plants, thereby shading the plants from the sunlight needed for photosynthesis. Several commercial dye products are available; they impart a blue color to the water. Best results are obtained when the dye is used early in the growing season.

Advantages of using water column dyes for aquatic plant control are that no special equipment is needed for application (the dye can be poured into the water by hand from shoreline or boat) and the dyes are non-toxic to aquatic organisms, livestock, and humans. However, there are several drawbacks. Water column dyes can be used only in shallow water bodies with no outlet and are less effective when aquatic plant growth is within two feet of the lake surface. Repeat dye treatments may be necessary throughout the plant-growing season. Water column dyes should not be applied to lakes used for drinking water.

Another method for reducing aquatic plant biomass in a lake is rotovation (bottom derooting or underwater bottom tillage). Rotovators use underwater rototiller-like blades to uproot aquatic plants. The rotating blades turn seven to nine inches deep into the lake bottom to dislodge plant roots. Plants and roots may then be removed from the water using a weed rake attachment to the rototiller head, a harvester or manual collection. This technique is most suitable for use in larger lakes because of the larger size and high costs of the equipment.

Rotovation can be used year-round but is most effective in the winter and early spring when plants have died back. Depending on plant density and sediment type, two to three acres per day can be rotovated. Rotovation is particularly effective in controlling Eurasian watermilfoil and can produce a high level of milfoil control for two to three growing seasons.

Advantages of rotovation are that it can remove the entire plant, can decrease density of undesirable plants, and may stimulate the growth of some desirable native plants. Rotovated areas in the state of Washington and the province of British Columbia have shown increases in species diversity of native plants, with resulting benefits to fisheries. Fish are not removed through rototilling as they are by mechanical harvesters. Since rotovating takes place during winter and early spring, there is no interference with peak summertime lake recreational activities.

On the other hand, rotovation is limited to areas with few bottom obstructions and should not be used where water intakes are located. Bottom sediments are disturbed which can result in short-term impacts on water quality and benthic infauna. Plant nutrients or toxic chemicals in the sediments may be released into the water. Since rotovation is not species-selective, beneficial plants may inadvertently be removed. Rotovation may also interfere with fish spawning or migration. Some rotovators are difficult to maneuver around docks and in shallow water. Plant fragmentation resulting from rotovation may increase the spread of invasive weeds like milfoil.

Another physical method for controlling aquatic plant growth in lakes is diver-operated suction dredging. This method was used in the late 1970s in British Columbia as an improvement to hand removal of sparse colonies of Eurasian watermilfoil. Scuba divers operate portable dredges with suction heads to uproot and remove individual plants from the lake sediment. After the divers physically remove the plants with sharp tools, the plant/sediment slurry is then suctioned up and carried back to a barge through hoses operated by the diver. Plants parts are sieved out on the barge and retained for later off-site disposal. The water-sediment slurry can be discharged back to the water or piped off-site for upland disposal.

Efficiency of plant removal is dependent on sediment condition, density of aquatic plants, and underwater visibility. Diver-operated suction dredging is best used for localized infestations of low plant density where plant fragmentation must be minimized. Therefore, this technique has great potential for milfoil control and can remove 85 to 97 percent of milfoil from a lake.

Advantages of diver-operated suction dredging are that it is species-selective and site-specific. Disruption of sediments and plant fragmentation are both minimized. The method can be used to cover areas larger than practicable for hand pulling or cutting, and can be conducted in tight places or around obstacles that would preclude use of larger machinery.

Drawbacks to this method are that it is labor-intensive and expensive. The usefulness of this method may be reduced in dense plant beds. Returning the water-sediment slurry directly to the lake may result in some loss of plant fragments. If the dredged slurry is disposed of upland, more specialized equipment and materials are required and the process is much more costly. Short-term environmental impacts include localized increases in turbidity and release of nutrients and other contaminants from the sediments. Some sediments, benthic infauna, and non-target plants may also be inadvertently removed during this process.

A new mechanical method of controlling aquatic plant growth is weed rolling. The method uses a commercially available, low-voltage power unit that drives an up-to-25-foot long roller set on the lake bottom through an adjustable arc of up-to 270°. A vertical drive head mounts to a dock. A reversing action built into the drive automatically brings the roller back to complete the cycle. The 25 feet of roller includes flexible couplers to follow the contour of the lake bottom. The device operates on the principle of the "well-worn path." Whole plants or plant stems and leaves can be removed. Fins on the rollers detach some plants from the soil, while the rollers force other plants flat, gradually inhibiting growth. Detached plants and plant fragments need to be removed from the water with a rake or net.

Once plants are cleared from the area, the weed rolling device can be used weekly or less often to prevent regrowth. The device requires little maintenance, lasts for at least five years, and can be moved from dock to dock which allows sharing by a lake association.

A weed rolling device was tested on three lakes in King County, Washington, during the summer of 1995 and was found to be easy to operate and effective in reducing the density of both milfoil and water lilies. Weed rolling has also been effective in removing unwanted aquatic plants from many Midwestern lakes.

Advantages of weed rolling are that it creates and maintains areas of open water adjacent to docks, installation is simple, operating costs are low (similar to the costs of using a 75 watt lightbulb), and the treatment area can be modified by varying the number of roller tube sections used and adjusting the roller tube travel arc.

There are several disadvantages. Although it is easy to operate a weed rolling device, collection of the resulting plant fragments is labor-intensive. If the plant fragments are not collected, invasive plants may be spread from one area of the lake to the other. Weed rolling may disturb some bottom-dwelling animals and may interfere with fish spawning. There are also some safety considerations. People cannot be allowed in the water near where the weed rolling device is operating. After each use, the rollers must be unplugged from the power source, moved, and stored under or alongside a dock.

Chemical Methods Chemical control of aquatic plants in lakes is accomplished through the use of aquatic herbicides. Systemic herbicides are absorbed by and translocated throughout the entire plant, and kill the entire plant. Contact herbicides kill the parts of the plant with which they come in contact, leaving roots alive and capable of regrowth. To be most effective, herbicides must be applied to plants during the period when they are growing most rapidly. Because of environmental risks from improper application, aquatic herbicide use is regulated. Some expertise in using herbicides is necessary in order to be successful and avoid unwanted impacts. Generally, applicators must be licensed by and obtain permits from a local, state or federal agency. A certain percentage of all aquatic plants in each lake must usually remain untreated in order to provide food and habitat for fish and wildlife.

Some aquatic herbicides that are registered with the U.S. Environmental Protection Agency include glyphosate, fluridone, endothall, and copper compounds. Glyphosate (commercial name Rodeo) is a non-selective, broad-spectrum herbicide used to control floating-leafed plants like water lilies and emergent plants like purple loosestrife and cattails. It is generally sprayed or painted in liquid form on plant leaves. Symptoms of herbicidal activity are apparent after seven days and include wilting and yellowing of plants, followed by complete browning and death. Since glyphosate is a systemic herbicide, it is effective in long-term plant control. It dissipates quickly from natural waters, is low in toxicity to benthic infauna, fish, birds, and mammals, and does not prohibit use of a lake for swimming, fishing, and irrigation while treatment of plants is underway. The major drawback of glyphosate is that it can affect non-target plant species. Careful application to prevent drift will minimize this possibility.

Fluridone (commercial name Sonar) is a slow-acting systemic herbicide used to control Eurasian watermilfoil and other submersed plants. Applied to the lake water as either a liquid or pellet, this chemical begins to show effects 7–10 days after application, with full control of target plants often requiring 6–12 weeks. Fluridone acts by damaging plant chlorophyll and preventing photo-synthesis; hence dying plants exhibit retarded growth and bleached out leaves before falling to the lake sediments and decomposing. The best time for fluridone application is early in the growing season for the target plant, usually spring or early summer, because of the amount of time required for full plant control.

As is the case with glyphosate, long-term plant control can be achieved with the use of fluridone. Fluridone has a very low order of toxicity of aquatic animals and humans. Disadvantages of its use are that it is very slow-acting and therefore not effective in flowing water. Fluridone can drift out of the treatment zone, thereby affecting non-target plants. Consequently, it is most suitable for whole-lake treatments, not for treating a defined area within a large, open lake. As the affected plants decay, they may consume dissolved oxygen from and release nutrients to the water column. Furthermore, fluridone-treated water may result in injury to irrigated vegetation. Therefore, use of lake water for irrigation is delayed following treatment. To protect drinking water sources, fluridone should not be applied within 1/4 mile of a lake water intake.

Endothall (commercial name Aquathol) is a fast-acting contact herbicide which is applied in either a granular or liquid form and destroys the plant stems and leaves but does not kill the roots. It is used for short-term control of aquatic plants, i.e., a few weeks to a few months, with no carryover to the next growth season. Advantages of endothall are that plant death occurs in one to two weeks and there is little or no drift impact from proper application of this chemical. On the other hand, non-target plant impacts are possible because many aquatic plants are susceptible to endothall. Following treatment, there is often a period of time when lake waters cannot be used for swimming, fishing, or crop irrigation.

Copper sulfate and chelated copper compounds are generally used to control algal blooms. Copper is an essential element, required in small amounts, for plant growth. High concentrations of this metal can lead to inhibition of photosynthesis and to plant death. The effectiveness of copper chelates is enhanced by warm temperatures and sunlight, conditions that stimulate copper uptake by sensitive plants. The effect of copper treatment can be observed within 10 days, with full effects manifested in four to six weeks. Several treatments may be needed each season. An advantage of using copper compounds for aquatic plant control is that there are no restrictions in lake uses following treatment; copper compounds can even be used in lakes that are drinking water supplies. However, copper compounds must be used with extreme care because copper is persistent in the environment. Yearly application of copper to lakes can result in elevated copper concentrations in sediments, where the copper can then be taken up by benthic infauna and fish. The toxicity of copper to fish is higher in soft than in hard water.

Biological Methods The development and use of biological methods of aquatic plant control is in its infancy. Interest in biological control agents was stimulated by a desire to find more "natural" means of long-term control of nuisance aquatic plants as well as reduce the use of expensive equipment or chemicals. The biological control method that is used in many states is stocking the lake with sterile grass carp (white amur), a vegetarian fish native to large rivers of China and Russia. The fish are sterile because they are triploid, i.e., they have an extra set of chromosomes. Grass carp were first introduced into the United States in 1963 by the state of Arkansas and are now legal in most states. The objective of grass carp use is to end up with a lake that has 20 to 40 percent plant cover, not a lake that is totally devoid of plants.

Grass carp are rapidly growing fish that live for at least 10 years. In general, they reach at least 10 pounds in weight and have been known to reach 40 pounds in the southern U.S. They feed from the top of the plant down and therefore do not stir up lake sediments. They are most appropriately used for lakewide, low-intensity control of submersed plants such as water weed, water celery, and certain pondweeds. Milfoil is less preferred and water lilies and watershield are not eaten at all. Grass carp are dormant during the winter and start intensive feeding when water temperatures reach 68°F.

The appropriate stocking rate is usually determined by the fisheries department of the particular state in which the lake is located and will depend on the amount and type of plants in the lake as well as spring and summer water temperatures. Survival rates of the fish will vary depending on fish disease and presence of predators such as ospreys and otters.

Grass carp are inexpensive compared to some other aquatic plant control methods and offer long-term control. However, there are numerous disadvantages to their use. There is no control over how much the grass carp will eat. Overstocking of grass carp could result in eradication of beneficial as well as nuisance plants or eradication of all plants in the lake. Removing excess fish is difficult and expensive. If native plants are not growing in a lake, other plants such as invasive, non-native plants or algae will move in to fill the void. Furthermore, the grass carp will add nutrients to the lake through their wastes and through the decay of their bodies when they die; this will contribute to the growth of other plants in the lake.

In addition to no control over how much the grass carp will eat, there is no control over where they will graze. The fish may avoid areas of the lake experiencing heavy recreational use (e.g., swimming docks), resulting in less plant removal than desired for those areas. Substantial removal of vegetation by grass carp may not become apparent until three to five years after introduction. Furthermore, all inlets and outlets to the lake must be screened to prevent grass carp from migrating out of the lake into streams, rivers, and other lakes with potential impacts on downstream, non-target plants.

REFERENCES

Carlson, R.E. 1977. "A Trophic State Index for Lakes." *Limnology and Oceanography* **22**, 361–368.

Cobbossee Watershed District. 1988. *Cochnewagon Lake Restoration Project: Hypolimnetic Alum Treatment.* Preliminary Report No. 1: 1985–1988. Cobbossee Watershed District, Winthrop, ME.

Cooke, G.D. and R.E. Charlson. 1986. "Water Quality Management in a Drinking Water Reservoir." *Lake and Reservoir Management* **2**, 363–371.

Cooke, G.D., E.B. Welch, S.A. Peterson, and P.R. Newroth. 1986. *Lake and Reservoir Restoration.* Butterworths, Boston, MA.

Cooke, G.D., E.B. Welch, A.B. Martin, D.G. Fulmer, J.B. Hyde, and G.D. Schrieve. 1993a. "Effectiveness of Al, Ca, and Fe Salts for Control of Internal Phosphorus Loading in Shallow and Deep Lakes." *Hydrobiologia* **253**, 323–335.

Cooke, G.D., E.B., Welch, S.A. Peterson, and P.R. Newroth. 1993b. *Restoration and Management of Lakes and Reservoirs.* Lewis Publishers, Boca Raton, FL.

Cooke, W. 1996. *Crary Weed Roller Pilot Project Report.* King County Surface Water Management Division, Seattle, WA.

Dominie, D.R. 1978. *Cobbossee Watershed District Lake Restoration Report.* Progress Report No. 4. Cobbossee Watershed District, East Winthrop, ME.

Garrison, P.J. and D.R. Knauer. 1984. "Long-Term Evaluation of Three Alum-Treated Lakes." *Lake and Reservoir Management,* pp. 513–517. EPA-440/5-84-001, U.S. Environmental Protection Agency, Washington, D.C.

Gibbons, M.V., H.L. Gibbons, and M.D. Sytsma. 1994. *A Citizens' Manual for Developing Integrated Aquatic Vegetation Management Plane.* First Edition. Washington State Department of Ecology, Olympia, WA.

Gulland, J.A. 1970. "Food Chain Studies and Some Problems in World Fisheries." Pp. 296–315 in J.H. Steele, ed., *Marine Food Chains.* University of California Press, Los Angeles.

KCM, Inc. 1994. *Cottage Lake Nutrient Budget.* Prepared for King County Surface Water Management Division, Seattle, WA.

Kenefick, S.L., S.E. Hurdey, E.E. Prepas, N. Motkosky, and H.G. Peterson. 1992. "Odorous Substances and Cyanobacterial Toxins in Prairie Drinking Water Sources." *Water Science and Technology* **25**(2), 147–154.

Kennedy, R.H. and G.D. Cooke, 1982. "Control of Lake Phosphorus with Aluminum Sulfate. Dose Determination and Application Techniques." *Water Research Bulletin* **18**, 389–395.

Kozlovsky, D.G. 1968. "A Critical Evaluation of the Trophic Level Concept. I. Ecological Efficiencies." *Ecology* **49**, 48–60.

Lorenzen, M.W. and A.W. Fast. 1977. *A Guide to Aeration/Circulation Techniques for Lake Management.* EPA-600/3-77-004, U.S. Environmental Protection Agency, Washington, D.C.

Murphy, T.P., E.E. Prepas, J.T. Lim, J.M. Crosby, and D.T. Waltz. 1990. "Evaluation of Calcium Carbonate and Calcium Hydroxide Treatment of Prairie Drinking Water Dugouts." *Lake and Reservoir Management* **6**(1), 101–108.

Narf, R.P. 1990. "Interactions of Chironomidae and Chaoboridae (Diptera) with Aluminum Sulfate Treated Lake Sediments." *Lake and Reservoir Management* **6**(1), 33–42.

Nurnberg, G.K. 1987. "Hypolimnetic Withdrawal as a Lake Restoration Technique." *Journal of the Environmental Engineering Division of the American Society of Civil Engineers* **113**, 1006–1017.

Porecella, D.B., S.A. Peterson, and D.P. Larsen. 1980. "Index to Evaluate Lake Restoration." *Journal of Environmental Engineering* **106**, 1151–1169.

Smelzer, E. 1990. "A Successful Alum/Sodium Aluminate Treatment of Lake Morey, Vermont." *Lake and Reservoir Management.* **6**(1), 9–19.

Solomon, F., S. Walton, K. Easthouse, D. Bouchard, and P. Noppe. 1996. *Cottage Lake Management Plan.* King County Surface Water Management Division and KCM, Inc., Seattle, WA.

United States Environmental Protection Agency (U.S. EPA). 1990. *The Lake and Reservoir Restoration Guidance Manual.* Second Edition, EPA-440/4-90-006. Prepared by the North American Lake Management Society for the U.S. EPA, Office of Water Assessment and Watershed Protection Division, Nonpoint Sources Branch, Washington, D.C.

Uttormork, P.D. and M.L. Hutchins. 1980. "Input-output models as decision aids for lake restoration." *Water Research Bulletin* **16**, 494–500.

Vollenweider, R.A. 1975. "Input-Output Models with Special Reference to the Phosphorus Loading Concept in Limnology." *Schweizerische Zeitschrift fur Hydrologie* **37**, 53–84.

Washington State Department of Ecology. 1994. *Aquatic Plant Control.* WQFA-94-107. Olympia, WA.

Water Environmental Services, Inc. and KCM, Inc. 1995. *Lake Twelve Integrated Aquatic Plant Management Plan.* Prepared for the Lake Twelve Association and the King County Surface Water Management Division, Seattle, WA.

Welch, E.B. 1992. *Ecological Effects on Wastewater.* 2nd Edition. Chapman and Hall, New York.

Welch, E.B. and C.R. Patmont. 1980. "Lake Restoration by Dilution: Moses Lake, Washington." *Water Research* **14**, 1317–1325.

Wetzel, R.G. 1983. *Limnology.* 2nd Edition. Saunders College Publishing, New York.

FRANCES PAULA SOLOMON, PH.D.
King County Department of Natural Resources
Seattle, Washington

FIFTH EDITION

ENCYCLOPEDIA OF ENVIRONMENTAL SCIENCE and ENGINEERING

VOLUME 2 M-Z

EDITED BY
JAMES R. PFAFFLIN
EDWARD N. ZIEGLER

Taylor & Francis
Taylor & Francis Group

Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the
Taylor & Francis Group, the academic division of T&F Informa plc.

M

MANAGEMENT OF RADIOACTIVE WASTES

RADIOACTIVE WASTE

Radioactive waste may be defined as solid, liquid, or gaseous material of negligible economic value containing radionuclides in excess of threshold quantities. High level wastes (HLW) are produced in the first cycle of reprocessing spent nuclear material and are strongly radioactive. Intermediate level wastes (ILW) can be divided into short lived, with half lives of twenty years or less, and long lived, in which the half lives of some constituents may be thousands of years. Low level wastes (LLW) contain less than 4 GBq/ton of alpha emitters and less than 12 GBq/ton of beta and gamma emitters. Very low level waste (VLLW) contains activity concentrations less than 0.4 MBq/ton.

ACTIVITY AND EXPOSURE

The Becquerel (Bq) is the activity of one radionuclide having one spontaneous disintegration per second. One Curie (Ci) is defined as 3.7×10^{10} disintegrations per second. The Becquerel is the more commonly used unit. The unit of ionizing radiation which corresponds to energy absorption of 100 ergs per gram is the rad (roentgen-absorption-dose). The newer unit is the Gray (Gy), which is equal to 100 rads. The amount of radiation which produces energy dissipation in the human body equivalent to one roentgen of X-rays is the rem (roentgen-equivalent-man). One Sievert is equal to 100 rems and is the commonly accepted unit.

Philosophy

The group of people engaged in management of radioactive wastes has evolved from a small body of operators who, originally with little or no expert knowledge, were engaged in day-to-day solution of unpleasant problems. They now form a recognized profession, extending from whole-time research scientists to field workers who in some countries are conducting a profit-making industry.

The members of the profession came mainly from Health Physics and brought with them the caution and "conservative" attitude to radiation hazards characteristic of Health Physicists. They regard their mission as being to ensure that members of the public, as well as workers in the field of nuclear energy, will not be harmed by the radioactive material for which they are responsible. With their Health Physics background this sometimes leads to an attitude which industry regards as overrestrictive, although recent controversies have tended to cast them ironically in the role of particularly dangerous polluters of the environment.

It is clear that any human activity that involves conversion of something into something different must produce waste. Conversion of energy from one from to another is no exception. It is sometimes possible for an industry to recycle its waste products and to convert part of them to a useful form, but there is always some minimal residue which cannot be retained within the system. This must find some place within the environment. Usually the cheapest procedure is to discharge it in some way that will ensure a sufficient dilution to make it innocuous. If this is impracticable for technical or political reasons it must be confined, but usually the more effective the confinement, the higher the cost. To say that a process must be conducted with now waste is equivalent to saying that the process may not be conducted at all, and to demand a certain level of confinement or restriction of wastes implies an acceptance of the cost of the waste management system as a necessary part of the cost of the process.

Discharge of potentially noxious materials into the environment involves some risk, which may or may not be measurable. Within very broad limits research in nuclear hazards enables us to forecast the effects of exposure of large groups of people, for extended periods, to low doses of radiation. We can also estimate, with less accuracy, the probability that an individual will suffer some harm from such exposure, and we can say with much greater confidence what will happen if an individual is exposed to larger doses—say 50 rem and upwards—in a single dose. The nuclear industry, then, can provide some information on the probable consequences of environmental contamination extended over a lifetime, and better information on the probable consequences of a major nuclear accident which leads to high radiation exposure. In other words we can, within rather broad limits, estimate the risks.

The situation is different in most other industries. The consequences of acute doses of cyanide, lead, fluoride or carbon tetrachloride are well known, and there is some evidence for the effects from low doses received over a lifetime, but who knows what effect to expect in humans from benzpyrene or nitrous oxide emitted from smoke stacks or from the low levels of polychlorodiphenyls and mercury compounds that are liberated into the environment? They affect every age group in the population and are a potential lifelong hazard. But nothing is known about the probability that they will eventually do harm, and it is difficult to see how such knowledge could be obtained in a human population.

Every human activity is associated with some risk, however small. Normally we do not solemnly calculate the risk, weigh it against the benefit we expect to obtain, and then decide for or against the activity. Yet to decide to do something—such as driving a car, getting up in the morning, or going mountain climbing—must involve some sort of conscious or unconscious weighing of risk against benefit.

In deciding upon a particular waste management system, or in deciding to license a particular kind of nuclear power station, a much more deliberate weighing of cost vs. benefit must be undertaken. There is, however, a fundamental difficulty which up to now has made it impossible to express such a judgment in numbers. It is characteristic of a ratio that the numerator and the denominator must be in the same units. It should be possible to express most of the benefits of nuclear power, for example, in dollars, but if we regard part of the cost of nuclear power as an increase in the probability that people will develop cancer or that they will experience a shortened lifetime, how can that be expressed in dollars?

One benefit of nuclear power is the difference between death and injury among uranium miners and processors and the corresponding figure for equivalent energy production by the coal mining industry. This, again, cannot be expressed in dollars. To work out a true COST/BENEFIT ratio is thus little better than a dream, and the people responsible for approving a waste management system or a new power station are therefore faced in the last analysis with a value judgment, which is at least to some extent subjective. It is not a scientific decision. In the broadest sense, the decision is political.

Controls

The responsibility for making decisions on matters related to "dealing in"—i.e. having anything to do with—radioactive materials, machines capable of producing electromagnetic radiation (expect for medical purposes) and certain scheduled materials such as heavy water, usually rests with a national atomic energy authority. Typically, regulations are issued by the authority that have the force of law. Assistance is given to the authority in assessing hazards of reactors and other installations—including waste management systems—by an independent advisory committee which can call on the services of an expert staff.

In most countries regulations lay down the maximum permissible exposure to radiation for workers in nuclear industry and also for the general population. Maximum permissible doses (MPDs) have been recommended by the International Commission on Radiological Protection (ICRP), which have received worldwide acceptance as the fundamental basis for national regulations. The ICRP has derived from the MPDs a list of maximum permissible concentrations (MPCs) in air and water on the basis that if workers were to breathe air, or drink water, at the MPC for any particular radionuclide over a lifetime they would not suffer any unacceptable harm.

"Unacceptable" means "detectable", in the sense that it could reasonably be regarded as caused by the radiation. The ICRP has also laid down rules for calculating the MPC for mixtures of more than one radionuclide.

The MPDs are constantly under review by the ICRP, which consists of people who have devoted their professional lives to assessment of radiation hazards. They drawn upon the work of large numbers of scientists throughout the world, many of whom are actively engaged in research on somatic and genetic effects of radiation. Changes have been made from time to time in details of the ICRP recommendations but it is remarkable that in such a rapidly developing field the necessary changes have been so few.

The ICRP has consistently emphasized that the MPD and its associated MPs are *maximum permissible* figures. The Commission has made another recommendation equal in force and status to those on maximum permissible doses. This states that exposure to radiation must always be held down to the lowest PRACTICABLE dose. The world "practicable" was carefully chosen, after considerable debate. If "possible" had been used it could have been claimed that a single contaminated rat must be buried in a platinum box. It is our mission to see that all practicable steps are taken to protect mankind from exposure to radiation, and we can do that very effectively.

SOURCES OF WASTES

Uranium Mining and Milling

Apart from the normal hazards associated with hard-rock mining, the workers in uranium mines are exposed to radon and the decay products which arise from the radium content of the ore. These hazards can be controlled by sealing old workings and general "good house-keeping", but more particularly by installation of an efficient ventilation system and, where necessary, the use of respirators. The ventilation air contains radioactive material and dust, some of which can be removed if necessary by filtration, but the radon remains. The large volume of air used for mine ventilation is ejected at high velocity from a stack, which ensures adequate dilution into the atmosphere.

The end products of the mill are uranium oxide and "tailings". The tailings, together with mine drainage water, contain most of the radium originally present in the ore. Radium is

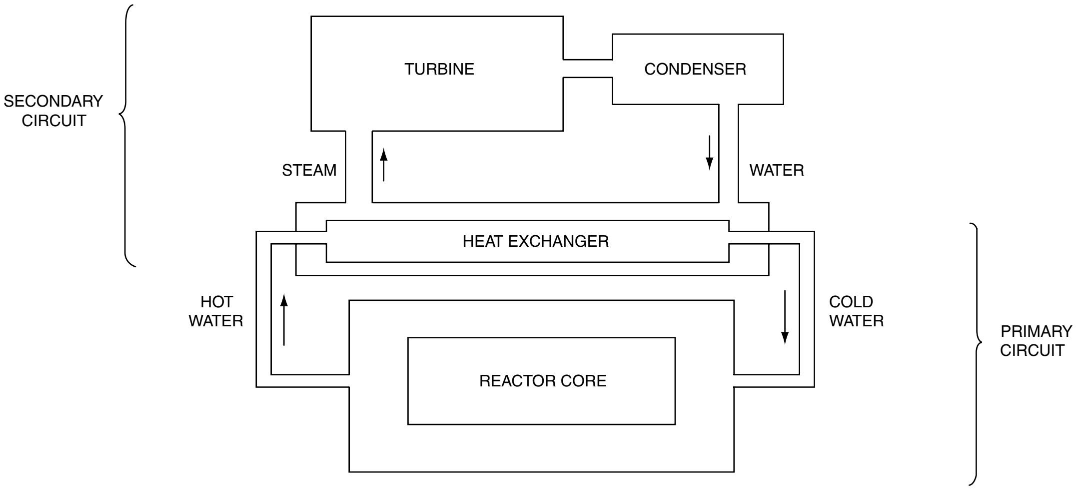


FIGURE 1 Schematic diagram of processes in nuclear power station. Nearly all radioactivity remains inside the fuel, which is inside the core, which is inside the primary circuit.

one of the most toxic of all radionuclides and presents a serious potential hazard. Various methods of treatment, such as co-precipitation with barium, render most of the radium insoluble. But the water draining from tailings ponds often contains more radium than is permissible in drinking water.

Proper design of outfalls into suitable bodies of water can ensure adequate dilution, but vigilance is necessary to prevent rupture of the tailings ponds or improper practices that will nullify or bypass the treatment system. A monitoring system for analysis of downstream water and fish is common today, but in the early days of the industry the dangers were little understood or ignored, with the result that lakes and streams in uranium mining areas became contaminated.

In Canada the existence of a problem was recognized in time to avert a public hazard, but the Report of a Deputy Minister's Committee showed that action was necessary to protect the environment in the Elliott Lake and Bancroft areas. This was particularly urgent as greatly increased activity in uranium mining was anticipated within a few years.

The size of the problem can be judged from the fact that a Congressional Hearing was told that 12,000,000 gallons of water containing nearly 10 g of radium was discharged daily to the tailings ponds of American uranium mills.

Processing of Uranium Oxide

The crude (70%) U_3O_8 produced by the mills may be converted to metal, to UO_2 or to UF_6. The hexafluoride is used in separation of ^{235}U from ^{238}U. A serious waste problem would result from nuclear fission if a critically large amount of ^{235}U were to accumulate accidentally in one place. This is a rare event, but is not impossible. Otherwise, the wastes consist of uranium chips and fines, contaminated clothing and respirators and dust accumulated in air-cleaning systems. The uranium at this stage is practically free from radium so it is hardly a radioactive hazard. The toxicity of natural uranium or ^{238}U is that of a toxic metal rather than of a radionuclide.

Uranium metal is produced by converting the dioxide to tetrafluoride which is then reduced to the metal at high temperature with magnesium. The waste form this process—magnesium fluoride slag and uranium metal fines from trimming the ingots—is a normal slag disposal problem since it is sparingly soluble in water.

Fuel Fabrication

There are many different kinds of fuel elements, but their manufacture produces little waste beyond dust and faulty pellets or fuel pins. This material is usually recycled, particularly if it contains added ^{235}U.

Reactor Wastes

An operating reactor contains a very large inventory of fission products. A 500 MW (thermal) reactor, after operating for 180 days, contains four hundred million curies for fission products, measured one day after shutdown. This is equivalent to the activity of about 400 metric tons of radium. The fission products decay rapidly at first, leaving 80 million curies at the end of a week, and more slowly later. After a month, the inventory is reduced to about 8 million curies.

Nuclear power stations rated at 1000 MW (electrical)—i.e. 3000 to 5000 MW thermal—are not unusual. At first sight it would seem that these plants would be enormous potential sources of radioactive wastes, but in practice this is not the case (Figure 1). In an operating power reactor the fuel is contained within a non-corrodible cladding—usually zirconium or stainless steel—and the fission products cannot get out unless the cladding is ruptured.

It is possible to operate the reactor with defects in a few fuel elements, but these sources of leakage make the primary cooling circuit radioactive. It is impracticable to operate a station in the presence of high radiation fields, so the primary coolant is continually purified by ion exchangers. Again, it is a practical necessity to renew the ion exchangers after they have developed a certain level of radiation. The net result of these considerations is that for reasons of operator safety and economics the presence of more than a small proportion of ruptured fuel in a reactor will require its removal.

Fuel removed from the reactor is normally stored on site for a considerable time to permit decay of shortlived radioactivity. Storage facilities are usually deep tanks filled with water, which acts simultaneously as coolant and radiation shield. If defective fuel is present the water will rapidly become contaminated, but even if there are no defects in the cladding the water in cooling ponds does not remain free from radioactive material. This is because the cladding and the reactor structure contribute neutron activation products (or corrosion products) to the cooling water and the cladding itself always contains minute traces of uranium, which undergoes fission in the reactor. Hence, the pond water must be purified, usually by resin ion exchangers, so these resins also become a waste.

If resins are regenerated, the regenerants (acids, alkalis, or salts) will appear as a liquid waste for disposal. Otherwise, the resin will be handled within its original container or as a powder or slurry.

The radioactive content of gaseous effluents from reactors depends upon the design of the reactor. If air passes through the core very large amounts of argon-41 may be emitted from the stack. Although ^{41}Ar is a hard gamma emitter it has a short half-life (about two hours) so its effects are only noticeable within or very near to the plant. Radioactive isotopes of nitrogen and oxygen decay so rapidly that they do not reach the stack in appreciable amount and the long-lived carbon-14 is not produced in sufficient amount to be hazardous at the present scale of nuclear power generation. Some concern has, however, been expressed that by the end of this century the buildup of ^{14}C in the atmosphere might become a significant source of radiation within the biosphere.

More concern attaches to radioactive krypton, ^{85}Kr, with a half-life of 10.4 years. This, in contrast with ^{41}Ar and ^{14}C, is a fission product. It is liberated via fuel defects and by diffusion through fuel cladding. It is not a hazard from any single plant, but with increasing numbers of nuclear power stations it might become an ubiquitous source of low-level radiation, though the source of most of the ^{85}Kr would be spent fuel processing plants rather than power stations.

Similar concern has been expressed regarding tritium, the radioactive isotope of hydrogen, which is produced within the fuel and by neutron activation of the heavy hydrogen in ordinary water or the D_2O coolant and moderator of heavy-water reactors. It is also formed by neutron activation of lithium, sometimes used as a neutralising agent in reactor coolants, or of boron which functions as a "poison" in some reactor control systems.

Sometimes the significance of a "source" of radioactive waste depends on whether one is considering the safety of people within the plant, or the public outside. For example, ruptured fuel elements or ordinary day-to-day type mechanical failures can produce air-borne radioactive iodines and other fission products which are a nuisance to operators because they have to work in plastic suits and respirators. The ventilation filtration system and the high dispersion capability of the atmosphere combine to make sources of this kind insignificant beyond the boundary of the exclusion area. However, they may reduce efficiency and disrupt work schedules within the station very seriously, and give rise to significant disposals in the form of clean-up solutions, contaminated clothing, mopheads and metal scrap.

A noteworthy source of this nature is the tritium which builds up in the coolant and moderator of heavy-water reactors. In a 1000 MW (electrical) power station the equilibrium tritium concentration in the moderator is about 50 Ci/litre. This leads to stack discharges which are quite negligible, but any leaks in pump seals, valves or pipe joints within the station would produce operating problems for those responsible for the radiation safety of the staff. On the other hand, material sent for waste disposal would be no problem, partly because heavy water is recovered for economic reasons and partly because the maximum permissible concentrations of tritium in air and water are much higher than those of most other radionuclides.

In summary, in spite of the enormous potential source of radionuclides within an operating power station the amount of waste generated is small compared with that arising from a research and development establishment, and minute in comparison with a plant fuel processing plant. This statement covers normal operation, including the ordinary accidents and malfunctions expected in any well-designed plant. It does not include the consequences of the "Maximum Credible Accident" which is, in fact, so improbable that designers of waste management systems do not normally make provision for it.

However, the accident at the Chernobyl Nuclear Power Station in 1986 was particularly sensational. A reactor exploded and caught fire, releasing an estimated 30 million Curies. Half of the resulting fallout was within 30 kilometers of the plant. The remainder spread over much of Europe. There was great economic loss and many cancer deaths were attributed to the incident.

Spent Fuel Processing

Wastes arising from processing of spent fuel account for more than 99.9% of the "waste disposal problem". Fuel which has been enriched with ^{235}U must be treated for recovery of unburned ^{235}U because the fission product load of spent fuel reduces its efficiency as a source of energy. It ceases to be economic as fuel long before the expensive ^{235}U is exhausted.

After removal from the reactor, and storage for sufficient time for decay of short-lived fission products, the fuel is de-sheathed and dissolved, usually in strong nitric acid (Figure 2).Uranium and plutonium are extracted into an organic solvent, and the acid solution of fission products left behind forms the high level or primary waste. Washing of the organic extractant produces Medium Level wastes, whereas Low Level waste consists of further washings, cooling water, scrubber water and liquids from other sources too numerous to catalogue.

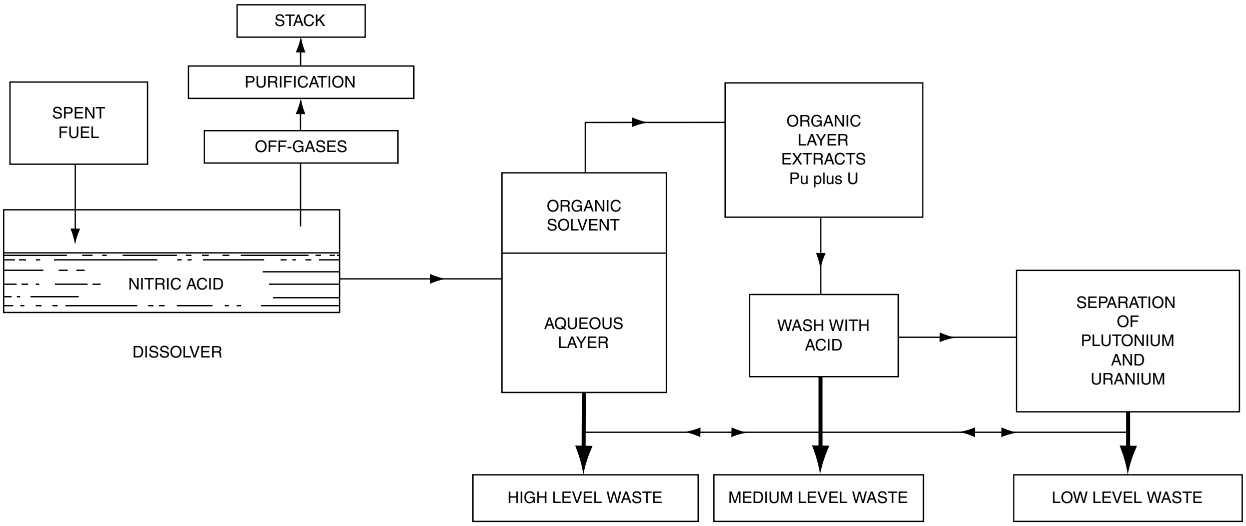


FIGURE 2 Schematic diagram of fuel processing plant. Showing origins of main waste streams. Reactor fuel contains over 99.95% of the total radionuclides eventually disposed of as waste.

As long ago as 1959 fifty million gallons of High Level wastes were stored in stainless steel tanks at Hanford (USA) alone. The radionuclides in solution generate so much decay heat that many of the tanks boil, making the provision of elaborate off-gas cleaning systems necessary. Some high level waste tanks have ruptured, but since they are constructed on a cup-and-saucer principle, with adequate monitoring for spills, and spare tankage is kept available, no unexpected contamination problems have arisen.

Gases from the dissolvers and storage tanks contain tritium, bromides, iodines, xenon, krypton and smaller amounts of less volatile elements such as ruthenium and cesium. After storage for decay, scrubbing and filtration, off-gases can be liberated from a tall stack. As mentioned in the section on reactors, proliferation of fuel processing plants in the future might conceivably lead to local or even eventual world-wide atmospheric contamination if improved containment is not provided in time at spent fuel processing sites.

Solid waste may include glasses or ceramics, used as a means for fixing the activity in high-level liquid wastes, and bitumen or concrete blocks containing less active material. Products of waste processing such as sludges, evaporator bottoms, incinerator ash, absorbers, filters and scrap fuel cladding are usually in the medium level category. Worn and failed equipment such as pipes, tanks and valves, unserviceable protective clothing, cleanup material and even whole buildings may have a variety of levels of contamination, by numerous different radionuclides, which defies quantitative assessment. This is not a serious difficulty, except for administrative and recording purposes when quantitative reports have to be made, because most of these wastes have to be contained in some way and none of them are dumped into the environment.

The most difficult problem for the fuel processing industry is not high or medium level waste, offgases or heterogeneous contaminated scrap. The real problem is very low level liquid waste, because it arises in such enormous volume. Coming from numerous different sources—e.g. cooling and final wash waters, laundry and decontamination center effluents, floor drainage from cleanup operations, personnel shower drainage and effluent from the final stages of liquid waste purification plants—low level and “essentially uncontaminated but suspect” waste adds up to billions of gallons per year. Although some countries (Sweden and Japan, for example) evaporate such effluents on a large scale they are usually discharged by some route into the environment.

Research and Development

A wide variety of wastes arises in such research establishments as Brookhaven (USA), Chalk River (Canada) or Harwell (UK) and the include many of the types mentioned under the heading of fuel processing. In addition the research reactors usually produce very large quantities of radioisotopes which may be processed onsite. However, the quantities involved are very much lower, especially in the high level category, and elaborate waste processing systems are seldom needed even at large research centers unless they are situated in built-up areas or immediately over important aquifers.

Hospitals and Biological Laboratories

Organic material and excreta makes wastes from these institutions difficult to handle. The radioactive content is usually small, and limited to a restricted list of radionuclides. Those used as sealed sources seldom appear as waste, and the rest are practically confined to ^{131}I, ^{32}P, ^{59}Fe, ^{51}Cr, ^{35}S and ^{24}Na. Other nuclides may be used in small amounts for special purposes such as specific location in certain organs. The nature and amount of radionuclides used in these institutions are such that a high proportion of the waste can be handled safely by the municipal sewage and garbage systems.

Sealed sources are, however, a very difficult matter. While they remain sealed they are usually within heavy shielding in teletherapy machines, which are only operated by competent people, or they are in the form of needles and plaques for implantation, or instrumental standard sources used by specialists. However, the time comes when such sources have decayed to the point where they are no longer useful. Sufficient activity remains for them to be highly dangerous to the unwary, so they are dealt with in special ways, usually after return to the supplier.

Isotope Production Plants

These facilities are often associated with large reactors, and wastes are similar to those generated in Research and Development plants. Processing of very large sources of volatile elements such as iodine and tellurium necessitates an elaborate ventilation cleaning system. Manufacture of large sources of ^{90}Sr, ^{137}Cs or the trans-uranic elements as power sources may call for sophisticated remote handling equipment in heavily shielded cells. But the waste problems are difficult only in scale from those encountered in an *R* and *D* plant.

Some people have considered the separation of ^{90}Sr and ^{137}Cs from fuel processing wastes as a helpful step in their management. Removal of these nuclides leaves a mixture which, during 20 years' storage, would decrease in activity by a factor of about 30,000. However, an industry handling the fission products from 50 tons of ^{235}U burned in one year would have to deal with 500,000,000 Curies of separated ^{90}Sr and about the same amount of ^{137}Cs. It might be difficult to find a market for sources of this scale unless they were cheap, and it must be remembered that they would eventually come back as "waste."

Industrial Applications

Use of radioisotopes in industry is not a significant source of wastes. Most industrial sources are sealed, and nearly all unsealed sources are short-lived.

Transportation

Ships are the only form of transportation using nuclear reactors as a source of power. They include naval ships, ice breakers and merchant vessels. They contain large amounts of fission products within the reactors, but as a source of waste they are not important, except possibly in some harbours and inshore waters.

During start-up of the reactor the secondary coolant expands and the limited space in submarines necessitates the dumping of this expansion water. In common with landbased reactor coolant it contains radioactive corrosion products and tritium. The coolant is maintained at a low level of activity by means of ion exchangers, which become waste eventually. Normally this material is disposed of on land, although it has been shown by the Brynielsson Panel of the International Atomic Energy Agency that resin from a fleet of as many as 300 nuclear ships could be dumped safely if this were done only on the high seas.

Apart from these sources wastes from nuclear shipping consist of clean-up solutions, laboratory wastes, laundry effluent and other minor sources common to all reactor operations. Except in submarines, practically all wastes can if necessary be retained on board for disposal ashore.

DISPOSAL PRINCIPLES

There are two main procedures available for disposal—Concentration and Confinement: or Dilution and Dispersion.

a) If wastes are truly confined, in the sense that in no credible circumstances could they be liberated into the environment, then the only additional requirement is "perpetual custody" to ensure that the confinement is never broken. This is easier said than done. In the field of high level wastes when we say "perpetual" we are speaking in terms of thousands of years. Few private firms go back for 100 years, political regimes have seldom lasted for as long as 500 years, and there are few civilizations that have survived for 2000 years. In our own day forecasters tend to regard dates beyond 2000 AD as being in the distant future. What, then, can we do about "perpetual custody" of wastes containing, for example, plutonium with a half-life of 24,000 years?

This is not a fanciful dilemma. A story from Chalk River will illustrate the point. When the Canadians decided to concentrate on natural uranium heavy water reactors for power production it became apparent that processing of spent fuel would be uneconomic until the price of uranium or plutonium rose considerably. Processing was therefore stopped, but the wastes accumulated during the pilot plant operation had to be disposed of. A considerable volume of medium level waste was mixed with cement in steel drums and enclosed within solid concrete monoliths below ground in the waste management area (Figure 3).The question then arose "What if some archeologist digs this structure up 1000 years from now and thinks it is an ancient temple or tomb?" Eventually someone suggested that its true nature should be inlaid in non-corrodible metal on the top of the monolith. Dr. A. J. Cipriani, who had listened to the debate in silence, then asked "In what language?"

The implications of this question are profound. Some of the wastes for which we are responsible will still be radioactive after our present civilization has disappeared and perhaps been forgotten. So far as we know there is no practicable solution to the problem. The best we can do is ensure that the nature, amount and location of all major disposals

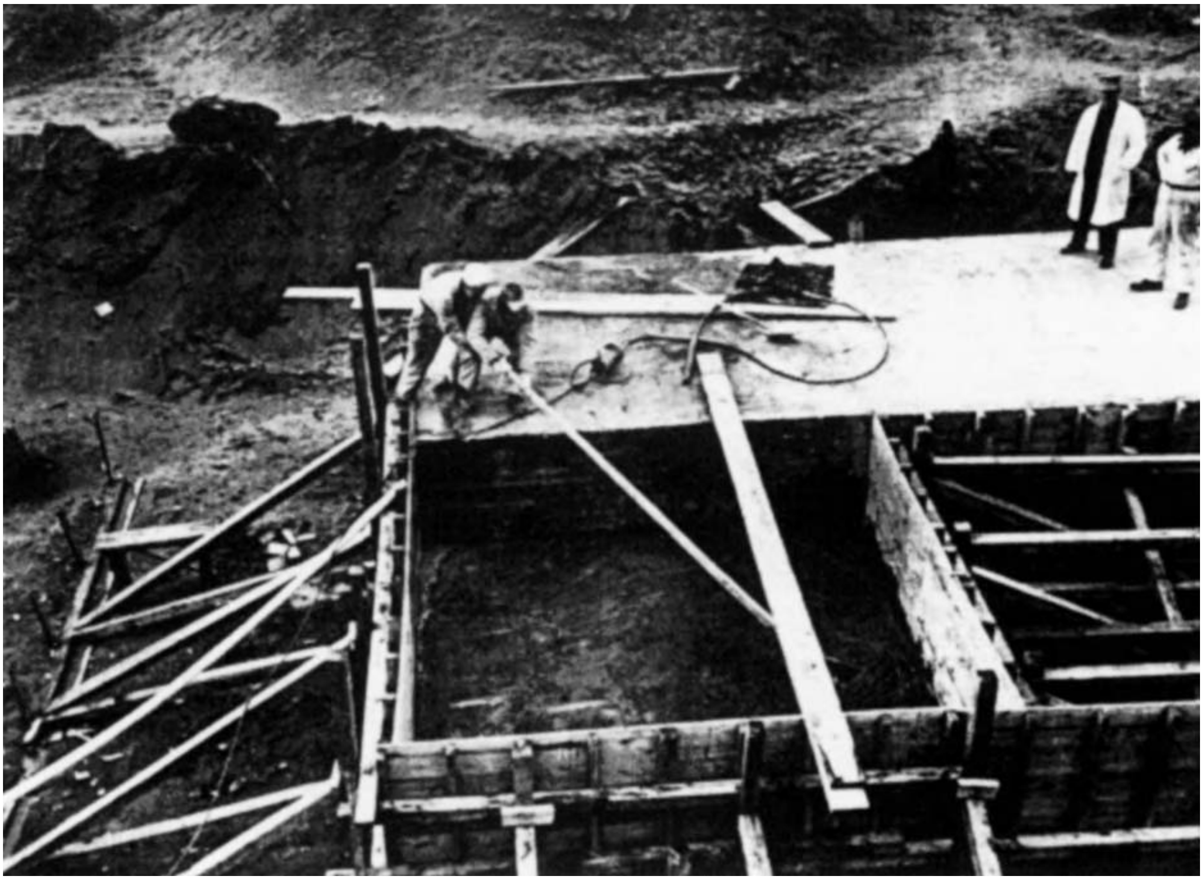


FIGURE 3 Pouring a concrete monolith. Steel drums filled with waste, solidified by mixing with cement, were stacked on concrete slabs surrounded with forms. The forms were filled with concrete. The monoliths were about 2 m below ground level.

are recorded in the nearest approximation we have to a perpetual repository of archives—a government department. Beyond that we can only rely on folk memory. After all, farmers in Europe have been ploughing around Neolithic tumuli and prehistoric roads for thousands of years for no good reason known to them, except that it was accepted to be the right thing to do.

b) Dilution and dispersion is the traditional method that men have always used for dealing with their wastes. Until recently it seemed to work fairly well unless populations became very concentrated, but it is now becoming clear that there are so many people that the system is showing signs of breaking down. It depends upon the capacity of the environment to dilute or detoxify the wastes to a level that is innocuous to man and to organisms of interest to man. We are still a very long way from contaminating our environment with radioactivity to a point where radiation effects are observable, even in close proximity to nuclear enterprises, but we must maintain vigilance to ensure that slow and subtle changes do not occur which escape our notice until it is too late.

Safety in discharge to the environment depends upon three factors—(1) Dispersion by such means as atmospheric dilution, mixing into big bodies of water, or spreading through large volumes of soil. (2) Fixation of radionuclides on soil minerals and organic detritus. (3) Decay of radionuclides, dispersed or fixed, before they are able to affect man.

The principle of dispersion has one logical trap into which regulatory bodies have sometimes fallen. In some countries the discharge of liquid and gaseous wastes is limited by the concentration in the effluent pipe or the concentration at the stack mouth. This is based upon the assumption that if the concentration is limited to the maximum permissible value, all will be well. However, the “dilution capacity” of a river is a function of the number of Curies per day put into the river, divided by the daily flow of water. If an operator wishes to dispose of double the amount of waste, and he is limited only by the concentration in the effluent pipe, he need simply double the amount of water flowing in the pipe. But the downstream effect will be a doubling in the concentration, unless he has doubled the flow in the river.

> For this reason, limitations must be made in Curies per unit time, not in micro-curies per millilitre, and account must be taken of volume of river flow if this is seasonally variable. Regulations set on the basis of concentration at the point of discharge only protect people close to the discharge point.

DISPOSAL PRACTICES

Gases

Radioactive gases arise mainly in reactors, spent fuel processing, isotope production, and research and development facilities. The general principles are the same for all procedures that depend upon dispersion into the atmosphere.

If we have a stack that is emitting Q Curies/sec., the concentration C at a given distance downwind will be KQ. The parameter K is a very complex function which depends upon wind speed and direction, weather conditions, stack height, topographical features, variability of temperature with height, velocity and buoyancy of the effluent and other conditions. Values of K for a range of conditions can be calculated from equations proposed by Sutton (1947), Pasquill (1961) and Holland (1953). These equations have been used to calculate the permissible emissions from stacks by inserting appropriate numbers and parameters applying to unfavourable weather conditions likely to obtain at the site. The permissible emission rate has been set at a value which would ensure that populations downwind would not be exposed to more than an agreed maximum radiation dose rate.

The classical equations have been based on statistical theory with empirical values for the diffusion parameters being obtained from experimental work which has sometimes had little relation to real emissions from actual stacks.

Returning to the superficially simple equation $C = KQ$, it is apparent that if we could observe, over a long period of time, the maximum value of C ever attained per unit emission rate, we could define a figure $K_{\max}$ which was not likely to be exceeded. With a sufficient number of observations of C and Q, extended over a sufficient variety of weather conditions, we could estimate the probability that our value $K_{\max}$ could ever be exceeded.

When a maximum permissible concentration is set for a noxious substance the decision really depends upon a belief that the probability of damage is so low that it is acceptable. If, then, C is set at the *MPC* at a given distance from the stack, and $K_{\max}$ is known for that distance, then Q_p, the maximum permissible release rate, is determined.

It has been shown by Barry that $K_{\max}$ is not very dependent upon topography or climate, because it depends mainly on rather large-scale behaviour of the atmosphere, and the frequency of most adverse conditions normally experienced do not vary grossly from one place to another.

The maximum permissible emission rate—or in some cases the *MPC* at the stack mouth—is given in the regulations governing the plant or laboratory. It is then the responsibility of the operator to ensure that emissions are kept as far below the permissible level as may be practicable. Numerous methods are available, other than variation of stack height, for achieving this end (Figure 4).

Filtration It is advisable to filter contaminated air near to the source of the activity. This reduces the amount of air to be filtered and also cuts down the "plating-out" of radionuclides on the duct-work, which can be a source of radiation fields with the plant.

Filters must be suitable for the job they are supposed to do. They should be made of non-flammable material such as glass or other fibre and should be tested before and after installation. If fine (e.g. "Absolute") filters are used it is often necessary to precede them with a coarse filter to avoid rapid clogging with dust.

Filters must be very efficient to be adequate for fuel processing plants and incinerators burning highly active waste. For example, a sand filter at Hanford capable of passing 10,000 m^3/min had an efficiency of more than 99.5%, but this was inadequate. The necessary efficiency of 99.99% was attained with a bed of glass fibers 100 cm thick.

Electrostatic Precipitators Small airborne particles are usually electrically charged. The charge can be increased by passing the air through a corona discharge, or through a charged fabric screen. The particles are attracted to a surface carrying the opposite charge, from which they can be removed mechanically. It is possible to use the same principle by imposing a charge on filters.

Steam Ejector Nozzles The most efficient air cleaning device other than "Absolute" filters consists of a nozzle in which the air is mixed with steam and expelled into an expansion chamber where the steam condenses on the particles. After passing through a second construction into another expansion chamber, where the air is scrubbed with water jets, removal efficiency for 0.3 micron particles is 99.9%.

Incinerator Off-gases The hot gas from an incinerator carriers with it fly ash, tars and water vapour as well as particles. Tars may be removed and the gases cooled by water scrubbing devices. Water droplets must then be eliminated by reheating or passage through a "cyclone". This is a cylinder with a conical bottom. Gas injected tangentially at the top sets up a vortex which causes deposition of particles on the sides.

In smaller incinerators the gases are cooled and some fly ash is removed by passage through a cooling chamber fitted with baffles. After this stage a roughing or "bag" filter is used, followed if necessary by Absolute or charcoal filters.

Processing Plant Gases The devices required for cleaning gaseous effluent depend on the nature of the process. Off-gas from boiling high level wastes must be passed through condensers and scrubbers to recover nitric acid as well as to remove volatile radionuclides. However, these and other air cleaning equipment previously mentioned will not remove gases such as ^{85}Kr, nor hold back all of the radioactive halogens.

Radioactive iodine in molecular form is fairly easily absorbed by alkaline scrubbers and copper or silver mesh filters, but in the form of methyl iodine it can only be arrested by an activated charcoal filter. These filters have to be kept cool, not only to remove the decay-heat of adsorbed halogens

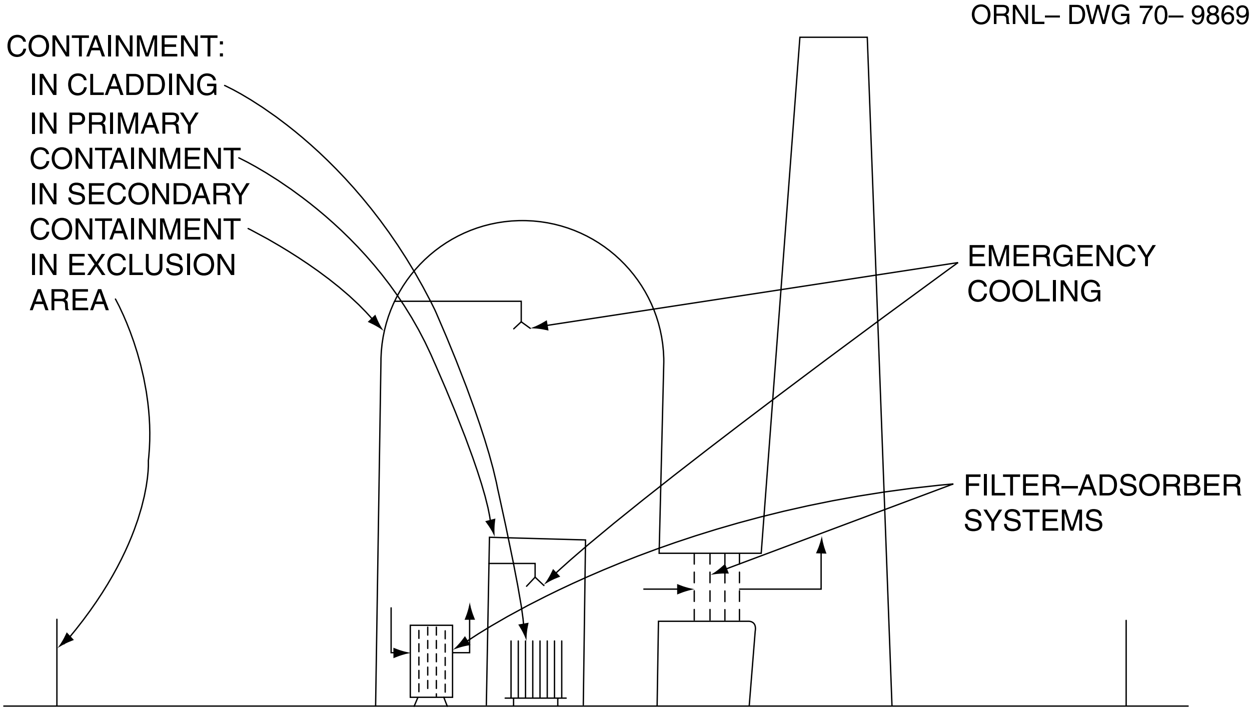


FIGURE 4 Reactor containment system. Any leakage from fuel must pass through the cladding, the primary containment, and either the secondary containment or the stack filters. Contamination within the building can be removed by sprays and/or filters.

but also because ^{85}Kr is absorbed much more powerfully by cold charcoal. This is the only practical means we have for removal of radioactive noble gases.

The very large dispersive capacity of a high stack usually makes it unnecessary to remove ^{14}C (as $^{14}CO_2$) or tritium (mainly $^3H^1HO$) because their toxicity is very low. However, the coolant CO_2 in a gas-graphite reactor does contain enough ^{14}C to require alkaline scrubbing, which removes radioiodine as well.

Liquids

Storage The necessity for long-term storage of very large quantities (many millions of gallons) of high level, strongly acid waste has led to the development of tankage and pipeline systems which have stood up to severe conditions for many years. Failures have occurred, but good design and carefully selected materials have prevented environmental contamination.

Tanks are constructed from material, often stainless steel, which will not be corroded by the solutions to be stored. Secondary containment is provided by catch tanks or drip trays and sufficient spare tankage is kept available for rapid emptying of a ruptured tank. Leakage is detected by a monitoring system which alarms immediately if radioactive liquid appears in the catch tank (Figure 5). Movement of active liquid is effected by pumping rather than by gravity to ensure that it is the result of deliberate action rather than accident.

Evaporation The most straight-forward and apparently the simplest method of treatment for radioactive liquid wastes is evaporation. In a carefully designed evaporator with an efficient droplet de-entrainment system the radionuclide content of the distillate can be about one millionth of that in the pot. There is little about the design that is specifically related to radioactivity except that shielding may have to be provided for the operator, and off-gases must be monitored and possibly treated in some way. Unfortunately, evaporation is expensive because it consumes a large amount of energy and the end product—the concentrate—is still a radioactive liquid waste. Evaporation to dryness or to the point of crystallization has been practised, by the residue is so soluble in water that without further processing it is not suitable for disposal.

Where discharge of a large volume of low-level waste into the environment is unacceptable the cost of evaporation may be justified by its many advantages. Practically all liquid wastes are treated by evaporation in Denmark and Sweden, and it is also widely used in Japan.

Residues from evaporation may be mixed with cement, fused with glass frit or various ceramic mixtures, or incorporated with melted bitumen. The product is then handled as a solid waste.

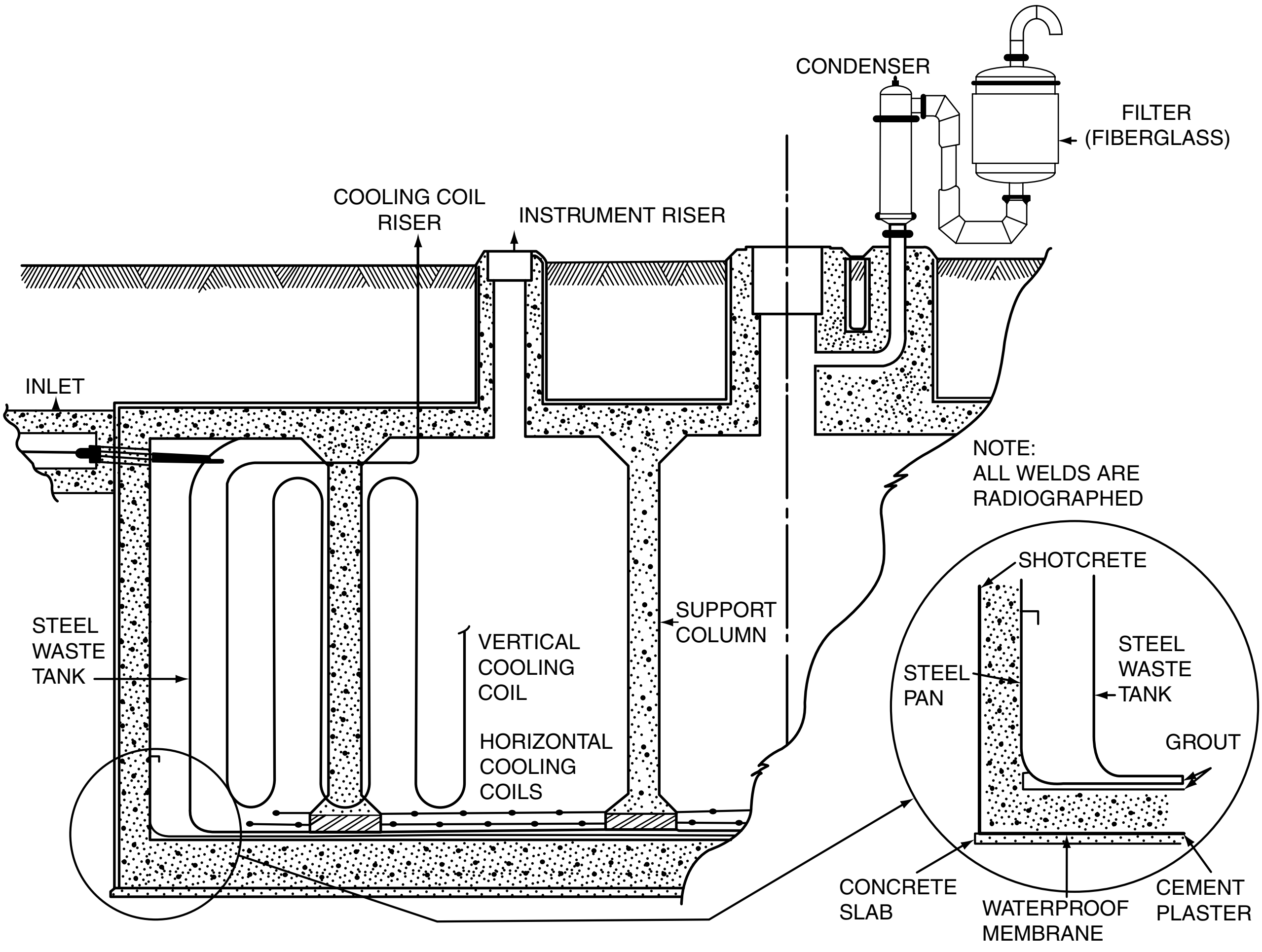


FIGURE 5 Structure of high level waste tank at Savannah River.

Flocculation and Precipitation The cheapest and simplest process for treatment of radioactive liquids is removal of the activity on some kind of precipitate, either as an integral part of the precipitated material, or adsorbed on its surface.

In most waste tanks a sludge settles out which may contain up to 90% of the activity, and a copious precipitate of metallic hydroxides is formed on neutralization which may carry down up to 90% of the remainder. Further purification of the clear effluent after separation of these sludges can be achieved by addition of lime and sodium carbonate. Up to 99% of the remaining activity can sometimes be removed by this treatment. Treatment with lime and sodium phosphate is also very effective (Figure 6).

The treatment used depends upon the particular radionuclides present in the waste, and also its gross composition for example, the pH and salt content of the solution. In some cases ferric chloride, clay or other additives are introduced at carefully chosen points in the process. The selection of the process, and modifications introduced as the composition of the waste changes, require constant analysis and control by specialized chemists.

One problem common to all flocculation processes is how to deal with the sludge. The floc settles very slowly and after it has been drained through filters or separated by centrifugation it is in the form of a thick cheese-like solid which, in spite of its appearance, still contains 80 to 90% of water. In a successful British process the sludge is repeatedly frozen and thawed. The separation of pure ice crystals leaves behind a concentrated salt solution which coagulates the small particles of floc into a form which settles more rapidly and is less likely to clog vacuum filters.

Ion Exchange The effluent from a flocculation process may still contain too much activity for discharge to public waters. It can then be passed through ion exchangers, which are expensive but very efficient. They cannot be used economically on a solution with a high salt content because their ion-exchange capacity would rapidly be exhausted by absorbing the dissolved salts.

The effluent from a well-controlled flocculation process has a low total-solids content and after filtration to remove traces of floc it can be passed through a cation exchanger or mixed-bed resin suitable for removal of the radioactive

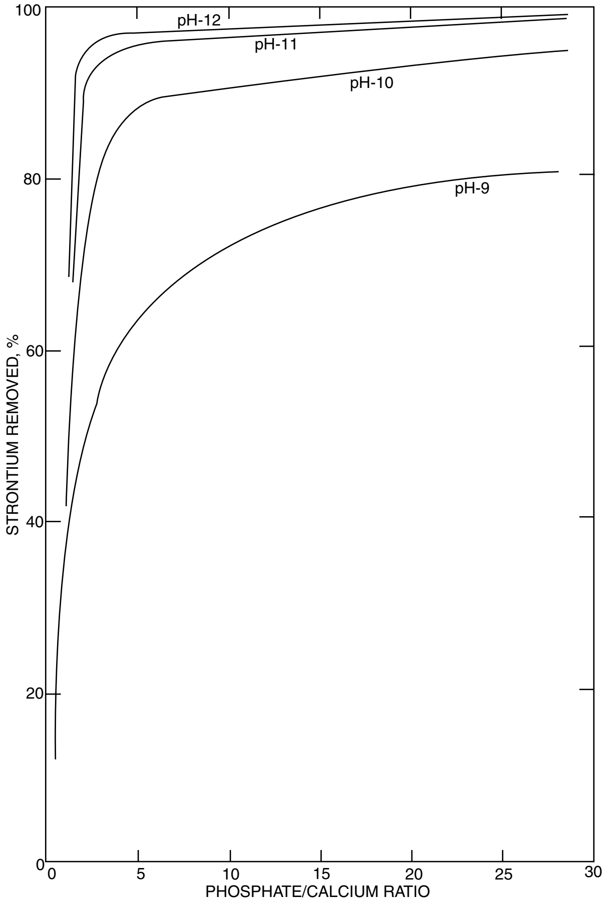


FIGURE 6 Decontamination with lime and phosphate—effect of pH and lime/phosphate ratio on removal of strontium-90.

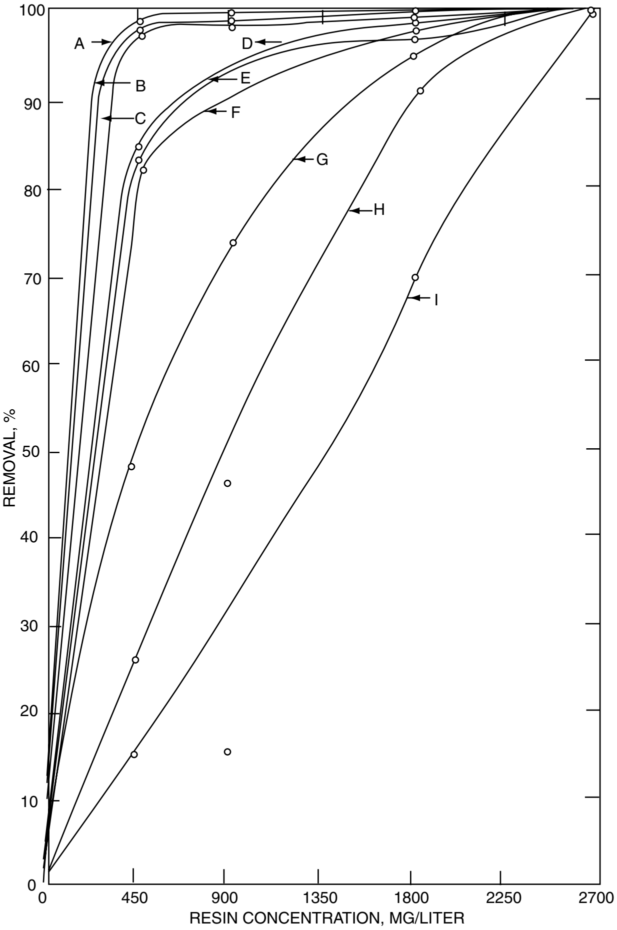


FIGURE 7 Decontamination with ion exchange resin—efficiency for various radionuclides; A, ^{182}Ta; B, ^{144}Ce$-^{144}$Pr; C, ^{95}Cr$-^{95}$Nb; D, ^{140}Ba$-^{140}$La; E, ^{131}I; F, fission product mixture; G, ^{32}P; H, ^{115}Cd; I, ^{137}Cs$-^{137}$Ba.

contaminants. If properly chosen such a resin will remove 99.9% of most radionuclides (Figure 7). Certain minerals—clinoptilolite, greensand and vermiculite are examples—are also efficient ion-exchangers. They are much cheaper than synthetic resins but they require longer contact times for maximum effectiveness.

Glass The very high-level "self heating" wastes—the primary wastes held in stainless steel tanks—are too active to be treated by flocculation or ion exchange. Storage in liquid form is seldom regarded as a permanent solution—somehow these wastes must be fixed in a nonleachable solid form which can be stored safely without danger of leakage or constant maintenance costs. One of the most promising ways to fix high level waste is to incorporate it into a glass.

Glass is a leach-resistant material which can be made from simple ingredients. Its quality varies with composition but it is not usually sensitive to changes in minor constituents. Its low melting point makes it convenient for casting in various shapes and sizes for different disposal procedures. Glasses are supercooled, very viscous, liquid solutions of silicates. Soda glass is made by melting together silica, calcium carbonate and sodium carbonate. Other varieties contain potassium or potassium plus lead instead of sodium, and phosphate or borate in place of part of the carbonate. Metallic oxides are incorporated to form coloured glasses. Such a mixture might well be suitable for fixing the radioactive metallic oxides which form the major proportion of "mixed fission products", after the nitric acid has been removed and the residue ignited.

Successful fixation of radionuclides in glass has been reported from the USA, UK and Canada. British and American practice has concentrated on borate and silicate glasses, or fusing the waste oxides with glass frit, whereas the Canadians have used a natural silicate, nepheline syenate, instead of a glass mix (Figure 8).

Glass fixation is now being done on quite a large scale at the Pacific Northwest Laboratory (Hanford) Washington, USA. By the end of July 1970 nineteen million Curies had been solidified, representing waste from about ten tons of irradiated fuel.

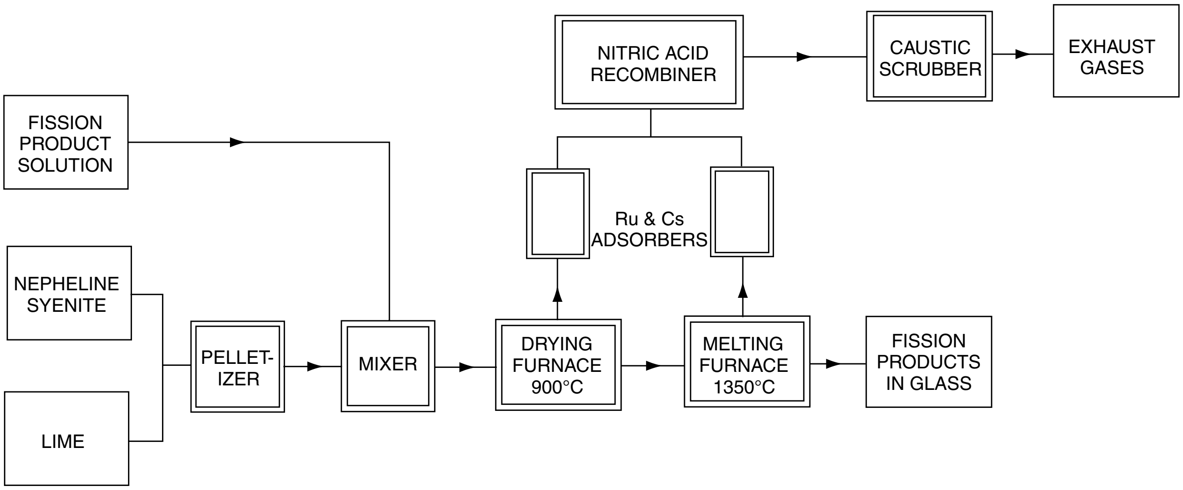


FIGURE 8 Fixation of fission products in glass. Fission product solution is added to pelletized nepheline syenite + lime, dried and melted at 1350°C.

In Canada nepheline syenite glass containing 1100 Ci mixed fission products has been buried in a swamp, below the water table, since 1960. Careful studies of leaching, by sampling of the soil and ground water downstream from the disposal, have shown that fusion products equivalent to the dissolving of 10^{-10} g of glass per cm^2 per day are being removed from the disposal. Less than 1 mCi has been dissolved in ten years from 1100 Ci. This suggests that burial of active glass in dry soil, or even disposal into a big body of water, would be acceptable for quite large quantities of wastes.

Calcination Several methods have been developed for evaporation and subsequent calcination of wastes. Oxides are often soluble in water, so materials are usually added that will bind the oxides into soluble complexes. The calcination process is done in a heated steel container, a fluidized bed or a spray calciner.

The pot calciner is essentially an expendable piece of steel pipe heated in an electric furnace. The waste, mixed with flass-forming fluxes such as borax or lead oxide, is heated to about 9000°C.

The spray calciner is a heated steel cylinder with a nozzle at the top through which the waste is sprayed. At a temperature of 875°C a fine powder is produced which must be stored in a dry place as it is leachable by water.

It is characteristic of all waste fixation methods involving evaporation, sintering and fusion that elaborate off-gas treatment systems are required to prevent environmental contamination by dust and volatile radionuclides. The concentrating equipment itself is essentially simple and often not expensive to build, but the glass purification plant is always sophisticated, complex and expensive. However, it is also very effective.

Rock Fracturing The oil industry has developed methods for creating fissures in rock in order to encourage movement of oil to gas through a formation towards a well. This process has been adapted to disposal of medium-level wastes.

A horizontally bedded formation—shale has been used up to now—is drilled to several thousand feet. A high pressure jet of sand and water cuts through the well casing and penetrates between the strata near the bottom of the hole. The well is then sealed and water forced down under very high pressure, splitting the rock between the bedding planes. The water is followed up by the waste, mixed with cement, sugar and other additives. The mixture spreads out in a thin horizontal sheet, which solidifies after several hours. Typically the sheet is about a half inch thick.

The method has been used for disposal of a very large volume of waste at Oak Ridge, Tennessee. The equipment, including large bins for ingredients of the cement mix, mixing apparatus, a drilling rig and a very powerful pump is expensive, but the method is suitable for large-scale operation because successive sheets can be injected at intervals of a few feet through the depth of the bedded rock formation.

Salt Mines The hazard that must be met by most radioactive waste management systems is contamination of public waters leading directly or indirectly to intake of radionuclides by man. An ideal situation for disposal would therefore be one where public access was impossible and contact with water incredible. The nearest approach to these conditions is found in a deep salt mine. The presence of the salt guarantees that water has been absent for millions of years, and geological study can produce assurance that water is not rapidly penetrating into the salt bed. The excavated galleries of salt mines are large and stable tunnels, suitable for storage and roomy enough for safe work with active loads.

Major disposals of “solidified” waste are being made in a salt mine in Kansas, where detailed investigation has shown that eventually the creep characteristics of the salt will seal the disposal sites of heat-producing wastes. Work on disposal of liquids into cavities cut in the salt suggests that hot liquid waste could also be placed in such a site.

Solids

As with liquid wastes, the most intractable problem is the safe management of the high-volume, low activity waste. The high activity waste is at first sight more dangerous, but although safe custody may be expensive it is not technically difficult.

Low-level waste consists mainly of "garbage"—contaminated clothing, equipment and structural material; broken glassware, cleanup materials such as cloths and mops; and a large amount of "potentially contaminated" material such as packing and paper which must be treated as active simply because it originates in an active area.

Much of this material can be reduced in volume by incineration or baling under high pressure. Fumes and smoke from incinerators and the dusty air from baling plants are cleaned up by methods dealt with under Gases (pp. 717–718), but the ash and baled waste remain to be dealt with.

In some countries geographical or legal circumstances restrict the possibility of burial of radioactive material in the ground. Elsewhere, ground burial is regarded favourably. In the latter case bales and non-combustible waste are likely to be buried in sparsely populated regions. Where land is cheap, low-level wastes may be buried without any volume reducing process.

Conditioning Pre-treatment of waste before final disposal is called "conditioning". The aim is usually immobilization of radionuclides together with, if possible, volume reduction. There is very wide variation in practice from one country to another. For example, in France quite low level solid wastes are put into concrete containers which are then filled with cement mixture so that the end product is a large concrete block. These blocks are stored, under a roof, on a concrete floor. In Canada, on the other hand, similar wastes are put into open trenches at Chalk River and covered with the local sandy soil. Practical measurements seem to show that both procedures are equally safe in the local circumstances.

A very effective conditioning process is fixation in bitumen or asphalt. Bitumen is very resistant to radiation, has a low melting point, is impermeable to water and has some mechanical flexibility. Radionuclides enclosed in, or even mixed with, bitumen leach very slowly into water. Sludges are dewatered when mixed with melted bitumen, which helps considerably in restricting the volume of the disposals. In general, bitumen is beginning to be favoured over concrete as the method of choice for "fixing" otherwise mobile waste radionuclides.

Ground Disposal In some countries direct burial of contaminated material in the ground is forbidden at any level, whereas in others the amount and nature of ground disposals is left to the discretion of the operator.

Nearly all cations move through soil more slowly than the ground water although some anions—ruthenate and iodide for example—are retarded very little. In the case of the average "mixed fission products" usually of concern in waste management the fastest moving radionuclide is ruthenium, usually followed by Sr, Cs and Ce in that order. Relative rates of movement are affected by the nature and pH of the soil and the ground water, but even in acidic sandy soil ^{90}Sr moves through the soil at only 1/25 to 1/100 of the rate of movement of the ground water.

If the site of the waste management area is selected with care in relation to potable water supplies, so that the time of transit between the point of disposal and the point of human consumption is prolonged in relation to the half-life of the critical radionuclides, direct ground disposal of low level waste is effective and safe. There are a great many places where knowledge of the rate and direction of movement of the ground water, together with the distribution coefficients of radionuclides between water and soil, make it apparent that no significant discharge into the environment would be credible as a result of direct disposal into the ground.

When simple burial is unacceptable, disposal trenches and areas can be drained, with processing of the drainwater, or the area can be covered with asphalt and protected from encroachment of ground water by circumferential drainage.

A further step in the direction of safety is the "engineered enclosure." This is a structure built like a concrete house basement. It usually takes the form of a long concrete-lined trench divided into sections by concrete cross-walls. The section in use is covered by a temporary roof (Figure 9). The object of the structure is to prevent the ingress of water, so joints in

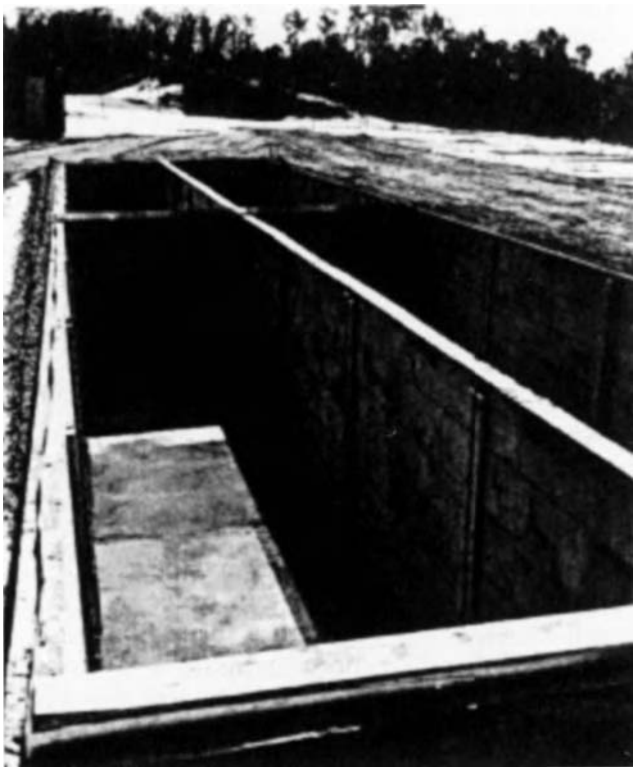


FIGURE 9 Concrete trench. Double trench, for medium-level solid wastes, is covered with a light roof when in use. The filled trench is levelled with sand and a concrete roof is poured. Note galvanized steel seals for joint.

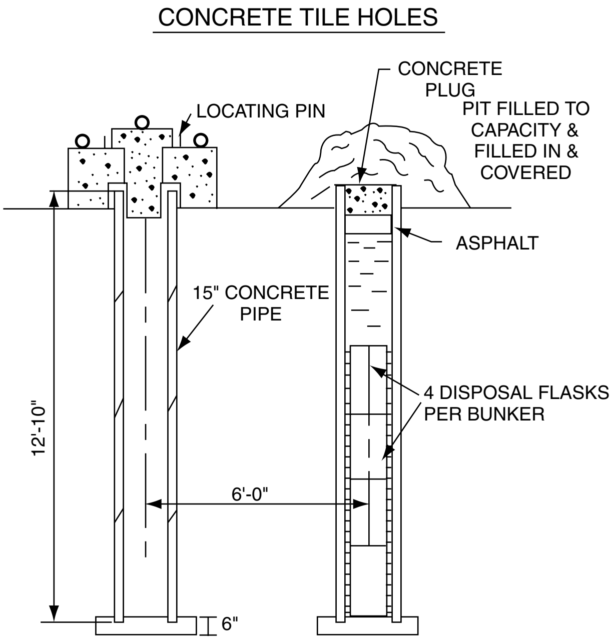


FIGURE 10 Concrete Holes, constructed from drain pipe painted with bitumen, provide will-shielded receptacles for high-level solid waste.

the concrete are water-sealed and the base of the work is laid well above the maximum height of the water table. Since the facility will be used for reception of quite high-level waste the approaches must be suitable for trucks and mobile cranes.

Concrete trenches are unsuitable for reception of small, intensely radioactive objects such as spent teletherapy sources because of the inconvenience of scattered radiation fields. They can be accommodated on concrete-lined holes fitted with removable shielding plugs. Canadian practice is to construct these from sections of concrete drain pipe, painted on the outside with bitumen, which is also used to seal joints between sections (Figure 10).

There are many different versions of the types of disposal facility just described. Some are in the open, some within buildings, but all are designed to prevent access of water to the contents.

It is often convenient to delay the passage of radionuclides contained in high volume low-level wastes before discharge into the environment in order to take advantage of radioactive decay. If the local soil and ground water regime are suitable this can often be done by discharge into seepage. In 1969 the Hanford (USA) laboratories discharged 5000 million gallons of low-level waste into the ground, containing nearly 4000 Ci of radionuclides. More than 99.9% of this activity is held on the sediments immediately below the disposal facilities.

METHODS FOR "SMALL USERS"

The International Atomic Energy Agency has issued a code of practice on management of radioactive wastes by hospitals, research institutes and industry when no special facilities are available on the site. It gives a review of the scope and nature of the necessary control, particularly in the establishment of permissible limits for discharge into the environment. These institutions rely heavily upon the public sewers and garbage disposal systems, depending upon the fact that in practice there are levels of radioactivity below which things are not regarded as radioactive. Sometimes this is the level at which measurement becomes practical and sometimes legal limits exist.

The ICRP and the IAEA both recognize 10^{-4} μCi/ml as the concentration in the sewer of an institution below which no restrictive action is required, irrespective of the nature of the radionuclide. This assumes that since large dilutions will occur before ingestion by the public, the individual at risk is the sewer worker. Plumbers can also encounter hazards in traps and filters, and they must be made aware of the situation. Similarly, although very low-level discharges to public disposal areas are usually acceptable, the waste management authorities must know of the practice so that they can warn their staff or undertake special procedures such as tip-and-fill operations. It is particularly important that scavenging should be prevented, because a very small source, normally innocuous, can be hazardous if carried for a long time in a pocket.

Apart from the use of public facilities, radioactive waste disposal for the "small user" does not differ in principle from the methods available to the larger producer of waste. A good deal of common sense and sense of proportion are required in dealing with the problem, aided by technical advice such as that in the IAEA report. It is also useful to remember that an ordinary illuminated wristwatch gives a count of several thousand per minute on a Geiger counter from the face side and almost zero from the back.

SEA DISPOSAL

In 1983 the London Dumping Convention passed a non-binding resolution which imposed an international moratorium on sea dumping of radioactive wastes. In 1993 a binding resolution against sea disposal of these wastes was passed by the LDC. A scientific evaluation of this ban will be conducted about the year 2019.

THE FUTURE

The accidents at Three Mile Island and Chernobyl have focused public attention on the potential for disaster. After the Chernobyl incident the WHO (World Health Organization) European Office set up a Working Group to consider harmonization of response to any similar incident which might occur in the future. The accident has proved conclusively that some nuclear accidents will have consequences for removed from the accident sites. Thus, it makes little sense for one nation to take measures in isolation. The report produced by the WHO European Office sets forth not only the effects of Chernobyl but also presents an excellent discussion of the foundations for international cooperation in case of a future accident.

With the reduction of tensions between the major nuclear powers the questions of nuclear weapons destruction and clearup of weapons producing plants have become the dominant issues. The technology for accomplishing these tasks is available. However, the massive sums necessary are not readily available. Nevertheless, these sums must be expended.

REFERENCES

Hilliard, T.R., *et al., Radiological Water Pollution in the Elliott Lake and Bancroft Areas, Government of Ontario, Toronto,* 1965.

Ontario Water Resources Commission, *Interim Report on Radiological Water Pollution in the Elliott Lake and Bancroft Areas,* OWRC, Toronto, 1966.

Holaday, A.D., *Hearings on Industrial Radioactive Waste Disposal,* **1,** p. 82. U.S. Gov't Printing Office, Washington, D.C. 1959.

Rodger, W.A., Report ANL-6233 (1960).

Sutton, O.G., *Quart. J. Roy. Meteorol. Soc.* 73, **252,** and **426,** (1947).

Pasquill, F., *Meteorol. Mag.* **90,** 1961.

Holland, J.L., Report ORO-99, 1953.

Barry, P.J., *Atmospheric Env.,* **3,** 1969. See Reference 3 above.

Schauer, P.J., *Ind. Eng. Chem.*, **43,** 1532, 1956.

DeBruyn, J. and K. W. Pearce, Report AERE-M 713, 1960.

McElroy, J.L., J.N. Hartley and K.J. Schneider, Report BNWL-1185, 1970.

Elliott, M.N. and J.R. Grover, Report AERE-R 4844, 1965.

Watson, L.C., A.M. Aikin and A.R. Bancroft, *Disposal of Radioactive Wastes,* Vol. 1, pp. 373–399, IARA, Vienna, 1960.

Merritt, W.F., *Disposal of Radioactive Wastes into the Ground,* pp. 403–408, IAEA, Vienna, 1967.

Elliott, M.N., R. Gayler, J.R. Grover and W.H. Hardwick, *Treatment and Storage of High Level Radioactive Wastes,* pp. 465–506, IAEA, Vienna, 1963.

Lakey, L. T., G.E. Lohse and W.P. Palica, Report IDO-14608, 1965.

Hartley, J.N. and W.R. Bond, Report BNWL-1494, 1970.

De Laguna, W., *Nuclear Safety,* **11,** 391, 1970.

Galley, J.E. (ed.), Subsurface Disposal in Geologic Basins, American Assoc. Petrol. Geologists, Tulsa, 1968.

Bradshaw, R.L., *et al.,* Report ORNL-4584, 1970, pp. 16–36.

Cerre, P., *Disposal of Radioactive Wastes,* **1,** pp. 226 to 234.

International Atomic Energy Agency, Technical Reports Series No. 116, *Bituminization of Radioactive Wastes,* IAEA, Vienna, 1970.

Mawson, C.A., *Management of Radioactive Wastes,* pp. 117–124, Van Nostrand, Princeton, N.J., 1965.

Beard, S.S. and W. L. Godfrey, *Disposal of Radioactive Wastes into the Ground,* pp. 123–134, IAEA, Vienna, 1967.

International Atomic Energy Agency, Safety Series No. 12, *Management of Radioactive Wastes Produced by Radioisotope Users,* IAEA, Vienna, 1965.

Nuclear Accidents—Harmonization of the Public Health Response. WHO, Geneva, 1989.

Krauskopf, K.B., *Radioactive Waste Disposal and Geology,* Chapman and Hall, London, 1988.

Chapman, N.A. and I.G. McKinley, *The Geological Disposal of Nuclear Waste,* Wiley, Chichester, 1987.

Office of Technology Assessment, U.S. Congress, *Managing the Nation's Commercial High-Level Radioactive Waste,* U.S. Government Printing Office, Washington, D.C., 1985.

Low-Level Radioactive Waste; Building a Perspective, Atomic Industrial Forum, Bethesda, Maryland, 1986.

Severe Accidents in Nuclear Power Plants, International Atomic Energy Agency, Vienna, 1988.

Disposal of Radioactive Wastes in Deep Depositories. Report of a Royal Society Study Group. The Royal Society London, 1994.

COLIN A. MAWSON
Ottawa, Ontario

YUAN DING
New Jersey Institute of Technology

MANAGEMENT OF SOLID WASTE

INTEGRATED WASTE MANAGEMENT

The most recent comprehensive document produced by the federal government characterizes the materials commonly referred to as "municipal solid waste" ("MSW") as follows:

" … residential solid waste, with some contribution from commercial, institutional and industrial sources. In some areas, nonresidential wastes are managed separately, largely because industrial and some commercial sources produce relatively uniform waste in large quantities, which makes them more suitable for alternate disposal techniques or recycling. Hazardous wastes, as defined by Federal and State regulation, generally are managed outside the municipal solid waste stream. Exceptions are household hazardous wastes and hazardous wastes generated in very small quantities, which are often placed in the municipal solid waste stream by the generator."[1]

One of the most significant developments in municipal solid waste is the growing acceptance by citizens, all levels of government, and industries of a new overall philosophy concerning the management options available to address the problem of increased waste generation in the face of ever-decreasing land disposal sites. This philosophy is commonly known as "integrated waste management" and involves the reliance upon a hierarchy of options from most desirable to least desirable. The options are as follows:

Source reduction, limitation of the amount and/or toxicity of waste produced
Recycling, reuse of materials
Incineration, thermal reduction
Sanitary landfill, land disposal

While this hierarchy is little more than a common sense approach to municipal solid waste problems and the unit operations represented are not new, emphasis on the source reduction and recycling options as preferred represents a profound shift in attitudes toward municipal waste management. The traditional perspective that generators could produce discards without limit and depend on technological approaches to mitigate such wastes and any associated effects of treatment is no longer acceptable. This approach is not unique to the solid waste area but is a part of federal and state "pollution prevention" strategies, which emphasize avoidance of all types of pollution as preferable to "end of pipe" and other traditional methods of environmental regulation.

LEGISLATION

In 1984, amendments were made to the Resource Conservation and Recovery Act of 1976 ("RCRA"), the existing federal legislation covering solid waste management. Although the majority of these amendments were concerned with the regulation of hazardous waste as were the original RCRA mandates, some changes and additions were made to those provisions which were directed at nonhazardous waste.

The U.S. Environmental Protection Agency ("EPA") was directed to determine whether the existing criteria for land disposal of waste previously promulgated pursuant to Sections 1008(a) and 4004 of RCRA are adequate to protect human health and the environment from groundwater contamination and whether additional authorities are needed to enforce them. In addition, EPA must revise the criteria for those facilities which may receive hazardous household or small quantity generator waste. Furthermore, States were given three years to develop a program to ensure that municipal facilities met the existing criteria and the revised criteria when they are promulgated. Although enforcement is still largely a state matter, EPA is empowered, though not required, to enforce the criteria if states fail to comply with their obligations. As of this writing, revised criteria have been proposed but not yet adopted.[2]

Perhaps the most significant aspects of the federal law and its implementation involve initiatives with legislative roots in the original RCRA legislation which had historically received less attention than the Act's mandate to establish a hazardous waste management regulatory system. EPA has begun pursuing a number of activities such as conservation of virgin materials through guidelines establishing revised product specifications and similar initiatives.

State legislation has also witnessed a marked shift toward more conservation-oriented management schemes as well as stricter standards for processing and land disposal facilities. For example, at least twenty-four states have laws mandating the use of recovered materials in procurement processes. As of this writing, nine states had legislation requiring deposits on beverage containers and four states had mandatory recycling laws covering a wide range of materials. The scope of these new legislative initiatives and the myriad of options and alternatives they entail is beyond the purview of this analysis. What is apparent, however, is that source reduction and recycling represent an important part of modern waste management systems.

INTRODUCTION

Any discussion of solid waste neatly divides into three categories:

1) Source and composition,
2) Collection,
3) Disposal (or, hopefully, reuse).

Another natural division, resulting in part from the current regulatory states, is between hazardous and nonhazardous wastes. This section will deal primarily with nonhazardous wastes; specifically, with their source and composition and disposal. However, a brief discussion of hazardous wastes is included because of their importance in understanding the management of urban waste. More detailed discussion of hazardous waste is found in another section. The important problem of collection is also left to a special section on that subject.

Solid waste used to be considered any solid matter which was discarded as no longer being useful in the economy. During the last decade, this definition has been considerably broadened. For regulatory, and usually disposal purposes, solid waste is now defined as "any garbage or refuse, sludge from a waste treatment plant, water supply treatment plant, or air pollution control facility, and other discarded material, including solid, liquid, semi-solid, or contained gaseous material resulting from industrial, commercial, mining, and agricultural operations, and from community activities, but does not include solid or dissolved materials in domestic sewage, or solid or dissolved materials in irrigation return flows or industrial discharges which are point sources."[3] This definition is important because it indicates that all matter which is disposed of onto the land in any form is considered "solid waste." In addition that material which causes or significantly contributes to an increase in mortality or serious illness or poses a substantial hazard to human health or the environment, is considered a "hazardous waste." Hazardous wastes have been further defined by rulemaking to a limited set of materials and criteria such as toxicity, flammability, reactivity, or corrosivity.[4] The handling of hazardous waste requires special care and special permitting. Contrary to the management of normal refuse or solid waste, the generators, transporters, and disposers of hazardous wastes must meet stringent federal and state criteria and have considerable potential liability exposure. The disposers of solid waste which is not hazardous must meet state criteria that are not nearly as stringent as those for hazardous materials. Thus, while hazardous material in the past has been often disposed of along with all other refuse, today this is no longer the case. Industrial waste generators segregate their hazardous from their industrial waste so as to minimize their problems.

Solid wastes are one of the three major interacting *waste* vectors; the others are air and water pollutants. Solid wastes, if improperly handled, can be a source of land, air and water pollution. They are, also, at this writing, one of the most volatile public issues and a problem which is presenting many communities with significant institutional challenges.

Significant progress has been made in regulating the disposal of solid waste over the last decade. Open dumps which presented aesthetic as well as environmental challenges are for the most part closed. Regulations are in place for managing solid wastes in an acceptable manner. However, dumping into the ocean, which can create "dead" zones, hopefully will be eliminated. Nor have we eliminated the potential problems of leachate from landfills. Perhaps the most significant problem is the one of locating new landfills or substituting resource recovery, reuse and recycling capacity for landfill disposal. The technologies are available, but the economics still favor land disposal. In the early '70s there was great hope for massive resource recovery and recycle projects. Some of those, discussed later in this section, have not come to fruition because of economic and institutional barriers. Others have succeeded but the technology has not been spread, primarily because of economic barriers. Individual and community action to reduce the amount of wastes generated and collected has, in many areas of the country, been successful. For example, solid waste contains significant amounts of valuable material; 40% to 50% of urban waste is paper and, if recycled, can replace virgin stock equivalent to about 9 trees per person per year. In addition, the community and thus the taxpayer also saves in terms of lower collection and disposal costs. However, this is still of limited application because it is usually limited to newspapers, aluminum cans and perhaps glass. Both technology and institutional methodologies for recycling solid waste are still in their infancy and must gain momentum if we are to meet the challenge of solid waste management in the years to come.

REGULATION OF SOLID WASTE MANAGEMENT

Regulation of solid waste management has been scattered. The federal government, contrary to its prior policies in air and water, did not take a strong posture in solid waste management. It left regulatory initiative to the states and localities. These dealt with the solid waste management primarily through the licensing of collectors, through the "Utility Commissions" and adding to zoning ordinances regarding local landfills. Public health regulations also played a role with respect to reduction of rodents and pests at landfills. Air emissions from incinerators were regulated as were wastewater discharges. In the last several years a number of states have enacted and implemented legislation to regulate landfills. Probably the earliest and still among the most comprehensive is the regulatory effort of the State of California which has classified landfills which respect to underlying geological conditions in terms of what a landfill can and cannot accept.

A comprehensive solid waste law at the federal level was passed in 1976 as the "Resource Conservation and Recovery Act of 1976."[5] This act provides for federal assistance to states and regions developing and encouraging environmental sound disposal of solid waste and the maximum utilization of resources. It calls for state and regional plans and for federal assistance to develop these plans. It requires that each plan shall prohibit the establishment of open dumps and provides for the upgrading of open dumps that are currently in use. It also requires that criteria for sanitary landfills be established. However, it leaves enforcement to the states. At the same time, the Act under Subtitle C provides for federal regulation of the management of Hazardous Wastes. Many of these regulations have been issued but the critical ones covering treatment, storage and disposal facilities are still under review.

SOURCES OF WASTE

Solid waste differs from air and water pollutants in that it comes in discrete quanta and is very heterogeneous in nature. Both composition and rate vary significantly from day to day and from season to season as well as from otherwise similar sources.

The solid waste production in the United States is in excess of four billion tons/year and was expected to increase to five billion tons by 1980.[6] Table 1 breaks this down for the year 1967 by major source. However, waste generation appears to have stabilized despite increased loads from air and water pollution control facilities. How long this will last remains to be seen, if and when significant conversions to "coal as fuel" and more stringent air and water pollution control take place.

Urban Waste

Urban waste collected is between 4 and 8 lbs per person per day, with typical values lying between 4.5 and 5.5 lbs per day. This differs from the amount generated because of self and private disposal. The major wastes included in this category are tabulated in Table 2, which includes a summary of disposal trends. One should be careful in the terminology because often domestic and municipal are used interchangeably to indicate the total refuse picked up from residential (domestic), institutional, small business and light industrial sources.

Some further definition of terms may be useful at this point. In general usage many of the terms have been used interchangeably. However, an effort to standardize the terminology was made by the Institute for Solid Waste of the American Public Works Association and the Office of Solid Waste Management of the Environmental Protection Agency.[7] The standard usage of terms detailed by these groups is summarized here:

Refuse All solid waste matter.
Garbage The animal and vegetable waste resulting from the preparation of food.
Rubbish The waste from homes, small businesses, and so on excluding garbage.
Trash Used equivalent to rubbish.
Litter Street refuse.
Industrial Waste Specialized refuse from manufacturing plants, and usually excludes rubbish.

TABLE 1
Major sources of waste matter United States 1967[5]

	Solids generated	
Source	lab/cap/day	Million tons/yr
Urban		
Domestic	3.5	128
Municipal	1.2	44
Commercial	2.3	84
Sub total	7.0	256
Industrial	3.0	110
Agricultural		
Vegetation	15.0	552
Animal	43.0	1563
Sub total	58.0	2115
Mineral	30.8	1126
Federal	1.2	43
Total	100.0	3650

Domestic waste composition will vary seasonally, as well as with locale and economic status. Typical analyses for domestic plus municipal refuse are shown in Table 3. As can be seen in a comparison of the data, the composition has not changed drastically with time except for a significant reduction in ash because of the change from coal as a home heating fuel. Location variations noted are as great or greater. A study of seasonal variations made in 1939 for New York City also showed greater variations: the ranges were garbage, 44 to 3.5%; and metal, 11.6 to 3.1%.[8] Base data have been difficult to obtain because of the many variabilities in the base. The most significant variables include the economic level of the area, the ratio of commercial to residential property, the type of commercial establishments and the housing density and age. The entire picture on obtaining accurate data on urban and/or domestic refuse is further complicated by the sampling problem. A discussion of this problem is beyond the scope of this work; the reader is referred to some basic work in this area by Carruth.[9] An excellent review of sampling and testing has been prepared by the Institute of Solid Wastes.[10] Further work is being done in this area by ASTM's D-34 Committee.

The ultimate chemical composition of municipal refuse has been examined by a number of investigators. Table 4 gives the range of values to be expected. Recently 0.3 to 0.5% chloride has been found in refuse independent of the presence of polyvinyl chloride; this is due to the presence of salt primarily.[11]

Density of municipal refuse varies with the load applied to it. Typically household refuse has a density of 350–400 pounds per cubic yard. Transfer stations and/or landfill operations can compact it to between 500 and 800 lbs per cubic yard depending upon the material and conditions. The effect of compression on density for the Chandler, Arizona refuse is shown in Figure 1. High pressure compaction (see Compaction) can increase the density to 1200 to 1400 lbs per cubic yard.

Industrial Wastes

Industrial wastes amount to about 115 million tons annually. They include any discarded solid materials resulting from an

TABLE 2
Composition of wastes from urban sources[6]

Urban sources	Waste	Composition	Disposal, present
Domestic	Garbage	Wastes from preparation, handling and sale of food	
	Rubbish, trash	Paper, wool, excelsior, rags, yard trimmings, metals, dirt, glass, crockery, minerals	Landfill
	Ashes	Residue from fuel and combustion of solid wastes	Incineration
	Bulky wastes	Furniture, appliances, rubber tires	Dumping
Commercial	Garbage	Same as domestic	Landfill
Institutional	Rubbish, trash	Same as domestic	Incineration
	Ashes	Same as domestic	
	Demolition wastes, urban renewal, expressway	Lumber, pipes, brick masonry, asphaltic material and other construction materials	Dumping Landfill
	Construction wastes	Scrap lumber, pipe, concrete, other construction materials	Dumping Landfill Open burning
	Special wastes	Hazardous solids and semiliquids, explosives, pathological wastes, radioactive wastes	Burial, incineration Special
Municipal streets, incinerators, sewage treatment plants, septic tanks	Street refuse	Sweepings, dirt, leaves	Fill
	Dead animals	Cats, dogs, horses, etc.	Bury or incinerate
	Abandoned vehicles	Unwanted cars and trucks	Reclaim
	Fly ash, incinerator residue, boiler slag	Boiler house cinders, metal scarps, shavings, minerals	Landfill or dump Landfill
	Sewage treatment residue	Solids and sludge	—

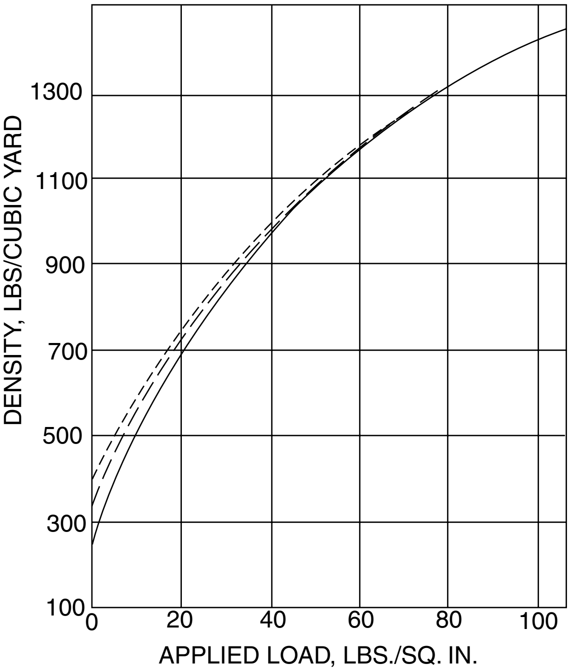


FIGURE 1 Refuse density. Household refuse, Chandler, Ariz., 1954. Credit: APWA, Municipal Refuse Disposal; 1966.

industrial operation or establishment with the exception of dissolved or suspended solids in domestic or industrial waste waters. The composition and quantity of industrial solid wastes vary significantly from location to location, as well as between industries and within a given industry. Table 5 lists the type of wastes to be expected from the various SIC Industrial Groups. A large fraction of the wastes are generally common to most industries and are listed on Table 6.

Data on the amounts of waste generated by or collected from various industries is very limited. Industry, quite naturally, has considered this type of data confidential in that it often reveals significant process and economic information. Average data, even if available, are of limited value because wide variations can result from process differences, process efficiencies and direct recycle, as shown in a study based on detailed interviews. The results of this study giving total waste by industry are summarized in Table 7. Industry waste production on a unit per employee basis vary widely and are summarized for large and small companies in Tables 8 and 9.

Increased efficiency as well as new uses for present industrial waste streams will alter both the quantity and composition of the material for disposal in the next decade. For example, saw mill waste is being reprocessed into composition board and this utilization could essentially eliminate this waste stream. Only limited projections can and have been made and these show only a reduction in saw mill wastes.[12] Conversely, enforcement of air pollution statutes will increase the amount of potential solid wastes significantly. Greater purification of industrial wastewater will also affect the solid waste load.

Agricultural Wastes

Agricultural wastes are principally organic as indicated in Table 10. The exceptions are chemicals used in various facets of farming such as pesticides, containers, and small amounts

TABLE 3
Urban refuse, typical compositions

Source time	Hempstead, NY 64		Hempstead, NY 6/66	 2.67	Chicago 56–58 average	Chandler, Arizona 1953	St Clara County, Calif. 1967	Berkeley, California 1952	 1967	Del. Co.Pa. 1980
Material										
Paper and paper prod.	56.01	53.5	32.71	53.33	56.5	42.7	50 compostable material	69.0[c]	69.7[c]	38
Wood	2.82	—	1.22	1.46	—	2.3	2	—	—	—
Grass, leaves, etc.	7.56	9.14	33.33	0.26	9.6	1.3	9	—	—	8
Rubber	0.42	0.38	—	—	—	0.7	1	—	—	12
Plastic	3.50	0.76	2.45	3.45	—	0.4	1	—	1.9	—
Oil, paint	0.84	0.76	—	—	—	—	—	—	—	—
Dirt	2.52	2.29	[a]	[a]	—	—	—	—	—	—
Rags	0.84	0.76	3.00	2.24		1.9	2	1.5	1.1	4
Miscellaneous	0.52	0.38	—	—	—	—	8	7.6	7.4	6
Rubbish	—	—	—	—	—	—	—	—	—	—
Garbage	9.24	6.11	9.58	16.70	4.8	21.8	12	[d]	[d]	9
Fat	—	2.29	—	—	—	11.3	—	—	—	—
Metal	7.53	6.85	7.96	10.60	14.8[b]	9.8	8	10.6	8.7	13
Glass, ceramics	8.50	7.73	9.75	11.87		7.8	7	11.4	11.3	10
Ash	—	—	—	—	18.7	—	—	—	—	—
Reference	(7)	(8)	(9)	(9)	(10)	(11)	(12)	(13)	(13)	—

[a] Included in glass and leads.

[b] Glass averaged 6.4% range 3.5–9.3%.

[c] Includes garbage.

[d] Included in compostable material.

TABLE 4
Municipal refuse[B] ultimate chemical analysis

Constituents	% by weight (as received)
Proximate Analysis	—
Moisture	15–35
Volatile matter	50–65
Fixed carbon	3–9
Noncombustibles	15–25
Ultimate analysis	—
Moisture	15–35
Carbon	15–30
Oxygen	12–24
Hydrogen	2–5
Nitrogen	0.2–1.0
Sulfur	0.02–0.1
Chloride	0.3–0.5 (16)
Noncombustibles	15–25
Heating values, Gross	3000–6000 Btu/1b

of miscellaneous waste matter resulting from maintenance and general housekeeping.

Most crop waste is either plowed back into the soil or composted. Some open burning takes place. In some special cases such as bagasse (sugar cane stalks) industries have been established to utilize the waste material. Essentially none of this material finds it way into the usual disposal facilities.

Animal wastes pose a different problem because much is produced in very concentrated areas such as feed lots or poultry farms. The disposal of these wastes is posing a greater problem than crop waste, but may be more easily solved because it is concentrated and therefore susceptible to processing without collection. Average waste yields for a variety of domestic animals are summarized on Table 11.

Mineral Wastes

Mineral wastes including solids generated in mining, milling and processing industries are expected to reach between two and four billion tons per year in 1990. In 1965 this waste amounted to 1.4 billion tons, as summarized in Table 12.

Hazardous Wastes

Hazardous wastes as defined by the federal government and in many cases similarly by the states, must be receiving special handling. These wastes generally include materials that are injurious to human health, toxic, can cause irreversible environmental damage, such as high concentrations of pesticides, are corrosive, reactive (form toxic gases), or highly inflammable. These wastes are defined in Federal Regulations (40CFR261). They require special management from generation through treatment and disposal as defined again by Federal Regulations. A detailed discussion of Hazardous Waste Management is covered in a section on Hazardous Waste.

Processing Methods

A variety of processing methods, as summarized in Table 13, are available at present for handling solid wastes. Most have been in use in some modification for at least the last 50 years. The choice of processing method will depend not only on the type of waste but also on location, sources, quantity of waste, method of collection, public opinion, and ultimately economics.

Solid waste management was a 4.5 billion dollar industry in 1968. It is only in recent years that the public has begun to worry about disposal of solids. Prior to that it was "out-of-sight, out-of-mind." With ever growing amounts of solid waste as detailed in the discussion on sources, and concerns about pollution of ground and drinking water as well as release of hazardous materials, public pressure is becoming a major factor in any decision on waste management.

The major disposal methods in use are landfill and incineration. Of potential interest in the United States are high pressure compaction and reclamation by recycling. Recycling is being used, but requires solution of institutional and technological barriers before becoming a major factor. Compaction is utilized in at least one major facility in the Meadowlands in New Jersey. Composting is practiced in Europe, but also has not been successfully applied in the United States although it does have potential. There are new processes and techniques appearing for waste disposal and for the first time an organized research and development effort was mounted in the early '70s to look at solid waste disposal; it has slowed down but there is ample opportunity for further progress.

Disposal methods could be discussed from the point of view of source: a brief summary of the most used methods for a variety of sources may be found in Table 14. This discussion will instead focus on the disposal methods most commonly in use today, landfill and incineration, followed by discussion of compaction, composting, and some of the newer disposal techniques.

The oldest method of disposal is dumping either on land or sea. Here dumping in distinguished from Sanitary Landfill (see below). Dumping costs between $6 and $10 per ton and has been used for all waste materials. It is totally unsatisfactory for putrescible materials such as food wastes and unsatisfactory from a public health as well as aesthetic and land use viewpoint, even for inert material such as demolition waste.

Open burning is often used for demolition waste, tree branches and stumps, and similar items; it is unacceptable because of the air pollution it creates. Neither dumping nor open burning have a place in the modern waste disposal scheme and are illegal.

Sanitary Landfill

Landfill is the most widely used method of waste disposal. There are 8900 authorized sites (about half publicly operated) used by the 6300 communities surveyed in 1968.[14] There appeared to be an equal number of unauthorized dumps. Unfortunately only 6% of the sites were considered to be "truly" sanitary. The remainder fell either into Category B or C on the US Public Health Service Classification Scale, summarized in

TABLE 5
Sources and types of industrial wastes

SIC group classification	Waste generating process	Expected specific wastes
Plumbing, heating, air conditioning Special trade contractors	Manufacture and installation in homes, buildings, and factories	Scrap metal from piping and duct work; rubber, paper, and insulating materials, miscellaneous construction and demolition debris
Ordnance and accessories	Manufacturing and assembling	Metals, plastic, rubber, paper, wood, cloth, and chemical residues
Food and kindred products	Processing, packaging, and shipping	Meats, fats, oils, bones, offal, vegetables, nuts and shells, and cereals
Textile mill products	Weaving, processing, dyeing, and shipping	Cloth and fiber residues
Apparel and other finished products	Cutting, sewing, sizing, and pressing	Cloth and fibers, metals, plastics, and rubber
Lumber and wood products	Sawmills, mill work plants, wooden container, miscellaneous wood products, manufacturing	Scrap wood, shavings, sawdust; in some instances metals, plastics, fibers, glues, sealers, paints, and solvents
Furniture, wood	Manufacture of household and office furniture, partitions, office and store fixtures, and mattresses	Those listed under Code 24, and in addition cloth and padding residues
Furniture, metal	Manufacture of household and office furniture, lockers, bedsprings, and frames	Metals, plastics, resins, glass, wood, rubber, adhesives, cloth, and paper
Paper and allied products	Paper manufacture, conversion of paper and paperboard, manufacture of paperboard boxes and containers	Paper and fiber residues, chemicals, paper coatings and fillers, inks, glues, and fasteners
Printing and publishing	Newspaper publishing, printing, lithography, engraving, and bookbinding	Paper, newsprint, cardboard, metals, chemicals, cloth, inks, and glues
Chemicals and related products	Manufacture and preparation of organic chemicals (ranges from drugs and soups to paints and varnishes, and explosives)	Organic and inorganic chemicals, metals, plastics, rubber, glass, oils, paints, solvents and pigments
Petroleum refining and related industries	Manufacture of paving and roofing materials	Asphalt and tars, felts, asbestos, paper, cloth, and fiber
Rubber and miscellaneous plastic products	Manufacture of fabricated rubber and plastic products	Scrap rubber and plastics, lampblack, curing compounds, and dyes
Leather and leather products	Leather tanning and finishing: manufacture of leather belting and packing	Scrap leather, thread, dyes, oils, processing and curing compounds
Electrical	Manufacture of electric equipment, appliances, and communication apparatus, machining, drawing, forming, welding, stamping, winding, painting, plating, baking, and firing operations	Metal scrap, carbon, glass, exotic metals, rubber, plastics, resins, fibers, cloth residues
Transportation equipment	Manufacture of motor vehicles, truck and bus bodies, motor vehicle parts and accessories, aircraft and parts, ship and boat building and repairing, motorcycles and bicycles and parts, etc.	Metal scrap, glass, fiber, wood, rubber, plastics, cloth, paints, solvents, petroleum products
Professional, scientific controlling instruments	Manufacture of engineering, laboratory, and research instruments and associated equipment	Metals, plastics, resins, glass, wood, rubber, fibers, and abrasives
Miscellaneous manufacturing	Manufacture of jewelry, silverware, plated ware, toys, amusement, sporting and athletic goods, costume novelties, buttons, brooms, brushes, signs, and advertising displays	Metals, glass, plastics, resins, leather, rubber, composition, bone, cloth, straw, adhesives, paints, solvent
Stone, clay, and glass products	Manufacture of flat glass, fabrication or forming of glass: manufacturer of concrete, gypsum, and plaster products; forming and processing of stone and stone products, abrasives, asbestos,and miscellaneous nonmineral products.	Glass, cement, clay, ceramics, gypsum, asbestos, stone, paper, and abrasives
Primary metal industries	Melting, casting, forging, drawing, rolling, forming, and extruding operations	Ferrous and nonferrous metals scrap, slag, cores, patterns, bonding agents
Fabricated metal products	Manufacture of metal cans, hand tools, general hardware, nonelectric heating apparatus, plumbing fixtures, fabricated structural products, wire, farm machinery and equipment, coating and engraving of metal	Metals, ceramics, sand, slag, scale, coatings, solvents, lubricants, pickling liquors
Machinery (except electrical)	Manufacture of equipment for construction, mining, elevators, moving stairways, conveyors, industrial trucks, trailers, stackers, machine tools, etc.	Slag, sand, cores, metal scrap, wood, plastics, resins, rubber, cloth, paint solvents, petroleum products

TABLE 6
Solid wastes common

Packing materials	fiber
	metal
	paper
	plastic
	wood
Maintenance materials	paints
	metal
	grease
	plastic
	rags
General housekeeping	waste paper
	fires
	glass
	solvents
	industrial chemicals

TABLE 7

Industry	Waste for disposal thousand tons/yr
Saw mills	33,000
Demolition	20,000
Food	7,200
Paper	5,000
Automobile and aerospace	1,600
Rubber	1,500
Chemical	1,400
Printing and publishing	1,300
Glass	1,400
Electronics	1,000
Wood products	3,000
Tanning	400
Paints	160

Table 15. There are additional classifications with respect to use in force in California and suggested in the new Federal Regulations.[15] There is an increase in "Sanitary Fills" and an elimination of "Dumps."

Sanitary landfill is an acceptable method of disposal of solids and provides for the ultimate disposal of many types of waste; exceptions are non-degradable materials such as plastic or aluminum which are placed in landfills. Other items material, toxic chemicals, and hazardous materials, are not allowed in landfills for safety. Where land is plentiful, or marginal areas are available for reclamation, sanitary landfills offer a number of advantages over other disposal methods including low initial and operating costs. Other advantages and disadvantages are summarized in Table 16. Sanitary landfill is basically the dumping of wastes followed by compaction and the daily application of an earth cover. This situation has improved in the last decade and by the mid-1980s—all landfills will be sanitary. Several techniques are available, some of which are depicted in Figure 2, depending on the type of site available. The one constant in all operations is the daily earth cover, preferably a sandy loam, amounting to, usually, one part earth for every four parts refuse. Another, which is being required in new landfills, is leachate collection and treatment. In addition these types of waste disposal are limited to "non-hazardous" materials unless the landfill is especially constructed, licensed and managed.

Proper site selection is as critical to a satisfactory landfill as is sound operation. Selection criteria include proper ground and surface water drainage and isolation as well as leachate collection and treatment, to prevent pollution of the ground water table. Location in a drainage basin near streams or lakes and in or close to the ground water table present special problems and should be avoided, where possible. Placement in the 100 year flood plain is prohibited. Accessibility of cover material is an important consideration. The use of tidal areas and marshes is prohibited. Dry pits, abandoned quarries and certain types of canyons of depressions are often satisfactory landfill sites.

The size of landfills is often restricted by the amount of land available. The capacity can be estimated with a fair degree of accuracy. Refuse on arrival may vary in density from 300 to 800 pounds per cubic yard, depending on the delivery method. Typically the density in the "fill," of the initial compaction with a typical crawler tractor will be 1000 lbs/yd for a single lift (layer) with a depth of 20 feet of less. For multiple lifts the initial density can reach 1250 lbs/yd. This initial loading increases by as much as 50% over a period of time as further compaction and decomposition takes place.[16]

Much of the material in the sanitary landfill decomposes over a period of between three and ten years depending on climate, permeability of the cover, composition of the refuse and degree of compaction. The decomposition in sanitary landfills is anaerobic as compared to aerobic degradation often found in other types of fill. Temperatures typically reach 120°F in the fill as a result of the degradation. The principal gas products are carbon dioxide and methane. The greatest gas production takes place in the first two years, according to a study made at the University of Washington. Ammonia and hydrogen sulfide are not problems in sanitary landfills although small amounts of these gases are produced. Odors resulting from the decomposition of putrescible material can be controlled by observing good operating practice; that is, covering the fill continuously and sealing surface cracks. Fire hazard and insects and vermin are not a problem, as compared to dumps, in a properly operated sanitary landfill although chemical control of the latter two is sometimes required.

Completed landfills are suitable for use as recreational facilities, airfields and parking areas; light industrial buildings may be erected on landfill. Building of residential structures on fill requires special precautions because of the potential hazards associated with the evolution of methane and other decomposition gases.

The cost of operating a sanitary landfill makes it an attractive means of disposal where land is available. Costs for a

TABLE 8
Waste generation for large firms[13]

Industrial classification			
Title	Employment 1[a]	Annual wastes vol. Cu yd 2[b]	Annual wastes per employee cu yd 3[c]
Ordnance and accessories	29,356	131,404	4,476
Canning and preserving[d]	11,389	102,238	8,977
Other food processing (except 203)	2,012	17,545	8,720
Tobacco	[e]	[e]	[e]
Textiles	[e]	[e]	[e]
Apparel	601	1,248	2,077
Lumber and wood products	[e]	[e]	[e]
Furniture and fixtures	[e]	[e]	[e]
Paper and allied products	250	9,360	37,440
Printing, publishing and allied	968	7,020	7,252
Chemicals and allied	[e]	[e]	[e]
Petroleum refining	[e]	[e]	[e]
Rubber and plastics	481	9,069	18,854
Leather	[e]	[e]	[e]
Stone, clay, glass, and concrete	1,258	6,617	5,260
Primary metals	[e]	[e]	[e]
Fabricated metal products	3,565	47,078	13,206
Nonelectrical machinery	8,872	101,153	11,401
Electrical machinery	7,807	57,252	7,333
Transportation equipment	4,100	100,776	24,580
Instruments	[e]	[e]	[e]
Miscellaneous manufacturing industries	[e]	[e]	[e]

[a] Column 1: Data on employment were obtained for those large firms which were surveyed and included in the wastes calculation from the research department of the Association of Metropolitan San Jose (Greater San Jose Chamber of Commerce).
[b] Column 2: FMC report, Solid Waste Disposal System Analysis (Preliminary Report), Tables 10 and 11, 1968. [5]
[c] Column 3: Column 2/Column 1.
[d] For Canning and Preserving (SIC 203), no individual firm data were available. The industry total developed for the county as a whole was divided by the total employment in the industry (specially tabulated) to arrive at the multiplier. See text for further explanation.
[e] Data not available.

sanitary fill will vary between $3 and $10 per ton, depending on location and size of the fill. Small fills, handling less than 50,000 tons per year, will have a unit cost of $5 to $10 per ton. A large urban fill more typically shows costs of $3 to $6 per ton. The wide variation is a result of location differences, which include differences in land acquisition costs, labor costs and operating differences due to local surface conditions and requirements.

The use of landfill will continue; however, its future, particularly in densely populated urban areas, is in doubt. Land is at a premium for this type of application close to urban centers. What land is available must be preserved for non-combustible material and ashes. For examples, one urban county in New Jersey has less than three years landfill capacity available and in portions of Long Island no more land for landfill is available. Hauling costs too, as well as public resistance in more rural areas is making landfill less attractive for urban areas such as metropolitan New York. Finally, landfill does not provide for maximizing the value of refuse as a source of raw materials.

Recent studies to find alternatives to traditional landfill practices include a demonstration of shredding prior to filling. Only domestic refuse was shredded; the product was a superior fill compared to “raw” refuse. It could be left uncovered with satisfactory sanitary and aesthetic results and was easier to dump and compact. Flies and rats did not breed on the shredded refuse.

The compacted, uncovered fill also had better weathering and load bearing characteristics. This can be achieved at a cost of about $5.00 per ton in a 65,000 ton per year operation.[17] The method has some attractive features, and some commercial facilities including one in Monmouth Country, NJ, which incorporates some recycle, use this principle. However, operating and investment costs do appear to be higher than the more traditional method of filling “raw,” as collected, refuse.

Baling of refuse may be particularly attractive where landfill sites are not locally available. A feasibility study was carried out in Chicago which showed that this method overcomes many of the present objections to landfill. The

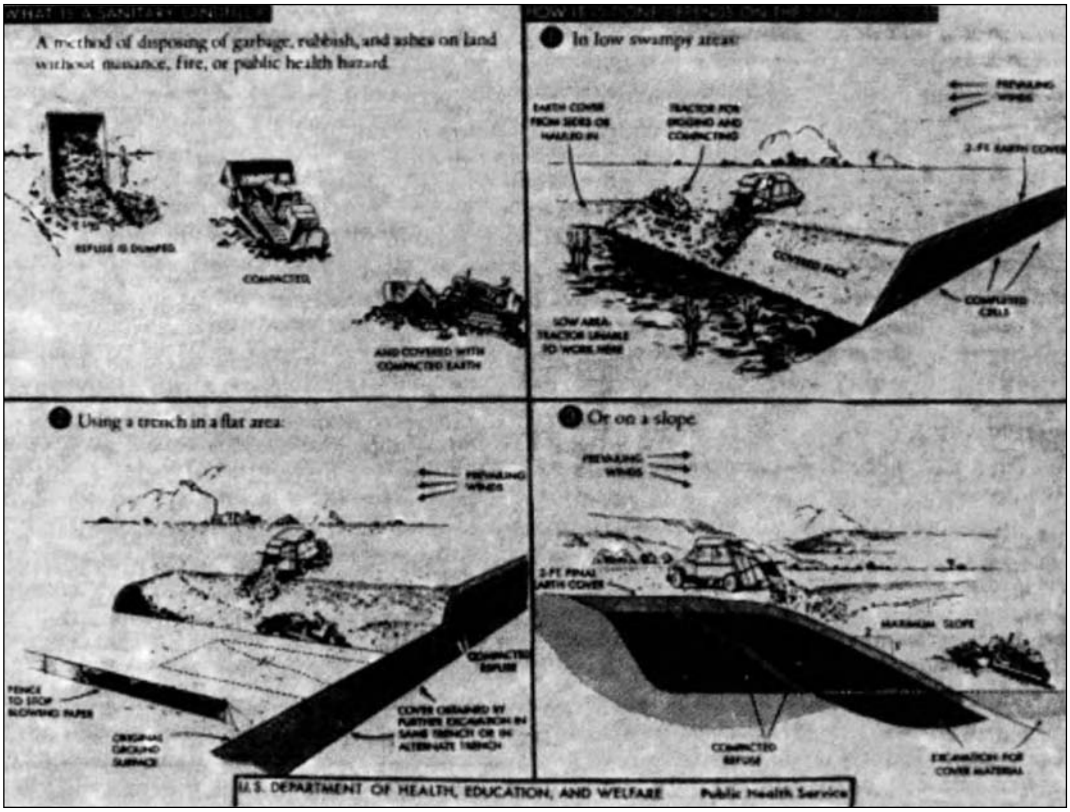


FIGURE 2 Sanitary land fill operations: Credit: US Public Health Service.

Japanese have been leaders in this area using high pressure presses to provide solid cubes suitable for use in building new land in tidal areas. A facility is being successfully operated in New Jersey. More details may be found in the discussion of compaction.

Incineration

Incineration is essentially a method for reducing waste volume and at the same time producing an inert, essentially inorganic, solid effluent from material which is largely organic. Typical feed analyses are shown in Table 4. In addition to the solid product a gas is produced consisting mainly of CO_2, H_2, O_2 and N_2 but containing other gaseous components in tract quantities depending on the type of material burned and the operating conditions. Incineration is not an ultimate disposal method in that the solid residue which is primarily an ash containing some metal must still be disposed of, usually as landfill. The primary advantage is that it reduces the volume to be disposed of and results in a "clean" inert fill. For every 100 tons of material fed to the incinerator approximately 20 tons of residue result. The volume reduction is even more significant, often resulting in a 90% lower solids volume for organic materials.

The theory of incinerator operation is very simple. A unit is designed to expose combustible material to sufficient air at high temperature to achieve complete combustion. Combustion is usually carried out in fuel beds to ensure good contact of air and refuse. Several types of configurations are used to achieve contact; these include concurrent flow of fuel and air-underfire, countercurrent flow of fuel and air-overfire, flow of fuel and air at an angle to each other—crossfeed; and combinations of these. The combustion is basically the same for all methods in that at the ignition front oxygen is rapidly consumed in the reaction $O_2 + C \rightarrow CO_2$ and if oxygen is depleted $CO_2 + C \rightarrow 2CO$. Therefore, sufficient oxygen must be available to obtain complete combustion; usually this is provided by adding additional air in the chamber above the fuel. Incinerators are typically operated with about 50 to 150% excess air in order that the gas temperatures do not drop below that required for good odor-free combustion; this is usually in the 1700–2300°F range. Recent trends have been to go to the higher part of this range while old units often operate at 1600°F or below. The effect of excess air on gas composition is summarized in Table 17 for a typical refuse. A detailed discussion of typical air requirements and their effect on the thermal balance may be found in Principles and Practices of Incineration.[18]

Trace components in the incinerator-start gas include some SO_2 and NO_x. The former depends on the sulfur in the refuse and is typically around 0.01 to 0.02%. Nitrogen oxide is generally formed in combustion processes and depends on the amount of excess air and to some degree the operating temperature of the incinerator. Typical values of two pounds of equivalent NO_2 per ton of refuse have been reported.[19,20]

TABLE 9

Waste generation typical for small firms[14]

Industrial classification				
Title	Weekly wastes vol. per firm cu/yd 1[a]	Annual wastes vol. per firm cu/yd 2[b]	Average employment per firm 3[c]	Annual wastes vol. per employee cu/yd 4[d]
Ordnance and accessories	2.500	130.00	[e]	[e]
Canning and preserving[4]	—	(not surveyed)	—	
Other food processing (except 203)	10.875	565.50	26.979	20.961
Tobacco	—	NA	—	—
Textiles	—	NA	—	—
Apparel	4.000	208.00	5.882	35.360
Lumber and wood products	16.083	836.33	17.247	48.492
Furniture and fixtures	23.000	1,196.00	13.767	86.877
Paper and allied products	44.650	2,321.80	35.479	65.442
Printing, publishing and allied	6.448	335.29	13.289	25.230
Chemical and allied	6.506	338.31	18.439	18.438
Petroleum refining	[e]	[e]	[e]	[e]
Rubber and plastics	5.275	274.30	9.596	28.583
Leather	[e]	[e]	[e]	[e]
Stone, clay, glass, and concrete	9.415	489.60	16.747	29.235
Primary metals	2.000	104.00	23.409	4.443
Fabricated metal products	5.284	274.65	12.951	21.214
Nonelectrical machinery	4.450	231.40	12.921	17.909
Electrical machinery	6.733	350.13	21.036	16.645
Transportation equipment	4.550	236.60	16.490	14.348
Instruments	3.600	187.20	20.1933	8.943
Miscellaneous manufacturing industries	1.250	65.00	10.931	5.946

[a] Column 1: Data obtained and calculated for each SIC on the basis of small firm questionnaire response supplied by FMC.

[b] Column 2: Weekly average in Column 1 multiplied by 52.

[c] Column 3: Average size of small firm estimated from the distribution of firms by employment size, supplied by the California Department of Employment (Research and Statistics), San Francisco Office.

[d] Column 4: Column 2/Column 3.

[e] Data not available.

TABLE 10[12]

Agricultural waste (1966)

Waste	Composition	Amount (million tons/yr)
Crop residue	Corn stalks, grain stubble, cull, fruit and vegetable, vines, rice hulls, bagasse, tree prunings, etc.	552
Animal manure (paunch manure)	Organic matter, protein, fat, carbohydrates, nitrogen, phosphorus etc.	1.532[a]
Poultry manure	Same as animal manure	30[b]
Animal carcasses	—	—
Forest operations	—	25
Pesticides, insecticides, etc. residue and containers	Chlorinated hydrocarbons, organophosphorus compounds, other organics and inorganics, e.g. sulfur, lead arsenate, etc.	—

This is equivalent to 500 to 1000 ppm of NO^x in the off-gas depending on the refuse composition and the amount of excess air. Other trace components can be found in the off-gas and air summarized in Table 18. Their presence or absence is very much dependent on the type of refuse incinerated and the operating conditions.

Particulate matter is also present in the stack gas and is removed by the usual techniques discussed in the section on Air Pollution. Particulate loadings of 3 to 25 pounds per tonne of refuse burned have been reported.[21,22] Typically, particles range from 5 to 350 microns in size with 30% by weight under 10 microns and 75% less than 200 microns in size.

Solids residue from incinerators will vary widely with the type of feed and incinerator operating conditions. Typical residues have been examined by the Bureau of Mines. The results of this work are summarized in Table 19. A typical ash and slag chemical analysis may be found in Table 20. This residue can be utilized in road fill or separated (see Reclamation).

Incineration can effectively be divided into local, onsite and central methods. The basic principles are the same but the applications vary considerably. Central incineration facilities handle refuse from many sources and a wide variety of feeds. Local incinerators handle either special feeds, onsite, such as industrial or hospital wastes, or serve a particular small location such as an apartment house. Size is not necessarily a criterion although generally central incineration facilities have capacities in excess of 100 tons per day.

At the present time there are about 200 central incineration facilities in use (making this type of waste reduction facility the most prevalent one). Central incineration handled about 15 million tons of waste annually and is concentrated in the northeastern part of the United States. It is also widely practiced in Europe. The practice of incineration of wastes was growing as land for fill, particularly in urban areas, becomes scarcer and technological improvements provide more efficient and cleaner systems.

A typical incineration facility will have a capacity ranging from 100 to 1200 tons per day with individual furnaces usually limited to a 300 ton per day rating. Most large incinerators today are continuous-feed rather than batch design because operation is more controlled and easier. In addition the absence of the heating and cooling cycle results in lower maintenance and a higher capacity per investment dollar. Air pollution control is improved significantly in continuous-feed incinerators are compared to batch plants.

A large central incineration facility is schematically shown in Figure 3. It can be divided into five areas: (1) the receiving section which includes the weight station, storage hopper and bucket crane; (2) the furnace—which includes the charging hopper, stokers, furnace chamber and air feed system; (3) the effluent gas treating facilities; (4) the ash handling system; and (5) the cooling water system. The particular system shown does not have provision for waste-heat recovery; only a few systems incorporate this at present.

For mixed refuse, a typical refractory-wall incinerator will have 12.7 cubic feet in the primary furnace chamber and 18.5 cubic feet in the secondary chamber per ton of refuse per 24 hours with a grate loading of 77 pounds per square foot per hour. Volume and loading requirements will vary with the type of feed as well as furnace configuration. Typically the values quoted correspond to a 12,500 Btu per hour per cubic foot heat release. A detailed discussion of furnace design is

TABLE 11
Unit generation rates

Animal	Waste (tons unit yr)
Cattle	12.0
Cows, milk	10.6
Hogs	8.0
Sheep	3.0
Chicken, broilers	0.0045
Turkeys	0.025
Chicken, layers	0.047

TABLE 12
Generation by type of solid wastes from the mineral and fossil and fuel industries (1965)

Industry	Mine waste	Mill tailings	Washing plant rejects	Slag	Processing plant wastes	Total (thousands of tons)
Copper	286,600	170,500	—	5,200	—	466,700
Iron and steel	117,599	100,579	—	14,689	1,000	233,877
Bituminous coal	12,800	—	86,800	—	—	99,600
Phosphate rock	72	—	54,823	4,030	9,383	68,308
Lead-zinc	2,500	17,811	970	—	—	20,311
Aluminum	—	—	—	—	5,350	2,350
Anthracite coal	—	—	2,000	—	—	2,000
Coal ash	—	—	—	—	24,500	24,500
Other	—	—	—	—	—	229,284
Total	419,571	288,900	144,593	23,919	40,233	1,146,500

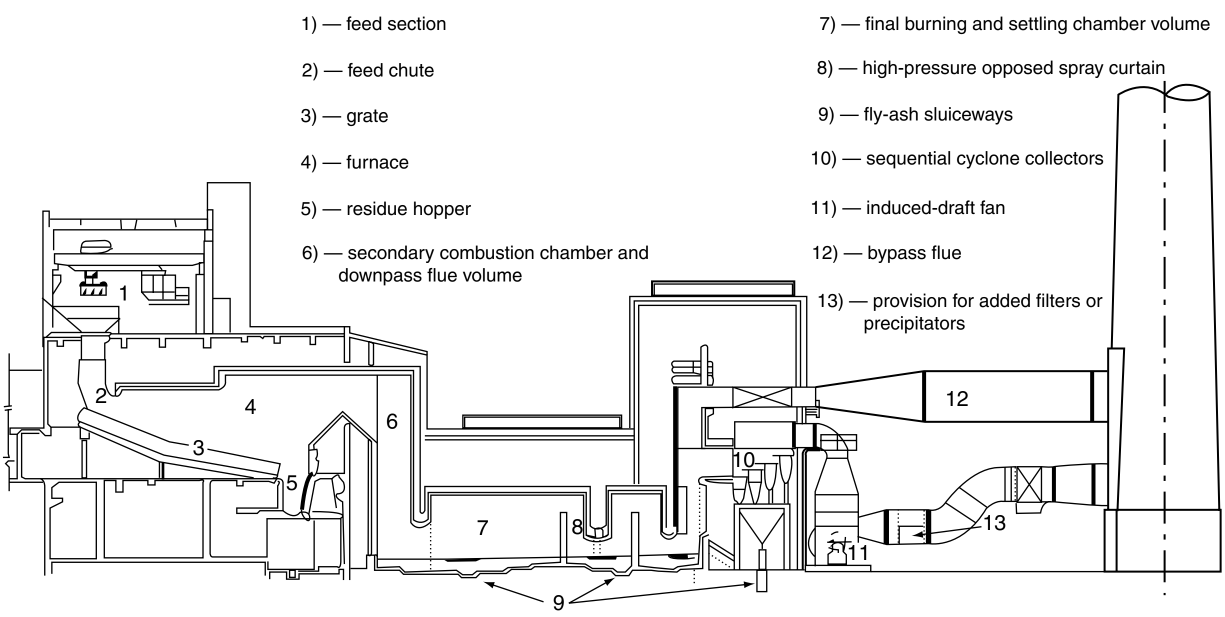


FIGURE 3 Large incinerator schematic.

TABLE 13
Solid waste management methods

Type	Present usage	Relative cost	Items disposed of	Principal benefits
Sanitary landfill	Most used (80%) decreasing[9]	Low[b]	All except hazardous materials	Low initial cost, takes all items may reclaim land
Central incineration	Increasing second largest method (4%)	High	All burnable except special items and over-sized items	Reduces volume, clean product can produce by-product items
Open burning	Illegal	Low	Construction wastes, leaves, agricultural waste	
Compaction, high pressure	Two plants in operation	Medium-high	All except hazardous materials	Produces dense, essentially inert blocks for fill
Composting	Very few	Medium-high	Organic only. No tires, large pieces	Provides soil conditioner
Garbage grinding	Large number home units	High	Organic only	Reduces domestic collections
Dumping	Not legal	Lowest	Non-putrescibles	
Recycling	Only for selected materials and areas, increasing	High	Selected. Depends on process	Reduces quantity for ultimate disposal

[a] Many landfills are not sanitary but are included in this classification.

[b] Low under \$10/ton; Medium \$10 to \$30/ton; High \$30 + ton.

beyond the scope of this work and the reader is referred to an excellent work by Richard C. Corey.[22]

Incineration in the past has received a bad reputation because of poor control of gaseous effluents and sloppy handling of solid and liquid effluents. With proper design and operation an incinerator can meet or exceed requirements on all effluent discharges. A modern central incinerator is a more complex operation than a large commercial steam boiler. It therefore requires skilled operating, maintenance and supervisory personnel to ensure efficient operation.

At the present time control of particulate matter in the effluent gas is the most critical problem in incinerator design

TABLE 14
Waste source and disposal methods

Source	Methods recently used	Methods for future consideration
Municipal	Landfill (80%) Incineration (10%)	Compaction Composting Recycle and reclamation Chemical processing
Industrial	Landfill Incineration Recycle	Recycle and reclamation Chemical processing
Demolition	Dump	Reclamation Incineration
Construction	Dump Open burning	Reclamation Incineration
Sewage	Landfill	Incinerate Compost Chemical processing
Agriculture	Landfill (plowback) Incinerate Open burning Dump	Compost Chemical processing

and operation. A typical modern facility will include either a wet scrubber or a spray chamber followed by solid separation in a baghouse filter or electrostatic precipitator. These methods can achieve up to 99% removal of particulate matter, which will meet the code requirement of 0.0 pound 0.03 grams per DSCF particulate in flue gas at 50% excess air in almost all cases. The costs for this required cleanup are significant and can be as high as $2000 per daily ton of refuse capacity (see Air Pollution Control). Odor control is achieved by providing adequate time (0.5 sec) in the combustion chamber at temperatures above 1500°F. As incineration temperatures in modern units are between 1800 and 2200°F this poses no problem.

Waste heat recovery has been practiced to only a very limited degree in the United States. Less than 10% of the incinerators surveyed in 1966 had waste heat recovery facilities. Presently, there are six major resource recovery mass burning units in the United States and about a dozen are in the design or construction stage. This is in contrast to the practice in Europe where waste heat recovery is practiced in a large number of units. Those US plants that do recovery of heat incorporate the water wall principle used in furnaces; in contrast, a few refuse-derived fuel units have been built modifying existing boilers previously used as suspension units; again this compares to the European practice where water wall incinerators are common.

One water-wall incinerator (600T/D) has been in operation since 1967 at the US Navy Base, Norfolk, Va.[23] Several reasons have been advanced for the lack of heat recovery in the United States. These included adverse economics. In addition corrosion problems and much slagging of the walls (due apparently to differences in waste composition) seems to have held back the use of water walls. With new technology, and a tighter fuel picture, waste heat recovery in incinerators will become commonplace in the next decade. In general municipal service, a 1200 per day ton facility in Montreal went on stream in 1970 and produces 100,000 pounds of steam per hour and a 1600 ton per day facility in Chicago started up in 1971. Cogeneration facilities that generate steam and electricity are now being designed.

The investment and operating costs for incinerators are high and to date have been one of the major deterrents to wider use. Typical installations of the 600 to 1000 ton per day range require an investment of $20,000 per ton or installed daily capacity depending in part on the air pollution control devices which can account for 20% of the total cost as well as size. Water-wall installations typically will run more than refractory lined incinerators; the 1600 ton per day Northwest Chicago plant cost about $16,000 per ton. Present costs are about $40–45,000 per installed ton.

Operating costs including amortization of the investment will vary between $50 and $200 per ton. Actual values for generated steam have not been published but estimates indicate

TABLE 15
U.S. Public health service landfill classification

A: *Sanitary*
Landfill operated without public nuisance or public health hazard; covered daily and adequately, no deliberate burning practiced.

B: Operated without public nuisance or public health hazard, but location permits modification of "A" such as burning of certain types of waste at site, or covering of fill only three times weekly.

C: Operating techniques permit development of public nuisance and potential health hazards, such as fly breeding, rodent substance, and odors.

TABLE 16
Sanitary landfill—advantages and disadvantages

Advantages

1) Most economical method when land is available.
2) Low initial investment.
3) Complete and final disposal.
4) Short period of time from need to full operation.
5) Flexible daily capacity with same working force.
6) Reclamation of marginal land for recreational and other uses.
7) All types of waste are acceptable.

Disadvantages

1) Lack of close-by suitable land in urban areas may make uneconomical.
2) Public opposition in, or near, residential areas.
3) Settling after completion means continued maintenance.
4) Public nuisance and health hazard is not properly operated.
5) Products of decomposition, methane and other gases, may create hazard.
6) Require special practices for construction on completed fill.

Source: US Dept. HEWPHS. Pub. No. 1792.

TABLE 17
Incinerator effluent gas composition[d]

			Combustion	Products
	Feed		Product	Gas composition vol. %
Component	Wt. %	Excess air component	0%	50%
Carbon	30.0	CO_2	16.00	9.86
Hydrogen[a]	6.1	H_2O	19.5	12.04
Oxygen[e]	43.3	O_2	—	7.94
		NO_x	—	~700 ppm[b]
Nitrogen	0.5	N_2	64.5	70.15
Sulfur	0.1	SO_2	0.02	0.01
Noncombustible	20.0[c]	—	—	—

[a] Includes moisture of 20%.

[b] Not computed.

[c] No combustion of metal assumed, in actuality same takes place.

[d] Excluding particular matter.

[e] Based on theoretical required less O_2 contained in feed.

TABLE 18
Incinerator stack gas contaminants

Component	Amounts reported
Organic Acids	
Formic	25–133 ppm (31)
Palmitic	0.6 lbs/ton of refuse (32)
Acetic (all organics)	40–600 mg m^3 (33)
Esters	
Methyl acetate	5–137 ppm (32)
Ethyl acetate	—
Aldehydes	
Acetaldehyde	2.8×10^{-4} (33)
Formaldehyde	1.1 lbs/ton refuse (32)
Hydrocarbons	4 μgm/gm of particulate (29[a])
Halogenated Hydrocarbons	
(depends on plastics and aerosols)	6–120 mg m^3 (33)
	0.44–10 ppm (31)
	0.3 lbs/ton refuse (32)
Ammonia	0.15–1.5 ppm (31)
Nitrogen Dioxide	0.15–5 mg m^3 (33)
Nitrogen Trioxide	4–100 mg/m^3 (33)
HCl	300–1200 mg/kg refuse (33)
	30–350 mg/m^3 (33)
SO_2	0.25–1.2 ppm (31)
	1.9 lbs/ton refuse (32)

[a] When burning rubber.

a potential recovery of $10 to $15 per ton or refuse for the steam sold.

In addition to the installation of units with waste heat recovery facilities the trend will be to longer installations to take advantage of the economics of size. However, small skid-mounted units with capacities in the range of 20 ton to 50 tons per day are available at a cost which will make central incineration for smaller communities economical possible. The first of these units that met most air pollution codes was marketed by Combustion Engineering in the late ’60s. It was ahead of its time and did not enjoy success. However, similar size units with energy recovery are now finding a good acceptability. Another development will be the installation of more combined sewage sludge-refuse incineration facilities using the Nichols multiple earth rotating grid or similar installations.

Rotary kiln incinerators have been successfully utilized in handling mixed wastes, often predominantly industrial wastes. They are particularly useful where long residence times are required to insure complete combustion. Their disadvantages center primarily around high maintenance costs. One of these incinerators has been in operation at Dow Chemical, Midland, Michigan for almost 30 years.

Incineration in fluidized beds was demonstrated by Black and Clawson in a facility in Franklin, Ohio, as well as by Combustion Power Co. which combined a fluid bed combustor with a gas-turbine generator to produce power from 400 tons per day of refuse. The latter demonstrations had technical difficulties and fluidized bed technology has not been commercialized.

Numerous on-site incinerators are operated with satisfactory results for the reduction of industrial wastes. These facilities are usually specially designed to handle one type of refuse. Typical materials that are incinerated include plastics, rubber, wood scrap and paper. The economics of waste recovery are changing so that often these materials are no longer burned. For example wood chips and sawdust at sawmills are often sent to paper-mills or composition board producers as feed, where formerly they were burned. Insulated copper wire and automobiles are incinerated to remove the organic components prior to recovery of the metal. Special liquid

TABLE 19
Incinerator residue composition ranges[23]

	Wt.%
Moisture	24–40
Components, Dry Wt. Basis	
Tin cans	16–22
Iron, all types	9–14
Nonferrous metals	0.1[a]–3.7
Stones and bricks	0.8–1.9
Ceramics	0.6–1.5
Unburned paper and charcoal	4–12[b]
Partially burned organics	0.1–1.3[b]
Ash	12–18
Glass	37–50

[a] After hand picking.

[b] High temperature operation will decrease this markedly.

TABLE 20
Incinerator ash and slag analysis[24]

SiO_2	46%	Na_2O	3%
$Al2O_3$	21%	K_2O	1%
$Fe2O_3$	8%	P_2O_5	2%
TiO_2	3%	BaO	0.6%
CaO	10%	SO_3	0.3%
MgO	3%	ZnO	0.5%

wastes such as water containing organics are incinerated. The special applications are numerous, including the incineration of radioactively contaminated wastes.

In addition to the more traditional incinerators, whether rectangular or cylindrical, special designs are employed in industrial waste disposal. For example shredded plastic[24] as well as “white water” from paper-mills[25] is incinerated in a fluidized-bed combustion chamber. Industrial sludge is being burned in a rotary kiln[26] by Kodak. Dow Chemical has had a rotary kiln on line for over 20 years handling a mix of refuse and industrial waste. Solid cyanide waste in automobile plants is put into solution and then burned while aluminum chloride sludge from a petrochemical operation can be burned to produce HCl and alumina.

Although hauling to landfill sites is the present disposal method for many industries, on-site incineration of industrial wastes will receive wider use for waste disposal where recycle is not possible and volume is sufficient, in excess of 500 lb/day, to justify an installation.

Hospital wastes are now commonly being disposed of in onsite incinerators. To eliminate the possibility of spreading infection, wastes should be promptly incinerated. This is best done in an onsite facility. The average load for hospital incinerators is about 20 pounds per day per patient with a very high fraction of garbage and paper and plastic throw-away products.[27] Provision must also be made to handle pathological wastes, therefore combustion temperatures should be in excess of 2000°F and adequate residence time for the gases at 1500°F should be provided.

Refuse-Derived Fuel

In the past there has been some objection to direct firing of refuse. Partly these are aesthetic in nature and partly they result from the high variability of raw refuse. At one time, it was thought that firing of coal and refuse might overcome a number of these problems. Indeed it did, but not the institutional problems of handling raw refuse. As a result several processes were developed to produce refuse-derived fuel (RDF). These processes have been in development for the past ten years and have not found, to date, wide commercial application. Essentially, raw refuse is separated into the organic and paper portion, and the recoverable, recyclable components, such as ferrous metal, aluminum, glass. This separation is carried out after shredding, as discussed under the section on Reclamation, Reuse and Conversion. The shredded material can then be fed as is; and that form is the lowest grade of RDF. Some cases it is palletized, and fired as pellets. Palletizing reduces handling problems and increases storability at the expense of an additional processing step. RDF has been successfully co-fired with coal and it is anticipated that over the next ten years a number of RDF fired power boilers will be installed either for steam generation or electric power generation.

Compaction

The reduction of waste volume is receiving considerable attention in an effort to reduce collection costs; compaction is one of the favored methods to achieve this reduction. High pressure compaction has been developed by Tezuka Kosan of Japan to provide a high density product suitable as an essentially inert fill or even as a building material.[28] Using this product as a base covered with a minimal earth over, the Japanese have reclaimed land from tidal areas having a water depth of 10 feet.[29]

The Japanese process shown in Figure 4 collects refuse and subjects it to three stages of compression with the final main press exerting 3000 psi on the refuse. The resulting bale is usually wrapped in chicken wire and coated with asphalt for ease of handling and to prevent crumbling and/or leakage. The bales have a density of between 1900 and 2300 pounds per cubic yard and result in a volume reduction of about 90%. This compares to densities of about 1200 to 1500 pounds per cubic yard achieved in lower pressure compaction. The product bale is inert and such bales have survived exposure in Tokyo Bay for three years to date without visible signs of degradation.

Studies by the Japanese indicate that the high pressure squeezing and resultant elevated temperatures decrease the BOD from 6000 in the raw refuse to 200 in the product. Similarly the COD of 8000 (which compares to about 14,000 in US refuse) was reduced to about 150. Inspection of the interior of the bale shows a homogenous, plastic like mass. The bale will not support vermin, rodents, or insects and is essentially odor free even if it is not protected by an asphalt coating. The only other product of the compaction is waste liquor, which amounts to 5% of the feed in Japan and will probably be about 3% with US refuse because of its lower moisture content.

Composting

Composting is the biochemical degradation of organic material (see: COMPOSTING) to yield a sanitary soil supplement. Anaerobic composting has been practiced in Asia and is the process by which sanitary landfills degrade refuse. Modern composting has been practiced in Europe for over 50 years using aerobic microorganisms. The practice has been attempted on a commercial scale in the United States but with very limited success. The unattractiveness of composting is primarily the result of American agriculture's orientation to chemical fertilizers. This has made the large scale marketing of compost difficult. Composting, while not economical now, could prove more attractive as public opinion moves toward an attitude which requires that wastes returned to the earth be compatible with the environment.

Composting has one overriding advantage; it is the only process which provides for recycling of organic residue. The process can handle garbage and other organic refuse (but not plastics) as well as sewage sludge and industrial waste from certain operations such as saw and paper mills. The primary disadvantages are cost, the need for fairly large areas for final outdoor curing, a slight odor associated with a composting plant, and lack of a market for the product.

Composting is practiced in several forms. Traditionally rows of refuse, shredded or ground, four to six feet high, are exposed to the environment and turned regularly. This is known as the "windrow" method and is still used. Complete composting can be achieved in 10 to 14 days, if seeding with compost is employed, but often four to six weeks.

Mechanically aided aerobic composting is carried out in a number of processes. Among the more prevalent are the Dano process, the Earp-Thomas Multi-Bactor compost tower, and a number of cell-type stated tower systems. Decomposition takes place under aerobic conditions with the microorganisms supplied by seed compost. Typical operating temperatures reach 130 to 140°F. Material is held in the unit from one to six days depending on the process. This is usually followed by an open air curing. A new plant at Sehweinfurt, Germany, using the Caspari-Brikollare process, produces briquettes in which from the compost is stored until it is to be used.

Raw materials suitable for aerobic composting will be finely ground (coarse for windrowing) and have a maximum carbon to nitrogen ratio of 50 to 1. It is important that good dispersion of air can be achieved and that the moisture level be maintained between 50 and 60%. Recycling of between 1 and 10% of active compost enhances the composting process by minimizing the time required for sufficient microorganisms to develop. The yield from composting is about one volume for every three volumes of feed; the weight yield is between 30 and 40%.

In Europe compost is utilized as an organic soil conditioner in luxury agriculture such as vineyards, hotbed vegetable farming, flower and seed production, fruit farming and the improvement of recreational land. It has found little or no use in basic agriculture, nor is it used for erosion control. In Germany less than 1% of the domestic refuse is composted and in Holland only about 15% is so treated. There appears to be no increase in composting operations because of a lack of additional marketing opportunities.

The investment for a composting facility varies widely depending on size and process. Investments of between \$7500 and \$12,000 per daily ton of capacity have been reported; no valid average figures can be reported because there are so few operational plants. Operating costs, including the cost of capital, will vary between \$8 and \$12 per ton of refuse (assuming labor at \$15,000 per man year) on a US basis. European investments appear to be as low as \$1000 per daily ton and operating costs in 1964 were between \$3.20 and \$6.60 per ton, with an average of \$4.50 per ton of refuse.[30] Part of this cost was recovered by sale of salvage (16¢ per ton) and compost; the average recovery amounted to \$1.17 per ton resulting in a net average cost of \$3.38.[31]

The Economics of Waste Disposal

Economic considerations have and will, of course, continue to play a significant role in the choice of waste management method. Table 21 summarizes both operating and investment costs for principal waste processing methods. It must be realized that values can vary widely depending on local conditions. Technological improvements can also alter the price structure. Most important, however, is the fact that long range implications of waste management, environment and resource considerations cannot readily be reduced to a quantitative cost, and these factors should weigh heavily in the choice of a waste processing method.

Reclamation, Reuse and Conversion

There has been a salvage industry as long as there has been waste. The intensity of this effort has been limited however and varies very much with location as well as the economic situation at the time. No concentrated effort, except perhaps in wartime, has ever been made to recover and recycle a high percentage of waste. True landfill can be considered a form of reclamation but it is a very low grade use of refuse. This is not to say that the salvage industry is small: the latest figures indicate that it has sales in excess of \$10 billion per year.

TABLE 21
Waste processing investment and operating costs (1980 basis)

Method	Investment $/ton daily capacity	Operating costs[a] $/ton
Sanitary landfill	5–20	10–40
Hazardous landfill	10–40	30–100
Central incineration		
No waste heat recovery	40,000	30–50
Waste heat recovery	10,000–80,000	40–80[b]
Composting	12,000	8–12[c]

[a] Per ton of capacity, based on 2–20 foot lifts.

[b] No allowance for value of steam, which will be between $1.00 and $2.00 per ton.

[c] 3 shift operation 450 tons/day.

TABLE 22
Value of reclaimed waste, average 1980 prices

	$/ton
Paper	
Newsprint	3–25
Other sorted	Up to 50[a]
Glass	3–90
Cotton	4–20
Metals	
Iron and steel	10–60
Copper and brass	270–800
Aluminum	160–250
Lead	70–160

[a] Depends on type, color, etc.

In certain areas such as paper, with 11 million tons recycled in 1968 (25% of consumption), aluminum with 700,000 tons recycled (30% of production), copper with 1.5 million tons recovered, and iron and steel salvage are an important source of raw materials. Unit salvage values are summarized in Table 22. Automobiles are regularly recovered; it is economically a break-even operation once the auto body is delivered to the salvage operation.

Mixed waste such as municipal refuse has not been a major source of this salvage and contains large amounts of additional metals as well as other potentially valuable materials. However it is only recently that essentially total recovery has been considered. That is, "waste" is being looked at as a potential "natural resource." As an example, one company, Industrial Services of America, operated a separations plant for industrial solid wastes where odor is no problem and manual separation is feasible in the late '40s.

One of the major problems of recycle and reuse has been separation. Most ferrous metal is easily separated by magnetic means, but other waste separation requires hand picking, which is very costly. This method can also create public nuisance in that the odors from such operations can be significant and have resulted in the closure of several such operations. Studies in the late '60s by the Bureau of Mines, Stanford Research Institute and others provide potential alternates for waste separation, but these have not found wide commercial application.

The work at SRI is particularly interesting because it is intended as a total separation process; waste is shredded and then classified into components by density in an air classifier.[32] While the process has not achieved separation of materials with densities that are close together (and much refuse is in the situation) the concept as illustrated in Figure 4 is in the right direction and should lead to promising results.

Similar work has been piloted using a hydrapulper to separate organic, glass and cellulose from metal. Hydrapulping as developed by Black Clawson Company costs about $6 per ton of feed.[33] A demonstration unit (Figure 6) with a 70 T/day capacity was operating in Franklin, Ohio, using the Hydraposal system (of which Hydrapulping is a part). Installed costs were about $14,000/ daily ton for a 300 T/day plant with net operating cost of about $4.5/ton. About 400 pounds/ton of Fiber are reclaimed along with glass cullet, metals and steam. A material balance is shown in Figure 5. However a dry separation such as that proposed by SRI may have significant advantages over wet separation because the removal of water from cellulose is a very costly operation. The fiber from Black Clawson was fed to a roofing paper mill.

The aluminum industry, for example, as shown in Figure 6, has suggested a process for taking municipal refuse and separating it into aluminum, other nonferrous metals, ferrous metals, glass and other waste. Dr. James Etzel of Purdue piloted a process, based on hydrapulping, which handles sewage and solid waste and yields metals and a slurry containing fine particles of glass and organics which can be further treated or used as a soil supplement. Such techniques require considerable additional development and refining but will be one of the key waste management tools of the future.

Pyrolysis, once thought to be a promising process, has not yet found wide acceptance. However, it remains technically feasible. Pyrolysis is the thermochemical degradation of complex organic molecules into low molecular weight

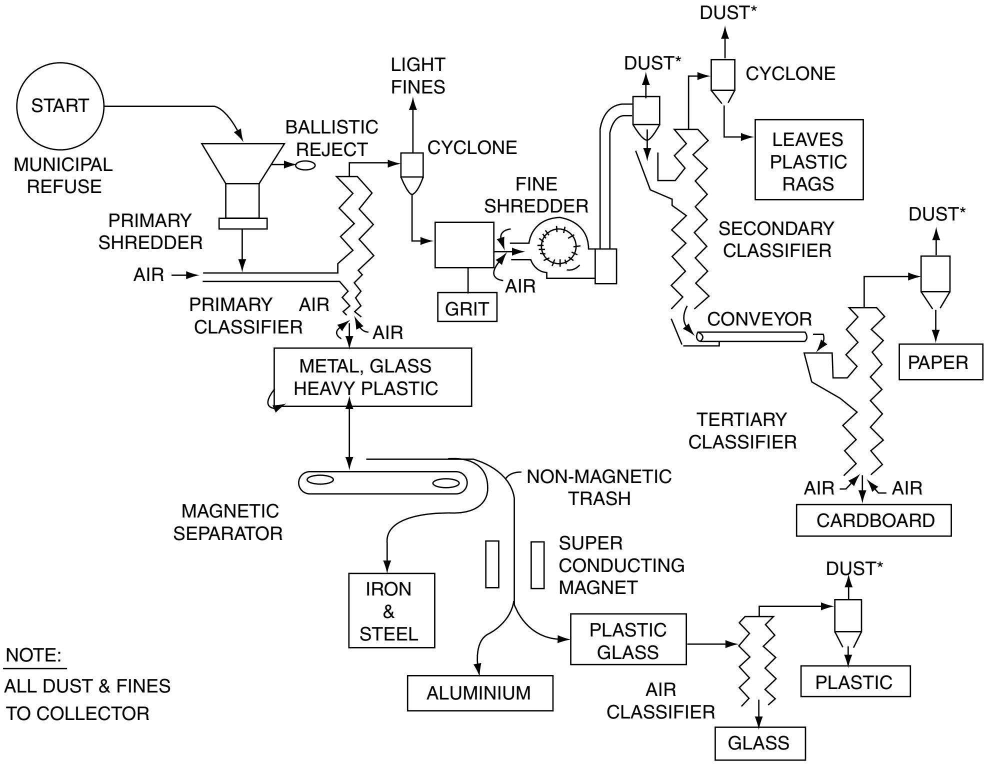


FIGURE 4 Proposed waste separation system, using SRI Air Classifier.

molecules. The process is carried out in the absence of added oxygen (or with very limited oxygen), and with the addition of heat, at temperatures in the range of 900–1700°F. The process will produce a fuel gas, oil and carbon. A study in San Diego indicated that about 50% of the refuse is susceptible to pyrolysis.[34] This study yielded low BTU gas, char and oxygenated hydrocarbons. Cities Services has evaluated a similar process and believed it to be potentially economic in plants with capacities of 5000 tons per day which are subsidized at the rate of \$2 to \$8/ton.[35] Pilot plant studies by the Bureau of Mines indicate that both industrial and municipal refuse yield large amounts of gas and solid, as shown in Table 23.

A process using pyrolysis has been piloted for tires, which present some particularly difficult disposal problems in incineration and landfill.[37] Similar processes can be effective in recovering chemicals from plastics but have not been developed because of separation problems.[51]

The US Bureau of Mines has also piloted a high pressure process where refuse free of glass and metal is reacted in the presence of water and carbon monoxide (hydro-oxynation) at 1400 psig and 500 to 700°F to yield oil, gas and carbon.[37] This process appears to have very favorable oil yields. Process variable studies have shown that conversions as high as 90% can be obtained with a 40% yield of oil; typical yield is shown in Table 24. These yields indicate a potential oil production rate of 2×10^8 tons annually as compared to US crude production of 5×10^8 tons per year. Similar studies using hydrogen showed lower yields and conversions. The use of CO does present some significant operating problems as well as economic debits.

Several systems using pyrolysis were ready for commercialization. Hercules in Delaware planned a unit to pyrolyze industrial waste. Monsanto Environchem built a pyrolysis unit using the LANDGARD process (Figure 7). This process emphasized waste reduction (with recovery of ferrous metal), rather than recovery of variable byproducts; it has been piloted in a 35 T/day semiworks facility. The process reduces the solid waste, typically, by 90%; a typical stack gas analysis is given in Table 25 and indicates the very low particulate matter in the effluent. Unfortunately the unit did not operate successfully at full scale because particulate removal did not meet expectations and costs became prohibitive.

In contrast to the Landgard system, a pyrolysis process emphasizing recovery of valuable products was developed by Garrett Research and Development Co. This process was piloted at a 4 T/day level. It consisted of shredding, air classification, pyrolysis and pyrolysis product separation steps. (A full scale unit was built in San Diego, but never operated a full capacity because of mechanical problems). Product recovery was similar to that obtained by the Bureau of Mines.

One may ask why solid refuse should be subjected to complex processes such as hydropulping or pyrolysis. Where actual wastes such as paper or cellulose fiber can be recovered, hydropulping is certainly attractive; on the other

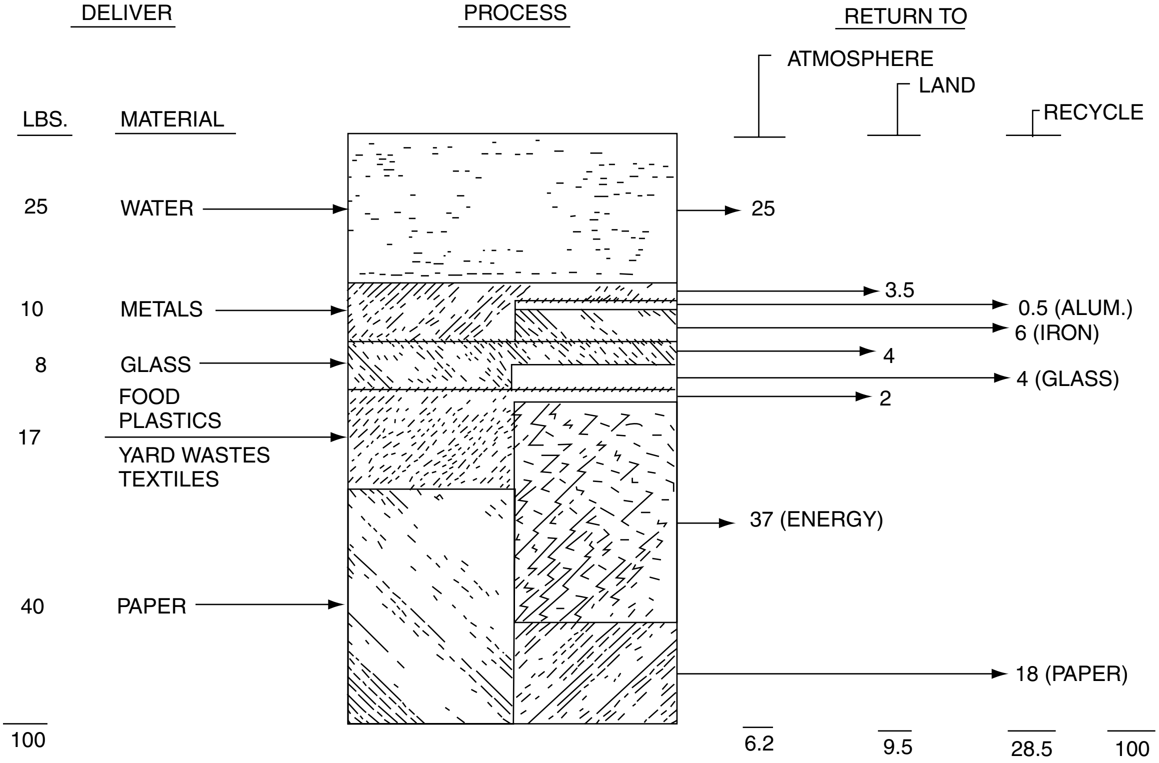


FIGURE 5 Black Clawson Hydraposal-Fiberclaim. Material balance. Credit: Black Clawson Co.

hand where waste is utilized as fuel the possibility of using it directly rather than converting it to fuel oil (Bureau of Mines process) must be considered. Pyrolysis, hydro-oxydation or similar processes should be competitive with direct fire waste-heat recovery incineration where clean fuel is desired or fuel is to be used at locations other than where the refuse is available. The recovery of chemical raw materials from refuse using these processes may also become attractive as the process technology is refined.

The recovery of valuable products from incinerator effluent is not, however, being neglected. The Bureau of Mines has demonstrated technology to recover metals and glass from incinerator residues using high-intensity magnetic sorting as well as chemical techniques at a cost of about $4 per ton.

ITT Research Institute has shown that refuse may be selectively separated into a ceramic and metal slag when incinerated at temperatures between 2800 and 3200°F. While technically feasible, this separation incurs some severe economic costs; it does yield some interesting products including tiles, pipe and structural shapes. In addition to the newer methods being studied, the use of incinerator fly ash and even slag for road fill and concrete aggregate appears to be attractive in some situations.

Specific process to convert refuse fractions are also receiving attention and will be utilized in limited situations where the economics are attractive. An engineering study has shown that the organic portion of raw refuse can be economically converted to sugar by acid hydrolysis. This can then be a raw material for alcohol production. One of the limitations to this and similar processes is a limited market for the product, and competition with other sugar sources as well as other sources of alcohol; for example conversion of 3% of the refuse to ethanol would saturate the normal market.

Though fuel use of alcohols is possible, only limited amounts of refuse can be converted using these processes unless major shifts occur in our economy. Process development to convert citrus waste to citric acid is being conducted at the University of Florida. Other conversion studies are being carried out, but all face the difficulties inherent in processing a heterogeneous, complex, often variable, mixture.

Of more than passing interest are studies being carried out of the University of Maryland to obtain a protein concentrate for human and animal consumption from food processing wastes. Similar studies at Louisiana State University, in the pilot plant stage, has shown that agricultural cellulose wastes can be broken down by selected microorganisms to yield a low cost, high protein food. Yeast can also be produced from cellulose wastes.

The use of waste as a "resource material" is still in its infancy. It is gathering momentum quickly and in the next decade should see significant changes in waste management.

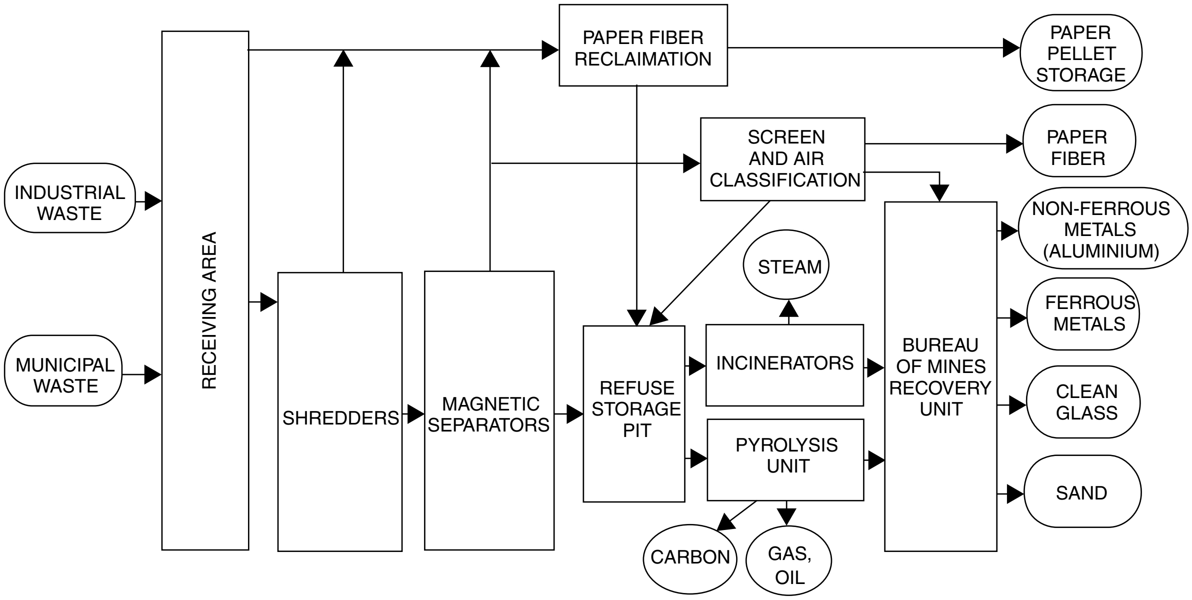


FIGURE 6 Refuse recycling plant. (Aluminum Association of America).

SOURCE REDUCTION

By way of background, it is important to identify the universe of municipal solid waste, or the “source” from which the volume of waste may be reduced. The total volume waste generated in the United States in 1986 was approximately 160 million tons.[40]

Source reduction, while related to recycling in some respects, is nonetheless unique as a waste management option which occurs “before the fact”. It has been said that “an ounce of prevention is worth a pound of cure”. In the area of solid waste management, a reduction of a single ounce in the current per capital generation rate of 3.5 pounds per person per day represents a decrease in annual volumes produced of almost 3 million tons! The importance of source reduction becomes at once apparent.

As noted above, source reduction involves a decrease in waste volume or toxicity. It is appropriate to offer some additional analysis of the nature of these two alternatives and the specific areas which provide the greatest opportunity for change and thus measurable progress in source reduction.

TABLE 23
Pyrolysis of refuse typical yields[36]

	Source	
	Raw material	Heil milled industrial
Feed, Million BTu		
Available per ton dry basis	17.09	11.29
Pyrolysis Temp.	900°C	900°C
Yield, Wt. % of Refuse		
Residue	7.7	38.8
Gas	39.5	29.4
Tar	0.2	0.2
Light oil	—	0.6
Free NH_3	0.3	0.04
Sour liquor	47.8	21.8
Yield Per Ton of Refuse		
Gas, cu ft	17,741	12,318
$(NH_3)SO_4$ lbs	25.1	21.7
Heating Value		
Gas btu/ft^3	447	498
Residue, btu/lb	5,260	2,180

Perhaps the greatest concern among solid waste management officials is the great number of single use or disposable products. Disposables are ubiquitous, from beverage containers to disposable diapers to food service containers and utensils. Furthermore, manufacturing technology has advanced to the point where even products which traditionally were never considered single use items may now be purchased at prices which make their use and replacement competitive with reliance on a far less frequent purchase of their durable antecedents. The availability of disposable razors and blades and such commonplace items as pens is not surprising. However, widespread marketing of disposable flashlights, electronic watches, and even cameras must be viewed as unexpected by all but the most optimistic technologists. Absent product bans which are unlikely at the federal level

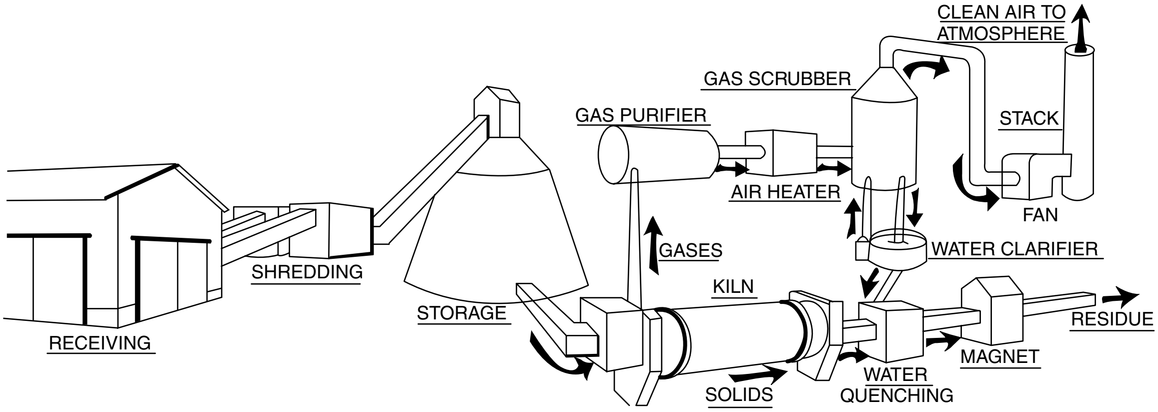


FIGURE 7 Monsanto-Landgard waste pyrolysis process. Credit: Monsanto Enviro-Chem Systems Inc.

TABLE 24[38]
Reduction of garbage by hydro-oxynation 52 1 BBL of oil ton of garbage (50% cellulose)

Conditions	
1500 psig initial pressure. 5000 psig operation pressure	
350°C.	
Addition of CO + H_2O	
Yield, wt%	
Oil	40%
Residue	10%
H_2O	27–36%
CO	15–20%
Other	2%
Oil Analysis C 83%, H_2 7.8%, O_2 7.8%, m N_2 1.9%, S 0.13%	

TABLE 25[39]
Langard pyrolysis system typical stack gas analysis

Component	Average value
N_2	40%
O_2	3%
CO_2	7%
H_2O vapor	50%
Combustibles	None
NO_4	50 ppm (vol.)
SO_2	100 ppm (vol.)
Chlorides	10 ppm (vol.)
Particulates	0.06 grains per SCF dry gas corrected to 12% CO_2

except in the case of palpable threat to human health and/or the environment, avoidance of these types of items is largely a mater of consumer choice. Nevertheless, some local governments have begun restricting the use of certain types of plastic, nondegradable or other materials.

Beverage container deposit legislation, mentioned above, is a good example of an issue which straddles the areas of source reduction and recycling. If a mandate to use only *refillable* containers is utilized, certainly waste generation will be decreased. This is clearly source reduction. However, the mandate that deposits be utilized to encourage the return of used containers is a recycling initiative, since the returned bottles and cans will be used as secondary materials rather than being refilled on a unit basis.

Disposable products may also be manufactured using less raw materials as another alternative to achieve measurable source reduction gains. Obviously the limiting factor here is the production of a product which contains less raw material but still serves its intended use.

Extending the useful life of non-disposable products is another legitimate source reduction alternative. The use of alkaline rather than lead acid batteries is a good example. The use of rechargeable batteries rather than disposable batteries is an even better example. As another example, consider the use of automobile tires which have a longer useful life. Tires are currently produced (and disposed) at a rate of 220 million unit per year. Any tangible extension of useful life for individual tires would decrease this generation rate and avoid any increase in the current estimated 2 to 3 billion discarded tires currently stockpiled.

The reduction in the toxicity of wastes is another important source reduction measure. Among the materials of greatest concern are lead and cadmium. The presence of these and other heavy metals in incinerator ash often results in the classification of such ash as a hazardous waste. Hence the desire to identify the likely sources of lead and cadmium and ultimately, to find ways to eliminate or replace these elements in the products which contain them.

The most current data available indicate a total of 213,652 tons of lead and 1,788 tons of cadmium in the municipal solid waste generated in 1986. The figure for lead does not include the lead present in the lead acid batteries which were recycled which amounts to some 80% of the batteries produced. The primary sources of lead entering municipal waste stream (again not including that contained in the batteries which are recycled) is as follows:

- 65% lead acid batteries (primarily car batteries)
- 27% consumer electronics
- 4% glass and ceramics
- 2% plastics

The primary sources of cadmium in the municipal waste stream (after recycling) are as follows:

- 54% household batteries
- 28% plastics
- 9% consumer electronics
- 5% appliances
- 4% pigments

To the extent that the lead and cadmium in the products listed above can be eliminated or replaced, substantial source reduction may be achieved.[41]

Recycling is the next preferred alternative in the hierarchy of integrated waste management options. As noted above, it generally involves the reuse of secondary materials as a supplement to or wholly in place of virgin materials in the production of various goods. Whether or not the material is used to produce the same goods which were recycled depends on a number of technical, economic, and aesthetic considerations. For example, plastics such as polyethylene tetraphthalate (or "PET") are currently being recycled from large relatively uncontaminated one and two liter beverage containers. They are not, however, used to produce new plastic containers which will be used in contact with food. This limitation is not a constraint where other types of beverage containers made from aluminum or glass are concerned. Aside from composting, which is addressed below, recycling generally takes place as an in-plant practice or after certain consumer products are used.

Recycling of process waste from a variety of industrial and/or manufacturing operations has been practiced for some time on a discretionary basis. The decision to use or dispose of the byproduct of a given process often involves technical considerations such as chemical or physical differences between the raw materials otherwise used and the scrap material available. Assuming that no technical impediments exist, the decision about whether to recycle is one of economics. There is no question that the higher disposal costs currently being experienced have driven more industries to recycle as a business decision. However, the types and amounts of materials and the number of industries potentially involved are beyond the scope of this document.

The recovery of materials from municipal waste streams or so called "post consumer" recycling has experienced substantial gains over the past several years due to diminishing disposal capacity and dramatic increases in tipping fees. At least 18 states have recycling goals established in legislation including the four mentioned above where recycling is mandatory. The materials most commonly recycled include newspapers, glass, aluminum cans and to a lesser extent, corrugated cardboard, various ferrous and non-ferrous scrap metals, and plastics. These materials are typically accumulated by private citizens and businesses and either collected curbside separately from the remainder of solid waste destined for disposal or brought to centralized collection facilities. As a matter of convenience and to maximize citizen participation, many systems collect commingled glass and cans or commingled glass, cans, and newspapers in a single container.

The value of the materials described has varied widely, depending on the quantity of material available, expected levels of contamination, and transportation distance to end markets. Furthermore, the rush to recycle by more and more communities has resulted in erratic markets for certain materials. For example, newspaper, which would demand approximately $60 per ton as recently as 1987, is now worth $20 per ton in some cases and is taken for no compensation in other cases. Projections by some waste newspaper exporters indicate a probable net cost of $25 per ton by 1990. While markets for other materials have been less volatile than this, considerable variations have occurred. As of this writing, typical values per ton for other recycled commodities are as follows:

- aluminum $800
- Plastic PET bottles $120
- glass $40
- steel cans $10

Composting of vegetative waste is also a form of recycling. It has become more widespread as landfill disposal costs have risen. Leaves and other yard waste are amenable to this process, which has proven more troublesome for grass clippings due to the anaerobic odors often associated with grass not mixed adequately to limit the development of such conditions.

Incineration of municipal solid waste has become more widespread with new plants almost always incorporating steam generation and electrical power production. Waterwall, or mass-burn incineration has become the technology of choice among those communities whose overall management strategies include large scale volume reduction processes after source reduction and recycling operations have been utilized to maximum advantage and prior to sanitary landfilling of residuals or noncombustible materials. Currently, 126 such facilities operate in 37 states. The average design capacity of these facilities is 814 tons per day. Total design capacity for all these facilities is 68,399 tons per day. This latter figure is projected to reach 107,832 tons per day by 1992 when all facilities currently under construction are completed and operational. The largest incinerators currently operating are a 3,300 tons per day plant in Michigan and a 3,000 tons per day plant in Florida. There are also 124 modular plants currently in operation. These modular plants are similar to mass burn plants but generally smaller and sold as prefabricated units. The average modular plant burns 124 tons per day.[42]

While the process description of incineration remains as described in the full text, significant design improvements and air pollution control methodologies have been applied to new facilities. The designs are all based on facilities established in Europe and are marketed by a number of United States companies. American ReFuel (Browning Ferris), Blount, Ogden Martin, and Waste Management currently hold the largest market share in terms of facilities operational and/or under construction.

Air pollution controls utilized include electrostatic precipitators or baghouses (sometimes both) as well as acid gas scrubbers where removal of HCl emissions is required. The newest plants will also incorporate thermal treatment to limit nitrous oxide emissions. The specific air pollution control equipment employed at individual facilities depends upon the requirements of the regulatory/permitting agencies (generally states with federal oversight) and the ambient air classification of the region in which the plant operates/will operate.

Nine plants continue to produce refuse derived fuel ("RDF") which supplements fossil fuel or is fired alone in dedicated boilers. The average RDF plant processes 953 tons per day of incoming solid waste.

Virtually all of the technology demonstration projects originally funded by the federal government have ceased to operate due to technical and or economic reasons. The only exception is the Delaware Solid Waste Authority facility in Wilmington, Delaware. This plant, originally designed and built by Raytheon, continues to operate as part of an integrated waste management system operated by the Authority. The writer is unaware of any proposed new installations of this particular technology.

A number of other *alternative technologies* have been offered to communities by private entrepreneurs. They are generally materials separation processes and a few more highly technical approaches such as laser destruction of raw waste and/or incinerator ash. These processes claim some success at laboratory or bench scale demonstrations for mixed municipal waste with some larger applications handling specific homogeneous waste streams. To the extent that larger operations (on the order of at least several hundred tons per day throughput) are built and evaluated over a number of years on mixed municipal waste streams, their viability may be determined.

Environmental controls required at *sanitary landfills* have become substantially more stringent over the past several years as states have revised regulations due at least in part to serious ground and surface water pollution problems arising from older sites without such controls. As a result of these stricter regulations many environmentally deficient sites were forced to close. This has resulted in a disposal capacity shortfall in many areas, particularly in the urbanized areas of Northeastern states.[43] Those sites which remain as well as the limited number of new sites being built must incorporate a variety of specialized controls which a few years ago were not even required for hazardous waste facilities. While it is impractical to list the many variations in individual state regulations, an overview of the proposed revised federal criteria for land disposal mentioned above will serve to provide a good indication of the minimum standards which will apply nationally if the final regulation is adopted as proposed. Obviously, there is no assurance that this will be the case but the proposal certainly reflects the federal government's best analysis of the degree of control necessary. As such it is worthy of some brief analysis. The discussion below highlights only the technical aspects of the proposed criteria and not the administrative concern such as facility registration and similar issues.

Location Restrictions: Areas of Special Concern

- landfills within 10,000 feet of an airport would be required to operate in a manner that precludes birds attracted by solid waste from creating a hazard to aviation
- landfills located in 100-year floodplain would be prohibited from restricting the flow of the 100-year flood, reducing the temporary water storage capacity of the floodplain, or resulting in the washout of solid waste so as to pose a threat to human health and the environment
- new landfills may not be sited in wetlands absent a demonstration that there is no practical alternative, no significant adverse environmental impacts, and that relevant discharge standards will be met.
- new landfills may not be sited within 200 feet of faults which have had displacement during Holocene time (i.e., within 11,000 years)
- new landfills in seismic impact areas would be required to be designed to resist ground motion from earthquakes
- landfills in unstable areas such as Karst terrain would be required to incorporate engineering safety design measures

Operating Criteria: Minimum Requirements

- procedures for excluding the receipt of hazardous waste
- application of daily cover material
- control of disease vectors
- monitoring and control of explosive gases
- prohibition of open burning
- limitation of site access
- control of storm water run-on and run-off
- limitation of surface water discharges
- prohibition of bulk liquids
- record keeping

In addition to the above requirements, the proposed criteria call for site closure and post-closure care criteria including a minimum of 30 year maintenance and monitoring, establishment of financial security to ensure that these activities are carried out. Finally there is a requirement that corrective actions be taken in the event of identification of groundwater contamination.[44,45]

One area of note is increasing interest in landfill mining as a source of combustible fuel, cover material for current landfills, and the creation of new fill capacity in the airspace vacated by the mined sections of a site. One such project was conducted in Naples (Collier County), Florida. In addition, several sites in New York State are being evaluated for possible research into this technology.

Some Special Problems

Radioactive solid wastes create special problems and are discussed elsewhere in a section on Radioactive Wastes.

Industrial wastes, as mentioned previously, has been dumped as a general rule. Because of the high specialized nature of industrial wastes, it is impossible to discuss them in a general way. Total recycling of many industrial wastes will become a more frequent practice. Mine tailings will find their way into construction material, fill, or may be recycled into the mines. Slag from steel mills should become less of a problem as different processes are used but will still remain a significant contributor; slag can be used in special concrete and efforts in this area will continue.

Sewage sludge (see: SEWAGE) presents some special problems. To date it has generally been dumped. Composting (see: COMPOSTING) should be a major process for handling sewage sludges. Several attempts to sell composted sludge have only been marginal because of the lack of markets; however this should change in the future. Material that, for economic reasons, cannot be composted, can be burned to recover waste heat in specially designed incinerators.

More of a problem will be spent solids from water treatment facilities. These solids are high in carbonates and often have a foul odor due to entrained organic material. They are now being filled and this practice will have to continue unless chemical recovery methods which produce a useful product are found; this is not likely as carbonates are in oversupply.

The management of solid wastes will undergo dramatic changes in the next decade. From a "cottage industry" it will emerge as a major process industry recovering many vital materials and converting others into valuable products. Landfill and incineration will continue to play a role with the former decreasing in importance and the latter coming into greater prominence. However new process technology, only some of which is now in development, will play an ever more important role in total solid waste resource management.

BIBLIOGRAPHY

General

1. Amer. Public Works Ass'n, Inst. for Solids Wastes, *Municipal Refuse Disposal,* Public Admin, Service, Chicago, 1970.
2. Eliassen, R., *Solid Waste Management,* Off. of Science and Tech., Exec. Off of the Pres., Washington, 1969.
3. Frey, D.N. (Chairman), *Policies for Solid Waste Management,* U.S. Dept. NEW, PHS Pub. 2018, 1970.
4. Train, R.E. (Chairman), *Environmental Quality—1st Ann. Report of the Council on Environmental Quality,* Washington, Chapter VI, 1970.
5. Hanks, T.G., *Solid Wastes/Disease Relationships,* U.S. Dept. HEW, PHS Pub. No. 999-UIH-6, Cincinnati, 1967.
6. Cooke, L.M. (Chairman), *Cleaning Our Environment, The Chemical Basis for Action,* ACS, Washington, 1969.
7. Pavoni, J.L. *et al., Handbook of Solid Waste Disposal,* Van Nostrand, New York, 1975.
8. Tchobanoglous, G, *et al., Solid Wastes—Eng. Prin. and Manag. Issues,* McGraw-Hill, New York, 1977.
9. Mantell, C.L., *Solid Waste,* Wiley-Interscience, New York, 1975.

Sources

1. Golueke, C.G. and P.H. McGauhey, *Comprehensive Studies of Solid Waste Management,* U.S. Dept. HEW, PHS Report No. 2039, Washington, 1970.
2. Amer. Public Works Ass'n, Inst. for Solid Wastes, op. Cit.
3. Copp, W.R. *et al., Technical-Economic Study of Solid Waste Disposal Needs and Practices,* **1**, *Municipal Inventory,* **2**, Industrial Inventory, U.S. Dept. HEW, PHS Pub. No. 1886, Washington, 1969.
4. Darnay, A. J. Jr. and W.E. Franklin, *The Role of Packaging in Solid Waste Management,* U.S. Dept. HEW, PHS Pub. No. 1885, Washington, 1969.
5. Black, R.J. *et al., The National Solid Wastes Survey, An Interim Report,* U.S. Dept. HEW, Washington, 1968.

Disposal methods, general

1. Amer. Pub. Works Ass'n, Inst. for Solid Wastes, op. cit.
2. Eliassen, R., op. cit.
3. Golueke, C.G. and P.H. McGauhey, op. cit.
4. Siebel, J.E. *et al., Technical-Economic Study of Solid Waste Disposal Needs and Practices,* **6**, Technical-Economic Overview, U.S. Dept. HEW, PHS Pub. No. 1886, Washington, 1969.
5. Jensen, M.E., *Observations of Cont. European Solid Waste Management Practices,* U.S. Dept. HEW, PHS Pub. No. 1880 Washington, 1969.

Disposal methods, sanitary landfill

1. Sorg, T.J. and H.L. Hickman, Jr., *Sanitary Landfill Facts,* U.S. Dept. Hew, HPS Pub. 1792, 2nd Ed., Washington, 1970.
2. Steiner, R.L. and R. Kantz, *Sanitary Landfill; a Bibliography,* U.S. Dept. HEW, PHS Pub. N. 1819, Washington, 1968.
3. Lambia, J.A. (Proj. Dir.), *Development of Construction and Use Criteria for Sanitary Landfills,* U.S. Dept. HEW, PHS Grant Do. 1-UI-00046, Cincinnati, 1969.
4. Brunner, D.R. and D.J. Keller, *Sanitary Landfill Design and Operation,* Rep. No. SW-65ts, USEPA,1972.
5. *Classifying Solid Waste Disposal Facilities,* SW-828, USEPA, 1980.
6. *Toxicity of Leachates,* EPA-600/2-80-057, USEPA, 1980.

Disposal methods, incineration

1. Corey, T.C. (ed.), *Principles and Practices of Incineration,* Wiley, New York, 1970.
2. DeMarco, J. *et al., Incinerator Guidelines,* U.S. Dept. HEW, PHS Pub. 2012, Washington, 1969.
3. Day and Zimmerman, *Special Studies for Incinerators—for the Government District of Columbia,* U.S. Dept. HEW, PHS Pub. 1748, Cincinnati, 1968.
4. Combustion Power Co., *Combustion Power Unit-400,* U.S. Dept. HEW, PHS, Rockville, Md, 1969.
5. Achinger, W.C. and L.E. Daniels, *An Evaluation of Seven Incinerators,* SW-51ts, USEPA 1970.
6. Ruble, F. N., *Incineration of Solid Waste,* Noyes Pub, New Jersey, 1975.
7. Sittig, M., *Incineration of Industrial Waste,* Noyes Pub, New Jersey, 1980.
8. Domalski, E.S. *et al., Thermodynamic Data from Waste Incineration,* ASME, Nat'l But. of Stds. Report NBSIR 78-1479, 1978.

Reutilization, Recycle and Resource Recovery

1. Drobny, N.L., H.E. Hull, and R.F. Testin, *Recovery and Utilization on Municipal Waste,* SW-10c, USEPA, 1971.

2. Darnay, A. and W.E. Franklin, *Salvage Markets for Materials in Solid Wastes,* SW-29c, USEPA, 1972.
3. Barton, A.J., *Resource Recovery and Recycling,* Wiley Interscience, New York, 1979.
4. Cherimissinoff, P. and A. Morresi, *Energy from Solid Wastes,* Deker, New York, 1976.
5. Vesillind, P. A. and A. E. Rimer, *Unit Operations in Resource Recovery Engineering,* Prentice Hall, Engelwood Cliffs, N.J., 07632.

Disposal methods, composting

1. Hart, S.A., *Solid Waste Management/Composting,* U.S. Dept. HEW, PHS Pub. No. 1826, 1968.
2. Gainesville Municipal Conversion Authority, Gainesville Compost Plant, U.S. Dept. HEW. Cincinnati, 1969.
3. Shell, G.L. and J.L. Boyd, *Composting Dewatered Sewage Sludge,* U.S. Dept. HEW, PHS, Washington, 1969.
4. Gotass, H.B., *Composting,* Mon. Ser. No. 31, World Health Org., 1956.

Bibliography

1. Golueke, C.G., *Solid Waste Management,* Abstracts and excerpts from the literature, **1** and **2,** U.S. Dept. HEW, PHS Pub. 2038, Washington, 1970.
2. Black, R.J. *et al., Refuse Collection and Disposal 1954–61,* U.S. Dept. Hew, PHS Pub. No. 91, Sup. B, C, D, E and F, Washington D.C., 1966.
3. Lederer, F. P., *Solid Waste Management,* Available information materials-interim catalog, SW-58.27, USEPA, 1977.

REFERENCES

1. United States Environmental Protection Agency, *The Solid Waste Dilemma: An Agenda for Action.*
2. PL 98-616, The Hazardous and Solid Waste Amendments of 1984, USC 6901, November 8, 1984.
3. PL 94-580, Resource, Conservation and Recovery Act of 1976, USC 6901, October 21, 1976. Section 1004.
4. 40 CFR 261.
5. Op. cit. PL 94-580.
6. Eliassen, R., *Solid Waste Management,* Off. of Science and Tech. Exec. Off. of the Pres., Washington (1969).
7. Amer. Pub. Ass'n Inst. for Solid Wastes, *Municipal Refuse Disposal,* Public Administration Service, Chicago, p. 11–20, 1970.
8. *Ibid.*, p. 52.
9. Amer. Pub. Works Ass'n, Comm. on Solid Wastes, op. cit. p. 47.
10. Carruth, D.E. and A.J. Klee, *Analysis of Solid Waste Composition—Sample Size,* U.S. Dept. HEW, PHS Report SW-19ts, Washington, 1969.
11. Kaiser, E. R., Private Communication.
12. Eliassen, R., op. cit. p. 32–33.
13. *Ibid.*, p. 221.
14. *Ibid.*, p. 59.
15. 40 CFR 264.
16. Amer, Pub. Works Ass'n, Inst. for Solid Wastes, op. cit., p. 94–133.
17. Reinhardt, J.J. and G. Rohlich, *City of Madison Pilot Plant Demonstration Project—An Interim Report,* U.S. Dept. HEW, PHS, Cincinnati, 1968.
18. Corey, R.C. (ed.), *Principles and Practices of Incineration,* Wiley New York, p. 9–33, 1969.
19. Hangebrauk, R.P., *et al., J. APCA,* **14**, (No. 7), p. 267, 1964.
20. Hovey, H.H. *et al.*, *J. APCA,* **16**, (No. 7), p. 362, 1966.
21. Lenehan, J.W., *J. APCA,* **12**, No. 9, 414, 1962.
22. Stenburg, R.L. *et al.*, *J. APCA,* **8,** No. 2, 297, 1959.
23. Kenahan, C.B. *et al.*, *Comp. and Characteristics of Municipal Incinerator Residues,* U.S. Bureau of Mines, R.I. 7204, Washington, 1968.
24. Kenahan, C.B. *et al.,* Municipal Incineration, M.H. Detrich Co., Chicago, 1967.
25. Overall, J.E., *Amer. Paper Industry,* **48**, p. 51, 1966.
26. Meile, R.L., *Proc. 1966 Nat'l Incinerator Conf.,* ASME, p. 202, 1966.
27. Amer. Pub. Works Ass'n, Inst. for Solid Wastes, op. cit. p. 231.
28. Rogers, C.A., Public Works, (June 1969), p. 85, 1969.
29. Caver, P.T., *Proc. 3rd ANERAC Conf.*, Univ. of R.I., Kingston, R.I., 1970.
30. Hart, S.A., "Solid Waste Management-Composting," U.S. Dept. HEW, PHS Pub. No. 1826, Cincinnati, 1968.
31. Kupchick, G.J., *Bull World Health Organ.*, **34**, No. 5, p. 798, 1966.
32. Boettcher, R.A., *Proc. 2nd Mineral Waste Utilization Symp.,* IITRI, Chicago, p. 271, 1970.
33. Abrahams, J. H. Jr., *Proc. 2nd Mineral Waste Utilization Symp.,* IITRI, Chicago, p. 363, 1970.
34. Hoffman, D.A., ASME Ann. Mtg., 1969, ASME Pub. No. 69-WA/PID-10, 1969.
35. Hoffman, D.A., *C. and E. News,* **48**, No. 10, p. 38, 1970.
36. Sanner, W.S. *et al.*, *Conversion of Municipal and Industrial Refuse by Pyrolysis,* U.S. Bureau of Mines, R. I. 7428, Washington, 1970.
37. Woldson, D.E. *et al., Destructive Distillation of Scrap Tires,* U.S. Bureau of Mines, R. I. 7302, Washington (1969).
38. Appell, H.R. *et al.*, *Proc. 3rd ANERAC,* Univ. of R. I., Kingston, R.I., p. 225, 1970.
39. Buss, T. F., Private Communication, Monsanto Enviro-Chem. Systems Inc.
40. Franklin Associates, Ltd., *Characterization of Municipal Solid Waste in the United States, 1960 to 2002 (Update 1988),* prepared for the United States Environmental Protection Agency, Washington, D.C., March, 1988.
41. Franklin Associates, Ltd., *Characterization of Products Containing Lead and Cadmium in Municipal Solid Waste in the United States, 1970 to 2000,* prepared for the United States Environmental Protection Agency, Washington, D.C., January, 1989.
42. Institute of Resource Recovery of the National Solid Wastes Management Association, *Resource Recovery Focus,* Vol. 1, No. 1, No. 2, Washington, D.C., Summer, 1989.
43. United States Environmental Protection Agency, *Report to Congress: Solid Waste Disposal in the United States,* Office of Solid Waste, Washington, D.C., October, 1988.
44. National Archives and Records Administration, Office of the Federal Register, *United States Code of Federal Regulations,* Volume 40, Part 257, United States Government Printing Office, Washington, D.C., 1988.
45. United States Environmental Protection Agency, "Solid Waste Disposal Facility Criteria; Proposed Rule", 53*FR* 3314, August 30, 1988.

PETER B. LEDERMAN
Peter B. Lederman Associates

MICHAEL. F. DEBONIS
FEMA

MARINE SPILLAGE—SOURCES AND HAZARDS

INTRODUCTION

Scope

Pollution of navigable waterways resulting from operation of commercial and naval vessels may be a consequence of normal service or from casualties such as collisions and groundings. Propulsion system fuel oil and liquid cargoes may be involved in any case and will be considered in this chapter. Waste disposal from shipping is of comparatively small magnitude compared to waterway pollution from shoreside sources and will not be considered. Emphasis will be on description of the pollution problem and on means for prevention. The subject of spill collection and disposal is considered elsewhere in this text.

Problem Definition

Normal operations A decade ago, the primary source of pollution of the world's waterways was the intentional discharge of oily ballast water during routine operations. At that time, for virtually all seagoing operations ballast water was taken aboard for a portion of the voyage in order to obtain sufficient draft and trim for propeller immersion, adequate steering ability, acceptable conditions of seaworthiness; and to satisfy mandated operational and regulatory requirements for intact and damaged stability. If assigned ballast tank capacity was inadequate to meet these requirements, it was the general practice to ballast empty fuel oil tanks, or empty liquid cargo tanks in the case of tank vessels. This procedure resulted in the necessity for pumping overboard large quantities of contaminated ballast water before taking on fuel oil or liquid cargoes. These procedures have been largely outlawed by international agreements developed by the International Maritime Organization (IMO) of the United Nations, and enforced by the national regulatory agencies of the member countries.

Casualties Spills resulting from casualties generally receive more attention in the world press than incidents involving operational discharges. Spills may occur from operational mishaps in the pumping of fuel oil bunkers and liquid cargoes. Incidents of spills that occur from collisions and groundings are accompanied by associated dangers to personnel and the environment and are likely to involve the largest quantities of pollutant discharged in a single incident. The magnitude of such spills is clearly far greater in the case of a loaded tank vessel than the grounding and rupture of the double bottom fuel tanks of a dry cargo vessel.

SHIP CLASSIFICATION AND DESCRIPTION

The arrangements and general characteristics of the various merchant ship types are well described in such standard texts as Reference (1) and in the technical literature, including the comprehensive discussion in Reference (2) covering U.S. shipbuilding during the 1936−1976 period. Accordingly, the following discussion will be restricted to characteristics pertinent to the pollution problem, for example, arrangement of tank spaces.

The following standard abbreviations have been used throughout for convenience:

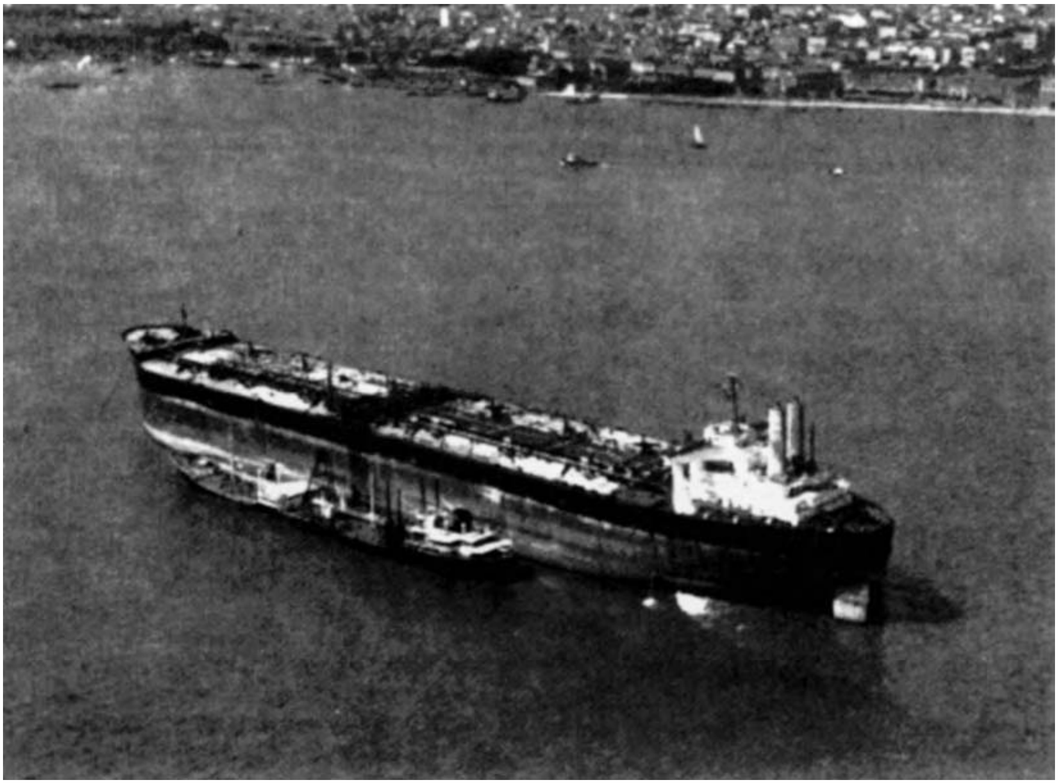
DWT = deadweight = total displacement − light weight
= cargo + consumables
mld. = molded
fbd. = freeboard
B.P. = between perpendiculars

Break Bulk Vessels

The greatest variety of seagoing vessels are in this category which includes the ordinary general cargo vessels carrying a great variety of dry products in raw material as well as finished and packaged form. An outline sketch of the cross-section through a typical cargo hold, showing hatchway, tween decks and fuel oil or ballast spaces, is shown in Figure 1. Fuel oil is commonly carried in the double bottoms, as indicated, but may also be carried in deep tanks, particularly outboard of shafting and in the vicinity of the machinery spaces. Except for settling and daily service tanks, all bunker spaces are normally piped for fuel oil or ballast. The availability of cubic capacity for tankage assigned only to ballast service is limited in such vessels and frequent use of fuel tanks for ballasting is likely in most operations.

Unitized Cargo Carriers

Ships in this category are usually designed for the exclusive transport of standard containers or wheeled trailer vans, and, to a lesser extent as hybrid carriers to handle container, wheeled vehicles and general break-bulk cargo. In the case of container ships, as illustrated in the typical hold section, Figure 2. the cellular nature of the cargo stowage requires some "squaring off" of the hold spaces, with the result that considerable wing space is available for ballast tanks. As a



16,000 DWT PARIA LIMPA, built in the 1940's alongside the 326,000 DWT UNIVERSE PORTUGAL
Courtesy, *The Motor Ship*.

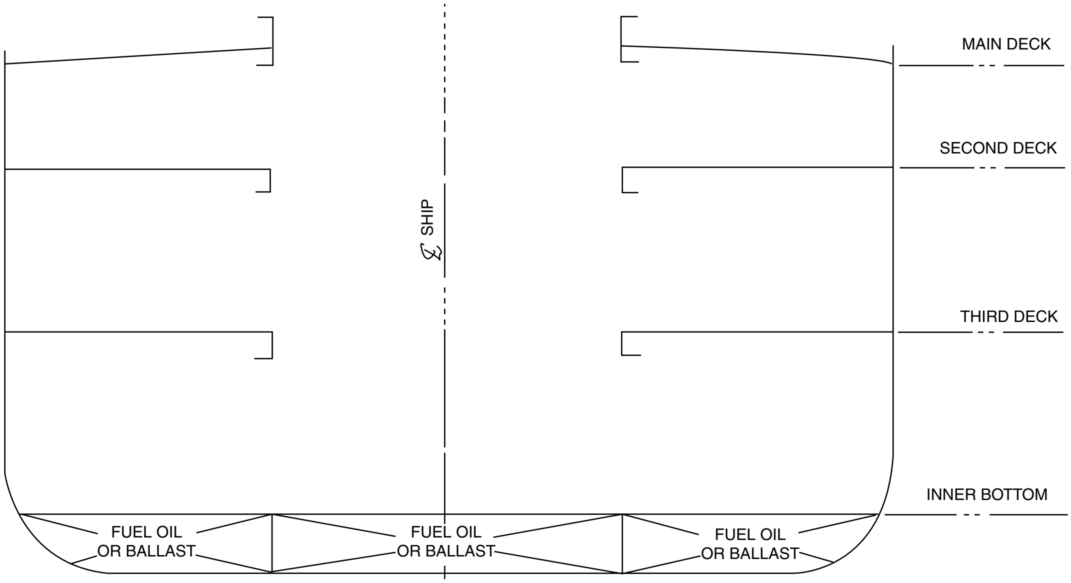


FIGURE 1 Outline midship section through cargo hold, typical break bulk dry cargo vessel.

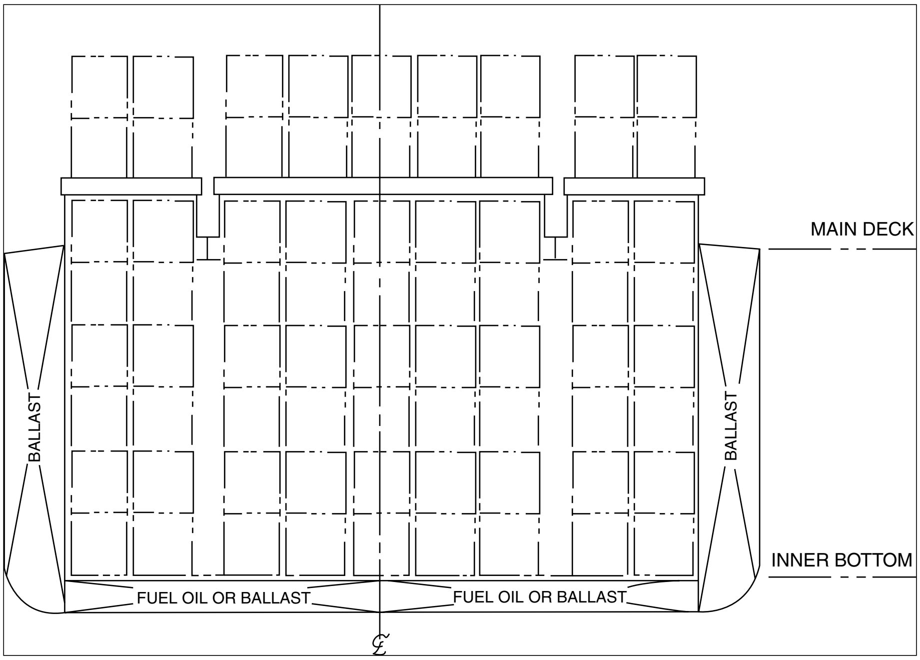


FIGURE 2 Outline midship section through container hold, typical container ship.

result such vessels are able to operate with clean ballast only and no ballasting of fuel tankage is normally required.

Roll-on/roll-off or trailer ships are similarly "squared off" internally and, in addition, may have extensive deep tank spaces available below the lowest vehicle deck. As with the container ships, tankage is likely to be available in sufficient quantity to permit full clean ballast operations.

Dry Bulk Carriers

Dry bulk carriers are engaged primarily in the transport of such commodities as coal, grain and ores. Two general configurations exist, as shown in Figure 3. For light weight, high cubic cargoes, such as coal and grains, the hold configuration is such that water ballast capacity, in the amount of 35 per cent to 40 per cent of cargo deadweight, is available for clean ballast service, as shown in Figure 3a and 3b. Fuel oil bunkers are generally confined to deep tanks within the machinery spaces or to portions of the wing and double bottom tanks adjacent to the machinery spaces. Clean ballast operation is generally feasible.

Vessels designed specifically for heavier cargoes such as ores are generally arranged with comparatively small cargo holds and large surrounding tank spaces as shown in Figure 3c. Clean ballast operation is readily accomplished under all loading conditions.

Liquid Bulk Tank Vessels

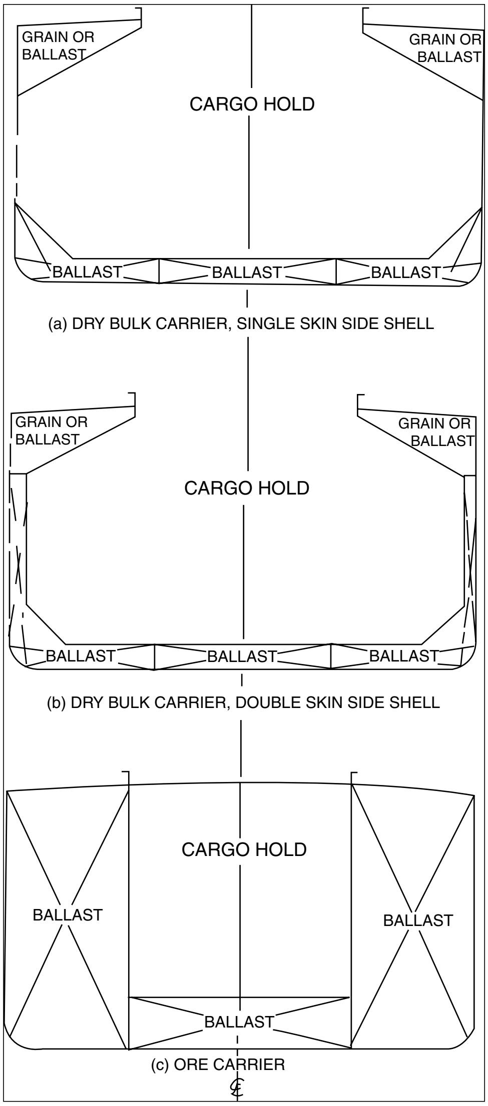
The modern tank vessel has evolved from the standard 16,000 deadweight ton (DWT) "T2" tanker of World War II to modern tank vessels exceeding 500,000 DWT capacity. The transport of liquid cargoes, predominantly petroleum crudes and refined petroleum products is the single largest category of waterborne commerce and represents the greatest potential pollution hazard with respect to normal operations as well as casualties. Accordingly, characteristics of vessels in the liquid bulk trades will be considered in somewhat greater detail than other ship types.

Petroleum Crude and Products Carriers With the exception of the steam turbo-electric main propulsion machinery and electric drive cargo pumps, the World War II T2 tanker is, in general arrangement, a parent of the tanker designs developed during the early post-war years. Typical characteristics of these vessels include:

1) Cargo section divided by a pair of longitudinal bulkheads into port, center and starboard tanks.
2) Relatively short cargo tanks independent of ship size.
3) Poop, bridge and forecastle superstructures with navigating bridge located amidships.
4) Forward and after fuel bunkers.
5) Forward and after pump rooms.
6) Relatively long, single screw, main propulsion machinery, with separate boiler and engine rooms.

From the late 1950s until the present, tanker design evolved through the following changes, all directly related to reduced cost of construction and operation:

1) Increase in size to over 500,000 DWT capacity, with corresponding increases in dimensions and operating drafts.



(a) DRY BULK CARRIER, SINGLE SKIN SIDE SHELL

(b) DRY BULK CARRIER, DOUBLE SKIN SIDE SHELL

(c) ORE CARRIER

FIGURE 3 Outline midship sections through cargo holds, typical dry bulk carriers.

2) Simplification of arrangement, particularly by reduction in number and increased size of cargo tanks. Typical modern crude oil tankers are arranged with as few as five center tanks and ten wing tanks. Secondary arrangement changes have included elimination of superstructure and houses amidships, location of all accommodation and navigation spaces aft and elimination of forward pump room and fuel bunkers.
3) Speed has remained within the 14 to 16 knot range.
4) Crew size has been reduced substantially, averaging as low as 19 men on U.S. flag as well as foreign tankers.
5) Propulsion system power levels have increased with size, approaching 40,000 SHP on a single screw. Centralized pilot house control of all propulsion machinery is a state-of-the-art development available to operators of diesel and steam turbine machinery.
6) Cargo pumping systems are generally similar to those of the post World War II period, except for increase in pumping rate with ship size. Elimination of pump rooms and fitting of deep well pumps in each cargo tank is a recent trend, following the arrangements of special products carriers.

Regulatory effects on tanker design, imposed since the 1973 MARPOL Convention for the Prevention of Pollution from Ships, have been significant. These regulations, imposed progressively from 1973 through 1985, include the following requirements and constraints:

1) Limitation on maximum cargo tank size to 30,000 cubic meters.
2) Segregated ballast tanks (SBT) on all new tankers larger than 20,000 DWT capacity. SBT capacity must be sufficient to obtain the following conditions:

    - Capacity to obtain minimum mean ballast draft of 0.02L + 2 meters, where L = length between perpendiculars
    - Trim by the stern no greater than 0.015L
    - Draft at the stern sufficient to submerge the propeller
    - SBT located to so as to protect 30% to 40% of the side shell in way of the cargo tanks. (Actual requirements vary with ship size and geometry.)

Since the segregated ballast tanks are restricted to clean ballast service only, the net effect of these requirements has been to increase the ship dimensions to accommodate the required SBT volume. As a result, modern tankers that meet the SBT requirements are volume rather than weight limited, and will only load to the assigned draft marks when carrying very dense cargoes.

The SBT capacity requirements are considerable, amounting to about 25% to 40% of the deadweight. The optimum SBT and cargo tank arrangements, for minimum ship acquisition cost, vary with ship size and proportions. A common arrangement is to assign two pairs of wing tanks within the cargo tank section to SBT service. In some cases the preferred arrangement is the concentration of segregated ballast in double bottom tanks extending under the entire length of the cargo tank section of the ship.

Table 1 include a summary of principal characteristics of U.S. flag tank vessels built since 1977. All meet the MARPOL SBT requirements. The SBT arrangements are reviewed later in connection with protection from collision and grounding.

Special Products A great variety of liquid products are carried in specially constructed tankers. While the quantities carried are small compared to the volume of petroleum

TABLE 1
Representative modern U.S. flat tank vessels

NAME	EXXON CHARLESTON	EXXON BAYTOWN	ATIGUN PASS	B.T. SAN DIEGO	EXXON VALDEZ
Length, B.P., m	185.93	229.82	263.35	278.90	288.04
Breadth, mld, m	32.26	32.26	52.74	50.60	50.60
Depth, mld, m	18.29	18.29	22.86	23.78	26.83
Draft, keel, m	12.80	11.73	17.47	18.08	19.66
Displacement, tonnes	56,970	73,700	200,400	220,800	244,145
Deadweight, tonnes	42,800	58,645	176,160	191,100	214,860
Cargo capacity, m^3	59,200	62,660	184,300	209,980	240,890
Ballast capacity, m^3	18,500	32,000	57,400	59,600	69,600
Cargo	Chem. and prods.	Crude	Crude	Crude	Crude
Number of cargo tanks	42	14	13	15	13
Propulsion machinery	Dir. diesel	Dir. diesel	St. turbine	St. turbine	Dir. diesel
Horsepower, max continuous	17,000 bhp	17,000 bhp	26,700 bhp	28,000 shp	32,240 bhp
Service speed, knots	16	15.8	16.5	14.25	16.25
Year delivered	1983	1984	1977	1978	1986
Notes:					
1) Segregated ballast	yes	yes	yes	yes	yes
2) Double bottom	yes	yes	no	yes	no
3) Double hull	no	no	no	no	no

crudes and refined products, transport of these commodities may involve unique containment problems and associated hazards.

Special products carriers may be classified in the following manner, according to nature of cargo:

1) Liquefied natural gasses (LNG) and liquefied petroleum gasses (LPG).

    (a) Low temperature—ambient pressure containment—The most exacting containment requirements are in this category, with cargo carried at about −260°F for liquefied natural gas (LNG). The largest LNG carriers at this time have capacities of about 130,000 cubic meters.

    LPG transport includes the carriage of such gasses as propane, butane and ethylene, with propane, carried at about −50°F, the most common.

    In a typical LNG or LPG carrier, cargo is carried in an independent, insulated tank or membrane liner. Double bottom and wing tank spaces are normally assigned to salt water ballast and fuel oil is carried in a relatively small portion of the double bottom and in deep tanks within the machinery spaces. Geometry is similar to that of a container ship and clean ballast operation is accomplished with no difficulty.

    (b) High pressure—ambient temperature—LPG may be carried in pressure vessels, designed to the A.S.M.E. Code for Unfired Pressure Vessels. While this mode of containment has been generally superceded by low temperature transport for international trade, a considerable amount of LPG and similar cargoes is carried in this manner on the inland waterways of the United States and Europe and in smaller coastwise vessels and barges. The limiting design condition is usually for propane, in cylindrical tanks designed for 250 psig. In general, vessels carrying cargoes in pressure vessels have sufficient cubic capacity to permit clean ballast operation.

2) Miscellaneous liquefied gasses—Anhydrous ammonia is carried in significant quantities in U.S. inland and coastal waters. This commodity may be carried at low temperature or under pressure, in containment designed for the transport of propane.

    Chlorine gas is commonly transported by barges in U.S. waters, primarily in pressure vessel containment. Other commodities of importance are primarily petro-chemicals, including butadiene, ethane, ethyl chloride, prophylene and vinyl chloride.

3) High temperature commodities—The transport of molten sulfur at about 275°F in heated independent insulated tanks has become the most common high

temperature commodity carried on international and inland waters. Internal hull geometry resembles that of the low temperature LPG or LNG vessel, with double bottom and wing tank spaces available for clean ballast.

The transport of asphalt and bitumen in the molten state is less exacting than the case of molten sulfur transport and cargo is normally carried in conventional integrated tanks.

Sulfur and bitumen cargoes are relatively dense and clean ballast operation should be expected. The use of cargo tanks for ballast services is not feasible, except for emergency situations.

4) Toxic and corrosive chemicals—A great variety of hazardous cargoes are carried in relatively small quantities in a variety of containment systems. References (3) and (4) contain a hazardous cargo classification and data for specific commodities, with particular respect to marine transportation.

Virtually all hazardous cargo carriers will be built with sufficient ballast tank capacity, in the form of integral double bottom or wing tanks. It is unlikely that the use of cargo tanks for salt water ballast would be permitted, except for emergency conditions.

Combination Bulk Carriers

In order to improve the overall utilization of conventional dry or liquid bulk carriers, combination bulk carriers have been developed to permit transporting dry and liquid cargoes within the same cargo hold spaces. A typical voyage, for example, would involve carrying crude oil from the Persian Gulf to Maine, ballast from Maine to Hampton Roads, coal from Hampton Roads to Japan, Japan to Persian Gulf in ballast, etc. Cargo operations of this type involve unique cargo handling and hold cleaning problems, with associated potential pollution problems.

Two general configurations exist, the ore-oil carrier and the more common ore-bulk-oil (OBO) carrier. These are analogous in function and similar in geometry to ore carriers and general bulk carriers, respectively, illustrated in this section sketches in Figure 3. A common modification in the latter case is the provision of a double skin side shell to facilitate hold cleaning.

Referring to Figure 3, the ore-oil carrier is equipped to carry cargo oil in the wing tanks as well as the center cargo hold. The double bottom space is normally reserved for clean ballast. The degree to which an ore-oil carrier can maintain a clean ballast operation, when operating as a tanker, will depend on the relation of cargo density to cargo volume available. The OBO will be operated with dry and liquid cargoes restricted to the main hold spaces, hence such vessels will normally operate with clean ballast, as a conventional bulk carrier. In both cases, however, hold cleaning between cargoes is a major operational problem that will be considered in later discussions.

The largest dry bulk carrier in existence is believed to be the 365,000 DWT ore carrier BERGE STAHL, delivered in 1986. The largest combination carrier is the 280,000 DWT ore/oil carrier SVEALAND, delivered in 1972.

Miscellaneous Commercial Vessels

The great variety of miscellaneous and floating craft that could be sources of pollution are too numerous to consider here. In general, all can be considered, with respect to pollution, in one of the categories considered earlier. One particular case, of current interest, however, is the development of large, unmanned seagoing barges for the ocean transport of dry and liquid bulk commodities. Tank barges of 50,000 DWT are in service. The geometry of a tank barge resembles that of an austere crude oil tanker of comparable deadweight, with five center tanks and 10 wing tanks. Operational as well as casualty pollution hazards are comparable to those of a self-propelled tanker, with the added complication that no personnel are aboard when the vessel is underway.

POLLUTION FROM NORMAL OPERATIONS

Ballasting and Tank Cleaning

Break Bulk Vessels The major source of pollution from break bulk general cargo vessels is in the intentional discharge of dirty ballast. As consumables, primarily fuel, are expended, displacement, draft and stability changes and may reach the condition that the addition of water ballast may be required. Some tankage may be available for clean ballast, but, in general, ballasting of fuel tanks will probably become necessary at some point beyond the expenditure of one half the consumables on board.

Since the imposition of the MARPOL regulations, the use of clean segregated ballast tanks has been mandatory and the disposition of oily ballast at sea should no longer be a major problem.

Tank Vessels, Crude and Refined Petroleum Products Until the MARPOL agreements came into effect, the greatest source of intentional discharge of contaminated ballast into the sea was from the operation of tank vessels transporting petroleum crudes and products. Tankers are normally one way product carriers and return voyages to the cargo source are in ballasted condition. MARPOL segregated ballast requirements for tank vessels were summarized earlier in Section 2.4. The arrangement of the ballast tanks to meet operating requirements and to provide some collision and grounding protection is discussed later.

Tank Vessels, Special Products Carriers The special products carriers described in earlier discussions are predominantly clean ballast vessels. Sea water will rarely be pumped into cargo tanks and sufficient tankage is normally provided, in the form of double bottoms and wing tanks, to serve as cargo tank protection as well as clean ballast tankage.

Dry Bulk Carriers The typical dry bulk carrier operates in ballast over a significant portion of the operating life. Many trade conditions exist in which return cargoes are not available and ballasting is required for adequate propeller immersion and seaworthiness. Bulk carriers are inherently stable and ballasting is not required for this purpose. As discussed earlier, sufficient cubic capacity is normally available in the wing and double bottom spaces to permit clean ballast operation.

Combination Carriers The most common of the combination carriers, the OBO, has the same general configuration as the dry bulk carrier and, accordingly, is generally capable of clean ballast operation. When in petroleum crude or product service, the OBO operates as a tank vessel and must comply with all relevant regulations. Hold cleaning between voyages with incompatible cargoes, however, is an additional source of pollution. In the example given in earlier discussions, the OBO discharges crude oil at Portland, Maine, proceeds in ballast from Maine to Hampton Roads and takes on coal at Hampton Roads for delivery to Japan. During the ballast voyage from Maine to Hampton Roads, the holds are cleaned by conventional means and the dirty oil washings discharged into slop tanks located in a pair of wings immediately aft of the cargo holds. A more complex situation arises for the ballast voyage from Japan to the Persian Gulf, after discharging coal. Several days of manual labor are required to remove coal residue, followed by Butterworth cleaning to remove the fine coal powder remaining. The solids and wash water are discharged at sea in unrestricted zones, unless prohibited by environmental regulations.

In general, combination carriers can be changed over from liquid to solid cargoes in a comparatively short time, say a period of 18 hours to two days. The reverse procedure may require more time consuming cleanup procedures. Operators, accordingly, will tend to prefer maintaining a given ship in a single cargo trade for seasonal periods, if permitted by the economics of the trade. Operational pollution problems are reduced in complexity when the occasions for cleaning between incompatible cargoes are minimized.

Cargo Transfer, Loading and Unloading

Liquid Cargoes Some pollution inevitably occurs as a result of fuel and cargo oil transfer between ship and terminal or between ship and lighter alongside. The majority of such incidents results, directly or indirectly, from human failure. Typical incidents include overflow through tank vents and hose failures. In most cases, proper monitoring or automatic control of cargo transfer and fueling operations will minimize the probability of oil spillage. Normal inspection routines should permit anticipating most equipment failures.

The rapid development of the large crude oil tankers has been accompanied by the parallel development of offshore terminals to accommodate the deep draft vessels. The tankers moor to large "monobuoy" single point mooring buoys which are anchored permanently to the sea floor. Oil pipelines are led to the underside of the monobuoy along the sea floor, from the shore tanks. Flexible hoses are led from the buoy to the midship pumping station on the tanker, usually by a tending launch. Means for mooring the tanker and connecting up to the oil hoses are under constant development and are reaching a high level of reliability. It is not expected that operations of the offshore terminals, under proper control, will represent a major source of pollution.

The lightering of petroleum crude from deep draft tankers offshore to draft-limited ports is a major activity at U.S. coastal ports. The majority of the existing shuttle tankers are relatively small 40,000 DWT to 50,000 DWT vessels. Shuttle tankers operate between transit vessels and terminals over one-way distances generally less than 100 miles. In some cases the service is limited to a lightening operation to reduce the transit tanker draft to the allowable terminal draft.

It is anticipated that pending U.S. legislation will address the lightering issue by allowing existing single skin transit tankers to be served by double hull shuttle tankers discharging to U.S. coastal terminals. It is understood that the minimum allowable standoff distance between transit tanker and coastal terminal will be 60 miles.

Shuttle tankers operate on short voyages, with frequent encounters with large transit tankers and terminals, through heavily travelled shipping channels. It is anticipated that requirements for environmental protection for this class of tankers will be demanding, considering the nature of the service and proximity to environmentally sensitive coastal areas.

Dry Bulk Cargoes Earlier discussions of combination operations included mention of solids pollution from hold washings when converting from dry bulk to liquid bulk operations. Of far greater importance is the harbor pollution occurring at dockside from the simple transfer, by grabs or similar mechanical devices, of dry bulk products between ship and shore storage facility. Over a long period of time, dry bulks spilled between ship and pier accumulate and become a local, but significant cause of harbor pollution. While many of the commodities are inert, others, including coal and some ores have an adverse effect on the ecology. The cargoes involved are of low value and command low freight rates, hence there is little incentive to control spillage of small, but accumulative, quantities into harbor waters.

POLLUTION FROM CASUALTIES

The magnitude of a particular oil spill, or other pollution casualty, is a function of ship type, ship size and nature of the incident. Tank vessel collisions or groundings involve the greatest magnitude of pollution resulting from individual casualties and, accordingly, will be considered in some depth in these discussions. Other vessel types are considered briefly.

Break Bulk Vessels

Pollution resulting from rupture of fuel tanks, as a result of collision or grounding, is the only significant casualty of this class of shipping likely to result in pollution. Less important is the potential rupture of deep tanks carrying various special cargo oils, primarily edible oils. Figure 1, showing a typical section through a cargo hold, indicates that double bottom tanks are normally assigned to fuel oil or ballast service. The largest general cargo vessel operating under the U.S. flag has a maximum fuel oil capacity, including all double bottoms, deep tanks and settling tanks, of about 3700 tons. About 83 per cent of this tankage, or about 3000 tons, is in the double bottoms. In general, the operator will tend to carry a minimum weight of fuel oil in order to maximize cargo deadweight, hence somewhat less than the 3700 tons of fuel oil is likely to be aboard.

The largest double bottom fuel tank in this particular vessel holds about 280 tons. A one-compartment damage collision, assuming damage from the side to the centerline, could expose 465 tons of fuel oil to the sea. Two-compartment damage, again from the side of the centerline, could expose about 900 tons of fuel capacity to the sea. A grounding incident, in which the bottom shell is opened to the sea for a considerable portion of the ship's length, could expose as much as 2/3 the double bottom fuel capacity, or about 2000 tons, to the sea. These values represent the maximum quantities likely to be exposed following a casualty. A considerable portion of the fuel would be released to the sea in any of these incidents. Magnitudes of such spills are significant but small relative to the catastrophic effects of a comparable tank vessel incident.

Unitized Cargo

Container, roll-on/roll-off and unitized cargo combination vessels in liner service are larger and higher powered than the break-bulk vessels, hence carry greater quantities of fuel oil. Arrangements of fuel, ballast and cargo oil tanks are varied and fuel oil may be located in bottom, wing or deep tanks. In general, the mode of release of fuel oil to the sea would be as discussed for break bulk vessels, with the quantities somewhat greater. A considerable portion of the wing and bottom tankage of unitized cargo vessels is piped only for ballast, thus lessening the probability that only fuel tanks would be breached in the event of a collision or grounding.

Tank Vessels

Fuel tanks are of relatively minor importance in the case of tank vessels involved in casualties. Fuel is generally confined to two or three deep tanks and settling tanks and represents a small portion of total tankage exposed to the sea following a casualty.

The evolution of tanker design since the early 1950's, with respect to pollution from collision and grounding, was considered briefly in earlier discussions.

Collision and Grounding Protection The cargo section of a modern tank vessel is arranged with the minimum number of tank divisions to meet loading, trim and safety requirements. Crude oil tankers may have as few as five tanks along the cargo length, divided into port, center and starboard tanks by a pair of longitudinal oiltight bulkheads, resulting in a 3 × 5 matrix of cargo tanks. A sixth pair of wing tanks, designated "slop tanks", may be located immediately forward of the machinery spaces, to accommodate cargo oil or cargo tank washings. A typical cargo tank arrangement is shown in the outline arrangement, Figure 4, illustrating the tank arrangement of the EXXON VALDEZ, Table 1. This design meets the MARPOL requirements for segregated ballast and limitations on cargo tank volume prevailing at the time of the construction contract in 1984. Wing tanks numbers 2 and 4 are designated as segregated ballast tanks within the cargo spaces.

A great variety of segregated ballast tank arrangements have been adopted to meet the mandated protection of 30% to 40% of the shell in way of the cargo tanks. The most common arrangement consists of two pairs of wing tanks, as in the EXXON VALDEZ, Figure 4. A less common alternative is to provide a continuous double bottom in way of the cargo tanks to carry most of the ballast, while providing significant grounding protection. This arrangement, typical of tankers carrying refined products or chemicals, is illustrated in Figure 5.

By the time of this writing in the spring of 1990, a series of major casualties had occurred during the 1989–1990 period. These events were followed by a period of intense investigation and legislative activity directed to development of improved means of minimizing the consequences of collisions and groundings. The widely publicized grounding of the EXXON VALDEZ in Prince William Sound, with an estimated outflow of 11 million gallons of crude oil, resulted in extensive environmental damage and massive cleanup efforts by EXXON and state and federal agencies. Figure 6 shows, diagrammatically, the extent of damage, involving eight of the 11 cargo tanks. It is estimated that 60% or more of the cargo outflow would have been retained had the EXXON VALDEZ been designed and built with a continuous double bottom. It is ironic to note that the EXXON VALDEZ design was based in part on the earlier design of the 188,700 DWT B.T. SAN DIEGO class of tankers which were built with double bottoms.

The recently enacted Oil Pollution Act of 1990, discussed further in Section 5, establishes requirements for double hulls for tank vessels operating in U.S. waters. Requirements include specific minimum values for depth and breadth of double bottoms and wing tanks, respectively. In anticipation of these requirements, designers and builders have developed designs of "environmental" tankers and a significant number of building contracts have been let for construction of these vessels.

A variety of cargo tank configurations have been developed for double hull tank vessel designs. The most widely proposed is a variation of the conventional arrangement, Figure 4, wherein the longitudinal bulkheads are located well outboard to form relatively narrow segregated ballast wing tanks, in association with a continuous double bottom. Two innovative concepts recently developed are the Japanese EPOCH design, Figure 7, and the Danish product tanker design, Figure 8. The latter evolved from a successful series of bulk carrier designs.

It should be noted that double hull design to meet anticipated regulatory safety requirements does not require new

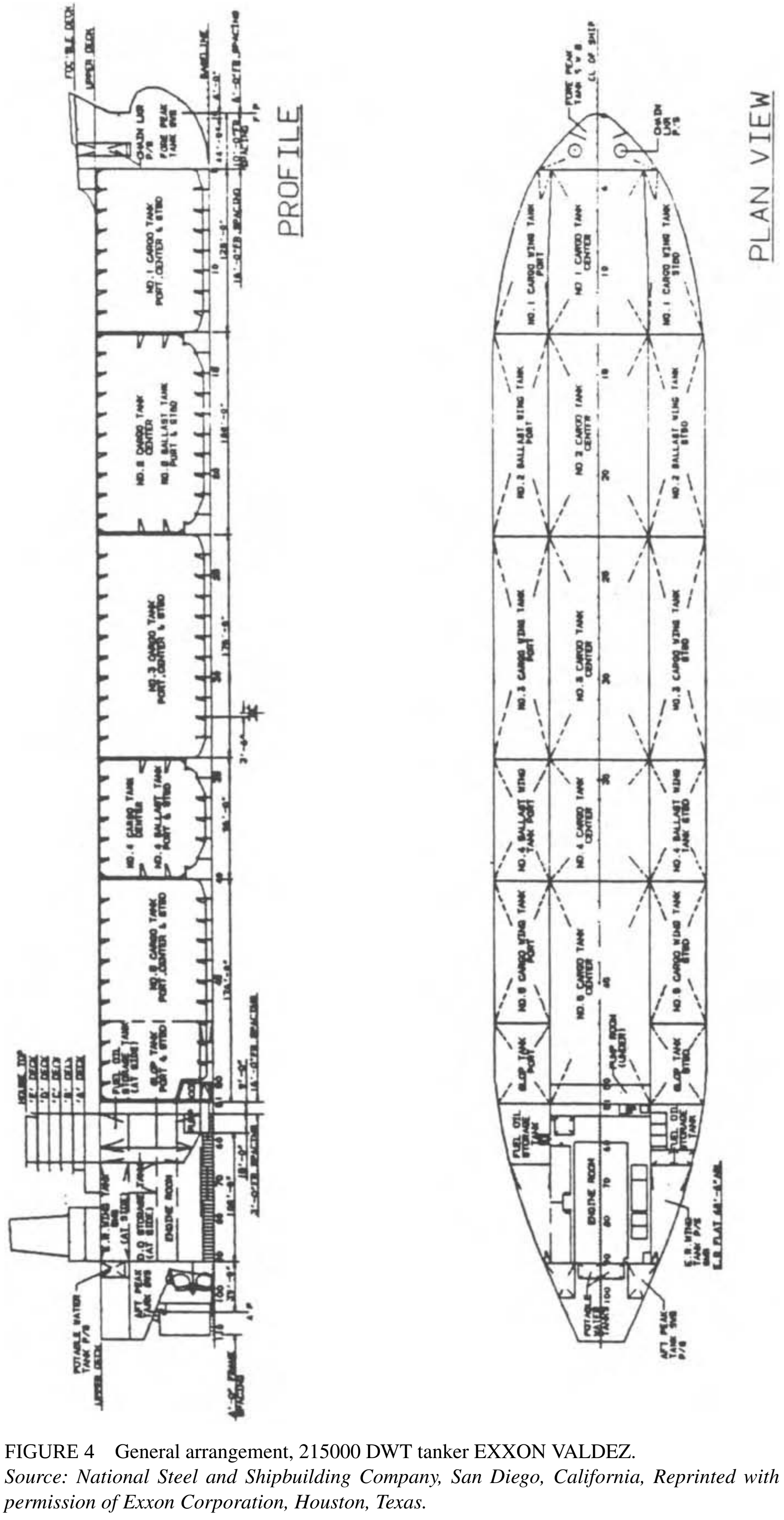


FIGURE 4 General arrangement, 215000 DWT tanker EXXON VALDEZ.
Source: National Steel and Shipbuilding Company, San Diego, California, Reprinted with permission of Exxon Corporation, Houston, Texas.

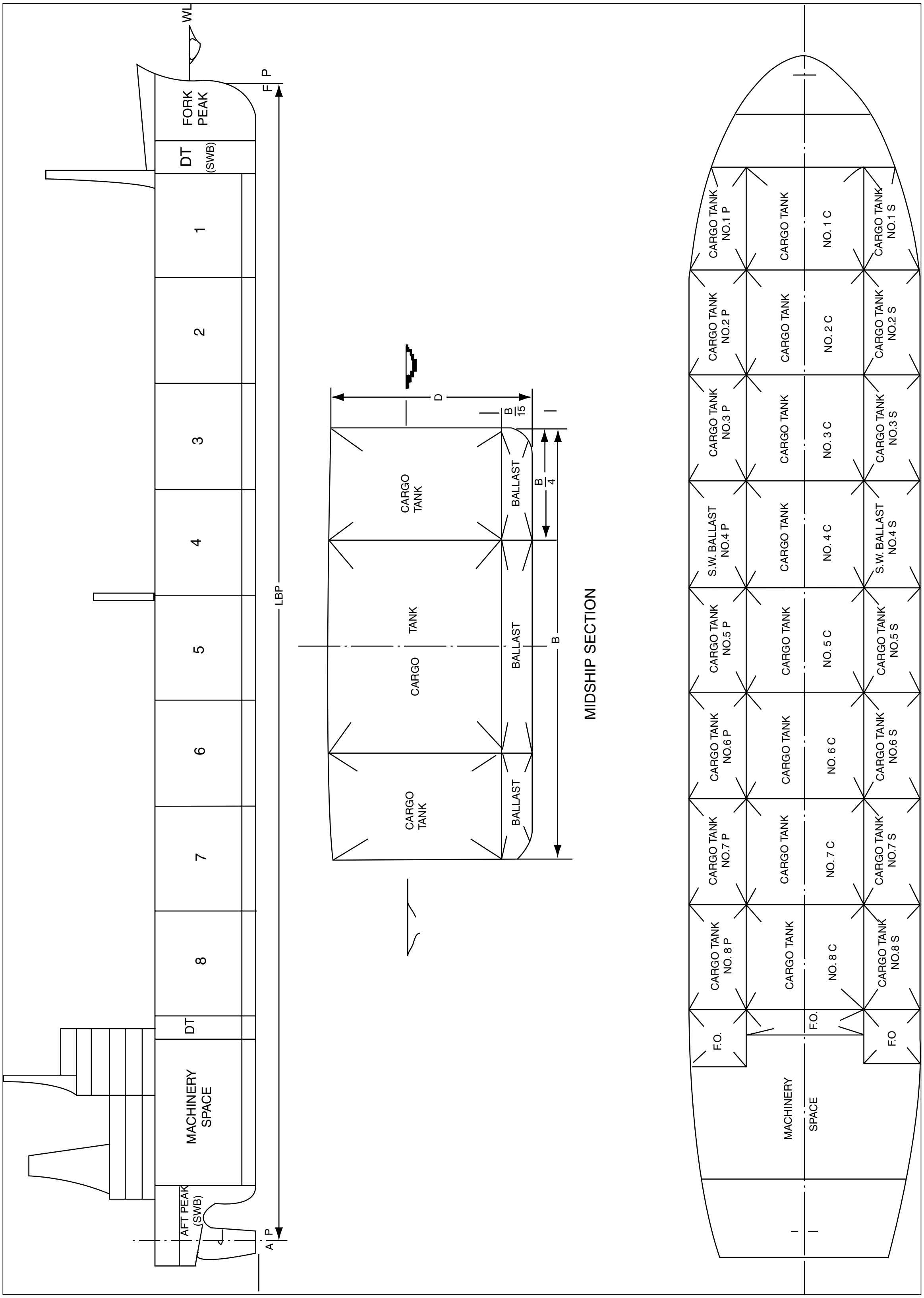


FIGURE 5 Outline arrangement of typical product carrier.
Source: U.S. Department of Commerce, Maritime Administration.

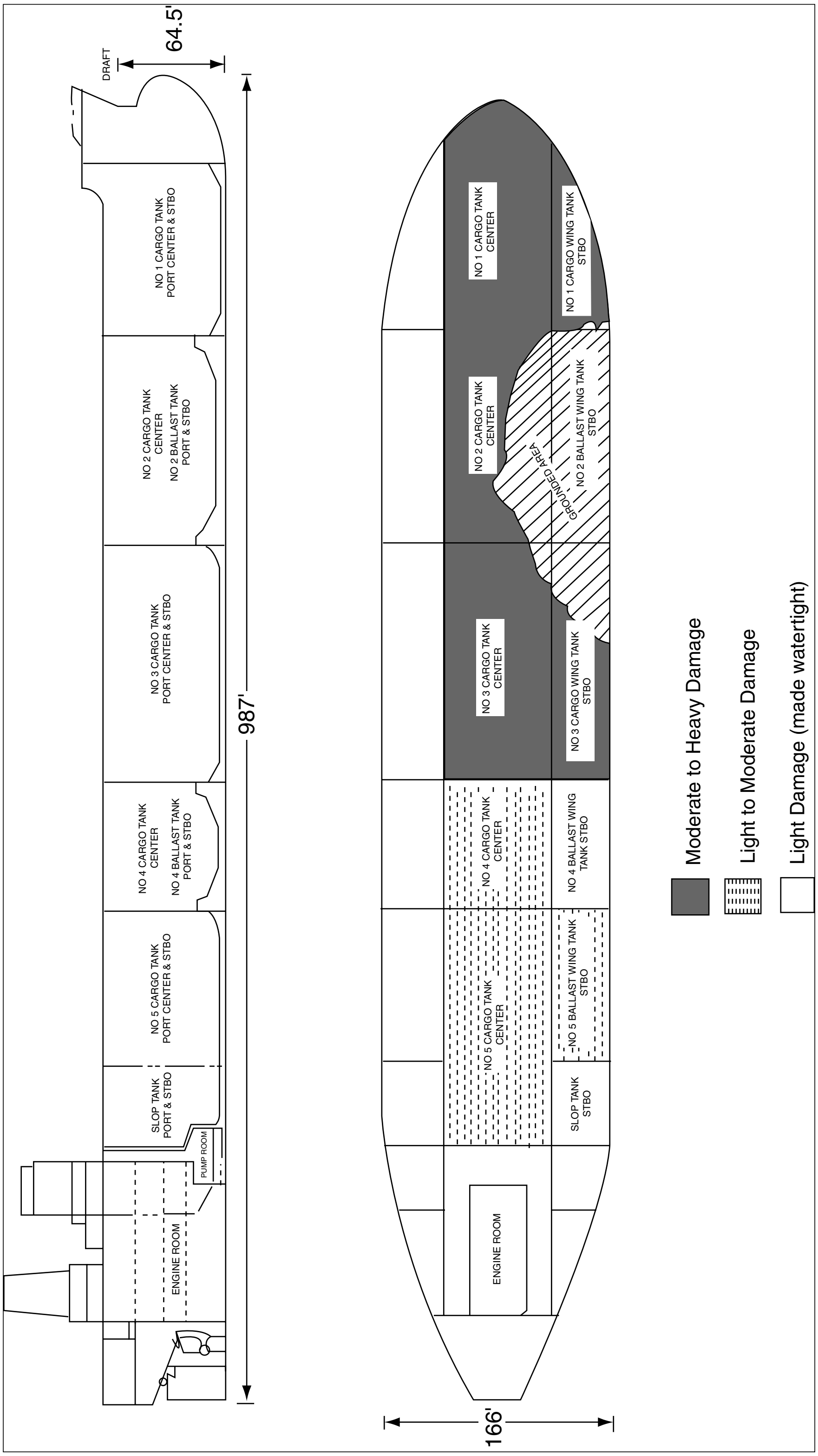


FIGURE 6 Estimated grounding damage to the EXXON VALDEZ. *Permission for use granted by Exxon Corporation, Houston, Texas.*

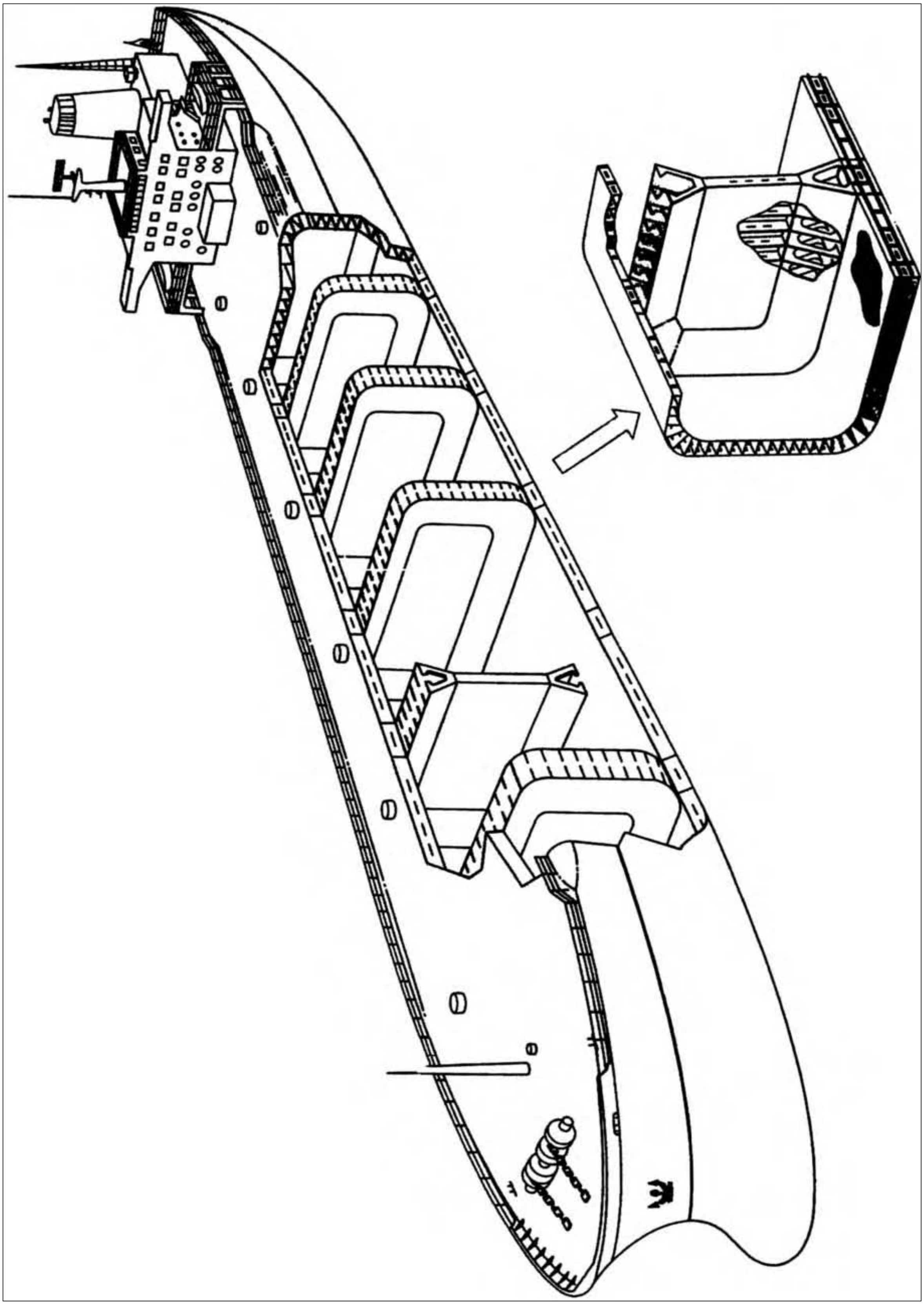


FIGURE 7 Hitachi Zosen EPOCH double hulled tanker design.
Source: Hitachi Zosen Corporation.

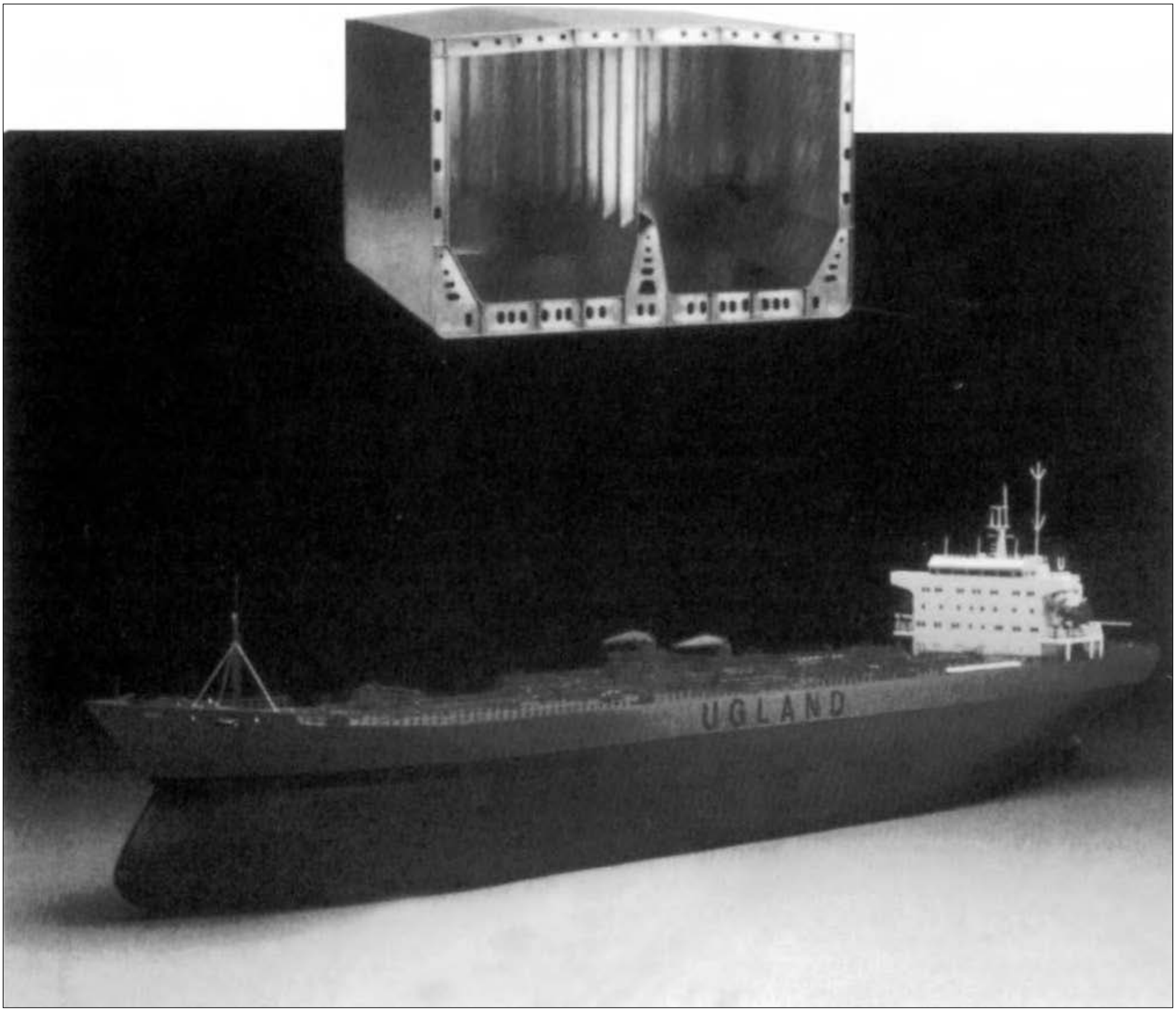


FIGURE 8 Burmeister & Wain double hull product tanker design. *Source: Burmeister & Wain Skibsvaerft A/S.*

technology. As noted earlier, LNG, LPG and a variety of special products carriers have all been built with double hulls, as mandated by regulatory requirements.

Special Hazards in Tanker Operations In addition to collision and grounding, the following hazards to tank vessel operation should be considered as potential sources of pollution:

a) Fire—Of the many possible disasters which may occur to a tank vessel in petroleum trades, fire is among the most common. Should rupture of a tank occur incidental to the fire, the cargo may be burned, thus preventing extensive pollution. However, gross structural failures may occur as a consequence of fire damage, resulting in uncontrolled release of cargo that may not be consumed by the fire.
b) Explosion—Despite the many regulations and operational precautions, explosions occur when an obscure source of ignition takes place in the gaseous atmosphere above the liquid cargo level. Some of the most disastrous incidents have taken place when tankers in ballast with empty tanks not gas-free have been involved in minor collisions.

It should be noted that the MARPOL requirements include the installation of inert gas systems (IGS) on all tank vessels of 20,000 DWT or greater. IGS systems consist of means for pumping flue gas or treated gas from an inert gas generator into the cargo tanks to displace air with a combustible oxygen content.

Special Products Carriers Casualties to the special products carriers described in earlier discussions involve unusual hazards and pollution problems. The responsible regulatory agencies were active in studying and regulating various special hazardous cargoes long before oil pollution was recognized as a major hazard. Accordingly, safety codes and regulations have been formulated and effectively enforced for a long period of time. Regulations of the U.S. Coast Guard, for example, which apply to various hazardous cargoes, as well as liquefied petroleum and natural gasses, are covered in References (9), (10) and (11).

Hazardous cargo transportation is given special treatment by the designer and the cognizant regulatory agencies by consideration of the following:

a) Materials—Where indicated by the corrosive or otherwise restrictive nature of the cargo, special materials, subject to rigorous testing requirements, are specified. Liquefied gas containment, for example, requires special steels or other materials to avoid brittle fracture at low temperature. Similar restrictions hold for piping and cargo pumps.
b) Containment—As noted earlier, in all cases where special hazards exist, it has generally been the practice to require transport of cargoes inboard of shell plating, that is, cargo is contained within an inner hull formed by an inner bottom and longitudinal bulkheads. Further restrictions in some cases require that the cargo containment be structurally independent of the ship's main hull structure. This is the case with pressure vessels and rectangular tanks for carrying liquefied gasses at low temperature and molten sulfur at high temperature. The requirement for secondary containment, that is, the ability to contain the cargo for indefinite periods of time in the event of rupture of the primary tank, has long been established as a basic requirement for the transport of liquefied gases at low temperature.
c) Cargo handling—Elaborate safety precautions to avoid spillage and sources of ignition are inherent in established design practices and regulatory requirements.

The above brief discussion indicates a comparatively high level of design for safety in the case of hazardous cargoes. In the event of a low energy collision or stranding there is a reasonable probability that the primary containment system will remain intact.

For the liquefied gasses, the greatest tonnage category of special hazardous cargoes, pollution hazards are limited, or at least of short duration, since the cargoes are volatile and would be dissipated in a short period of time in the event of primary tank rupture. For the other cargoes, toxic materials in particular, the pollution hazard remains. Cargoes are shipped in relatively small quantities, however, and effective containment is readily achieved with modern construction methods.

Dry Bulk Carriers

Fuel oil is likely to be carried primarily in deep tanks, and, to a lesser extent, in selected double bottom and wing tanks. Accordingly, fuel bunkers comprise a relatively small portion of total tank space and do not represent an unusual in the event of collision or grounding.

Combination Bulk Carriers

When transporting oil cargoes, combination carriers function as tankers. The internal arrangement of ore/oil tankers, Figure 3, resembles the arrangement of a tanker, with respect to location of longitudinal bulkheads at least 20% of the breadth from the side shell. Accordingly, the extent of cargo oil pollution from side collision is comparable to the case of the tanker, assuming comparable compartment lengths.

The ore/bulk/oil, or OBO, carrier, however, is designed with high capacity cargo holds in order to transport low density dry bulk cargoes as well as oil. As a result, the boundary of the center holds is well outboard the 20% breadth penetration line, for either the single skin or smooth skin side shell arrangement. In the event of a side collision in way of a single compartment of a nine-hold vessel about 1/9 the oil cargo will be exposed to the sea, compared to about 1/20 the total cargo in the case of a modern crude tanker. In the case of a low energy collision, the OBO with double skin side shell will offer a somewhat greater resistance to penetration than the more common single skin arrangement.

The double bottom arrangement of combination carries does provide for some grounding protection, particularly in way of the center holds. The lower outboard sloping wing tanks of some vessels, however, are regularly used as cargo oil tanks.

PREVENTION

Improvements in maritime safety standards have historically followed some major disaster at sea. This was the case with respect to loss of the TITANIC, the 1960 SOLAS convention with respect to the sinking of the ANDREA DORIA, and now the accelerated IMO activity in the wake of the 1988−1989 casualties.

It must be emphasized that economic incentives are the fundamental basis for the increased potential for pollution due to collision and stranding. The trend to larger tank vessels and toward simplification of the ship design, all directed to substantial economies in construction and operation, are not easily reversed. Yet this trend necessarily increases the pollution hazard occurring from an individual incident. In the light of the overwhelming short term economic incentives, it is clear that all improvements must necessarily come from the imposition of requirements from outside the industry; that is, from the deliberations of regulatory bodies at an international level and perhaps just as effectively, from the pressures of the insurance industry.

Some suggested directions for study in improvement of design and operation are included in the following discussion.

Operational Measures

Dirty Ballast This source of pollution should be largely eliminated with the imposition of MARPOL regulations requiring segregated ballast and the general prohibition against discharging oily waste at sea.

Fuel and Cargo Oil Transfer Fuel and cargo transfer operations are, to a large extent, under control of shore personnel. The ship's crew is responsible for tank loading sequences and monitoring the loading operation to avoid overflow through tank vents. The general adoption of cargo automation systems involving existing and developmental hardware should minimize this source of pollution.

Ship Design, Arrangement and Structure

The recently enacted Oil Pollution Act of 1990, Reference (12), mandates that all tank vessels of 5000 gross tons or larger, operating in waters subject to jurisdiction of the United States, must be of double hull construction by the year 2015. The law includes explicit phase-out time tables for retiring existing tank vessels, depending upon year of delivery and type of cargo tank arrangement, that is, single skin, double bottom, or double side configuration. The law also mandates the establishment of a research and development program, directed to improvements in ship design and operations. For example, the following are among the structural design measures which warrant study:

a) Structural design to increase collision protection. Wing tanks could be designed to absorb collision energy and limit penetration. Techniques are available to study various arrangements empirically. See References (5) and (6).
b) Structural design to improve grounding resistance. Double bottom structural systems could be optimized to enhance grounding protection. Empirical and analytical approaches to this design objective are reported in References (7), (8), and (9).
c) Mandatory installation of stress measuring instrumentation at strategic locations in the hull structure, with indicators located on the bridge. The objective here is to monitor cargo, fuel and seaway loading effects that may cause unacceptable stresses.

Ship Design, Performance and Maneuverability

Maneuverability is a primary means for avoiding collision and grounding in emergency conditions. Accordingly, the following observations and studies are pertinent to the overall problem.

Backing Power The conventional crash stop maneuver for the 550,000 DWT twin screw, twin rudder tanker BATILLUS, Reference (10), required 34 minutes to stop in 22 ship lengths. With both rudders turned outward, these values were reduced to 28 minutes and 16 ship lengths, respectively. Data given in Reference (11) demonstrates that steam turbine propelled single screw U.S. flag tankers, with displacement ranging from 25,000 to 77,000 tons, can be expected to stop in 8½ to 9 ship lengths from the full ahead condition and periods of 5½ to 9 minutes will be required to execute this maneuver from the time the crash stop order is given.

The geared steam turbines installed in the above ships are fitted with astern elements with a capability of about 80 per cent full ahead torque and 50 per cent full ahead RPM, i.e., about 40 per cent full ahead power. The time required to obtain full astern revolutions, from the full ahead condition, is on the order of one minute. Crash stop characteristics of the diesel ship may be somewhat better, because full astern power is available.

It has been shown that tanker sizes have grown at a far greater rate than the installed power, hence, the stopping ability has been reduced significantly, with the obvious reduction in ship maneuverability.

Steering and Directional Control Turning circles of conventional single screw tankers, expressed as turning circle diameter/LBP, vary from 3 to 4. Current U.S. Coast Guard regulations require the posting of turning circle and crash stop trial data on the bridges of ships over 1600 gross tons. While no minimum standards are included in international or federal regulations at this time, it is understood that establishment of such standards is under active consideration.

Miscellaneous Safety Considerations

Regulatory agencies such as the U.S. Coast Guard regularly upgrade national shipping regulations in compliance with changes in international regulations, and in response to changes in the nature of particular transportation systems. Current examples (mid 1990) include upgrading of safety standards for coastal passenger vessels, and the comprehensive Oil Pollution Act of 1990. Other current design and operational trends that are being addressed by regulatory agencies and classification societies include the following:

a) Manpower reduction and automation—Crew sizes on large oceangoing tank vessels average about 19 to 21 persons. Trends to further reductions to 12 to 15 persons within the next few years are conservatively anticipated. These crew reductions must be supported by integrated command and control systems and condition monitoring of main and auxillary machinery systems and components.
b) Maintenance reduction—To maintain a safe operating condition, inspection and maintenance procedures must be maintained at acceptable levels, while recognizing the constraints of reduced crew size.

c) Enhanced navigation systems—Long term objectives include the integration of satellite and terrestrial navigational aids, electronic updating of charts, and automatic position plotting. Safety objectives include installation of grounding and collision avoidance systems and "dead man" alarms. Problem areas to be addressed include international standardization of data format and transfer, and barriers to electronic data transfer inherent in chart copyright status. Research and development is required in all areas noted above, to attain a "user friendly" state of development and to reach the degree of reliability necessary to satisfy regulatory requirements.

Detailed discussions of potential improvements to tank vessel design, construction and operation are included in Reference (13).

REFERENCES

1. Taggart, "*Ship Design and Construction*", The Society of Naval Architects and Marine Engineers, 1980.
2. Dillon, Hoffman and Roseman, "Forty Years of Ship Designs Under the Merchant Marine Act, 1936−1976", Transactions, The Society of Naval Architects and Marine Engineers, 1976.
3. National Academy of Sciences, National Research Council, "Evaluation of the Hazard of Bulk Water Transportation of Chemicals—A Tentative Guide," report to the U.S. Coast Guard, 1970.
4. "Chemical Data Guide for Bulk Shipment by Water", U.S. Coast Guard, CG-388.
5. Chang, *et al.,* "A Rational Method for the Prediction of Structural Response Due to the Collision of Ships", Transactions, The Society of Naval Architects and Marine Engineers, 1980.
6. Reckling, "Overall Structural Response of a Ship Struck in a Collision", Transactions, Society of Naval Architects and Marine Engineers. Spring Meeting/STAR Symposium, 1980.
7. Poudret, *et al.,* "Grounding of a Membrane Tanker—Correlation Between Damage Predictions and Observations", Transactions, Society of Naval Architects and Marine Engineers. Symposium on Extreme Loads Response, October 1981.
8. Poudret, "Collisions and Groundings—Practical Analysis Methods", Arctic Section, Society of Naval Architects and Marine Engineers., March 1982.
9. Wierzbicki, Shin and Rady, "Damage Estimates in High Energy Grounding of Ships", presentation to Committee on Alternatives for Tank Vessel Design, National Research Council, 25 March, 1990.
10. Laredo, Beghin and Garguet, "Design of the First Generation of 550,000 DWT Tankers", Transactions, The Society of Naval Architects and Marine Engineers, 1977.
11. Long, Stevens and Tompkins, "Modern High Speed Tankers", Transactions, The Society of Naval Architects and Marine Engineers, 1960.
12. Oil Pollution Act of 1990, Conference Report to accompany H.R. 1465.
13. Tanker Spills: Prevention by Design, Committee on Tank Vessel Design, Marine Board, National Research Council; National Academy Press, Washington, D.C., 1991.

DONALD P. ROSEMAN
David Taylor Research Center

MICROBIOLOGY

INTRODUCTION

Microbiology is the study of organisms which are small enough to require the aid of a microscope to be seen. In a few cases, organisms are included in this group which can be seen by the unaided eye because these organisms are clearly related to the smaller ones. Microorganisms include viruses, bacteria including rickettsia, mycoplasma, fungi (yeast and molds), most algae, protozoa and, if one interprets “micro” broadly, certain tiny multicellular plants and animals. The study of cells and tissues from higher plants and animals (*tissue culture*) uses techniques common to the microbiologist and is frequently considered part of modern microbiology.

Cells in general vary greatly in size but have many similarities in internal organization. Among the most primitive type of cells, it is impossible to clearly distinguish whether they are distinctly “plants” or “animals” since they may have some of the properties of each type. Viruses, on the other hand, are *not cells* at all. Instead of arguing endlessly about whether a microorganisms is more plant-like or more animal-like and worrying how to assign viruses, many scientists have divided organisms in general into those which have (1) only animal characteristics, (2) only plant characteristics and (3) the *Protista* (Table 1), which have some characteristics of both plants and animals. Some protists, viruses, may have characteristics not shared by either plants or animals, that is, crystallizability and ability to reproduce only by infecting some cell and using the cell’s manufacturing machinery.

PHYSICAL CHARACTERISTICS OF MICROORGANISMS

Protists vary greatly in size, shape and internal architecture. Protists are subdivided into *prokaryotes,* and *eukaryotes. Prokaryotes* do not have their genetic material (chromosomes) separated from the rest of the cell by a membrane whereas eukaryotes have a true nucleus (*eu*—true, *karyo*—nucleus) separated from the rest of the cell by a nuclear membrane. Viruses (virions) are usually included among the prokaryotes. There are 9 types of prokaryotes.

Prokaryotes

1) *Viruses* are the *smallest protists.* They range in size from about 30–300 nm. The smallest viruses can only be visualized with an electron microscope while the largest can be seen with a light microscope. Viruses are composed of two general molecular types (1) *only one* nucleic acid, either *ribonucleic acid* (RNA) or *deoxyribonucleic acid* (DNA), and (2) a group of proteins also called *protein subunits* or *capsomeres,* which surround the

TABLE 1
Characteristics of the Protista

Virion (virus)	30–300 nm	icosahedron, hollow cylinder icosahedral head + tail	RNA, DNA	requires participation of host machinery
Mycoplasma	100–300 nm	pleomorphic prokaryotes	DNA	fission
True bacteria	250–3000 nm	spherical, rod, spiral rods, prokaryotes	DNA	fission
Higher bacteria	500–5000 nm	spherical, rod, spiral rods, filamentous, prokaryotes	DNA	fission, budding
Prokaryotic algae	500–5000 nm	spherical, rods in chains, spiral rods in chains	DNA	fission, internal septation, gonidia
Eukaryotic algae	500 nm to macroscopic	unicellular or multicellular, filamentous, leafy	DNA in nucleus, chloroplasts, mitochondria	asexual or sexual simple fission to complex life cycles
Protozoa	500–500,000 nm	unicellular or colonial various forms	DNA in nucleus, mitochondria	asexual or sexual simple fission to complex life cycles

nucleic acid and form a protective coat or *capsid.* The smallest viruses appear spherical but magnification in the order of 150,000–700,000 × reveals that they are *icosahedrons* (20 triangular faces and 12 corners) for example, wart virus. Other viruses, for example, tobacco mosaic virus (TMV), the first virus crystallized in 1935 by Wendell Stanley, is grossly rodlike. Tobacco mosaic virus is composed of a central, spirally-attached RNA to which capsomeres are attached to form the outside of a cylinder. The center of the RNA spiral of TMV is hollow. Structurally, the most complicated viruses are some which attack bacteria and blue-green algae.

These complicated viruses are composed of an icosahedral head, containing DNA, a protenaceous tail and sometimes accessory tail structures which are important for the attachment of the virus to its host cell.

2) *Mycoplasma* are prokaryotes which overlap viruses in size. They range from 100–300 nm in size. They are highly *pleomorphic:* they do not have one typical shape but rather can appear coccoid, filamentous, or highly branched. Unlike most other prokaryotes, they do not have cell walls external to their cell membranes. Their cell membranes usually contain sterols, which are thought to lend strength to these cell-limiting membranes (see also Table 2).
3) The *true bacteria* or *Eubacteriales* are prokaryotes which are built on three general geometric forms: spheres or cocci, *rods,* and *spirals* (including spiral helices). All true bacteria have rigid cell walls. They are either permanently immotile or move by means of one to many flagella. They may be aerobes or anaerobes. Some of the anaerobes are photosynthetic. Their sizes and shapes are usually constant except among the rods, in which rapidly-multiplying cells may be somewhat smaller than usual. When the cells divide, they often remain attached to each other and form characteristic, multicellular clusters. The shape of the cluster is determined by the number of division planes.

When *cocci* divide in only one plane, they form chains which may be as much as 20 cells long. *Diplococcus pneumoniae* forms chains only two cells long while *Streptococcus* is an example of the long-chain forming type. On the other hand, cocci which divide along two planes, at right angles to each other, form sheets of cells, and cocci which divide in three planes form cube-shaped packets. If there is no regular pattern of the orientation of successive division planes, a randomly-shaped cluster is formed. *Staphylococcus* is an example of a coccus which forms random clusters. A typical coccus is in the size range of 0.15–1.5 μm in diameter.

Rods always divide in only one plane. They may appear as single cells or groups of only two when they separate rapidly. The common intestinal bacterium *Escherichia coli* (size 0.5 × 2.0 μm) is an example of this type. Frequently rods form long chains or *streptobacilli. Bacillus megaterium* (size 1.35 × 3.0 μm), the organism responsible for the "bloody bread" of ancient times, is an example of a chain forming rod. Some basically rodshaped bacteria are either curved or helical rods. Their sizes range from almost as small as the smallest straight rod shaped form to close to twice the length of the largest straight rod.

True bacteria always divide by binary fission after their single circular chromosome replicates in a semiconservative fashion.

Some true bacteria have complicated life cycles which includes *spore-formation.* Spore-formers are all rods but belong to diverse genera. They are ecologically related in that they are found primarily in soil. Since that natural

TABLE 2
Some characteristics of prokaryotic and eukaryotic cells

Structure	Prokaryote	Eukaryote
Weight Chromosome	0.001–1.0 pg one, single circular DNA double helix not complexed with histones	10–10,000 pg several linear DNA double helices (several chromosomes usually complex with histones)
Nucleus	No true nucleus. Chromosomes not separated from cytoplasm by a membrane	True nucleus. Chromosomes enclosed in a nuclear membrane
Reproduction	Usually asexual, conjugation takes place rarely, no mitosis or meiosis	Asexually by mitosis or sexually after meiosis
Membranes	Only cell limiting membrane present. Usually lacks sterols (except for mycoplasma)	Cell limiting membrane plus membrane limited organelles present. Composition includes sterols
Organelles	None	Many including mitochondria, chloroplasts (plants only), Golgi apparatus, lysosomes, etc.
Apparatus for protein synthesis	Ribosomes, 70 S type usually not associated with membranes	Ribosomes, 80 S type in cytoplasm associated with endoplasmic reticulum. 70 S type in mitochondria and chloroplasts not associated with membranes

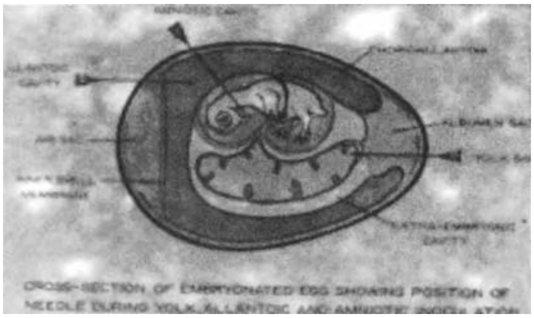


FIGURE 1 Animal viruses are often grown in embryonated eggs. The position of the hypodermic needles indicates three common inoculation places.

environment is rather variable in that it can range from very hot to very cold and from very wet to very dry, the heat- and cold-resistant dormant spores offer the bacteria a means of surviving adverse environmental conditions for months or even years. Many important pathogens and commercially important organisms are spore formers, e.g. *Bacillus anthracis* which causes anthrax, *Clostridium tetani* which causes tetanus and *Clostridium acetobutylicum* which can ferment corn or potato mash into acetone, ethanol and butanol.

Corynebacteria are also rod-shaped bacteria but they are *pleomorphic* and often look club-shaped. One of the best known members of the genus is *C. diphtheriae*, which causes diphtheria. Other members of the genus are commercially important as producers of the vitamin folic acid.

Arthrobacter species are found widely in soil and water. Depending upon the nutrients supplied, they can appear as cocci or pleomorphic rods.

4) *Spirochetes* are NOT true bacteria though they resemble Eubacteriales in that they are spirally curved, unicellular and multiply by binary fission. They differ from eubacteria by the *absence of a rigid cell wall* which allows them to be quite flexible. They are all motile by means of *axial filaments* attached at the cell poles and spirally wrapped around the cell. The smallest spirochete is 0.1 × 5 nm while the largest is 3.0 × 120 μm. One of the most important spirochetes is *Treponema pallidum*, which causes syphilis.
5) *Actinomycetes* are NOT true bacteria. Rather, they are naturally-branching, filamentous, spore-forming organisms which have a *mycelial structure* similar to that of filamentous fungi. Many actinomycetes, especially those from the genus *Streptomyces,* are commercially important sources of antibiotics.
6) *Mycobacteria* are rods which can form a rudimentary mycelium which resembles actinomycetes, but they differ in that their cell walls are particularly rich in waxes, which allows them to retain stain imparted by such dyes as basic fuchsin even after treatment with dilute acid. This property, called *acid fastness,* is characteristic of mycobacteria. Many species occur in soil but the best known are the human pathogens *M. tuberculosis* and *M. leprae*, which cause tuberculosis and leprosy respectively.
7) *Budding bacteria* are NOT true bacteria. They possess a complicated life cycle which includes *multiplication by budding* rather than binary fission. Their type of budding can be readily distinguished from that of true fungi such as yeast. The budding bacterium *Hyphomicrobium* exists for part of its life cycle as a flagellated, slightly curved rod. For multiplication, the flagellum is lost, the chromosome replicates, and one chromosome migrates to one end of the cell where a hypha-like lengthening takes place. When the hyphal extension ceases, it becomes a rounded bud which contains the chromosome. The bud grows in length and diameter until it reaches the size of the mother cell, grows a new flagellum, and separates from the hyphal extension.
8) *Gliding bacteria* are diverse group of prokaryotes which are motile without having flagella. They have very close affinities to blue-green algae although gliding bacteria are not themselves photosynthetic. They may be unicellular rods, helical or spiral-helical, or filamentous.
9) *Blue-green algae* or *Cyanophyta* are the only prokaryotic algae. They are a diverse group that include both unicellular and filamentous forms. They have cell walls that resemble Gram-negative bacteria but their photosynthesis more closely resembles that of eukaryotes in that it is aerobic rather than anaerobic (as in photosynthetic bacteria). They are among the most complex prokaryotes. Even though they lack defined organelles, e.g. they lack *chloroplasts,* many species have complex membranous or vesicular substructures which are continuous with the cell membrane.

Some filamentous forms contain specialized structures such as *gas vacuoles, heterocysts,* or *resting spores* (*akinetes*). Gas vacuoles are frequently found in *planktonic species,* i.e. those which live in open water. These vacuoles are thought to provide the algae with a means of floating and sinking to the depth most appropriate to support photosynthesis. *Heterocysts* arise from vegetative cells and are thought to function in N_2 fixation. Some blue-green algae show gliding motility. None are flagellated. They are very widely distributed either in terrestrial or aquatic habitats from the arctic to the tropics. Some forms are found in hot springs. Other *Cyanophyta* are *symbionts* in a variety of plants and animals. For example a species of *Anabaena* fixes N_2 for its host the water fern, *Azolla.* Many blue-green algae form especially luxuriant mats of growth called *blooms* which clog waterways and limit their use for navigation, etc.

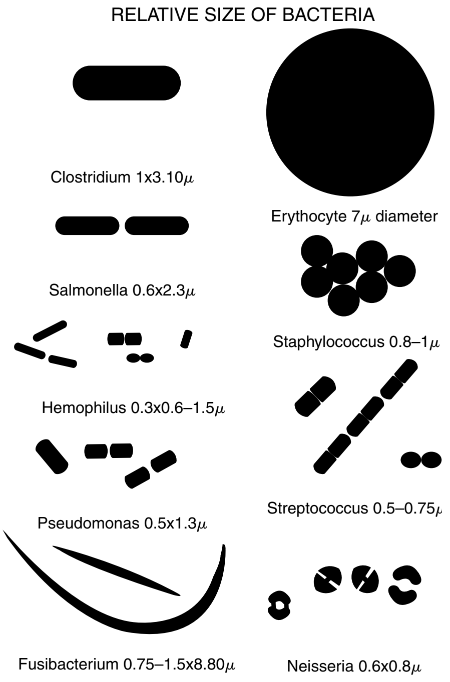


FIGURE 2 Relative sizes of bacteria.

The prokaryotic blue-green algae, *Cyanophyta,* are usually divided into 5 groups:

Chooccocales are unicellular. They sometimes occur in irregular packets or colonies. Cells multiply by binary fission.

Chamaesiphonales are unicellular, filamentous, or colonial epiphytes or lithophytes. Cells show distinct polarity from apex to base. The base usually has a holdfast which permits attachment to the substrate. Cells multiply by internal septation or by formation of spherical cells (*gonidia*) at the ends of filaments.

Pleurocapsales are filamentous with differentiation into aerial and nonaerial elements. Cells multiply by crosswall formation or by internal septation.

Nostocales are filamentous without differentiation into aerial and nonaerial elements. They are unbranched or falsely branched and frequently have pale, empty-looking cells called *heterocysts* and resting spores (*akinetes*). Reproduction is by liberation of a short filament only a few cells long, called a *hormogonium,* which then elongates.

a) *Nostacaceae* are unbranched and produce heterocysts. They frequently produce akinetes.
b) *Rivulariaceae* are unbranched or falsely branched. Filaments taper from base to tip. Heterocysts are usually present at the base. There is some akinete formation.
c) *Scytonemataceae* are false branched. Heterocysts are frequently found at branch points.
d) *Stigonematalis* are filamentous with aerial and nonaerial differentiation. Hormogonia and heterocysts are present. They often show true branching and have pit connections between cells. Akinetes are rare.

Eukaryotes

Eukaryotic microorganisms include all the algae (except the *Cyanophyta*), all the protozoa, and most fungi. All are microscopic in size.

The eukaryotic algae are separated into nine *divisions* based upon their pigment and carbohydrate reserves (Table 3). They are all photosynthetic and, like higher plants, evolve oxygen during photosynthesis. Many algae are *obligate phototrophs.* That is, they are completely dependent upon photosynthesis: they can not use exogenously supplied organic compounds for growth in either the dark or light. Some algae are *facultative phototrophs;* they are able to utilize organic compounds for growth in the dark but fix carbon dioxide photosynthetically in the light.

Occasionally algae, especially unicellular forms, permanently lose their chloroplasts by exposure to any one of several adverse conditions, e.g. heat or chemicals. If the organism had been a facultative phototroph, before the loss of the chloroplasts, it has the enzymatic machinery necessary to survive except that now, in its chloroplastless state, it is indistinguishable from certain other unicellular organisms more commonly called protozoa. The ease with which an organism at this primitive level of evolution may be interchanged between groups containing a preponderance of plant-like or animal-like attributes underlines the need for the term *protist* rather than plant or animal to describe them. Indeed both botanists and zoologists claim the protists. Some algae e.g. *Euglena* spp., normally only form chloroplasts when they grow in the light while others e.g. *Chlorella* spp. form chloroplasts regardless of the presence of absence of light. There is great diversity in size, shape, presence or absence of life cycles, type of multiplication, motility, cell wall chemistry, and chloroplast structure. Although these parameters are of great assistance in defining affinities among algae, there are still groups whose proper place is debated.

Many algae are important as sources of food, chemical intermediates of industrial and medical importance, and research tools. Others are nuisances which clog waterways or poison other aquatic life with their potent toxins.

Eukaryotic Algal Groups

The eight groups are:

1) *Chlorophyta* (green algae) are either marine or freshwater forms. This large and diverse group includes forms which are either unicellular, colonial, filamentous, *tetrasporal* (cells separated but held together in groups of four in a mucilaginous material), *coenobial* (cells more or less attached to each other in an aggregate), or *siphonaceous* (simple, nonseptate filaments). They frequently have life cycles which

TABLE 3
Divisions and characteristics of the eukaryotic algae

	Pigments							Reserve material	
	Chlorophyll					Biliproteins		Starches (α-1,4-glucans)	
	a	b	c	d	e	Phyco-cyanin	Phyco-erythrin	True starch	Floridian starch
Chlorophyta (green algae)	+	+	−	−	−			+	
Euglenophyta	+	+	−	−	−	−	−		
Xanthophyta	+	−	−	−	+	−	−		
Chrysophyta	+	−				−	−		
Phaeophyta (brown algae)	+	−	+	−	−	−	−		
Bacillariophyta (diatoms)	+	−	+	−	−	−	−		
Pyrrophyta	+	−	+	−	−	−	−	+	
Cryptophyta	+	−	+	−	−	+	+	+	
Rhodaphyta (red algae)	+	−	−	?	−	+	+		+

Reserve material (cont.)

	β-1,3 glucans			Sugars			
Division	Laminarin	Paramylon	Chrysolamainarin	Floridoside	Sucrose	Sugars alcohols Mannitol	Lipid
Chlorophyta (green algae)					+		
Euglenophyta							
Xanthophyta		+					+
Chrysophyta							
Phaeophyta (brown algae)	+ o					+	+
Bacillariophyta (diatoms)			+				+
Pyrrophyta							+
Cryptophyta							
Rhodophyta (red algae)				+			

include motile, flagellated stages. Both asexual and sexual reproduction occurs.

2) *Euglenophyta* differ from the other algae by possessing a rather flexible cell wall which allows considerable plasticity of form. They are either fresh water or marine forms. They all have two flagella but in some genera the second flagellum is often rudimentary. Many forms are *phagotrophic* (can ingest particles). Chloroplastless forms are fairly common. Multiplication is only by asexual means.
3) *Xanthophyta* are mostly freshwater forms. They may be unicellular, colonial, filamentous or siphonaceous. Some forms have life cycles which include both asexual and sexual reproduction. Motile anteriorly flagellated cells are found.
4) *Chrysophyta* are mainly freshwater forms but important marine forms are known. Most genera are unicellular but there are some colonial forms. Cell walls are often composed of siliceous or calcareous plates. Some form siliceous cysts. They are mainly found in fresh water but some important marine forms exist. Reproduction is asexual.
5) *Phaeophyta* (diatoms) are unicellular or colonial forms with distinctly patterned siliceous cell walls. Both asexual and sexual multiplication is found. Freshwater, marine, soil and aerial forms exist.
6) *Pyrrophyta* are unicellular flagellates with cellulose cell walls which are sometimes formed in plates. Reproduction is asexual. Sexual reproduction is rare.
7) *Cryptophyta* are unicellular, usually flagellated forms which produce asexually.
8) *Rhodophyta* (red algae) are unicellular, filamentous or leafy forms with complex sexual cycles. Most are marine but there are a few freshwater forms.

Fungi

The "true" fungi or *Eumycota* are eukaryotes which are related to both protozoa and algae. They are divided between

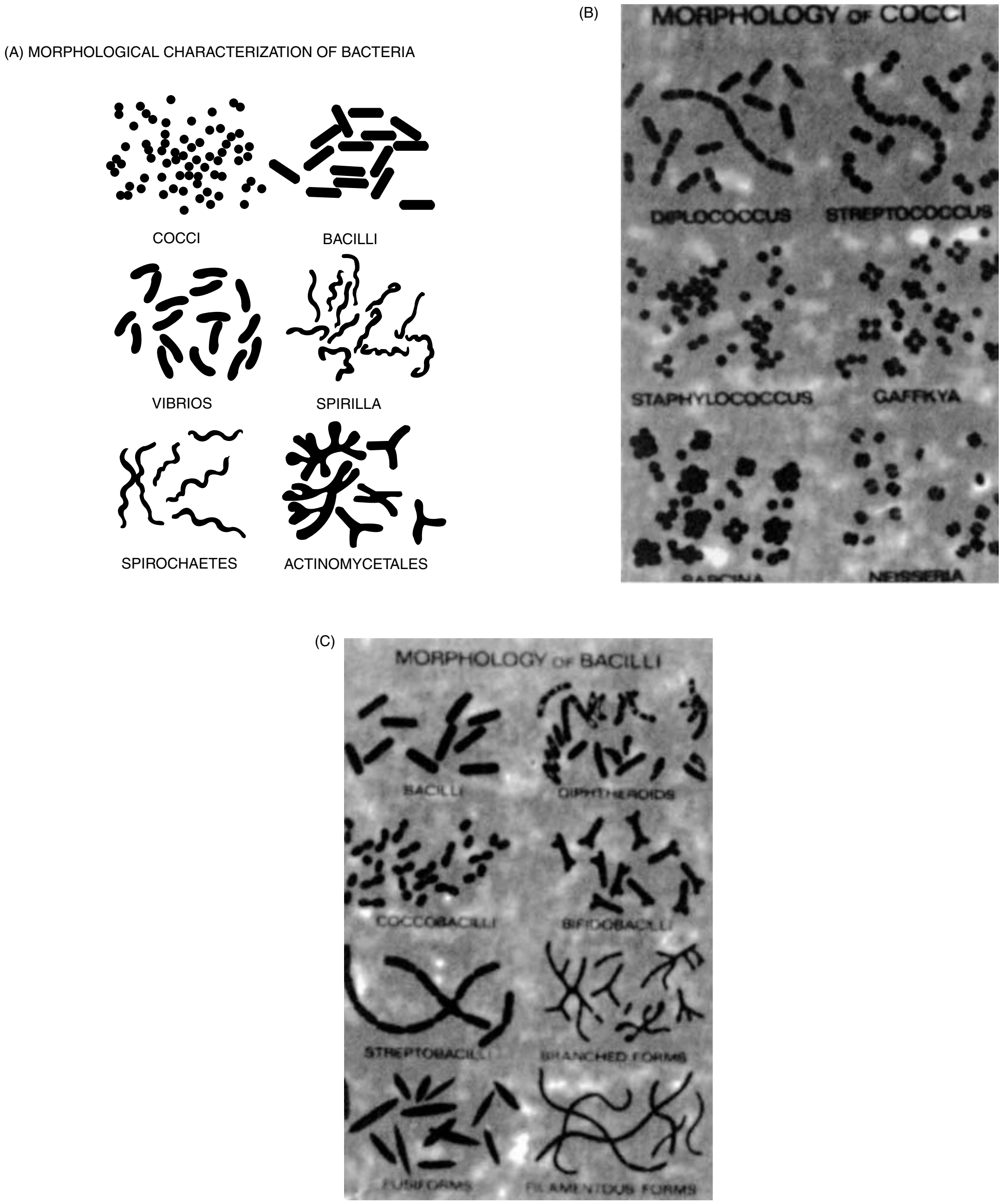


FIGURE 3 A. General morphological characteristics of bacteria; B. Variety of morphological types among the cocci; C. Variety of morphological types among the bacilli (rods).

microscopic and macroscopic and macroscopic groups. In general, they have rigid cell walls, lack chlorophyll, and are usually immotile. Most fungi reproduce asexually or sexually by means of spores though important budding groups such as yeasts are well known. Since fungi are classified by the pattern of their sexual structures, fungi whose sexual stages are unknown are placed into a group called *Fungi Imperfecti* and assigned genera on the basis of their asexual structures. They are further subdivided into the so-called lower and higher fungi. The lower fungi, *Phycomycetes,* are also called water molds but not all are aquatic (e.g. black bread molds). Some species multiply by means of flagellated gametes or flagellated spores i.e. more like certain green algae than other fungi; Most, but not all, Phycomycetes have

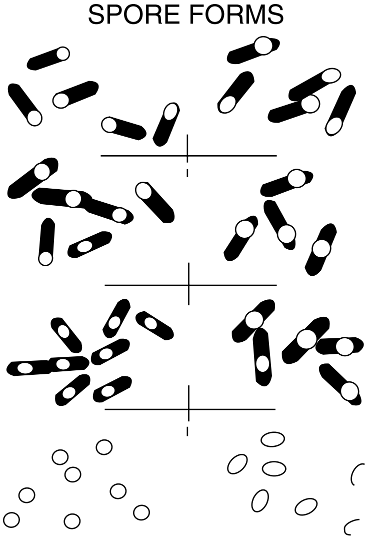


FIGURE 4 Diagrammatic representation of spores (clear areas) inside rod-shaped bacteria. Note (bottom row) that free spores may be ball or egg-shaped.

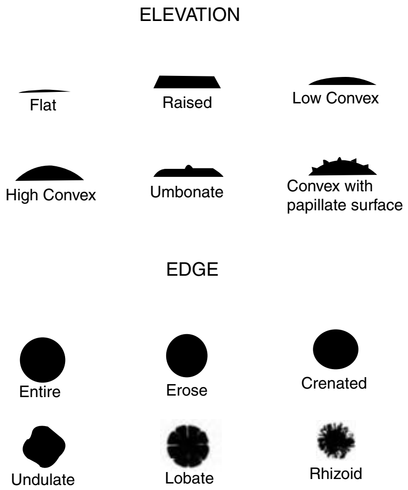


FIGURE 5 Diagrammatic representation of types of bacterial colonies. These shapes are specific for individual types and are therefore quite useful as a step in the process of identification of unknown organisms.

hyphae, microscopic cytoplasm-filled tube-like branches (lacking crosswalls), which together make a felty mat called a *mycelium.* Individual hyphae are microscopic but the mycelium, equivalent to a bacterial colony, is macroscopic. Growth takes place by extension of the hyphae. Specialized spore-containing bodies called *sporangia* can form at the ends of some hyphae. Sexual reproduction requires fusion of hyphae from two different mycelia to form a specialized *zygospore.*

It is more common now to discard the term Phycomyetes and instead subdivide the group into 4 classes in which affinities are much clearer. However, at present, the literature is divided in its use of the older and newer terminology. As with bacteria, chemical analyses of structures and metabolic pathways followed are important in defining the classes. These four classes are:

1) *Chytridiomycetes* lack true mycelia. They are aquatic, have posteriorly uniflagellated zoospores and cell walls composed of chitin.
2) *Hyphochytridiomycetes* have true mycelia. They are aquatic, have anteriorly uniflagellated zoospores and cell walls composed of chitin.
3) *Oomycetes* have true well developed mycelia and cell walls composed of cellulose.

   a) *Saprolegniales* are generally aquatic and have asexual spores on specialized mycelear structures. Only male gametes are motile.
   b) *Peronosporales* are generally terrestrial. Sporangia either produce asexual zoospores or may germinate directly to form hyphae. Both gametes are nonmotile.

4) *Zygomycetes* are terrestrial and have large and well developed mycelia and nonmotile spores. Asexual spores are produced in sporangia. Cell walls are made of chitosan or chitin.

There are two classes included in the higher fungi.

1) The *Ascomycetes* are the best known and largest class of fungi. Ascomycetes have hyphae divided by porous crosswalls. Each of these hyphal compartments usually contains a separate nucleus. Asexual spores called *conidia,* form singly or in chains at the tip of a specialized hypha. The sexual structure called *ascus,* is formed at the enlarged end of a specialized fruiting structure and usually contains eight *ascospores.* Some important microscopic members of this group include yeasts, mildews, the common red bread mold and many species which produce antibiotics. On the other hand macroscopic forms include *Morchella esculenta* or morels which are highly regarded as a delicacy by gourmets.
2) The *Basidiomycetes* are entirely macroscopic and are commonly known as mushrooms and toadstools.

Slime Molds

The slime molds, *Myxomycetes,* are at times classified with either true fungi or protozoa or, as here, treated separately. They produce vegetative structures which look like ameboid

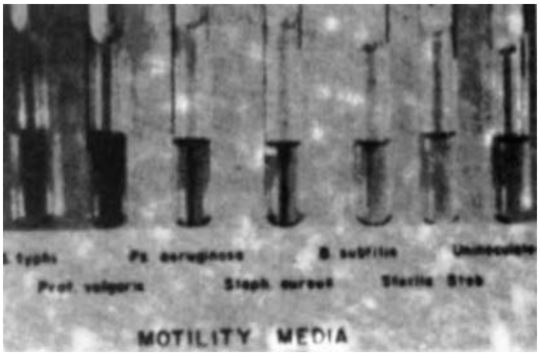


FIGURE 6 Bacterial motility. Motility is tested by stabbing an inoculated needle into a tube of very viscous growth medium. The motile organisms (*S. typhi* and *P. vulgaris*) grow away from the stab mark.

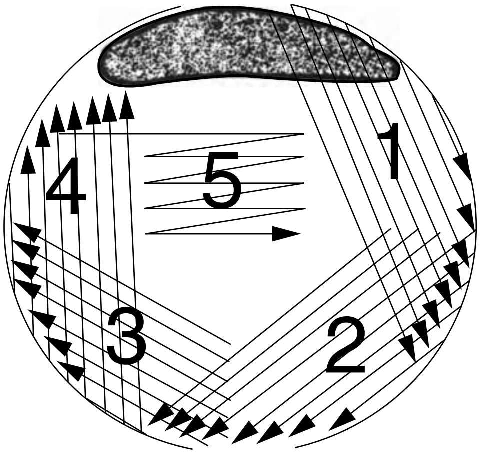


FIGURE 7 Isolation of single bacterial colonies on agar plates by dilution streaking. A diagrammatic representation of method of streaking inoculated needle across nutrient-containing plate. Stippled area is the primary inoculation. The inoculation needle is then flamed to sterilize and is then drawn across the stippled areas as indicated for area 1. The needle is then resterilized and drawn across area 2, etc.

protozoa and fruiting bodies which produce spores with cell walls like fungi. There are two major subdivisions (a) Cellular and (b) Acellular. They both primarily live on decaying plant material and can ingest other microorganisms, such as bacteria, phagocytically. Both have life cycles, but that of the acellular slime molds is more complicated.

Cellular slime molds have vegetative forms composed of single ameboid cells. Cyclically, ameboid cells aggregate to form a slug-shaped *pseudoplasmodium* that begins to form *fruiting bodies* when the slug becomes immotile. Spores are finally produced by the fruiting bodies.

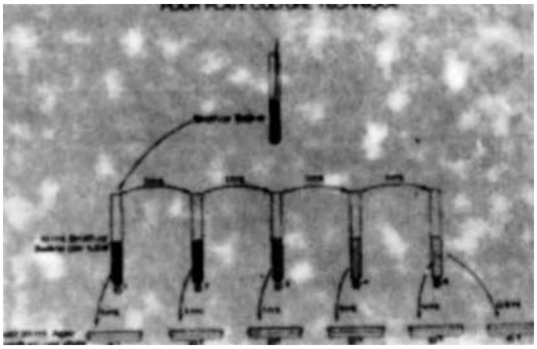


FIGURE 8 Isolation of single colonies by pour plate technique.

Acellular slime molds have vegetative forms called *plasmodia* which are composed of naked masses of protoplasm of indefinite size and shape and which travel by ameboid movement (protoplasmic streaming). Two kinds of nesting structures are produced: *fruiting bodies* (part of the sexual cycle) and *sclerolia.*

Protozoa

The last major group of microorganisms are the protozoa. As already stated, it is very hard to distinguish plants from animals at this primitive stage in evolution where organisms have some attributes of each. Most workers therefore are less interested in whether protozoa should be claimed by botanists or zoologists as they are in studying the group as the root of a phylogenetic tree which gave rise to clearly separable plants and animals. Protozoa range in size from that of large bacteria to just visible without a microscope. They have a variety of shapes, multiplication methods and associations which range from single cells to specialized colonies. They are variously found in fresh water, marine, terrestrial, and occasionally, aerial habitats. Both freeliving and parasitic forms are included. Most are motile but there are also important nonmotile forms. The protozoa are divided into four subphyla (I–IV).

I. *Sarcomastigophora* include forms which have either flagella, pseudopodia or both. Usually a single-type of nucleus (though opalinids contain multiples of this one type) is present except in development stages of a few forms. Asexual reproduction by binary fission is common. One whole class contains chloroplasts and are claimed by both protozoologists and algologists (they are considered here in detail with the eukaryotic algae). Many important parasites of diverse animal and some plant groups are found here. Sexual reproduction is present in a few forms.

The Sarcomastigophora are divided into three superclasses.

A. Mastigophora (*flagellates*) Are further sub-divided into *Phytomastigophorea* or plant-like flagellates (see eukaryotic algae) and *Zoomastigophorea* or animal-like flagellates which are divided into nine orders.

1) *Choanoflagellida* have a single anterior flagellum surrounded posteriorly by a collar. Some forms are attached to substrates. They are solitary or colonial and are all free-living.
2) *Bicosoecida* have 2 flagella (one free, the other attached to the posterior of the organism). They are free-living.
3) *Rhizomastigida* have pseudopodia and 1–4 or more flagella. Most species are free-living.
4) *Kinetoplastida* have 1–4 flagella and all have a kinetoplast (specialized mitochondrion). Many important pathogens (e.g. trypanosomes) and some free-living genera are included.
5) *Retortamonadida* have 2–4 flagella. The cytostome is fibril-bordered. All are parasitic.
6) *Diplomonadida* have 2 karyomastigonts, each with 4 flagella and sets of accessory organelles. Most species are parasitic.
7) *Oxymonadida* have one or more karyomastigonts, each with 4 flagella. All species are parasitic.
8) *Trichomonadida* have mastigont systems with 4–6 flagella. Some have undulated membranes. Many important pathogens (e.g. *Trichomonas*) are included.
9) *Hypermastigida* have mastigont systems with numerous flagella and multiple parabasal apparatus. All are parasitic. Some forms reproduce sexually.

B. Opalinata Are an intermediary group related to both ciliates and flagellates and are entirely parasitic. Opalinics have many cilia-like organelles arranged in oblique rows over their entire body surface. They lack cytosomes (oral openings). They have multiple nuclei (ranging from 2 to many) which divide acentrically. The whole organism divides by binary fission. Life cycles are complex.

C. Sarcodina Or ameboid organisms have Pseudopodia which are typically present but flagella may be present during certain restricted developmental stages. Some forms have external or internal *tests* or skeletons which vary widely in type and chemical composition. All reproduce asexually by fission but some also reproduce asexually. Most species are free-living (in both aquatic and terrestrial habitats) but some are important pathogens; for example, *Entameba histolytica*, which causes amebic dysentary. The sarcodinids are further divided into three classes.

1) *Rhizopodae,* a free-living, mostly particle-eating (phagotrophic) group which includes both naked and shelled species. The specialized pseudopodia are called lobopodia, filopodia, or reticulopodia.
2) *Piroplasmea.* These parasitic small, piriform, round, rod-shaped or ameboid organisms do not form spores, flagella or cilia. Locomotion is by body-flexing or gliding. They reproduce by binary fission or schizogony.
3) *Actinopodea* are free-living, spherical, typically floating forms with typically delicate and radiose pseudopodia. Forms may be naked or have membraneous, clutenoid, or silicated tests. Both asexual and sexual reproduction occurs. Gametes are usually flagellated.

II. *Sporozoa* typically form spores without polar filaments and lack flagella or cilia. Both asexual and sexual reproduction takes place. All species are parasitic. Some have rather complicated life cycles.
The Sporozoa are divided into three classes:

A. Telesporea Can reproduce sexually or asexually, have spores, move by body flexion or gliding and generally do not have pseudopodia.
B. Toxoplasmea Reproduce asexually, lack spores, pseudopodia or flagella, and move by body flexion or gliding.
C. Haplosporea Reproduce asexually and lack flagella. They have spores and may have pseudopodia.

III. *Cnidospora* have spores with one or more polar filaments and one or more sporoplasms. All species are parasitic. There are two classes.
IV. *Ciliophora* have simple cilia or compound ciliary organelles in at least one stage of their life cycle. They usually have two types of nucleus. Reproduction is asexually by fission or sexually by various means. Most species are free-living but parasitic forms are known.

ENERGY AND CARBON METABOLISM

All cells require a source of chemical energy and of carbon for building protoplasm. Regardless of whether the cell type is prokaryote or eukaryote or whether it is more plant-like or more animal-like, this basic requirement is the same. The most basic division relates to the source of *carbon* used to build protoplasm. Organisms which can manufacture all their carbon-containing compounds from originally ingested inorganic carbon (CO_2) are called *autotrophs* while those which require ingestion of one or several organic compounds for use in the manufacture of cellular carbon compounds are called *heterotrophs.* Some organisms are nutritionally versatile and may operate either as autotrophs or heterotrophs and are therefore referred to as *facultative-autotrophs* or *facultative-heterotrophs* (depending upon which mode of nutrition usually predominates).

Autotrophs are further divided according to the manner in which they obtain energy. *Chemoautotrophs* (also called chemotrophs or *chemolithotrophs* oxidize various inorganic compounds to obtain energy while *photoautotrophs* (also called phototrophs or *photolithotrophs*) convert light to chemical energy *via* the absorption of light energy by special pigments (chlorophylls and carotenoids). In both cases, chemical energy is stored in the form of chemical bond energy in the compound adenosine triphosphate (ATP).

When bonds of ATP indicated by ~ are broken, a considerable amount of energy is released. This ~ bond cleavage energy operates the biological engines: it is the universal chemical power which operates in all cells, autotroph or heterotroph.

Chemolithotrophic nutrition is only used by certain true bacteria. These bacteria are of ecological importance in that they are used to convert one form of nitrogen to another (i.e. in the nitrogen cycle) or industrially to oxidize low grade metallic or non-metallic ores. There are six bacterial groups which are chemolithotrophic.

1) The ammonia oxidizers such as *Nitrosomonas*, *Nitrosococcus*, *Nitrosocystis*, *Nitrosogloea* and *Nitrosospira*.
   One scheme for ammonia oxidation had hydroxylamine as an obligate intermediate and has been proposed for *Nitrosomonas.*
2) The nitrite oxidizers such as *Nitrobacter* and *Nitrocystis.* One proposed scheme for nitrite oxidation for *Nitrobacter* is:

$$NO_2^- \xrightarrow[\text{reductase}]{\text{cytochrome}} \text{cytochrome C} \xrightarrow[\text{oxidase}]{\text{cytochrome}} O_2$$

ATP ADP

3) Hydrogen oxidizers *Hydrogenomonas.* One proposed hydrogen oxidation scheme is:

   $H_2 \rightarrow 2H^+ + 2e \rightarrow$unknown$\rightarrow$flavor protein compound

   $\rightarrow$ubiquinone$\rightarrow O_2$
   cytochrome
   b compex

   Nicotinamide adenine$\rightarrow$menadione

   $\rightarrow$cytochrome C$\rightarrow$cytochrome a
   $\rightarrow O_2$ dinucleotide (NAD)
4) Ferrous compound oxidizing bacteria such as *Ferrobacillus* and *Thiobacillus ferroxidans.*

   One proposed ferrous oxidizing scheme for

   F. ferrooxidans is:

   $4FeCO_3 + O_2 + 6H_2O \rightarrow 4Fe(OH)_3 + 4CO_2$

5) Methane oxidizers such as *Methanomonas methanooxidans* and *Pseudomonas methanica* are common in the upper layers of marine sediments and soil.

   Methane is oxidized in the following manner:

   $CH_4 \rightarrow CH_3OH \rightarrow HCHO \rightarrow HCOOH \rightarrow CO_2$

6) The sulfur-compound oxidizing bacteria *Thiobacillus.*

Four pathways for oxidation of thiosulfate ($S_2O_3^{+2}$) by different *Thiobacillus* species are known. These are:

a) $6Na_2S_2O_3 + SO_2 \rightarrow 4Na_2SO_4 + 2Na_2S_4O_6$
   $2Na_2S_4O_6 + 6H_2O + 7O_2 \rightarrow 2Na_2SO_4 + 6H_2SO_4$
b) $Na_2S_2O_3 + 2O_2 + H_2O \rightarrow Na_2SO_4 + H_2SO_4$
c) $5Na_2S_2O_3 + H_2O + 4O_2$
   $\rightarrow 5Na_2SO_2 + H_2SO_4 + 4S$
   $2S + 3O_2 + 2H_2O \rightarrow 2H_2SO_4$
d) $2Na_2S_2O_3 + H_2O + 1/2O_2 \rightarrow Na_2S_4O_6 + 2NaOH$

Photolithotrophic nutrition is used by photosynthetic bacteria, blue green algae and eukaryotic algae. The general reaction in which both utilization of CO_2 (carbon dioxide fixation) and energy generation is summarized is:

$$CO_2 + H_2A \xrightarrow{nv} (CH_2O) + 2A + H_2O$$

Where A is either oxygen for all eukaryotic algae and the prokaryotic blue-green algae ($H_2A = H_2O$), or sulfur for green sulfur bacteria, *Chlorobacteriaceae,* and purple sulfur bacteria, *Thiorhodaceae* ($H_2A = H_2S$) or any one of several organic compounds for nonsulfur purple bacteria, *Athiorhodaceao* ($H_2A = H_2$-organic compound which is oxidizable).

Both green and purple sulfur bacteria are obligate anaerobes whereas the non-sulfur purple bacteria are facultative anaerobes (they are anaerobic when growing heterotrophically). In all cases, photosynthetic organisms operate by the initial transduction of light to chemical energy. In this transduction, chlorophyll + light quanta $Ch1^+$ (excited chlorophyll) + e^- (electron driven off of Ch1). Many such events take place simultaneously and electrons released during these reactions migrate through the photosynthetic unit to the reaction center and transfer energy to a special *reaction-center chlorophyll.* At the reaction center, a charge separation of the oxidant and reductant occurs. Electron flow after this event differs in photosynthetic bacteria as compared with algae and higher plants (Figures 9 and 10). In addition, differences in photosynthetic ability exist among organisms based upon the absorption maxima of their light-transducing pigments (primarily chlorophylls). The combination of light intensity, wavelength of available light, wavelength of operation of principal energy transducing pigment, degree of aerobiasis, and availability of oxidizable compound (H_2O, H_2S, or H_2-organic compound) all influence the efficiency of photosynthesis. These factors should be borne in mind when one looks for the ecological niche occupied by these various organisms.

Ecology of Microorganisms

One should understand the physiological requirements of microorganisms *before* investigating the effects of environmental changes on the distribution and activity of diverse

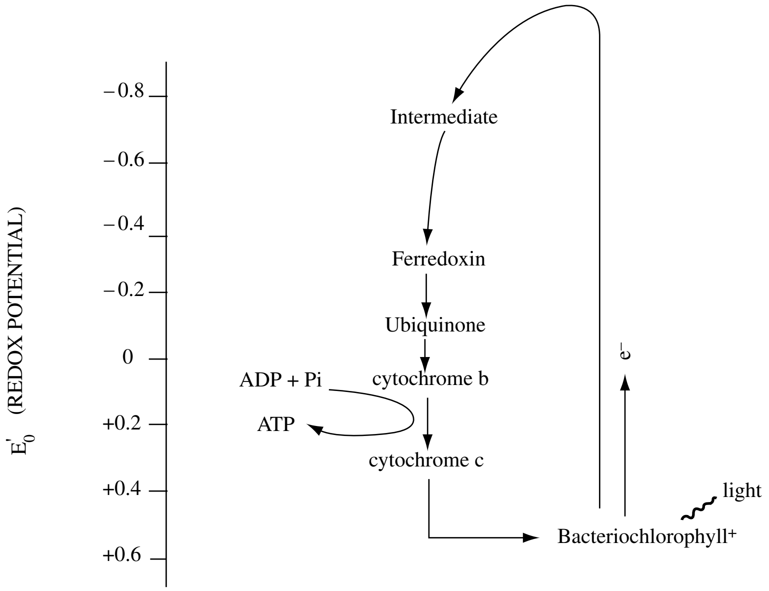


FIGURE 9 Electron flow in bacterial photosynthesis.

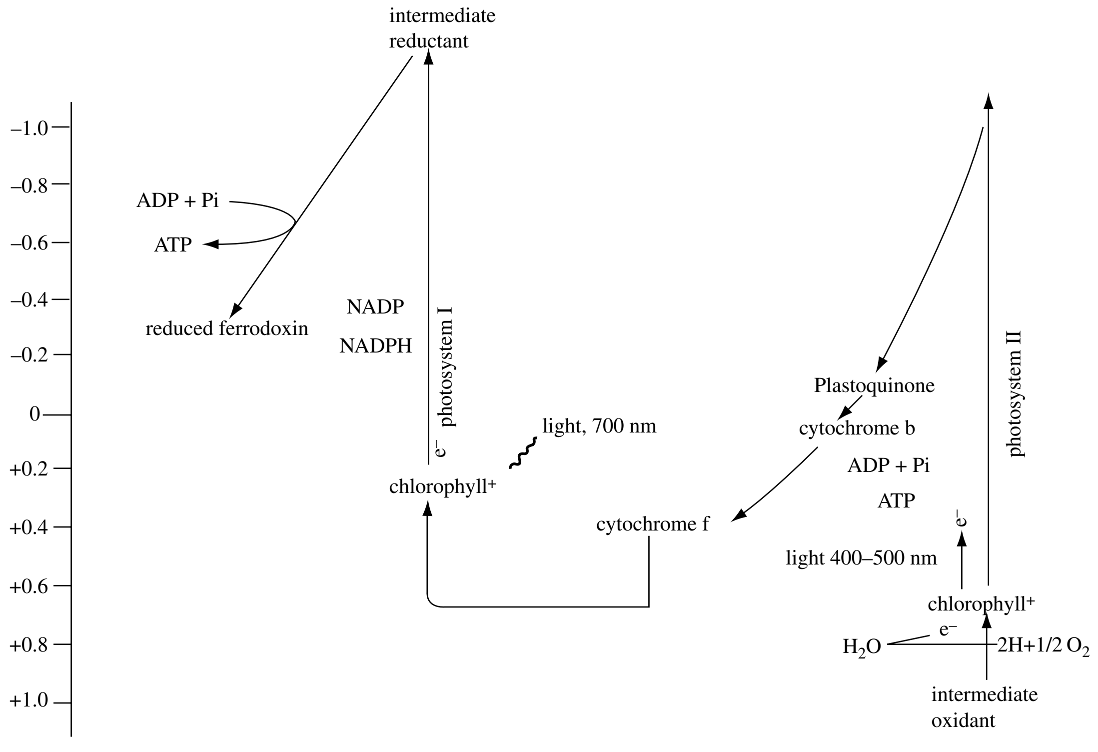


FIGURE 10 Electron flow in algal and higher plant photo-synthesis.

microbial types, their interactions and their relationships to higher plants and animals. It is important to note that the "natural" balance may be undesirable. Thus studied efforts to change these bal ances would be quite desirable. The most important precaution to observe relates to the ancillary consequences of these changes, i.e. do the changes produce side effects which may be as unappetizing as the original condition.

The development of techniques which form the bases for studying microbial ecology comprise an important chapter in

TABLE 4
Absorption maxima of chlorophylls from various sources

Organism	Chlorophyll type	Principal absorption maxima in nm
Green sulfur bacteria	Bacterial Chl_c	660
	Bacterial Chl_d	650
Purple sulfur bacteria	Bacterial Chl_a	820
	Bacterial Chl_b	1025
Non sulfur purple bacteria	Bacterial Chl_a	820
	Bacterial Chl_b	1025
Green algae and Euglenids, higher plants	Chl_a	683
	Chl_b	650
Diatoms, brown algae	Chl_a	683
Pyrrophyta	Chl_c	620
Xanthophyta	Chl_a	682
	Chl_e	
Cyano-, Chryso- and Rhodophyta	Chl_a	683

classical microbiology. These laboratory methods, pioneered by Winogradsky (1856–1953) and Beijerinck (1851–1931) and refined by others, utilize specialized, restrictive, physical and chemical conditions to select and cause to predominate one or few types of organisms from a highly diverse mixture. Hence this method is termed *selective enrichment culture.* Once one understands how to manipulate these laboratory systems, it is easier to analyze field observations in which specific conditions which result in microbial changes can be recognized and, if necessary, altered. We will first explain the principles of selective enrichment techniques and then look at the natural distribution of microorganisms and their relationship to higher plants and animals. Highlights of microbial characteristics which are useful taxonomically have been described in the various sections listed under *Physical Characteristics of Microorganisms.*

Selective Enrichment Methods

To determine whether a given sample of soil, water, or air contains microorganisms capable of living under a particular set of conditions, one prepares a growth medium which is selective for a property peculiar to those conditions. For example, if organisms which can fix atmospheric nitrogen are sought, all non-atmospheric sources of nitrogen (such as nitrites, nitrates, ammonia, amino acids) are eliminated from the growth medium. If organisms which obligately fix carbon dioxide are required, all nonatmospheric (organic) sources of carbon are eliminated from the growth medium and, frequently, additional CO_2 is bubbled through the medium.

On the other hand, it might be of particular interest to determine if a certain weed-killer is *biodegradable* before it is used under field conditions. Many different approaches to this important problem are possible. Thus more than one mode of attack is described. The approach which is closest to the general principle of revealing (selectively enriching for) a minor population of desired microbial type among a multitude of undesirable organisms makes use of an enrichment medium in which the weed-killer is used as either (a) the only source of organic carbon and nitrogen, (b) the only source of organic carbon though other sources of nitrogen are present, or (c) the only source of nitrogen though other sources of organic carbon are present. Subsequent microbial growth indicates biodegradability. The organisms may be isolated and used as seed cultures for a percolation system which is fed by the aqueous runoff from a field which was treated with the weed-killer. Thus the more public waterways fed by aqueous effluents from treated fields would not be polluted by potentially-toxic agricultural chemicals.

In order to prepare selective enrichment media, one needs to provide the microorganisms with all their nutritional requirements in proper proportions. Insufficient quantities will not support growth and excesses are frequently toxic. In addition conditions must be biased in some way to insure that most of the undesired organisms will not grow at all or will grow appreciably slower than the desired organisms. It should be recognized that it is rare for any single enrichment to select out only one microbial species. Thus further purification steps are required if one wishes to isolate only one species uncontaminated with other living things. An uncontaminated, single membered culture is called a *pure* or *axenic* (a = absence of, xenos = strangers) culture. The common nutritional requirements of microorganisms, the quantities in which they must be supplied, and the biological uses of each substance are shown in Table 5. The selective enrichment techniques to be described are most frequently used for the isolation of bacteria, yeast, and certain prokaryotic algae. An outline of selecting properties is given in Table 6.

A demonstration of factors involved during the natural selection which takes place under field conditions is shown in Figure 11. These models are called *Winogradsky columns*. The variety of microbial life which develops over a period of 2–10 days is determined by (a) the degree of acidity or alkalinity of these natural growth media, (b) the nutrients contained in the liquid and solid phases, and, of course, (c) the initial populations of microorganisms. In the examples shown, light is provided to ensure growth of photosynthetic organisms.

If the basic principles of the Winogradsky column are to be used to reveal the microbial population in a particular soil or water sample, then the column and all its components are first sterilized and then inoculated with a *nonsterile* soil or water sample. The microbial population of the sample will develop in the portions of the column which provide the proper physical and chemical conditions.

An important application of Winogradsky columns can be made for testing various chemicals for potential ecological changes. The chemical agent is percolated through a soil column or is simply added to a predominantly liquid column either when the column is started or after its microbial population has developed. Significant changes in the column's normal population (diversity or population density) is indicative of toxicity to one or more types of microorganisms. The profound changes in populations of higher plants and animals due to disruption of the balance of microbial life can be readily appreciated when the cyclic nature of nitrogen and sulfur dissimilation is considered (Figures 12 and 13). These interdependences emphasize the key role played by microorganisms in maintaining the balance of soil nutrients.

Use of the Winogradsky Column for Testing Biodegradability Capacity of a Natural Soil or Water Body

In the last section, an example was given in which one of these bodies with little capacity for biodegradation of a weed-killer can be controlled so that the body in question does not spread the potential pollutant to a bordering body. We now consider a method for pretesting the biodegradation test. One would hope that this or a parallel test would become standard before new agricultural chemicals are marketed. That is, chemicals which are not decomposed *before* they leave the immediate land or water body in which they are used would not be marketed or would only be marketed *after* controls against accumulation of the chemical had been worked out.

The Winogradsky column test for ability of a potential soil or water body to degrade a potenlially dangerous chemical such as a weed-killer consists of (a) preparing a standard column composed of soil or water from the body in question, and (b) after the column has been allowed to develop its natural population (c) an isotopically labelled version of the weed-killer can be added in the concentration (and 10 × the concentration that the weed-killer is to be used). After a time equivalent to that in which the weed-killer is expected to remain in the natural body (i.e. account for flushing time from rain or water currents), the column is tapped by elution with water or buffer, and the effluent is analyzed for per cent undergraded weed-killer as well as *the nature of the degradation products.* The latter point is particularly important

TABLE 5

Nutritional requirements of typical heterotrophic microorganisms with limited synthetic capacity

Type compound	Example	Quantity in typical medium (%)
Energy source	Glucose, sucrose, glutamic acid, succinic acid	0.1−2.0% 0.1−0.5%
Synthesis of protein and fat	sodium acetate	0.01−0.1%
	Lecithin amino acids	0.001−0.005% 0.002−0.1%
Synthesis of nucleic acids	purines and pyrimidines	0.0005−0.002%
Coenzymes	Vitamins	0.0001−0.005%
Major inorganic requirements	PO_4, Mn^{++}, Mg^{++}, Na^+, K^+	0.01−0.05%
Minor inorganic requirements	Ca^{++}, Co^{++}, Zn^{++}, Fe^{+++},Cu^{++}, Cl^- NH_4^+ or NO_3^-	0.001−0.01%
Water	All the above are prepared in aqueous solution	

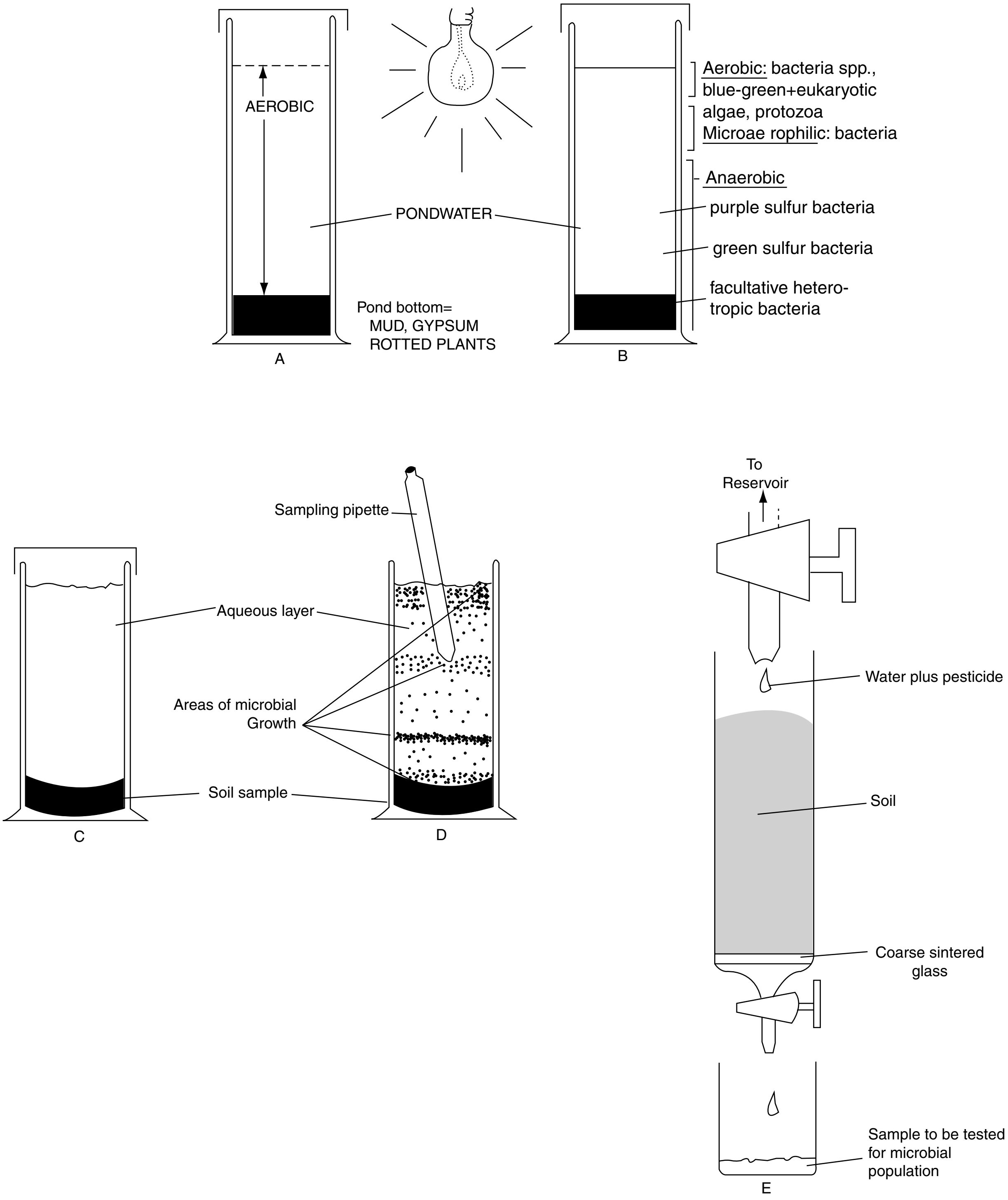


FIGURE 11 The Winogradsky column. A laboratory model for microbial population development in natural environments.

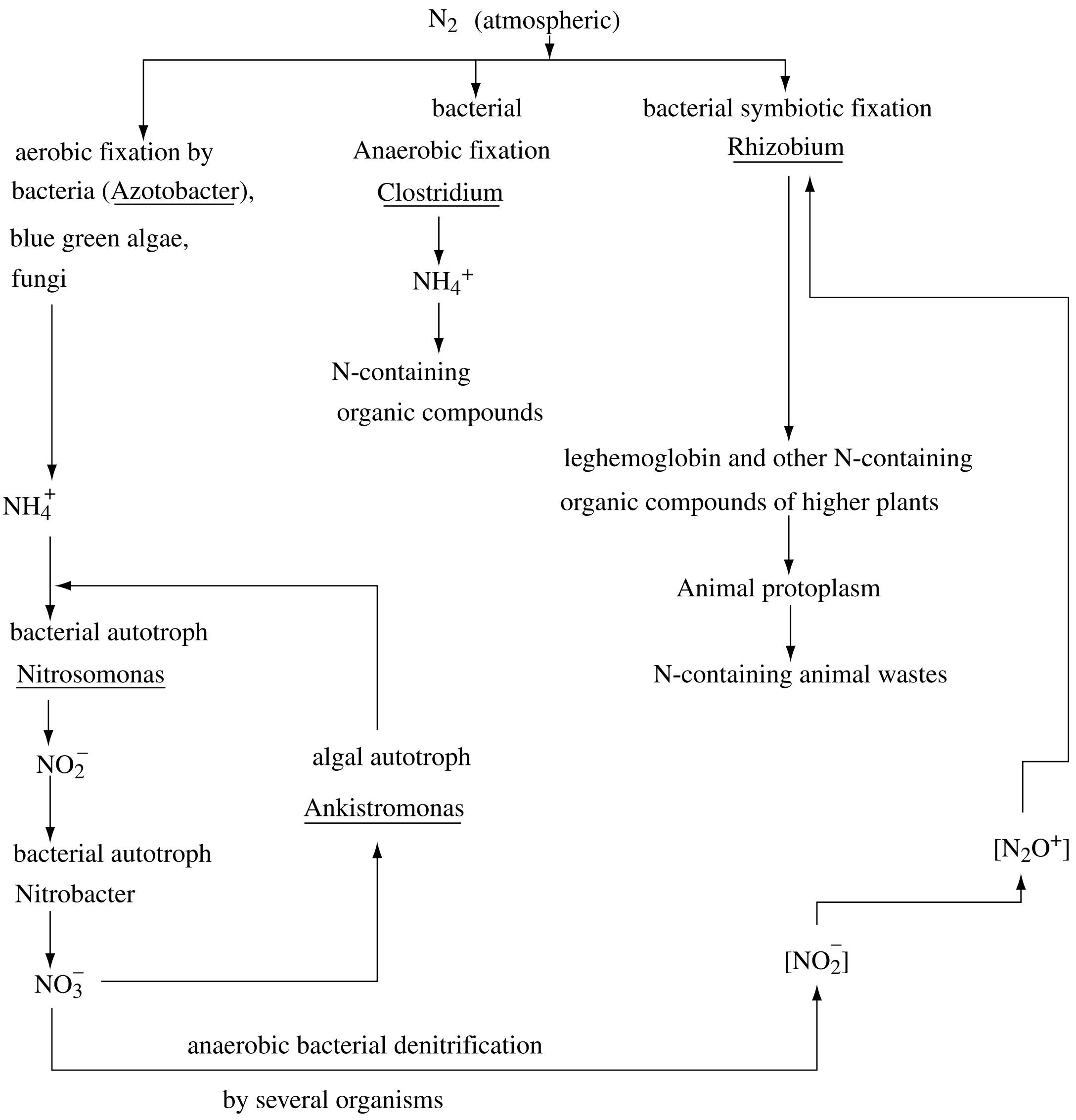


FIGURE 12 The nitrogen cycle presented in a generalized way so that the role of various microorganisms are indicated as well as the relationship to higher plants and animals. The precise arrangement will vary according to several physical parameters of the natural environment.

since the degradation products may themselves be noxious. Thus, the only degradation products which are acceptable are those which may enter normal metabolism or are labile enough to be further degraded to metabolizable compounds by the physical conditions in the body in question.

Determination of Gross Populations of Microorganisms

Quantitative sampling techniques are required for the determination of the microbial population in air, water, or soil samples. The method used for sample collection must ensure against (a) loss of more than a trivial number of microorganisms and (b) cross-contamination from other sources during sample transport and laboratory manipulation. Once the sample has arrived in the laboratory, selective enrichment techniques can be used to reveal the diversity of microorganisms or mixed populations can be counted by using some variation of the pour plate technique (see Figure 8).

A. *Air sampling techniques* Non-spore forming organisms are rarely found in air samples because they are too delicate to survive for long in the generally dehydrating conditions of atmospheric transport. A convenient sampler consists of a sterile membrane filter connected to a metering vacuum pump. At the beginning of the sampling period, the

TABLE 6

Selecting factors for the isolation of specific types of microorganism

- Organic compounds *present* in medium HETEROTROPHIC ORGANISMS (see 6B)
- Organic compounds *absent* from medium AUTOTROPHIC ORGANISMS (see 6A)

6A *Selecting Factors for Autotrophic Organisms*

- Anaerobiasis
  - Light absent, S or $S_2O_3^-$ oxidized for energy: anaerobic chemoautotrophic bacteria such as *Thiobacillus denitrificans*
  - Light present
    - grows in high sulfide concn: Green sulfur bacteria
    - grows in low sulfide concn: Purple sulfur bacteria
- Aerobiasis
  - Light and sulfide absent
    1. NH_4^+ is N source *Nitrosomonas* (bacterium)
    2. NO_2^- is N source, *Nitrobacter* (bacterium)
    3. NH_4^+ plus special atmosphere composed of 85% H_2; 10% O_2; 5% CO_2 at pH 7 Hydrogen bacteria
    4. NH_4^+ plus $S_2O_3^=$ at pH 7. Certain *Thiobacillus* spp. (bacteria)
  - Light present and sulfide absent
    - N_2 can serve as sole nitrogen source: Blue-green algae
    - Nitrogen must be present in combined form: Green algae

6B *Selecting Factors for Heterotrophic Organisms*

- Anaerobiasis
  - Light present, neutral pH: Non-sulfur purple bacteria
  - Light absent
    1. Acid pH, fermentable sugar present Yeast spp.
    2. Neutral pH
       - fermentable sugar
         - N_2 as sole nitrogen source *Clostridium* species (bacteria)
         - Presence of combined nitrogen source. Numerous microbial types.
       - non-fermentable substrate, electron acceptor is 1, 2 or 3
         1. NO_3 Denitrifying bacteria
         2. CO_2 Methane bacteria
         3. $SO_4^=$ *Desulfovibrio* (bacterium)
- Aerobiasis (see below)

- Aerobiasis
  - Light present Many algae will grow here, pH aids in further differentiation as does the precise selection of organic components
  - Light absent
    - N_2 is sole nitrogen source *Azotobacter* (bacterium)
    - Presence of combined nitrogen source. Depending upon the complexity of the medium and pH, most aerobic bacteria yeast, protozoa and some algae could grow. Specific texts should be consulted for further details.

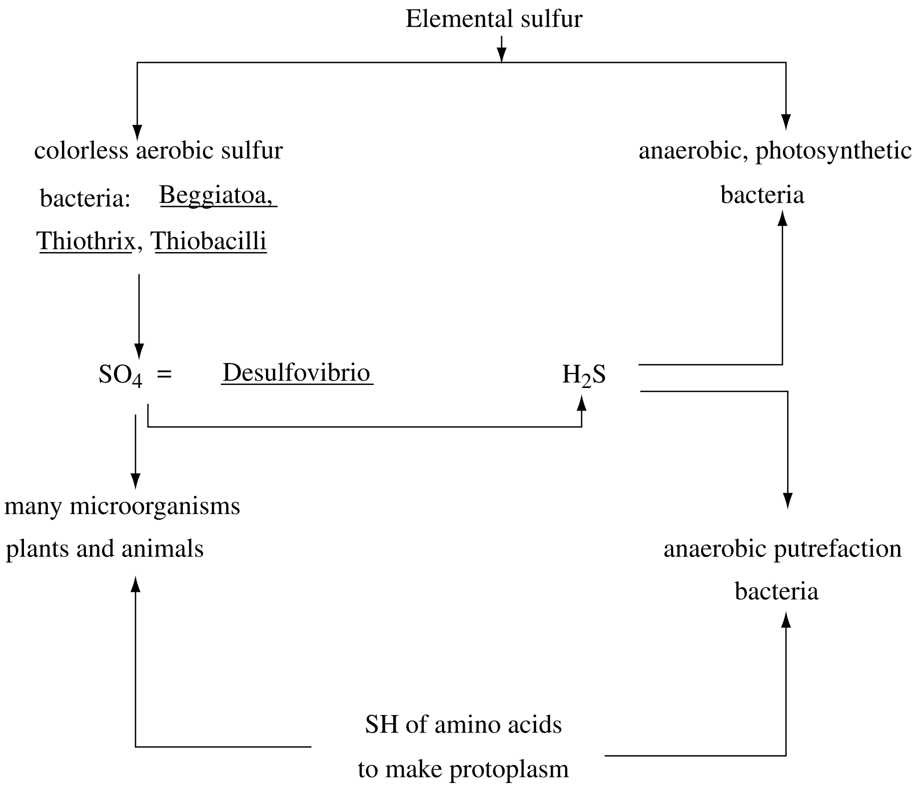


FIGURE 13 The sulfur cycle. Microorganisms are becoming increasingly useful for processing low-grade sulfur-containing ores. Sometimes overlooked is the natural cyclic distribution of sulfur-containing compounds in which microorganisms play principle roles.

membrane is exposed and the desired amount of air passed through it. Air-borne microorganisms are trapped on or in the membrane. Some discrimination among organisms can be made during sampling by using membranes with graded porosities. In this way organisms are segregated according to their cross-sectional dimensions. To reveal the organisms impinged upon the membrane, the membrane is treated as inoculum for the various selecting media.

B. *Water sampling techniques* Any container which can be sterilized and can be opened and closed by remote signal can be used. It should be recognized that pressure changes after retrieval may influence viability of organisms collected at great depths.

C. *Soil sampling techniques* A sterilized coring device is usually used to ensure against contamination with airborne organisms. The soil samples are handled aseptically and weighed amounts are tested for their microbial population.

REFERENCES

1. Environmental Microbiology, in The Natural Environment and the Bioecochemical Cycles, O. Hutzinger, ed., Springer Verlag, Heidelberg, 1985.
2. Saunders, V.A. and Saunders, J.R., Microbial Genetics Applied to Biotechnology, Croom-Helm, London, 1987.
3. Zehnder, A.J.B., Biology of Anaerobic Microorganisms, Wiley Interscience, New York, 1988.
4. Fleckenstein, L.J., Release of Genetically Engineered Microorganisms to the Environment, Resource Management and Optimization, Vol. 6, no. 4, 1989.

HELENE N. GUTTMAN
US Department of Agriculture

MINE DRAINAGE: see POLLUTION FROM MINE DRAINAGE

MOBILE SOURCE POLLUTION

Mobile sources, alternatively called transportation or vehicular sources, include cars, trucks, buses, ships and various aircraft. Air pollutants emitted will vary, depending on the fuel being combusted or reacted (in the case of fuel cells or batteries) and the engine design of each vehicle.

THE AUTOMOBILE

The automobile's discovery appears to satisfactorily combine a human desire for rapid transportation with the desire for independence and flexibility. However, the increasing vehicle population poses a series of threats to continued physical and psychological well-being and to the future of our environment.

As the automotive industry expands, other auxiliary industries such as petroleum production and concrete and tire manufacturing also grow, with additional potential for pollution problems. Additional roads also have to be built, with a negative impact on both ecology and landscape. A balance must be made between the right of an individual to use his own car when and where he drives and the harm brought upon society as a whole by his doing so. If it is accepted that society needs to be protected, a number of legislative and economic measures can be initiated to discourage automotive usage.

Legislation has been enacted to limit the emission of HC (hydrocarbons), NO_x (oxides of nitrogen), CO (carbon monoxide) and particulate matter including lead compounds. However, other waste materials such as tires and the automobile itself (including the repugnant abandoned cars) must be disposed of.

The broad approach to automotive pollution control is to encourage alternative means of transportation. This would include improvements in mass transit such as high-speed trains, moving sidewalks, increased links and modernization. In the city, bicycle riding and walking are low pollution, high exercise alternatives; in rural areas car pools might be formed.

Cars have been built using different energy concepts (ex. battery, turbine engine, sterling engine) and to run on different fuels (ex. natural gas, alcohol). Special control devices (ex. catalytic reactors, afterburners) placed at the exhaust of internal combustion engines are the primary means of reducing emissions of various materials. The total solution to the problem will probably combine technical and strategic methods.

The US EPA and US DOE produce the Fuel Economy Guide to help car buyers choose from the best fuel gallons per mile (mpg) ratings for both city and highway traffic modes. A pdf version may be found at http://www.epa.gov/cgibin/epaprintonly.cgi[1]

In the present discussion, we shall concentrate on the major moving source of pollutants, the internal combustion engine. In the automobile, evaporative losses of pollutants occur from the fuel tank and carburetor (ca. 5%), fumes from the crankcase (ca. 20%) and the exhaust system (ca. 75%). The major offenders are unburned hydrocarbons, carbon monoxide, nitrogen oxides and HC oxidants. Positive Crankcase Ventilation (PCV) on all modern cars reduces the emissions by pulling air and fumes into the engine by maintaining a vacuum at the engine. Of course a highly efficient combustion process will eliminate the partially oxidized substances. One reason for incomplete combustion is that as the "flame front" generated from the spark moves toward the relatively cool cylinder walls, a quenching action takes place preventing further reaction, the kinetics of which are very temperature sensitive. Other factors influencing incomplete combustion are improper dilution by poor cycle timing and less than the proper excess of oxygen admitted at the carburetor. Standard engine cycles have been developed so that pollutant guidelines might be drawn. The Federal Driving Cycle initiated in 1972 in considered the standard of vehicle testing. A federal short cycle Table 1 has been used for convenience in some instances. Measured emission values averaged for various test sites in representative US cities are presented in Table 2 for steady state at various miles per hour (mph) and for the short cycle. Carbon monoxide and NO_x levels increase as the mph level increases. The federal 1975 proposed standards for CO, hydrocarbons and NO_x are 11, 0.5 and 0.9 grams per mile, respectively, and 0.5%, 40 ppm and 225 ppm, respectively. Their extent of enforcement is increasing as of the early 1990s.

The CO standard is based primarily on tests which have shown that the capacity of blood to carry oxygen is hampered by CO absorption.[2] Brain function is retarded after an exposure of 10 to 15 ppm CO for several hours. The HC and NO_x levels are mainly based on their role in photochemical smog using ambient levels. Shy *et al.*[3] found in their Chattanooga study that "NO_2 alone and exposure to suspended particulate matter alone appear to be the most probable explanation for the observed excess in respiratory illness rates." The data are not completely convincing as pointed out by automobile industry researchers.[4] However, the U.S. government took the initiative for the welfare of the public in not waiting for all the details to be perfectly established. Certainly there are those among us (aged, with heart conditions, etc.) who will be more severely affected than the average by typical ambient pollutant concentrations and will need protection.

At least two approaches to the removal of internal combustion engine pollutants have come into wide acceptance. One is the improvement of the combustion process itself. This includes reducing manifold vacuum (and hence dilution of charge) between exhaust and intake steps; increasing the cylinder wall temperature; and designing for less maximum horsepower by minimizing the surface to volume ratio.

The other approach is controlling exhaust emissions by further reaction, either after-burning or catalytic. After-burning can be accomplished slightly downstream from the exhaust valve by additional oxygen injection converting HC and CO to CO_2 and water vapor. The after-burner has the disadvantage of not being able to remove NO_x. In fact, because of its high temperature, still more NO_x is formed. The catalytic approach is to find a catalyst or set of catalysts and temperature or set of temperatures which will completely oxidize both CO and HC, but also will reduce the NO_x. These two approaches will be discussed in detail below.

Meteorological and vehicle persistence factors have been developed for estimation of 1 hour carbon monoxide concentrations defined as worst-case total persistence factors.[33]

IMPROVEMENT OF THE COMBUSTION PROCESS

As mentioned above the explosion of fuel accounts for chemical pollutant formation. In the Rankine cycle engines, liquids may continually be vaporized and recondensed, and in the Sterling cycle engines a gas is repeatedly heated and cooled. Both cycles are accomplished in a sealed container and the heat used for the process comes from an open flame external to the engine. Very little pollution is generated in such a controlled rather than explosive flame.

Exhaust gas recirculation (EGR) is used on American Motors, Chevrolet and Chrysler automobiles. This includes a diaphragm-actuated flow control valve located between the exhaust and intake manifolds. The valve is operated by ported vacuum directed through hoses and a coolant temperature override. Permitting metered amounts of exhaust gases to enter the intake manifold, which are mixed with incoming fuel mixtures, lowers the combustion temperatures within the cylinders. Reducing maximum cylinder combustion temperatures minimizes the creation of Oxide of Nitrogen (NO_x). EGR operation does not take place until engine operating temperature has reached a preset level and engine load is sufficient to permit proper EGR operation.

Almost all autos now contain a PVC system which directs filtered air into the crankcase and channels vapors but toward the manifold leading to the combustion chamber. Fuel tank vapors are also concentrated by charcoal canisters in American Motors vehicles for recycled combustion.

An H.E.W. report[5] discusses engine modification systems, "Features shared by essentially all versions of the engine modification system include calibrated carburetors that provide (a) relatively lean air–fuel mixtures for idle and cruise operation and (b) higher engine idle speeds. Refined control of spark timing is also used, and, in some cases, regarded spark timing at idle is employed. In addition, many engines are fitted with special air cleaners and ducting designed to supply heated air at nearly constant temperature to the carburetor, to permit even leaner mixture settings. Most versions also incorporate high-temperature radiator thermostats to raise coolant temperatures, and thus improve mixture distribution and promote complete combustion. In some cases, higher capacity cooling systems are used to handle the additional cooling load at idle that results from wider throttle openings and retarded ignition timing during this operating condition. In addition, combustion chamber design attempts to avoid flame quenching zones where combustion might otherwise be incomplete, and result in high hydrocarbon emissions."

Hydrocarbon and CO emissions are reduced by adjusting the carburetor to a fuel-lean mixture during part throttle

TABLE 1
Federal short cycle

Mode					
No.	Type	Speed range	Time in mode (sec)	Average speed (mph)	Average acceleration rate (mph/sec)
1	Accel	0–16	6.0	8.00	2.67
2	Accel	16–29	23.0	22.50	0.57
3	Cruise	29	10.0	29.00	0.00
4	Accel	29–37	18.0	33.00	0.44
5	Accel	37–42	4.5	39.50	1.11
6	Decel	42–37	2.5	39.50	2.00
7	Decel	37–20	32.0	28.50	-0.53
8	Decel	20–0	7.5	10.00	-2.67
9	Idle	0	21.5	0.00	0.00

and idle operation. “Lean surge during cruise has been largely overcome through improvement in manifolding (better mixture distribution), better carburetor fuel-metering characteristics, higher coolant temperatures, increased heating of the air–fuel mixture, and, in some cases, provision for heating the incoming air to the carburetor.

Exhaust emissions of CO and HC are particularly difficult to control during engine idle and closed-throttle operation (deceleration). Considerable effort has gone into designing carburetor idle systems that will provide a lean air–fuel mixture and minimize emissions during these periods. To ensure that idle air–fuel mixture cannot be adjusted to be too rich (which would tend to increase CO and HC emissions appreciably), some means of limiting idle-mixture adjustment is used on most carburetors. Such devices allow idle mixture to be adjusted leaner than a predetermined value, but not richer.”

The effects of charge dilution on the exhaust emission of nitric oxide (NO) from a single-cylinder engine were evaluated over a range of engine design and operating parameters.[6] Nitric oxide emission decreased as much as 70% as charge dilution fraction was increased from 0.0065 to 0.164 due to increased valve overlap, external exhaust recirculation, and reduced compression ratio. NO emission was strongly dependent on charge dilution fraction, but was independent of the specific method used to change charge dilution. The combined effects of increased charge dilution and 10 degree spark retard reduced NO emission 90%. However, definite limits of operation were observed on the single-cylinder engine with high charge dilution.

The Ford Motor Company uses a system which reduces the hydrocarbon and carbon monoxide content of exhaust gases by continuing the combustion of unburned gases after they leave the combustion chamber. This is achieved by injecting fresh air into the hot exhaust stream leaving the exhaust ports. At this point, the fresh air mixes with hot exhaust gases to promote further oxidation of both the hydrocarbons and carbon monoxide, thereby reducing their concentration and converting some of them to carbon dioxide and water.

TABLE 2

Summary of results, steady state tests and federal short cycle vehicle exhaust emissions in grams per vehicle mile[a]

City	Effluent	Idle	15 mph	30 mph	45 mph	60 mph	FSC[b]
Los Angeles	HC	1.45	5.28	3.09	2.91	2.60	5.69
	CO	17.02	69.11	29.50	24.60	25.51	48.53
	CO_2	72.32	409.53	333.12	357.90	372.47	438.48
	NO_x	0.08	0.51	1.69	3.78	5.51	3.10
Denver	HC	1.47	6.06	3.74	3.94	3.87	7.07
	CO	17.20	86.49	52.13	56.43	71.13	94.68
	CO_2	60.54	345.65	297.92	330.09	368.91	348.85
	NO_x	0.12	0.83	1.53	3.07	4.52	2.13
Chicago	HC	1.30	5.47	3.48	3.59	3.62	5.74
	CO	15.49	66.57	29.77	30.53	32.63	59.74
	CO_2	63.85	358.70	318.61	357.65	417.11	392.28
	NO_x	0.10	0.87	2.36	4.37	6.26	3.42
Houston	HC	1.50	5.68	3.02	2.83	2.47	5.52
	CO	18.74	77.29	35.11	32.38	31.71	63.16
	CO_2	76.13	391.06	318.01	356.13	404.79	417.03
	NO_x	0.14	0.83	1.49	3.47	5.70	2.83
St Louis	CC	16.43	67.68	29.54	26.13	26.97	56.43
	CO_2	64.28	334.24	301.64	315.26	362.26	378.28
	NO_x	0.08	0.53	1.57	3.38	5.43	3.10
Washington	HC	1.12	4.23	2.59	2.57	2.88	4.46
	CO	13.25	56.10	26.19	25.40	25.79	48.11
	CO_2	65.04	380.44	343.54	390.83	452.50	417.79
	NO_x	0.13	1.04	2.88	6.04	8.90	4.23
All except	HC	1.34	5.11	2.99	2.90	2.85	5.34
Denver	CO	16.19	67.36	30.02	27.79	28.50	55.15
	CO_2	68.35	374.23	323.03	355.55	401.60	408.96
	NO_x	0.11	0.75	2.00	4.21	6.35	3.34

[a] Idle results in grams per minute. NO_x not corrected for humidity.

[b] FSC—Federal Short Cycle.

EMISSION CONTROL DEVICES

Additional Combustion

An afterburner is an additional baffled tubular reactor in which the gases are reignited and burned to completion. An air pump provides the necessary oxygen-rich mixture and the heat of reaction maintains a high temperature to speed its completion.

It is not necessary in all cases to have a separate afterburner. The aforementioned report[5] states, "An injection systems decrease exhaust CO and HC emissions by injecting air at a controlled rate and at low pressure into each exhaust port. Here, the oxygen in the air reacts with the hot exhaust gases, resulting in further combustion of the unburned hydrocarbons and CO that would otherwise be exhausted to the atmosphere. Optimum reduction of emissions by this method depends on proper air injection rates over a wide range of engine operating conditions, carefully tailored air-fuel mixture ratios and spark advance characteristics, and in some cases the use of heated carburetor air. Some engines also provide for retarded ignition timing during closed-throttle operation.

All air injection systems use essentially the same basic air pump, a positive displacement rotary-vane type.

To guard against excessive temperature and back pressure in the exhaust system resulting from high air delivery rates at full throttle and high speeds, a pressure-relief valve is installed in the pump housing. The valve opens to bleed off some of the pump flow at a predetermined pressure setting.

Output from the air pump is directed through hoses and an air distribution manifold (or two manifolds—one for each bank on V-8 engines) to the air injection tubes located in each exhaust port. A check valve between the air distribution manifold and the air pump prevents reverse flow of hot exhaust gases in the event that pump output is interrupted.

A vacuum-controlled antibackfire valve is used to prevent flow of air to the exhaust ports during the initial stage of closed-throttle acceleration. The high vacuum that occurs during deceleration causes rapid evaporation of liquid fuel from the intake manifold walls. The resulting rich mixture creates a potentially explosive vapor in the exhaust manifold if injected air is present.

As with engine modification systems, most air injection systems also employ spark retard during idle or idle and deceleration through use of 'ported' vacuum sources of dual-diaphragm distributor-vacuum-advance mechanisms."

Besides residual gas dilution and wall quenching, engine variables have the most effect on the amount of hydrocarbons in exhaust gases. Over sixty privately owned and operated automobiles fueled with commercial leaded gasoline have been tested and seven main engine variables were found which changed the hydrocarbon concentration in the exhaust gases. These results can be briefly summarized as:

1) Air-fuel ratio: Low value for gas mixture results in higher hydrocarbon concentration in exhaust gases.
2) Ignition timing.
3) Speed: Increase in speed of engine decreases the amount of hydrocarbon.
4) Air-flow rate: The effect of air-flow rate on the total hydrocarbon concentration depends on the air-flow ratio, ignition timing combination.
5) Compressor-ratio: Increasing the compression ratio, by decreasing head-to-piston distance, increases the total hydrocarbon concentration.
6) Exhaust Back Pressure: The amount of hydrocarbon in exhaust decreases with increasing exhaust pressure.
7) Coolant Temperature: Increasing the coolant temperature decreases the hydrocarbon concentration.

Catalytic Reactions

The engine exhaust gases may be passed through a cylindrical shaped canister packed with catalytic particles. Although this method has great potential, two problems may arise. The long term stability of the catalyst (50,000 miles) is difficult to maintain since lead and other chemicals in trace amounts poison the catalyst. The catalyst structural stability is difficult to maintain under the influence of varying gas flow rates and fairly high temperature. Also, removing three or four pollutants simultaneously can prove difficult for any single type of catalyst. The removal of NO_x, for instance, requires a reduction catalyst, whereas CO requires an oxidation catalyst. For this reason dual stage catalytic reactors have been proposed. The technical problems for this method are greater but the potential advantages are even greater.

Nitric Oxide Removal A review of some chemical reaction data which might be useful in automobile pollution control work has been presented by Shelef and Kummer.[7]

One approach to solving the stability problems has been to avoid leaded fuel in automobiles containing converters. The catalytic approach to conversion for American Motor cars is to use a pellet-type of catalyst with a monolithic-type warm-up feature for California and high altitude cars. The warm-up converter is separately mounted ahead of the catalytic converter. Chrysler Corporation and Chevrolet use a catalyst support coated with platinum, palladium and rhodium. Hydrocarbons, CO and NO_x are all reduced by this three component catalyst. An extensive literature exists on more economical active phases[11,13,16,22] which are not as effective converters. The air to exhaust ratio in catalytic converters is computer controlled in American Motor cars.

For this reason, reactions which involve combination with rather than decomposition of NO_2 are being studied very carefully. The equilibrium constants in terms of partial pressure are given in Table 3 for NO combination with hydrogen, CO and methane at various exhaust temperatures. The thermodynamic conditions (large K_p values) are generally favorable for conversion (reduction). A Monel (nickel–copper alloy) catalyst has been found reasonably successful for removing NO by combining it with residual CO in the exhaust stream. The Monel dissociates the oxygen from NO and then oxides CO to the harmless dioxide.

TABLE 3

The equilibrium constants for some reduction reactions of nitric oxide[17]

Reactions	log *Kp*				
	600	700	800	900	1000K
$NO + 5H_2 \rightarrow 2NH_3 + 2H_2O$	49.0	48.1	32.13	26.44	21.8
$NO + 2H_2 \rightarrow N_2 + 2H_2O$	51.7	43.4	37.1	32.2	28.3
$NO + CO \rightarrow ½N_2 + CO_2$	27.3	22.6	19.7	16.46	14.29
$NO + CH_4 \rightarrow ½N_2 + 2H_2$	19.59	18.1	17.6	16.2	15.5
NO + Hat → HNO	12.4	9.77	7.79	6.24	5.0
$NO + ½ H_2 \rightarrow HNO$	−3.7	−3.6	−3.52	−3.46	−3.42

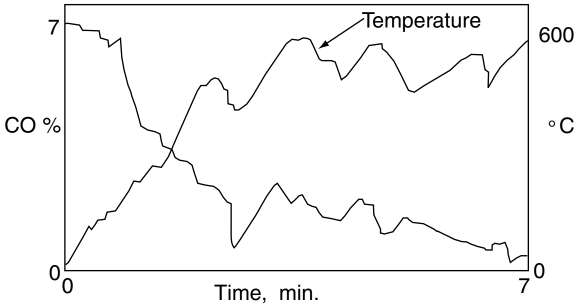


FIGURE 1

In this system, when the engine operation is too rich (too little oxygen) nitrogen oxide reduction is found to be excellent, but carbon monoxide conversion poor. The nitrogen oxide will readily combine with Monel, forming Monel oxide and nitrogen since there is little oxygen to compete with the nitrogen oxide for active sites on the Monel. However, since the CO:NO ratio is so large, there will be insufficient Monel oxide formed to give up its oxygen to the carbon monoxide.

On the other hand, if the engine operation is too lean, NO reduction will be poor and CO conversion excellent since the reaction becomes significant and oxygen will compete with the nitrogen oxide for sites on the Monel.

Some typical conversion data[1] at space velocities of 50,000 v/vhr indicated that conversion of NO in synthetic gas mixture (2% CO, 1000 ppm NO) was over 95% if the temperature was above 700°C, but fell off sharply at temperature below 650°C. Problems associated with this technique are the formation of NH_3 if any residual H_2 is present and back pressure buildups with the current catalyst structure. In addition, dusts form which are emitted as particulates. It is essential for efficient performance of this catalyst that it warms up to operating temperatures very rapidly. Lead in the fuel reduces the chemical activity and ultimately increased the rate of deterioration of the catalyst.

Another catalyst which showed promise for the same reaction but at lower temperature (200 to 350°C) has also been mentioned.[18] Despite the fact that conversions of better than 90% were reported equal amounts of CO and NO were used. Automotive exhaust have about 16 times more CO than NO. Activated carbon has been used[19] successfully with H_2 gas at 600°C. However, activated carbon lost a considerable portion of its activity after only 7 hours service.

CO and Hydrocarbon Removal Major automotive and petroleum companies have combined efforts in the development of an inexpensive multi-thousand mile catalytic package for reducing CO and HC exit concentrations. Since the input of gasoline and hence the effluent gases are rarely at steady state, any study of a catalytic reactor must consider the dynamic situation. To give some idea of magnitude, the exhaust gas flow in standard cubic feet per minute (SCFM) is roughly twice the miles per hour equivalent of an automobile and the temperatures of the exhaust gas change from that of the ambient to well over 600°C. Figure 1 shows typical CO and temperature levels in the exhaust stream after engine startup in a Federal cycle.

For balancing pollution problems with no catalytic or afterburner control an air–fuel ratio of about 16 is recommended. Above this level the NO_x level increases markedly and below it the amount of unburned HC and CO substantially increases. With catalytic devices this ratio may no longer be optimum, since the catalyst selectivity may be greater towards removal of one of the pollutants than any of the others.

Wei[20] has noted that the kinetics of CO oxidation over an egg-shell catalyst turn out to be first order for CO and 0.2 order for O_2 in the range of 1 to 9% oxygen. The curvature in the Arrhenius plot (Figure 2) is believed to be caused by a pore diffusional phenomena. As the catalyst ages and activity falls the reaction rate becomes controlling and the Arrhenius plot becomes a straight line. The catalyst of the figure is best above 350°C, but a lower operating temperature may be preferred for longer catalyst life. Wei[20] found that, "As the catalyst lost 90% of its activity, the emission rose by only 30%; but the last 10% of activity loss would result in a precipitous rise of carbon monoxide emission. A catalyst with 50% reduction in heat capacity performs much better; a reactor with 50% reduction in volume performs better when the catalyst is fresh and worse then it is aged." His philosophy is, "It is our engineering goal to produce a low-cost and convenient solution. However, any solution requires some inconvenience and cost from everyone. Quick warm-up is no problem if we are willing to sit and wait in the car for 2 min for an auxiliary heater to warm up the catalyst bed before the car moves. We have ninety million cars on the road, and a $100 device will cost us nine billion dollars. How much are we willing to pay for 90% cleaner air? These decisions belong to the public, not the engineers. For the sake of everyone we hope to be able to present to the public an economical and convenient solution in the near future."

Stein *et al.*[21] have evaluated the effectiveness of possible catalysts by a microcatalytic technique based on gas chromatography. The technique which is described in detail allows a large variety of hydrocarbons and catalysts to be rapidly tested over a wide range of temperatures (100–600°C). In general oxides of cobalt, chromium, iron, manganese, and nickel are the most effective catalysts. The higher molecular weight hydrocarbons are more easily oxidized than the lower

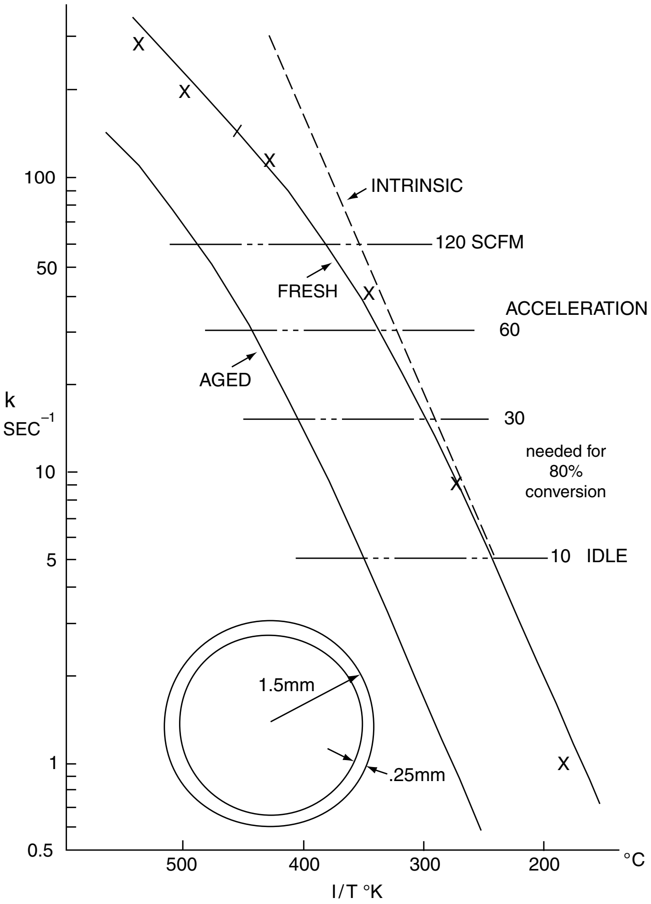


FIGURE 2 Carbon monoxide activity of an eggshell catalyst.

and hydrocarbons of a given carbon number increased in reactivity according to the series: aromatic < branched paraffin or alicyclic < normal paraffin < olefinic < acetylenic. The olefinic hydrocarbons, generally considered the most undesirable, are relatively easy to remove. Other results could be summarized in the following manner:

1) With most of the catalysts tested, some cracking occurs during the oxidation of hydrocarbons. Oxidation without CH_4 formation was possible with the oxides of Co, Cr, Cu, Mn and Ag only. Zirconium oxides are unique in that they produce only cracking products, mainly methane and smaller amounts of intermediate hydrocarbons.
2) Complete removal of all hydrocarbons was not attained with some of the catalyst, even at 600°C.

The oxidation is not a simple function of temperature.

The potential of copper oxide-alumina catalyst for air pollution control has been studied by Sourirajan and Accomazzo.[22] They stated that, "the simultaneous removal of hydrocarbons and carbon monoxide present in the auto exhaust gases has been tested making use of a six-cylinder Chevrolet engine run on leaded gasoline fuel. The hydrocarbon and carbon monoxide concentrations encountered in these studies varied in the range 170–17,000 ppm and 1–7%, respectively. It was found that the minimum initial temperature of the catalyst bed required for the complete removal of both hydrocarbons and carbon monoxide, simultaneously, was 226°C under no load condition, 342°C, under an engine load of 2.5hp, 400°C, under an engine load of 5.1hp or higher, and 236°C under deceleration conditions. The catalyst showed no deterioration in performance even after 100 hours of continuous service in conjunction with the above auto exhaust gases. The extent of removal of hydrocarbons from the exhaust gases was found to depend on the initial temperature of the catalyst bed and the engine load condition.

It is realized that a successful 100 hour run does not constitute a life test on the catalyst, but it does indicate the potential applicability of the catalyst in air pollution control devices. The engineering design of the suitable converter for any particular practical application of the catalyst should naturally take into account the heat liberated during oxidation. Instantaneous catalyst temperatures of the order of 900°C have been encountered in this work with no deleterious effect on the subsequent effectiveness of the catalyst . . . the heat liberated during the reaction can be advantageously used to maintain the full effectiveness of the catalyst under all conditions of engine operation encountered in normal practice."

When contaminants are passed through a Hopcalite (unsupported coprecipitate of copper and manganese oxides) catalyst burner, the results vary from almost complete oxidation of some organics to very slight oxidation of the lower molecular weight aliphatic hydrocarbons[23] at some 300°C. Nitrogen compounds form N_2O when oxidized and halogenated compounds indicate a strong acid reaction when the reactor effluent is tested with detector paper.

An interesting example of the use of exhaust gas recycle and catalysts has been presented by the Esso Research Group. In Figure 3, typical hot cycle traces of CO, O_2 and NO are presented for cases before the catalyst, after Monel catalyst and after a 2nd stage Platinum-alumina catalyst, for instance, with and without recycle. The major beneficial effect of recycle is on the NO concentration.

The combustion of gasoline is more or less incomplete regardless of the quantity of excess air used. About 1% of the exhaust gas is composed of harmful products chiefly carbon monoxide (CO), oxides of nitrogen (NO_x) and hydrocarbons (HC). A significant variable affecting each of these pollutant concentrations is the air to fuel ratio (ATFR). The stoichiometric value, $(ATFR)_{STO}$ is about 14.7:1.0 on a weight basis.

Using a catalytic three way converter, more than 90% of the pollutants can be converted to harmless substances. To avoid catalyst contamination lead free gasoline must be used. In the closed loop electro-mechanical control of $(ATFR)_{STO}$ described by Robert Bosch,[34] 1985, an oxygen sensor in the exhaust gas transmits a signal which is used to correct ATFR deviations. This control method is particularly effective on fuel injection engines because they do not have the additional delay times of carburetor engines. For catalytic converter operation, the optimum ATFR range is

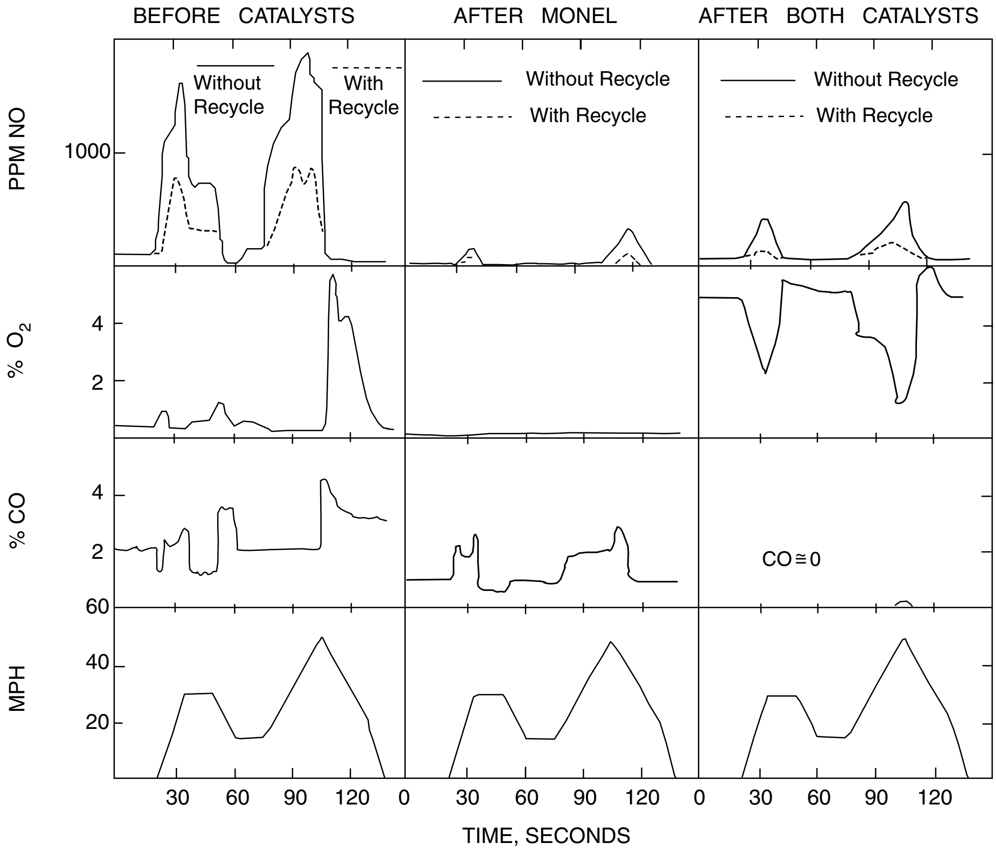


FIGURE 3

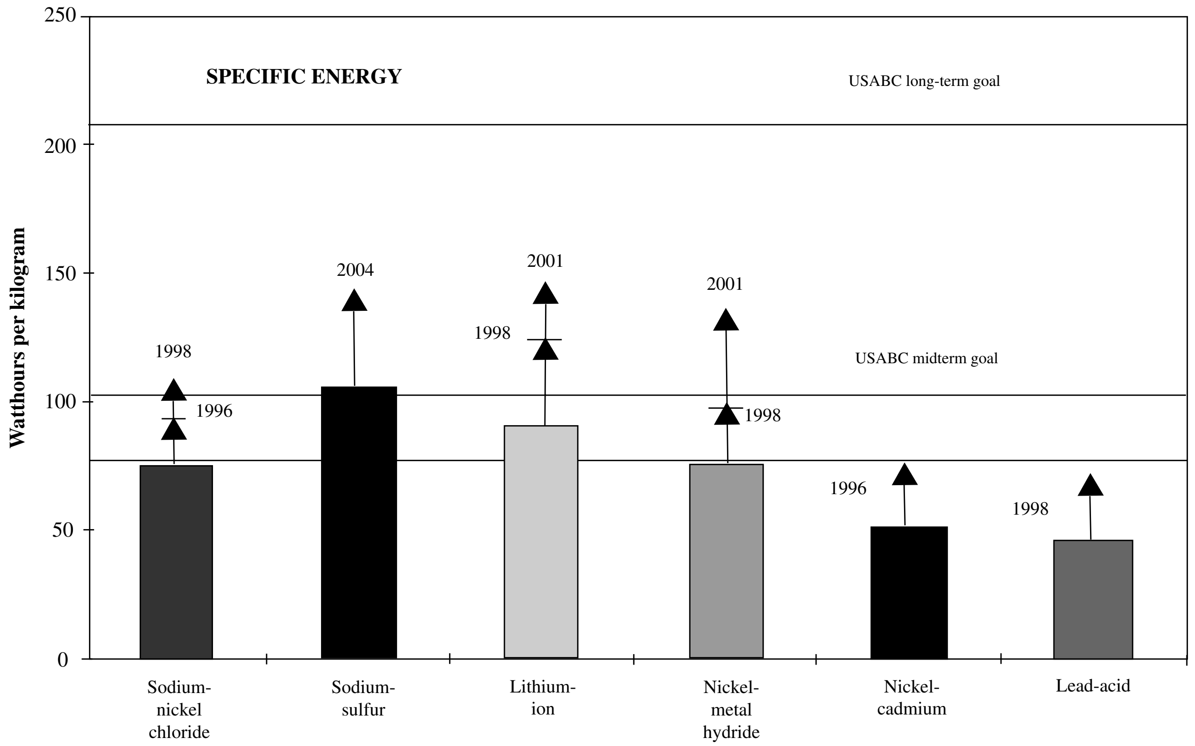
between 99 and 100% of $(\text{ATFR})_{\text{STO.}}$ Above this range NO_x levels increase markedly as ATFR is increased; below this range CO and HC levels increase as ATFR is decreased.

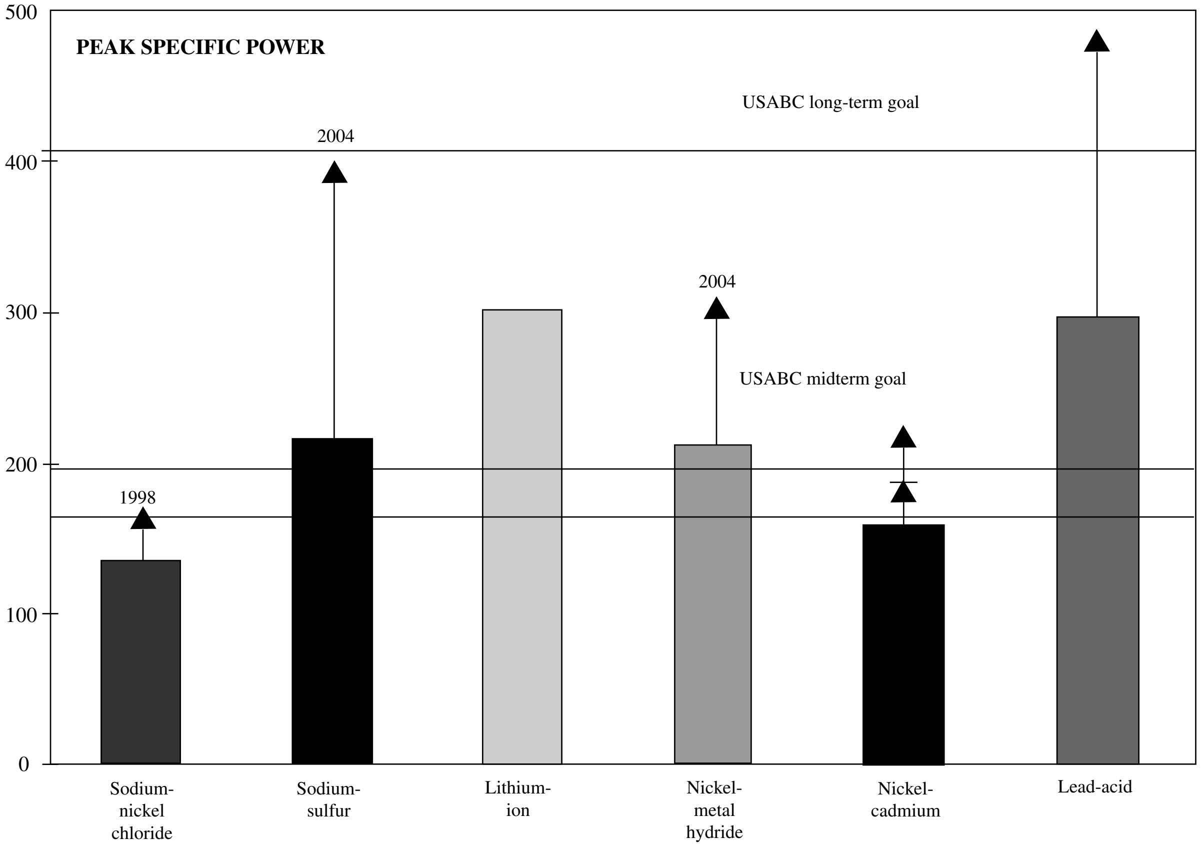
Electric vehicles (EV's) operated by high energy high powered batteries are making great strides toward commercialization. The near term goal is to provide over 100 miles per recharge at accelerations capable of matching the internal combustion engine. The California Air Resources Board has issued a technical document in December of 1995 supporting the concept of such vehicles. The key performance parameter for EV's is its specific energy (or energy density), measured in watthours per kilogram, with a near term goal of 80–100 wh/kg, suggested by the US Advanced Battery Consortium (USABC).[38] Another important measure of performance is the peak specific power (or power density), which gives us an idea of an EV's acceleration. It is measured in units of watts per kg with an USABC near term goal of 150 w/kg that can be sustained for 30 seconds during discharge down to 80% depth of discharge. A comparison of different chemical system performance parameters is presented in Figure 4. All of the batteries are expected to achieve the USABC's midterm goals for these EV parameters, which coincide with lasting about five years (-600 cycles) and costing no more than $150 per kwhr of battery capacity. Hybrid power electric-diesel engines became available in 1995. They alternate between diesel power operation at 2000–2600 RPM where its efficiency is best to battery power at other RPM.

RELATED TRANSPORTATION PROBLEMS

Diesel Exhaust Odors

Diesel engines are found in buses, trucks, off-road vehicles and power applicators and increasingly in automobiles. Public reaction to diesel-engine exhaust odors provides the impetus for controlling effluents of that type of fuel combustion.[24] A list of oily kerosene and smokyburnt odor compound identified by A.D. Little, Inc. is included in Table 4. Exhaust odor and smoke from diesel engines are more objectionable than those from spark ignition. CO emissions are generally less serious but NO_x is troublesome (4 to 10 g/mile). Improved fuel injection and afterburners are considered to be the most promising of the existing control methods. The injection differs from the internal combustion engine in that fuel does not enter the cylinder as a mixture with air, but is injected under high pressure into the chamber in exact quantities through low tolerance nozzles. For NO_x removal the basic approaches


SPECIFIC ENERGY
Watthours per kilogram
250
200
150
100
50
0
USABC long-term goal
USABC midterm goal
1998
1996
2004
2001
1998
2001
1998
1996
1998
Sodium-nickel chloride
Sodium-sulfur
Lithium-ion
Nickel-metal hydride
Nickel-cadmium
Lead-acid


PEAK SPECIFIC POWER
500
400
300
200
100
0
USABC long-term goal
USABC midterm goal
2004
2004
1998
Sodium-nickel chloride
Sodium-sulfur
Lithium-ion
Nickel-metal hydride
Nickel-cadmium
Lead-acid

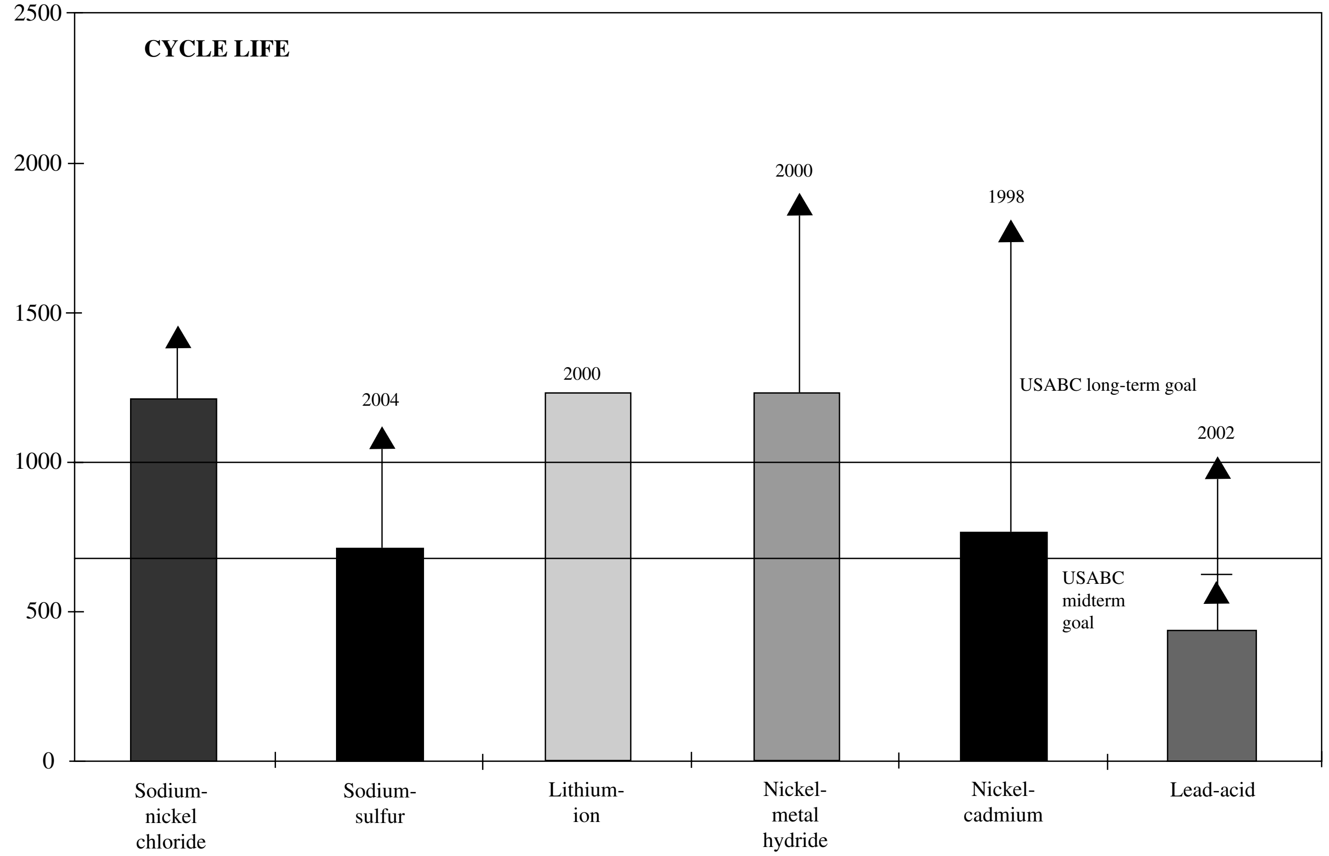


FIGURE 4: Performance Parameter Comparison for Electric Vehicles as a Function of Advanced Battery Type (After Moore[38]).

are engine refinement, fuel additives, catalytic conversion and reduction of peak combustion temperature.

Gas Turbines

The exhaust of gas turbines contains perhaps less pollutants than that of any other internal combustion process. The high airport traffic density makes the problem a real one, however. Sulfur dioxide emissions are low, but the smoke and odor producing compound levels are high. For automobile use, the total mass of gas turbine exhaust is many times greater than that for the gasoline engine of equivalent power. Hydrocarbon and CO mass emissions are known to be low and difficult to reduce. A greater deal of the pollution control work underway is in the engine modification area. In aircraft the turbofan engine is largely replacing the turbojet. Turbofans bypass some of the air past the engine and rejoin it with the burner gases at the exhaust tailpipe. Modern dry NO_x combustion systems can obtain emissions of 25 ppm at 15% O_2.[36,37]

The airplane is a much more efficient carrier (pollution wise) than the automobile, on a people × miles basis. Aircraft engine research has been concerned primarily with smoke reduction. Fuel type, fuel additives and combustion chamber design have been the primary variables studied.

PARTICULATE EMISSIONS

The particulate matter emitted from automobiles has been characterized[25] as consisting of "lead salts, alkaline earth compounds, iron oxides, soot carbonaceous material, and tars. This material ranges in size from large flakes to submicron particles and varies in consistency from hard and brittle particulate to oil mists. Some of the particulate material is generated in the engine combustion chamber and nucleated and agglomerated in the vehicle exhaust system before it passes out the tail pipe. On the other hand, a large portion of the particulate material generated in the engine subsequently deposits on various surfaces of the exhaust system. At some later time, this deposited material flakes off and becomes reentrained in the exhaust gas to be emitted from the vehicle. Obviously, opportunities exist for various types of chemical and physical processes to take place and, as a result, the overall particulate emission process for a vehicle is quite complex and difficult to define."

TABLE 4
Partial list of odorifous compounds

General classification	Compounds
Indans and tetralins	Methyl indan
	Tetralin
	Dimethyl indan
	Methyl tetralin
	Dimethyl tetralin
	Trimethyl tetralin
	Alkyl tetralin
Alkyl-substituted naphthalenes	Methylnaphthalenes
	Dimethylnaphthalenes
Indenes, acenapthenes, and benzothiophenes	Alkyl-substituted indenes
	Dimethylbenzothiophenes
	Acenaphthene
General class	Carbon range
Alkenone	C5 to C11
Furan	C6 to C10
Diene/	C9 to C12
Furfural	C6 to C7
Methoxy benzene	C8 to C9
Phenol	C7 to C12
Benzaldehyde	C7 to C10
Benzofuran	C8 to C9
Indanone	C6 to C10
Indenone	C9 to C12
Naphthol	C10 to C14
Naphthaldehyde	C11

TABLE 5
Lead particulate emissions as a function of size and mileage[25]

Average mileage	Average lead salt emissions, g/mile		
	>9 microns	<1.0 microns	<0.3 microns
5,000	0.04	0.07	0.05
16,000	0.06	0.05	0.04
21,000	0.06	0.05	0.03
28,000	0.10	0.03	0.02

Much of the particulate is in the form of polynuclear aromatic (PNA) hydrocarbons and tars. As for gaseous products, standard federal cycles have been designed to test for particulate emissions.

Lead particulate matter is emitted at a rate of about 0.16g/mile. The average particle size emitted is slightly less than a micron, but sizes are substantially distributed over the 0.10–10 micron range. In Table 5, the effect of mileage on particle distribution may be observed. As the number of miles driven increases, the segment of particles greater than 9 microns also increases, probably because of deposition buildup on exhaust pipes.

Using 500 microns as a divider, about 30% of the lead burned over the lifetime of a car is in the form of fines 40% in the form of coarse particulate.

Research studies[27] showed that a relationship exists between traffic volume, proximity to the highway engine acceleration vs. constant speed, wind direction and the amount of lead in the air. Lead values can be expressed as a quadratic function of time and a linear function of traffic volume. Ter Haar[28] and coworkers found that for Federal 7-mode cycle conditions (Z) cars vary widely in the amounts and composition of their particulate emissions; (2) cold cycle operation gives 2–8 times more particulate than hot engine operation; (3) lead compounds represent less than one-third of total particulates, the remainder being carbon compounds along with ammonium and nitrate ions and unknown materials; (4) carbon emission for stabilized cars using leaded gasoline varies widely but averages about 35% of the total; (5) suspended particulate emissions are nearly equal with new cars whether or not lead is present; (6) exhausted lead varies with the condition of the exhaust system and ranges between 7 and 30% of the lead consumed by the engine; (7) fuel additives affect the amount of emitted particulates; (8) probe sampling techniques underestimate by a large factor of amount of particulates emitted by vehicles; and (9) trapping systems offer potential for greatly reducing the emission of suspended total particulates. Nonleaded-fuel cars were found to exhaust 0.165g of particulate per mile while leaded-fuel cars exhausted 0.152g/mile. The reason for this is that nonleaded fuel is more aromatic in nature and that the percentage emission for aromatic hydrocarbons is considerably greater than that for paraffinic material. An increase in the volume of aromatics in the fuel from 10 to 70% doubles the amount of carbon particulates emitted. It has been claimed[29] that a reduction of 80–90% of the lead emitted is possible by trapping systems.

Two simple traps have been compared for particulate removal and the results presented in Table 6.

More complex traps are likely to be very successful in reducing both lead and total particulate matter.

A more detailed chemical analysis of particulates has been described under Federal emission test conditions.[30] This investigation found that a decrease in the total mass of particulates especially particles of very small size, occurred when using nonleaded fuel. The discrepancy with the aforementioned work[28] may be a result of differences in exhaust deposition for the two systems. Particles emitted in the vehicle exhaust varied in size from 0.01 to 5000 microns, the latter for such products as rust scale. Typical chemical compositions for the particulate exhausts are presented in Figure 5.

The ratio of Pb to Br is relatively constant 2.1:1. The ratio on impactor plates was somewhat lower, indicating that the chemical composition of lead salts may be related to particle size. Considering some of the differences in the findings for different particulate matter studies it is critical that such investigations be perfected and intensified. An interesting additional discussion of the characterization of particulate lead in vehicle exhaust may be found in the work of Habibi[31] and Mueller.[32]

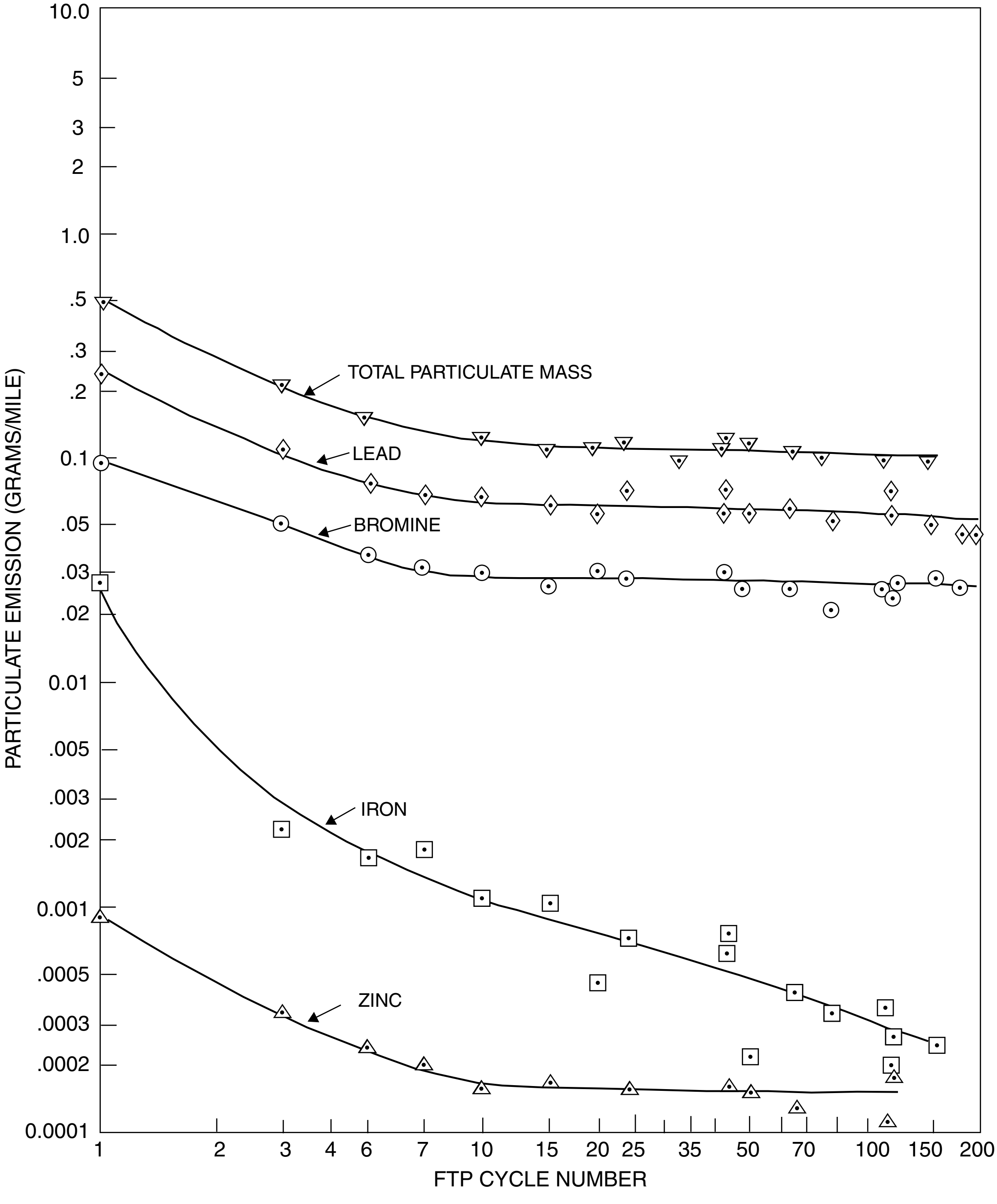


FIGURE 5: Particulate mass emissions from car operating on Indolene-30, as measured by the *Total Volume System,* versus FTP driving cycle number.[30]

In addition to the fuel and exhaust line, additional sources of particulates are to be found in catalysts for gaseous pollutant removal and from the disintegration of tires. The monel catalyst, described previously[1] for NO*x* and Co removal, deteriorates with time at about 0.019 g/mile. The presence of lead in the fuel substantially increases the rate of Monel deterioration.

Sheih[35] estimated that the increase in diesel emissions concentration in the Chicago area would be predominantly for particles having diameters smaller than one micron. At diameters in the vicinity of 0.2 microns the concentration would increase 10 fold.

One other source of vehicle particulate emissions which might be mentioned is that from jet aircraft. By weight, more

TABLE 6
Particulate trapping—federal 7-mode cycle[28]

Car, year, and make	Miles	Exhaust systems	Particulates, total	g/mile Pb
1970 make D	6,000	Standard	0.169	0.046
	6,000	Anchored vortex	0.139	0.039
		% Reduction	19	16
1970 make B	6,000	Standard	0.274	0.071
	6,000	Anchored vortex	0.227	0.056
		% Reduction	17	21
1970 make A	6,000	Standard	0.471	0.114
	6,000	Anchored vortex	0.189	0.043
		% Reduction	60	66
1970 make D	0	Standard	0.116	—
	0	Agglom.-inertial # 1	0.037	0.006
		% Reduction	68	
16 cars (1966–1969)	30,000+	Standard	0.339	0.060
1965 make D	24,000	Agglom.-inertial # 2	0.125	0.036
		% Reduction	63	39
1969 make D	12,000	Agglom.-inertial # 3	0.108	0.034
		% Reduction	67	43
1970 make D	12,000	Agglom.-inertial # 4	0.098	0.019
		% Reduction	71	68

than fifty per cent of the total particulate matter emitted from aircraft engines appears to be due to unburned and partially oxidized fuels. More detailed analysis of this constituent must await development of improved measurement techniques. About 90 per cent of the non-hydrocarbon material present is elemental carbon (i.e., smoke), while the remainder is material eroded from the engine, fuel contaminants, or dust ingested by the engine. Fuel additives containing metallic compounds are seldom used for normal aircraft operations. The emission of such metallic particles does not appear to constitute a problem.

It has been demonstrated that fuel-rich pockets in the primary combustion zone of the burner were the source of the smoke-producing carbon particles and that the smoke could be reduced by "leaning out" the fuel-rich areas. Redesign of the burner is believed to be the most effective approach to smoke reduction.

The author wishes to acknowledge the assistance of Mr. George Musser, Exxon Research and Engineering Company, for his discussion on particulate matter emissions, of Dr. William Peterson, formerly of the Polytechnic, for supplying test procedure information, and of Mr. John Pinto of the New York City Department of Environmental Protection for his review.

The particulate problem most often cited for diesel engines in PM-2.5 emissions.[39] The 2.5 refers to the particle aerodynamic diameter, d, in microns, μm and the PM refers to the concentration in $\mu m/m^3$ of particulate matter of $d \leq 2.5$ μm. An Environment Defense report estimated that in New York Harbor alone, ferries, tugboats, container ships and other commercial vessels produce as much smog-forming contamination over the course of a year as do 300,000 automobiles.[40] These sources were largely unregulated as of 2005. It has been noted by the International Maritime Organization (IMO) that in terms of emission intensity—emissions per ton of cargo transported—marine shipping is one of the most environmentally efficient methods of shipping. According to the IMO, between 85–90% of all commodities are transported by ships, and diesel engines power the majority of these vessels.[41] The diesel engines have become more efficient in recent years. New air pollution rules expected to be in effect will require diesel and gasoline engine emission level equivalence by 2010.

REFERENCES

1. US EPA and US DOE, "2005 Fuel Economy Guide", www.fueleconomy.gov
2. Beard, R.R. and G.A. Wertherm, *Amer. J. Public Health* **57**, 2012 (1967).
3. Shy, C.M. *et al., J. Air Poll. Cont. Ass.* **20**, 539 (1970).
4. Heuss, J.M., G.J. Nebel, and J.M. Colucci, *J.A.P.C.A.* **21**, 535 (1971).
5. Report AP-66, US Dept. of H.E.W., March, 1970.
6. Benson, J.D. and R.F. Stebar, Paper presented at S.A.E. Meeting, Detroit, Mich., Jan., 1971.
7. Shelef, M. and J.T. Kummer, *Chem. Eng. Prog. Symp. Ser.* **67**, 74 (1971).
8. Wise, H. and M.F. Frech, *J. Chem. Phys.* **20**, 22 (1952).
9. Fraser, J.P. and F. Daniels, *J. Phys. Chem.* **62**, 215 (1958).
10. Yuan, E.L. *et al., J. Phys. Chem.* **63**, 952 (1959).
11. Taylor, F.R., Air Poll. Foundation Rept. No. 28 (1959).

12. Sakaida, R.R. *et al., A.I.Ch.E.J.* 7, 658 (1961).
13. Sourirajan, S. and J.L. Blumenthal, *Actes du Deuxieme Congress International de Catalyse*, Vol. 1, p, 2521 (1960).
14. Peters, M.S., AEC TID-18423 (1963).
15. Wikstrom, L.L. and K. Nobe, *J. & EC, P.D. & D.* 4, 191 (1965).
16. Yur-eva, T.M. *et al., Kinet. Katal.* **6**, 1041 (1965).
17. Chkhubianishvili, M.G. and V.T. Chagunava, *Soobsheh. Akad. Nauk Gruz. S.S.R.* **41**, 591 (1966).
18. Ayen, R.J. and Y-S. Ng, *A. & W. Pol. Int. J.* **10**, 1 (1966).
19. Bedjai *et al., I. and E.C.* **50**, 1165 (1958).
20. Wei, J., *C.E.P. Monograph Series* **65**, 14 (1969).
21. Stein *et al., I. and E.C.* **52**, 671 (1960).
22. Sourirajan, S. and M.A. Accomazzo, *Can. J. of Ch. E.* **88**, April 1961.
23. Christian, J.G. and J.E. Johnson, *Int. J. of A. & W. Poll.* **9**, 1 (1965).
24. Somers, J.H. and G.D. Kittredge, *J.A.P.C.A.* **21**, 764 (1971).
25. Habibi, K. *et al.,* Paper presented APCA—5th West Coat Mtg. Oct 8–9 (1970).
26. Seizinger, D.E. and B. Dimitriades, *J.A.P.C.A.* **22**, 47 (1972).
27. Motto, H.L. *et al., Env. Sci. Tech.* **4**, 231 (1970).
28. Ter Haar, G.L. *et al., J.A.P.C.A.* **22**, 39 (1972).
29. Hirschler, D.A. and F.J. Marsee, NPRA Paper No. AM-70–5, April (1970).
30. Ninomiya, J.S., W. Berman, and B.H. Simpson, Paper EN-10G presented at 2nd Inter. Clean Air Congress, Dec. 6–11 (1970), Washington, D.C.
31. Habibi, K., *Env. Sci. and Tech.* **4**, 230 (1970).
32. Mueller, P.K., *Env. Sci. and Tech.* **4**, 248 (1970).
33. Cooper, C.D., *Jnl Air Poll Contr. Assoc.* **39**, 714 (1989).
34. Robert Bosch GmbH, Tech. Instr. Series Report ISBN-1–85226017–3, Dept. of Tech. Info., Stuttgart FDG, Sept., 1985.
35. Sheih, C.M., *Wat. Air & Soil Poll.* **32**, 417 (1987).
36. Jeffs, E., Gas Turbine World, May–June (1989), p. 20.
37. Angello, L. and P. Lowe, Gas Turbine and Aeroengine Congress and Exposition—June 4–8, 1989, Toronto, Canada, ASME Paper 89-GT-254.
38. Moore, Taylor. “The Road Ahead for EV Batteries.” EPRI Journal, March/April 1996, pp. 7–14.
39. Davis, W. “The PM-2.5 Challenge” Environmental Protection, June 1999, pp. 35–37.
40. DePalma, A., New York Times, June 17, 2004.
41. Betts, K.S., Environmental Science and Technology, Jan. 1, 1999, p. 9A.

EDWARD N. ZIEGLER
Polytechnic University

MODELING OF ESTUARINE WATER QUALITY

INTRODUCTION

Estuarine water quality is a term used to describe the quality of characteristics of the water in estuaries. Although the term implies quality in a physical-chemical sense, its use has been extended to include also the acceptability of water in a socioeconomic sense. The term "water quality," like environmental quality and air quality, has to do with the quality of the water or the environment wherever it is found and wherever it is used or encountered. The high chemical and bacteriological quality of the water supplies has become an almost matter-of-fact part of any American's life, but the quality of waters in which man recreates has become of greater concern with man's awareness of degradation of the quality of the waters around him.

Estuarine water quality has become a major focus of the U.S. Environmental Protection Agency with passage of the Water Quality Act of 1987 establishing the National Estuary Program with the goal of identifying nationally significant estuaries, protecting and improving their water quality, and enhancing their living resources. The original four estuaries selected in 1985 for study were Narragansett Bay in Rhode Island, Long Island Sound in New York and Connecticut, Buzzards Bay in Massachusetts, and Puget Sound in Washington. Within a year, Albemarle-Pamlico Sounds in North Carolina and the San Francisco Bay/Sacramento-San Jacinto Delta system in California were added, and most recently Galveston Bay in Texas, among others, has been added.

A thoroughly technical description of water quality would require several volumes to cover the physical, chemical, and biological characteristics of water and how these characteristics change in different environments, how they interact, and how they influence the many ways water is used by plants, animals, and especially man. This would be true even though this article is limited to estuaries, which are among the most complex natural systems known and which feel the impact of man perhaps more than any other natural aquatic system. Suffice it to say that estuarine water quality will be examined in a broad way only, and the reader is referred to the books and articles cited in the bibliography for further discussions on the topics covered herein.

Use Context for Water Quality

Quality of water may be discussed most usefully in the context of water use. That is, for certain uses of water, whether they be recreation, drinking, navigation, or some other use, some level of water quality is required or desired for that particular use. Some uses, such as drinking, will require a much higher level of water quality than will another use such as swimming, and swimming may require a higher quality of water than navigation. The important point is that for desired uses of bodies or areas of water, certain levels of water quality are desired, and if the quality of the water desired or needed to support that use is not present, the use may not be sustained. The concept of use applies not only to man's direct uses of the water, but applies also to biological uses of bodies of water such as fish spawning grounds, shrimp nursery areas, and so forth. Indeed, the history of setting levels of desired water quality for particular uses has shown that following the setting of levels for water quality for drinking and swimming, levels of water quality were set for protecting and enhancing the survival of fish and other organisms in streams. Levels of dissolved oxygen in streams, which are in state and federal water quality standards, are there to protect fish in those streams.

As uses for bodies of water become more numerous, a competition for use of the water begins to develop. Uses such as navigation, swimming, recreational fishing, fish and shellfish nursery areas, and other uses are not uncommon competing uses for a body of water. The quality of water required to support each of these uses is different as noted above, and because of this, some uses may or may not be sustained, depending on which use is the most "beneficial" of that particular body of water.

The Federal Water Quality Act of 1965 stated that water quality standards were to be adopted by all the states by June, 1967, and in preparation of these standard public hearings were to be held to determine the desired uses of all the waters of the state which were under federal jurisdiction. Although many states had already determined uses of their waters, particularly for streams, this was the first time that a nation-wide effort was made to determine desired uses of waters and to set water quality standards for them. The hearings, the water quality standards developed, and the subsequent implementation and enforcement of the standards showed the very real problems which arise when competing uses of the water resource become very strong and intense. One or two particular uses become dominant, and the water quality for a particular use is set to meet that use. Other uses which require a higher level of water quality may or may not be sustained, while levels of water quality required for other uses may be more than adequately met.

This use context has been supported in subsequent legislation, particularly the 1972 Water Pollution Control Act, which required that water quality criteria be updated periodically by the U.S. Environmental Protection Agency as well as by the states.

Estuaries

The estuary is one of those bodies of water which is the focus of intense competing uses. Estuaries comprise one of the most important resources of any country for the support of such uses as navigation, recreation, nursery and resting grounds for waterfowl and wildlife, nursery and spawning areas for fish and shellfish, and particularly sites for urban growth and the consequences or byproducts of urban growth. It is estimated that about 75% of the entire population of the U.S. lives within 50 miles of the nation's coasts (USEPA 1987), and such a large urban population presents heavy use pressures on coastal areas, particularly estuaries.

Estuaries are semi-enclosed coastal bodies of water having a free connection with the open sea and within which the sea water is measurably diluted with fresh water derived from land drainage (Pritchard, 1967). Along the coasts of the United States alone some 45,832 square miles of estuarine waters exist. Of this total, 17,058 square miles are found along the Atlantic Coast, while along the Pacific Coast south of Alaska but including the Pacific Islands some 14,353 square miles exist, 2760 square miles are found along the Coast of Alaska, and 11,661 square miles are found in the Gulf of Mexico and Caribbean Islands (National Estuarine Study, 1971). Of the total, less than 30% is water less than six feet deep, vulnerable to filling, as well as especially productive of fish, shellfish and wildlife. At least 6.8% of the latter have been obliterated through filling, most in the last 50 years (Stroud, 1971).

Development of estuary shorelines indicate some of the uses of the estuaries. Of the 89,571 statute miles of tidal shoreline in the United States estuaries, some 17,853 miles can be described as recreation shoreline, that is, accessible and useful for recreational pursuits. Of this shoreline, 16,559 miles are privately owned and 1,294 miles are publicly owned; however, only 770 miles may be considered recreation areas. Marine transportation terminal facilities are users of a portion of the shoreline estuaries. In 1966, there were 1,626 marine terminals providing deep water berths in 132 ports on the Atlantic, Gulf, and Pacific Coasts. Industries use estuarine waters for cooling and return a heated effluent. Industries and cities use estuaries as disposal sites for their wastes. With a third of the United States population located in the estuarine zone, the impact of man on estuaries must necessarily be quite high (National Estuarine Study, 1971).

Biological uses of estuaries are also quite high. It has been estimated that nearly 63% of the commercial catch on the Atlantic Coast is made up of species of fish believed to be estuarine dependent. Assuming that this applies equally to the combined catches by foreign nationals as to the US domestic catch, the fisheries yield to the US Atlantic continental shelf and present levels of development of the fishery is equivalent to about 535 pounds per acre of estuaries (McHugh, 1966). Similar but somewhat smaller estimates have been made for the Gulf of Mexico estuaries based on commercial catches in the Gulf of Mexico and for the Chesapeake Bay estuary based on catches within the estuary itself (McHugh, 1967).

Factors Influencing Estuarine Water Quality

What are the factors that control the quality of waters in estuaries? The predominant factors are the hydraulic (transport) characteristics of the estuary, the inputs or sources of materials which make up elements of the quality of the water, and the sinks present in estuaries—those physical, chemical and biological phenomena which cause materials in the water to change in concentration or to be altered chemically to a different form than when originally introduced.

The hydraulic regime of an estuary is dependent upon three particular factors: the physiography of the bay—its size, area in relation to volume, depth, and shoreline development; the amount and seasonability of river inflow to the estuary; and the wind and tidal mixing which takes place in the estuary on each tidal excursion. The latter factor is dependent upon the tidal range, the configuration of the entrance to the estuary, the volume of the river inflow and the periodicity of the tides. The impact of sources of material to an estuary are dependent upon the character and amount of the material and the location in the estuary where the material enters. Materials which enter with the river inflow will very likely reach broad areas in the estuary due to mixing within the estuary and the fact that the material will pass through the estuary on its way to the ocean. On the other hand, material discharged near the mouth of the estuary will travel only a short distance into the estuary and most likely be transported out of the estuary rather quickly. This generalization does not apply to all estuaries, particularly those which are strongly stratified. This type of estuary will be discussed in more detail later on. The size of the sinks for materials in estuaries is dependent on the conservative or non-conservative nature of the material, that is, whether the material can be broken down into by-products or whether it remains in essentially the same form throughout its history within the estuary. Conservative and nonconservative materials may both be removed from the water column due to flocculation or sedimentation within the estuary, in which case materials may become part of the bottom sediments and lost from the water column unless the sediments are disturbed.

Because of the intimate tie between water quality and estuarine hydraulics, they will be examined below as well as the sources and sinks for materials within estuaries, both natural and man-made materials, before discussing water quality-estuarine use interactions.

ESTUARINE HYDRAULICS

A spectrum of hydraulic types may occur or exist in an estuary. These may range from the situation in an estuary in which the river flow dominates to the estuary in which the river flow is negligible and the hydraulic regime is dependent on the tidal mixing. Naturally, in a river flow dominated estuary, the water quality in the estuary is most similar to that of the river, whereas in the tidal mixing dominated estuary, its water quality is more like that of the off-shore waters. The other factor greatly influencing the hydraulic regime in an estuary is the physiography of the estuary which is very greatly dependent on the origin of the estuary and the subsequent natural events which have taken place in geologic time and man-made events in contemporary time to modify the original shape of the estuary.

Origins of Estuaries

From a geomorphological standpoint, there are four primary subdivisions of estuaries: (1) drowned river valleys; (2) fjord type estuaries; (3) bar-built estuaries; and (4) estuaries produced by tectonic processes (Pritchard, 1967). Each of these types of estuaries is characterized by the fact that at some point in geologic time, it has been inundated with ocean water due to the rise in the sea level. During the last glacial stage, sea level was about 450 feet below its present level, and the shorelines of the continent were at or near the present continental slopes. Within the last 50,000 years, the sea level has risen from that stage to the present with the last changes in sea level occurring about 3,000 years ago (Russell, 1967).

As the name implies, drowned river valley estuaries are river valleys found along a coastline with a relatively wide coastal plain, which were inundated with ocean water as the sea level rose. The Chesapeake Bay is a prime example of this type of estuary. During the last glacial period, the Susquehanna River reached the ocean about 180 kilometers seaward of the present shoreline; the York River and the other rivers now entering the bay to the north of the York were then tributaries of the Susquehanna River. The rise in sea level flooded the valleys of these rivers to form the present Chesapeake Bay system. The drowned river valleys, or as they are more commonly called, coastal plain estuaries, extend up river to a point approximately where the floor of the river rises above sea level. This is also the point at which a major change in water quality occurs from the ocean and estuary type water quality to that of the river. This geographical point may be downstream from parts of the river which are still influenced by the oscillation of the tidal currents.

The fjord type estuary is that formed by glaciers. These estuaries are generally U-shaped in cross section, and they frequently have a shallow sill formed by terminal glacial deposits at their mouths. The basins inside these sills are often quite deep, reaching depths of some 300 or 400 meters. Most fjords have rivers entering at the head and exhibit estuary features in the upper water layers. The sill depths in Norwegian fjords are often so shallow that the estuarine features develop from the surface to the sill depth while the deeper basin waters remain stagnant for prolonged periods.

Bar-built estuaries are those formed in an offshore area where sand is deposited as a sand island and sand pit built above sea level, and they extend between the headlands in a chain broken by one or more inlets. Such bays often occur in areas where the land is emerging geologically. The area enclosed by the barrier beaches is generally parallel to the coast line. Frequently, more than one river enters the estuary, though the total drainage area feeding a bar-built estuary is seldom large. The lower valleys of such rivers have frequently been drowned by the rising sea level, and hence the bar-built estuary might be considered as a composite system, part being an outer embayment partially enclosed by the barrier beaches, and part being a drowned river valley or valleys. Tidal action is usually considerably reduced in such estuaries. These systems are usually shallow, and the wind provides the important mixing mechanism (Pritchard, 1967). Several of the North Carolina estuaries and most of those along the Texas Gulf Coast are examples of this type of estuary.

Estuaries produced by tectonic processes are those formed by faulting or by local subsidence, and they usually have an excess supply of freshwater inflow. San Francisco Bay is an example of such an estuary.

Circulation in Estuaries

Other than the physiography of estuaries, the dominant physical processes associated with movement of water and mixing in an estuary are the wind, tides, and the inflow of river water. Extensive analysis of these processes has been presented in Fischer *et al.* (1979), Fisher (1981), Thomann and Mueller (1987), and others. The composite actions of these processes produce a variable interaction or interfacing of fresh water from the river and salt water from the ocean. Because these two sources of water have very different densities, the less dense fresh river water will tend to float on top of the dense salt water, and the extent that the two types of water mix is dependent on the strength of the mixing mechanisms. In an estuary with no tides or wind and a steady river inflow, the fresh water inflow would ride on top of the salt water from sea level in the estuary or river bed to the ocean. Because in a real system friction is present, the fresh water will force sea water some distance downstream from the sea level point in the river and the interface between the salt and fresh water layers will tilt downward in the upstream direction in a wedge shape. The friction between the layers will also cause an exchange of water from one layer to another, generally from the salt water, or "salt wedge," to the fresh water. The amount of exchange depends strongly on the mixing mechanisms, wind, tides, and river inflow.

In a wind dominated estuary, wind provides most of the energy for moving and mixing the water. In a tide dominated estuary, turbulence associates with the tidal currents results in mixing between the salt and fresh water, which in turn produces the density gradients associated with the non-tidal circulation pattern. In a river dominated estuary, such as the Mississippi River estuary, water movement is predominantly related to riverflow and mixing is caused mostly by the breaking of unstable interfacial waves at the upper boundary between the fresh river water and the salt water from the ocean.

In an estuary in which a salt wedge occurs distinctly the river flow completely dominates the circulation. The salt water extends as a wedge into the river and the interface between the fresh and salt water slopes slightly downward in the upstream direction. The steep density gradient at the interface, amounting to a discontinuity, reduces the turbulence and mixing to a very low level. The effect of the Coriolis force causes the interface to slope downward to the right in the northern hemisphere looking toward the sea. In the moderately stratified estuary, the dominant mixing agent is turbulence caused by tidal action, rather than velocity shear at the interface between the salt water and overlying fresh water layer as in the previous case. With a tide of moderate amplitude, random water movements at all depths occur and turbulent eddies transport fresh water downward and carry salt water upward, in contrast to the dominantly upward advection of salt across the interface which constitutes the vertical flux of salt in the river dominated estuary. The result of this two way mixing is that the salt content of both the upper and lower layers increases toward the sea. At any given point the bottom layer is always higher in salt content than the lower layer. The boundary between the seaward flowing upward layer and the counter flowing lower layer occurs with a mid-depth region of relatively rapid increase in salt content with depth, compared to the vertical gradient in either the upper or lower layers. This type of mixing contributes a greater volume of salt water to the upper, seaward flowing layer than in the salt wedge estuary. The rate of flow in the upper layer of the moderately stratified estuary is therefore much greater in volume than in the highly stratified estuary, necessitating a correspondingly larger compensating up estuary flow in the lower layer.

When tidal mixing is sufficiently vigorous, the vertical salinity stratification breaks down, and the estuary approaches true vertical homogeneity. The type of circulation which exists in a vertically homogeneous system depends upon the amount of lateral homogeneity. Owing to the Coriolis force in the northern hemisphere, the water on the right of an observer looking seaward may be lower in salinity than the water to his left. A cyclonic circulation pattern is developed, with fresher, seaward flowing water concentrated to the right of center and a compensating up estuary flow of higher salinity water to the left of center. Although a vertical salinity gradient is absent in a vertically homogeneous estuary, vertical transfer of salt is not lacking. There is also a strong lateral transfer of salt which represents the dominant circulation pattern in this type of estuary.

Certain vertically homogeneous estuaries, particularly those which are relatively deep and narrow, do not exhibit these cyclonic circulation patterns. The direction of water movement is symmetrical about the longitudinal axis, and fluctuations in velocity are related to the tides and the net flow averaged over several tidal cycles is directed seaward at all depths. There is a tendency for salt to be driven out of the estuary by the action of the advective process. There must be a compensating non-advective longitudinal flux of salt directed toward the head of the estuary (Pritchard, 1967).

It is very important to note that the quality or character of the water at any point in the stratified, partially stratified, or vertically homogeneous estuary will be strongly correlated with the salinity content of the water. For example, the high salinity, bottom water in a stratified estuary will have a quality much like that of the offshore ocean water. The water at the geographical midpoint of a vertically homogeneous estuary will be a mixture of river and ocean waters. Also, materials introduced into an estuary will be influenced at any point in time or space by the circulation patterns in the estuary. Estuaries which have not felt man's influence either in the estuarine zone or the fresh waters which flow into them have biological systems adapted to whatever water quality patterns exist. Since these water quality patterns are strongly influenced by the circulation patterns and/or introduction or removal of materials, they will have a beneficial or deleterious effect on the biota of the estuary depending on the extent of the change or the nature of the material introduction or removal. Thus, it is important to examine the circulation patterns of estuaries as well as the material introduced or removed to understand the water quality and biota of the estuary and the uses which may be made of the estuary.

Estuarine Circulation Models

Numerous attempts have been made to model the hydraulic processes which occur in estuaries. Originally, these models were developed to determine circulation modifications which might occur because of physical modifications to the estuary. These models have been extended in recent years to include constituents of water and the prediction of their transport and fate in estuaries.

One of the first type of models developed for estuaries was the hydraulic model. This type of model is a physical representation of an estuary on a small scale. Such models are usually distorted in the vertical direction so that water depth may be represented on a larger scale than a lateral dimension. For example, if an estuary were modeled on a scale of 1:100, the width of the estuary, if it were 10 miles, would be 0.1 miles in the model, but the depth of the water, if it were 10 feet, would be 0.01 feet which would be not much more than a film of water in the model. To avoid this situation which would make the model unusable, the vertical scale is reduced to a lesser extent than the horizontal scale such that the 10 foot depth of water mentioned above would be about 1 foot. While the hydraulic models are capable of representing tidal currents, momentum entrainment, and gravitational circulation, they are not able to represent local currents and turbulent eddies. For this reason, there is considerable distortion of diffusive processes in the physical model that makes its utility in quantitative concentration distribution studies dubious (Ward and Especy, 1971). From a qualitative standpoint, the physical model possesses an excellent demonstration capability for the visualization of flow patterns in resultant concentration distributions, and this capability should not be under-rated.

The other types of models developed for estuaries are mathematical models which may be intended to model tidal currents, net advective movement, or tidal stage in an estuary, or they may be intended to model the transport of salt in the water or other chemical forms. Such models may be three dimensional to represent transport of material down the estuary as well as laterally and vertically, or they may be two dimensional to represent transport of material down the estuary and laterally or vertically, or they may be one dimensional to represent transport of material down the estuary. Because the complexity of developing and solving mathematical models decreases as the number of dimensions included are decreased, the one dimensional model has received the widest attention in terms of development and use. This type of model is most advantageously applied to linear type estuaries, that is, estuaries which have little or limited variation in cross sectional area and depth with distance down the estuary. Examples of such models include the model of the Thames River in England, the Delaware River in New Jersey, the Potomac River in Maryland, and the Hudson in New York (see Thomann and Mueller 1987 for an excellent introduction to such models).

Water quality models are usually derived from the following basic three dimensional continuity equation:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left(E_x \frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial y}\left(E_y \frac{\partial C}{\partial y}\right) + \frac{\partial}{\partial z}\left(E_z \frac{\partial C}{\partial z}\right) - \frac{\partial}{\partial x}(uC) - \frac{\partial}{\partial y}(vC) - \frac{\partial}{\partial z}(wc) \pm \sum S$$

where

E = dispersion coefficient along each of the three axes x, y, and z
u,v,w = velocity in x, y, or z direction respectively
S = source or sink for material
C = concentration of material.

This equation expresses a relationship between the flux of mass caused by circulation and mixing in the estuary and the sources and sinks of mass. In the one-dimensional form in which the assumptions have been made that concentrations of some material are of homogeneous concentration laterally and vertically (the y and z directions, respectively) and that the net transport of the material through the estuary is of concern, then the following equation has been used (O'Connor and Thomann, 1971; Thomann and Mueller 1987):

$$\frac{\partial C}{\partial t} = \frac{1}{A}\frac{\partial}{\partial x}\left(EA\frac{\partial}{\partial x}\right) - \frac{1}{A}\frac{\partial}{\partial x}QC \pm S$$

where

A = cross sectional area of estuary
Q = freshwater inflow
E = dispersion coefficient in x direction

and other terms are the same as above. Such models may be used to determine changes in material concentration with time for materials whose rate of entry to the estuary and/or loss from the estuary in a sink are steady or only slightly variable. A further assumption is to select the steady state situation, the condition in which the concentration of the material does not change with time. For this condition $\partial C/\partial t$ in the above equation is to set to zero and the equation solved.

Recently two dimensional models have been developed. These models often assume that vertical stratification does not occur in the water column and that lateral stratification does occur. Such models are most appropriately applied to estuaries with large surface areas and shallow waters. Such models have been developed for many estuarine systems. Feigner and Harris (1970) describe a link-node model developed specifically for the Francisco Bay-Delta Estuary, but applicable elsewhere. It models the two-dimensional flow and dispersion characteristics of any estuary where stratification is absent or negligible. Hydrological parameters of tidal flow and stage are computed at time intervals ranging from 0.5 to 5.0 mins and at distance intervals ranging from several hundred to several thousand feet. Predictions of quality levels are computed on the same space scale, but on an expanded time scale, ranging from 15 to 60 mins. The model is thus truly dynamic in character. It predicts fluctuating tidal flows and computes tidally varying concentrations of constituents, in contrast to a non-tidal model based on the net flow through the estuary such as that developed for the Delaware estuary. It can also accommodate both conservative and non-conservative constituents.

First, the hydraulic behavior of the estuary is modeled. Having established channel directions both in the actual prototype channels and (artificially) in the bay areas, the authors use one dimensional equations based on the following assumptions:

a) Acceleration normal to the x-axis is negligible.
b) Coriolis and wind forces are negligible.
c) The channel is straight.
d) The channel cross-section is uniform throughout its length.
e) The wave length of the propagated tidal wave is at least twice the channel depth.
f) The bottom of the channel is level.

Equations of motion and continuity are, respectively

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} = Ku|u| - g\frac{\partial H}{\partial x}$$

and

$$\frac{\partial H}{\partial t} = -\frac{1}{b}\frac{\partial}{\partial x}(uA)$$

where

u = velocity along the x-axis
x = distance along the x-axis
H = water surface elevation

g = acceleration of gravity
K = frictional resistance coefficient
t = time
b = mean channel width
A = cross-sectional area of the channel.

The terms on the right hand side of the equation of motion are, in sequence, the rate of momentum change by mass transfer, the frictional resistance (with the absolute value sign to assure that the resistance always opposes the direction of flow), and the potential difference between the ends of the channel element. In the continuity equation the right hand side represents the change in storage over the channel length per unit channel width. To minimize computation, the equation of motion is applied to the channel elements and the continuity equation to the junctions.

Both equations are rendered into partial difference form and solved for each channel element and junction, using a modified Runge-Kutta procedure. The results comprise the predicted channel velocities, flows, and cross-sectional areas and the predicted water surface elevations at each junction for each time interval. These data are then input to the water quality component of the model. The equations are put into finite difference form and solved to give the concentration of the substance at each junction.

Ward and Espey (1971) and Masch and Brandes (1971) describe a segmented hydrodynamic and water quality model which has been applied to Texas estuaries. Each segment is a square one nautical mile on each side, and the estuary is divided into these segments. Hydrodynamic transport across segment boundaries is represented much as the equations given above and occurs in response to forcing flows from river inflow at the head of the estuary and tidal exchange at the lower end. The model is able to simulate water stage change within each segment and flows between segments with change in tides, and the averages of the flows are used in conjunction with the water quality portion of the model to forecast concentrations of conservative and nonconservative constituents.

A third type of two-dimensional model is that of Leendertse (1970) who developed a water-quality simulation model for well-mixed estuaries and coastal seas (i.e., no stratification) and applied it in Jamaica Bay, New York. Leendertse and Gritton, 1971, have extended the model to include the transport of several dissolved waste constituents in the water, including any interactions among them. The changing tide level influences the location of the land-water boundaries in the shallow areas of coastal waters. To simulate this process, procedures were developed in the model to allow for time-dependent boundary changes. Large amounts of numerical data are generated by the computer program developed from the simulation model. To assist the investigator in extracting important and meaningful results from these data, machine-made drawings were used to graphically present the results of the computation.

The basic mass-balance equation for 2-dimensional transport of waste constituents in a well mixed estuary (uniform concentration in the vertical directions) is given in Leendertse (1970) as:

$$\frac{\partial}{\partial t}(HP) + \frac{\partial}{\partial x}(HUP) + \frac{\partial}{\partial y}(HVP) - \frac{\partial}{\partial x}\left(HD_x \frac{\partial P}{\partial x}\right) - \frac{\partial}{\partial y}\left(HD_y \frac{\partial P}{\partial y}\right) - HS_A = 0$$

where

P = integrated average over the vertical of the waste constituents mass concentration
U and V = vertically averaged fluid velocity (components in the x (eastward) and y (northward) directions respectively)
S_A = source function
D_x and D_y = dispersion coefficients
H = instantaneous depth at a point.

The generalized mass-balance equation for n constituents is written in matrix notation as

$$\frac{\partial}{\partial t}(H\bar{P}) + \frac{\partial}{\partial x}(HU\bar{P}) + \frac{\partial}{\partial y}(HV\bar{P}) - \frac{\partial}{\partial x}\left(HD_x \frac{\partial \bar{P}}{\partial x}\right) - \frac{\partial}{\partial y}\left(HD_y \frac{\partial \bar{P}}{\partial y}\right) + [K]H\bar{P} + H\bar{S} = D$$

where

$\bar{P}$ = mass-concentration vector with n elements
$[K]$ = reaction matrix
$\bar{S}$ = source and sink vector.

The reaction matrix $[K]$ in its most general form can give rise to a non-linear transport equation. This occurs because the individual elements of the matrix can be defined as functions of their own concentration, or that of other constituents or both. Since the elements of $[K]$ are multiplied by the elements of the concentration vector, such non-linear terms imply kinetics of an order higher than first.

Point sources, such as occur at the location of sewage discharges into the estuary, are simulated by adding delta-function source terms to the source vector.

For two-dimensional flow in a well-mixed estuary, vertical integration of the momentum and continuity equations yields the following basic equations for the flow model

$$\frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x} + V\frac{\partial u}{\partial y} - fV + g\frac{\partial \zeta}{\partial x} + g\frac{U(U^2+V^2)^{1/2}}{C^2H} - \frac{1}{\rho H}\tau_x^s = 0$$

$$\frac{\partial V}{\partial t} + U\frac{\partial V}{\partial x} + V\frac{\partial V}{\partial y} + fU + g\frac{\partial \zeta}{\partial y} + g\frac{V(U^2+V^2)^{1/2}}{C^2H} - \frac{1}{\rho H}\tau_y^s = 0$$

$$\frac{\partial \zeta}{\partial t} + \frac{\partial}{\partial x}(HU) + \frac{\partial}{\partial y}(HV) = 0$$

where

f = Coriolis parameter

g = Acceleration of gravity

C = Chezy coefficient

τ_x^s = Component of the wind stress in the x direction

τ_y^s = Component of the wind stress in the y direction

ρ = Water density

ζ = water level elevation relative to the reference plane.

The wind stress components are given by

$$\tau_x^s = \theta \rho_a W^2 \sin\varphi$$
$$\tau_y^s = \theta \rho_a W^2 \cos\varphi$$

where

θ = wind stress coefficient ≈ 0.0026

ρ_a = atmospheric density

W = wind velocity

φ = angle between the wind direction and the y axis.

In the finite difference approximations of these equations, the discrete values of the variables are described on a space-staggered grid. The position and time coordinates (x, y, t) are represented on the finite grid by (jΔx, kΔy, nΔt), for j, k, n = 0, ±1/2, ±1, ±3/2, K.

Water levels and pollutant concentrations are computed at integer values of j and k (x and y directions). Water depths, obtained from a field survey, are given at half-integer values of j and k. The velocity component U (x directed) is computed at half integer values of j and integer values of k, and the velocity component V (y directed) is computed at integer values of j and half-integer values of k.

The set of finite different equations used to approximate the momentum and mass-balance equations are then presented at two adjacent time levels, n and (n + 1/2). Numerical computation of the reaction matrix terms in the mass-balance equations is accomplished by a sequential use of forward and backward information. If M constituents are transported, then for constituent i(1 < i < M), in the first operation at time level n (going from t to t + 1/2Δt), information is used in the reaction matrix terms on the level (t + Δt) for all constituents for which the sequence numbers m is smaller than i. Information at the level t is used for which m > i. In this step, the constituents are computed is ascending order, from 1 to M.

In the second operation, at time level n + 1/2 (going from t + 1/2 Δt to t + Δt), the constituents are computed in descending order, M to 1. Information on the level t + 1/2Δt is used for all constituents whose m < i, and information on the level t + 1/2Δt is used for all constituents whose m > i. This procedure centers the reaction matrix information of the mass-balance equations within the time interval t to t + Δt. The reaction matrix terms which involve the ith constituent itself are taken centered over each half time step.

The sequential use of finite-difference approximations for the continuity equations at n and n + 1/2 results in alternating forward and backward differences. This means that over a full time step the terms are either central in time or averaged over the time interval. In the first operation at time level n (going from t to t + 1/2Δt), the momentum and continuity equations are solved first for the water levels and x-directed velocities at time level n + 1/2. The information generated is then used in the mass balance equations to obtain the constituent concentrations at time level n + 1/2.

The results of this first operation are then used at time level n + 1/2 to determine the unknowns in the second half timestep, going from t + 1/2Δt to t + Δt. Again, the momentum and continuity equations are solved first, but this time the water levels and y-directed velocities at time level n + 1 are obtained. This new information is then used in the mass balance equation to obtain pollutant concentrations at time level n + 1.

This procedure is then repeated for each succeeding full time step. The model can be used to investigate the influence of wind on low and circulation in the area covered, together with its effect on water levels and distribution of pollutants. This was the first time that real wind effects were investigated in detail.

The need for three dimensional models has been recognized for salt wedge type and moderately stratified estuaries, and three dimensional mathematical models of real estuaries have been developed. Leedertse and Liu (1975) developed a three dimensional code for water movements, salinity, and temperature which was applied to San Francisco and Chesapeake Bays and later to the Bering Sea, Chukchi Sea, the Beaufort Sea, and the Gulf of Alaska (Liu and Leendertse 1987). Other three dimensional models include that of Oey (1985) who modeled the Hudson-Raritan estuary.

SOURCES AND SINKS

In addition to the hydraulic regime of an estuary, the other factors which have great influence on the water quality of estuaries are the sources and sinks of the materials. The circulation patterns and water movement in estuaries will dictate the distribution of fresh and salt water in the estuary. Superimposed on this distribution is another pattern made up of materials introduced by sources and lost to sinks.

In all these equations, the source and sink terms become zero for conservative substances. For non-conservative substances, reactions that take place may usually be represented by first-order kinetics, i.e., the rate of reaction is proportional to the concentration of the material. In some cases the reaction term defines the fundamental reaction mechanism, whereas in other uses it is an empirical approximation to the phenomenon.

While as a general rule the modeling of the hydrodynamic transport of a constituent in an estuary is much further advanced than the modeling of its reaction kinetics, the most commonly unsatisfactory aspect of present water quality models is the specification of the source and sink terms. Many of the physical-chemical processes affecting the concentration of parameters lack adequate formulation. These include sedimentation and deposits of particulate matter, non-linear reaction kinetics, surface exchange of gaseous constituents, and chemical and biological reactions. Modeling of the relation of water quality and estuarine biota is not well advanced. Models of phytoplankton production, of nitrogen cycling, and of gross ecological parameters have been attempted with limited success.

Sources

Sources for the materials which are found in the waters of estuaries include two major sources, river inflow and ocean water inflow. The concentrations (or ranges) of selected chemical constituents of fresh and ocean waters are given in Table 1. Fresh waters may have large ranges of concentrations of the lands which they drain. These ranges are quite different from those of oceanic waters. In fresh waters, calcium is usually the most abundant cation and sulfate is the most abundant anion although carbonate may also be quite high in concentration. In sea water, on the other hand, chloride is the most abundant constituent and anion followed by sulfate and bicarbonate. Sodium and magnesium constitute the majority of the cations. Depending on the relative balance of river inflow and the incursion of seawater brought in by tidal action, the quality of the water in the estuary assumes a composition in proportion to the two sources. However, the location of constituents from the various sources either laterally in the estuary or vertically in the water column is highly dependent on the circulation patterns existing in the estuary which were discussed earlier.

Although in relation to river and tidal flows, direct precipitation is a small hydraulic input to an estuary, its water quality cannot be ignored. In shallow bays with little river inflow and a restricted opening to the ocean such as bar-built estuaries, rainfall directly on the estuary may be an important source of fresh water.

Waste discharges may exert a dominant influence on the water quality of estuaries depending on the amount of material discharged and its character. Because urbanization typically occurs around estuaries, waste discharges are usually directed to the estuaries since they are the most convenient waste disposal site. Domestic wastes, wastes derived from municipalities and ultimately humans, contains large amounts of organic and nutrient (nitrogen, phosphorus, trace) materials. Some typical concentration values are given in Table 2. The relative amounts of material discharged to estuaries or other bodies of water may be estimated by knowing the population served by a sewerage system and mass discharge coefficients. These coefficients indicate the amount of material discharged per person per day. Such coefficients are also given in Table 2.

TABLE 1
Quality of fresh and ocean water
(Concentration units are mg/L)

Constituent	Fresh	Ocean
Chloride	1.0–1,000	18,980
Sodium	1.0–1,000	10,560
Sulfate	1.0–1,000	2,560
Magnesium	1.0–1,000	1,272
Calcium	1.0–1,000	400
Potassium	0.01–10.0	—
Bicarbonate	1.0–1,000	142
Carbonate	0.01–10.0	—
Bromide	0.0001–0.1	65
Strontium	0.01–10.0	65
Boron	0.01–10.0	4.6
Fluoride	0.01–10.0	1.4
Aluminum	0.0001–0.1	0.16–1.9
Iodide	0.0001–0.1	0.05
Silicate	1.0–1,000	0.04–8.6
Nitrogen	0.01–10.0	0.03–0.9
Zinc	0.0001–0.10	0.005–0.014
Lead	0.0001–0.1	0.004–0.005
Iron	0.01–10.0	0.002–0.02
Phosphorus	0.0001–0.1	0.001–0.10
Mercury	—	0.0003

Industrial wastes also reach estuaries either as a direct discharge to the estuary, as spills from vessels carrying materials to or from the industries, as the result of dredging activities, as the discharge of heated effluents from power plants and heated effluents from nuclear power plants which also carry radioactive materials, and in other forms. Most industrial activities involve the use of and/or the disposal of water. Such waters usually contain the by-products of the industrial process and are characteristic of the process. For example in manufacturing steel, a certain amount of water is required for cooling and washing purposes. The amount of water used to produce a ton of steel by a given process is fairly consistent and the quality of the water resulting from the process is also fairly consistent. Table 3 lists various types of industrial activities, the amount of water used in the activity, and the pounds of oxygen required to oxidize the organic material in the wastewater as well as the pounds of suspended solids produced in making some unit amount of product.

Another source of waste material is urban and rural runoff. Urban runoff may consist of storm water runoff from the streets and gutters which is routed to the nearest waterway by storm water pipes, or it may consist of a mixture of storm water runoff and sanitary sewage in what is called a combined sewer system. Such systems are typical of older cities in the United States and other countries which built one pipe to carry both sanitary wastes and storm water wastes.

TABLE 2
Quality of domestic wastes (Concentration units are mg/L)

Constituent	Strong	Medium	Weak	Mass discharge coefficients (9 lbs/person/day)
Solids, Total	1,000	500	200	—
Volatile	700	350	120	—
Fixed	300	150	80	—
Suspended, Total	500	300	100	0.23
Volatile	400	250	70	—
Fixed	100	50	30	—
Dissolved, Total	500	200	100	—
Volatile	300	100	50	—
Fixed	200	200	50	—
BOD (5-Day 20°C)	300	200	100	0.2
Dissolved Oxygen	0	0	0	—
Nitrogen, Total	86	50	25	0.06
Organic	35	20	10	0.035
Ammonia	50	30	15	—
Nitrites (NO_2)	0.10	0.25	0	0.025
Nitrates (NO_3)	0.40	0.20	0.10	—
Chlorides	175	100	15	—
Alkalinity	200	100	50	—
Fats	40	20	0	0.03
Phosphorus	—	—	—	0.012
Flow	—	—	—	135 gal/person/day

Source: Water Encyclopedia, 1971.

TABLE 3
Industrial wastes characteristies

Industry (Unit)	Flow (gal/unit)	BOD (lb/unit)	Suspended solids (lb/unit)
Brewery (Barrel)	370	1.9	1.03
Cannery (Case)	75	0.7	0.8
Dairy (100 LB.)			
Butter	410–1,350	0.34–1.68	—
Cheese	1,290–2,310	0.45–3.0	—
Ice Cream	620–1,200	0	—
Milk	200–500	0.05–0.26	—
Meat Packing (100 LB. live wt. killed)	1,294	14.4	—
Poultry Proc. (1000 birds)	10,400	26.2	—
Petrol, Ref. (Barrel)	100	0.1	—
Pulp and Paper (ton)			
Bleached kraft	45,000	120	170
Bleached sulfite	55,000	330	100
Steel Mill (Injet Ten)	10,000	—	100
Tannery (100 LB.)	660	6.2	13.0
Textile (LB. Cloth)			
Wool	63	0.30	—
Cotton	38	0.16	0.07
Synthetics	15	0.07–0.10	0.02–0.07

Source: Malina, 1970.

TABLE 4

Combined sewer overflow and urban storm runoff characteristics

Constituent	Flow wtd. conc. (mg/l)[b]	Mass discharge (lb/acre/in. runoff)[c]	Coefficients (lb/acre/ year)[d]
BOD	150	30	125
SS	325	70	600
VSS	200	45	180
HEM	50	10	38
TKN	12	3	13
NH_3N	5	—	—
NO_3N	0.2	—	—
Total P	5.0	1	3
Total coli[a]	5×10^5	—	—
Fecal coli[a]	0.5×10^5	—	—

Urban Runoff

Constituent	Flow wtd. conc. (mg/l)	Mass discharge (lb/acre/year)[b]	Mass discharge (lb/acre/in. runoff)[b]	Coefficients (lb/acre/in. runoff)[b]
BOD	18	33	2.0	4.1
SS	77	730	45.6	17.4
VSS	25	160	10.0	5.7
HEM	2.8	—	—	—
TKN	1.2	9	0.56	0.26
TP	0.3	2.5	0.05	0.07
TC[a]	9×10^3			
FC[a]	3×10^3			

[a] Units are MPN/ml.

[b] Data from Weibel *et al.*, 1964.

[c] Data from Spring Creek Project, 1970.

[d] Data from San Francisco, 1967.

TABLE 5

Quality of rural runoff

Source	Total nitrogen (lb/acre/year)	Total phosphorus (lb/acre/year)
Forest Runoff	1.3–3.0	0.3–0.8
Surface Irrigation		
Return flow	2.45–24.0	0.92–3.88
Subsurface Irrigation		
Return flow	38.0–66.0	2.5–8.1
Urban Runoff	8.5	0.8

Source: Fruh, 1968.

Because of economics the pipe could be built just so big, and at the size it could carry all the domestic wastes during dry weather but only a portion of the wastes during wet weather. During a large storm, the pipe would fill to capacity and the flow would have to be diverted to a waterway to insure that backups did not occur in the sewage system. For such drainage systems, each large rainfall results in a certain amount of material being washed into the nearest waterway. The amount of material produced is highly dependent on the drainage system itself, on the use of land in the drainage area, and for some systems on the stage of the tide when the overflow occurs. The quality of such wastewater is given in Table 4 as well as some mass discharge coefficients. A more detailed discussion of this type of wastewater is given in another part of this Encyclopedia (see *Urban Runoff*).

Rural runoff, though less innocuous than urban runoff, cannot be disregarded as a wastewater source. Mass discharge coefficients relating to quality of the runoff water to the land use and amount of rainfall are given in Table 5. Miscellaneous waste discharges occur into estuaries which cannot be quantified in the way done for other types of waste discharges. These include such waste as oil spills, spills of toxic or hazardous materials waste from houseboats or larger vessels with quarters for crew for living purposes, dredging spoils, heat from power or nuclear plants, and other sources.

Another source of material in estuaries is biological recycling. Although biological recycling may also be considered as a sink and will be discussed as such later, recycling of material is extremely important in transforming waste materials from man-made waste discharges (or waste discharges from other lower animals or even dead organisms) to a chemical form in which it may be used again by the biological system. For example, organic material in a domestic waste discharge is oxidized at least partially by bacteria to carbon dioxide and water. Carbon dioxide is a necessary constituent for the growth of plants in conjunction with light. The process of photosynthesis using light reduces carbon dioxide in an organic material which is incorporated in the tissues of the plant. Since plants comprise the basic food stuff for all organisms, the recycling of carbon from the organic form to the inorganic form as carbon dioxide is vitally important. Similarly, nitrogen, which may be in an organic form in a waste discharge, is oxidized by plants for growth. In both cases the organic form acts as a sink while concurrently the inorganic form becomes the source.

Sinks

There are several types of water quality sinks in estuaries. Water withdrawal from an estuary is one type of sink; withdrawals may be made for industrial uses such as salt production or for cooling purposes. In the latter case, some water may be returned to the estuary in a heated condition, but the water withdrawn which is lost in the cooling process as evaporation is lost to the estuary. Sedimentation becomes another sink as molecules of material which were in the water column sorb into the particulate matter which settles out and becomes part of the bottom sediment. Unless the bottom sediment is disturbed, eroded or dredged, this material is essentially lost to the water column. Precipitation may also occur to tie up several minerals which then settle out on the bottom and as with sedimentation are lost from the water column. Precipitation quite commonly occurs in estuaries which have little fresh water inflow but a high evaporation rate. These estuaries are known as hypersaline estuaries because the salinity content rises to levels above that of normal sea water. In the Laguna Madre Bay of Texas, salinity levels reach two or three times that of normal sea water. In this bay, crystals of gypsum are found on the shores.

Another type of sink which is tied intimately to the biological system of the estuary is the degradation of materials, that is, the change from one chemical form to another by biological action. Above, the process of oxidation of organic material was mentioned. This is one form of oxidation in which organic material is oxidized to smaller molecular material, which is further oxidized to carbon dioxide and water as the ultimate inorganic products. Oxygen is consumed in the process and is lost to the water. Another oxidation process is nitrification. In this process, nitrogen in the organic form is oxidized to ammonia which is then further oxidized by the bacterium, *Nitrosomonus* to nitrite and further to nitrate by the bacterium, *Nitrobacter.* Under low oxygen or anaerobic conditions, the nitrate or nitrite may be reduced by bacteria to elemental nitrogen gas which may then be removed from the system according to the solubility of the nitrogen gas.

QUALITY NEEDED FOR ESTUARY USE SUSTENANCE

Now that some of the hydraulic, biological, and man-made mechanisms which influence water quality in estuaries have been described, the use concept for estuaries and the water quality needed to sustain uses should be considered.

As an example, the San Francisco Bay-Delta Study Report (1969) listed some of the uses which can be made of estuaries, and these are given in Table 6. These uses are quite varied and in some cases would be competing uses if applied to the same part of an estuary. For combinations of uses for a given area in an estuary, usually one use will require a water quality higher than the other uses, and this use will dictate the water quality needed in that particular area. Some of the uses listed are not really dependent on water quality such as shipping, unless the quality is particularly adverse (very acid water or large floating material).

The general philosophy of estuary use in particular or resource use in general is that a range of uses may exist bounded by two extreme views. These views are: (1) a resource may be used indiscriminately without regard to the consequences (e.g., total consumption, contamination, etc.) of that use; and (2) a resource should be preserved with no interaction with man. Both views, despite being held by some are recognized as incompatible, and the concept of sustained use, the designation of some use or uses for a resource area and management of the system to support that use best, is favored. As mentioned at the beginning of this article, the designation of uses for estuarine areas in all states through public hearings was a recognition of the sustained use concept. However, subsequent attempts to manage waste inputs to estuarine systems and the estuarine systems themselves have shown how difficult management is. The difficulties lie in the technical aspects of management (the interrelatedness of water quality with estuarine circulation and sources and sinks), the economic aspects (the cost of management systems directly and indirectly),

TABLE 6
Beneficial uses to be protected in San Francisco Bay

- Municipal Supply (seasonal)
- Industrial Supply
  - Boiler
  - Cooling
  - Rinsing
  - Processing
- Agricultural Supply (some seasonal)
  - Irrigation
  - Liverstock
- Fish and Wildlife Propogation and Aquatic Growth
  - Fish habitat, migration, spawning
  - Shrimp and crab habitat
  - Shellfish habitat
  - Waterfowl habitat
  - Mammal rookery
  - Kelp
- Comercial Fishing and Shellfishing
- Recreation
  - Swimming, waterskiing, skindiving, picnicking
  - Surface, beachcombing, sunbathing
  - Pleasure boating
  - Fishing
  - Shellfishing
  - Hunting
- Enjoyment of Esthetic Values
- Tidepool and Marine Life Study
  - Navigation
  - Resource Extraction

and the socio-political-legal aspects (the interrelatedness of social desires and political decisions and the lack of adequate pollution control laws).

For each of the uses listed in Table 6, a list of physical, chemical and/or biological properties or constituents may be compiled. This list would indicate the levels of the properties or constituents which have to be provided or cannot be exceeded if that use is to be sustained. These constituent lists have changed as they have been found to be unimportant or found to be needed, and as the levels have changed new knowledge has been gained about how the levels more specifically affect the desired use. The extension of water quality criteria to toxic materials in recent years (U.S. Environmental Protection Agency 1976, 1986) indicates how emphasis has shifted from one set of criteria for which waste discharge problems have been solved essentially to another set which is comprised of substances of major concern today.

Unfortunately at this time, very little is known about the bases for many levels set. For example, good epidemiological evidence relating fecal or total coliform organism concentrations in swimming areas to disease is still lacking. A larger gap, however, is the criteria list for the fish and wildlife of estuaries and the organisms they feed on. For these organisms the best information available pertains to temperature, dissolved oxygen, pH and salinity. Beyond these criteria, the levels set for other materials such as toxic substances are being developed (U.S. Environmental Protection Agency 1986).

The reader may obtain more specific information about water quality in general and estuarine water quality in particular by referring to the bibliography references and to other related material. Of special interest might be the reports of studies on major estuaries such as San Francisco Bay, Galveston Bay, Delaware Bay, and Jamaica Bay. The studies include not only the theory of water quality but the practical techniques for and applications of its management.

REFERENCES

1. Characterization and treatment of combined sewer outflows, report to Federal Water Pollution Control Admin. by City of San Francisco. Prepared by Engineering-Science, Inc., November, 1967.
2. Feigner, K.D. and H.S. Harris, “Documentation Report—FWQA Dynamic Estuary Model,” U.S. Department of the Interior, FWQA, 1970.
3. Fischer, H.B. Transport Models for Inland and Coastal Waters, Academic Press, 1971.
4. Fischer, H.B., E.J. List, R.C.Y. Koh, J. Imberger, and N.H. Brooks, Mixing in Inland and Coastal Waters, Academic Press, 1979.
5. Fruh, E.G., in *Advances in Water Quality Improvement,* Eds. E.F. Gloyna and W.W. Eckenfelder, Jr., Univ. of Texas Press, Austin, 1968.
6. Geyer, R.A., in *The Biological Significance of Estuaries,* Sport Fishing Institute, March, 1971.
7. Leendertse, J.J., RM-6230-RC, The Rand Corporation, February, 1970.
8. Leendertse, J.J., Rand Corp. Publication RM-6230-RC, Febr. 1970.
9. Leendertse, J.J. and E.C. Gritton, The New York City Rand Institute, R-708/6-RC, July 1971.
10. Leendertse, J.J. and S-K. Liu, A Three Dimensional Model for Estuaries and Coastal Seas: Vol. II, Aspects of Computation, R-1764-OWRT, The Rand Corporation, June 1975.
11. Liu, S-K. and J.J. Leendertse, Modeling the Alaskan Continental Shelf Waters, R-3567-NOAA/RC, October 1987.
12. Malina, J.F., Jr., Report No. CRWR 69, Center for Research in Water Resources, University of Texas, October, 1970.
13. Masch, F.D. and R.J. Brandes, Tech. Rept. HYD 12-7102, Hydraulic Engineering Laboratory, University of Texas, August 1971.
14. McHugh, J.L., in *A Symposium on Estuarine Fisheries,* Special Publ. No. 3, 133154. Amer. Fish. Soc., Wash., D.C., 1966.
15. McHugh, 1967, in *Estuaries,* Pub. No. 83, AAAS, 1967.
16. National Estuarine Pollution Study, U.S. Gov’t. Printing Office, 1971.
17. O’Connor, D.J. and R.V. Thomann, in *Estuarine Modeling: An Assessment,* Environmental Protection Agency, Project 16070 DZV, February, 1971.
18. Oey, L-Y., A three-dimensional simulation of the Hudson-Raritan estuary, J. Phy. Ocean. 15(12), 1985.
19. Pritchard, D.R., in *Estuaries,* Pub. No. 83, AAAS, 1967.
20. Russell, J.J., in *Estuaries,* Pub. No. 83, AAAS, 1967.
21. San Francisco Bay-Delta Water Quality Control Program, final report to the State of California by Kaiser Engineers and Associated Firms, March 1969.
22. Stroud, R.H., in *The Biological Significance of Estuaries,* Sport Fishing Institute, March, 1971.
23. Thomann, R.V. and J. Mueller, Principles of Surface Water Quality Modeling and Control, Harper & Row, 1987.
24. U.S. Environmental Protection Agency, Quality Criteria for Water, 1976.
25. U.S. Environmental Protection Agency, Quality Criteria for Water, 1986.

26. U.S. Environmental Protection Agency, The National Estuary Program, 1989.
27. Ward, G.H. and W.H. Especy, in *Estuarine Modeling: An Assessment,* Environmental Protection Agency, Project 16070 DZV, February 1971.
28. Water Quality Criteria, Report of the National Technical Advisory Committee of the Federal Water Pollution Control Administration to the Department of the Interior, April, 1968.
29. Weibel, S.R., R.J. Anderson, and R.L. Woodward, *J. Wat. Poll Contr.* 36(7), 914, 1964.
30. *Water Encyclopedia,* David K. Todd, Water Information Center, Port Washington, N.Y., 1970.

NEAL E. ARMSTRONG
University of Texas at Austin

MONITORING OF STACK EFFLUENTS: see STACK SAMPLING

MUNICIPAL WASTEWATER

Sewage is the spent water supply of a community. Because of infiltration of groundwater into loose sewer pipe joints, the quantity of groundwater is frequently greater than the quantity of water initially consumed. Sewage is about 99.95% water and 0.05% waste material.

A weaker (more dilute) sewage results from greater per capital water consumption. Industrial wastes will contribute to sewage strength. Sewage flow varies with time of day and will be higher during daylight hours. Larger cities will exhibit less variation with time than smaller cities. Many small communities will have a flow in the late night hours that is due almost entirely to infiltration.

Per capita production will vary from less than 100 gallons per day for a strictly residential community to about 300 gallons per day for a highly industrialized area. The concept of "population equivalent" is frequently applied in evaluating industrial waste contributions to sewage flow and planning for hydraulic, solids, and biochemical oxygen demand loadings.

Wastewater treatment facilities have high initial capital costs due to the need for large tanks, equipment and land area. The designed life of a treatment plant is normally equal to the life of the bonded indebtedness of the project. It is expected that capacity will be just reached at the end of this period. In a community with combined sanitary and storm sewers it is often found necessary to bypass wastewater flows during periods of heavy rain or, in low lying areas, during high tides. Excess water may be held in detention basins until normal flow resumes and then treated in the plant. Bypassed flow can be screened and chlorinated before discharge. This subject is receiving increasingly close attention.

Strength of sewage is usually expressed in terms of the following parameters: total solids (sometimes called volatile suspended solids, VSS), filterable solids, and biochemical oxygen demand (BOD). Many methods are available for determination of the foregoing. However, in order that operating and research data reported by workers in widely separated areas be comparable, it is necessary to have analytical methods which are simple, give consistent results and are widely accepted. This need is filled by Standard Methods for the Examination of Water and Wastewater. Contained in each edition are methods for analysis of water and wastewater which have been accepted by committees of experts in various fields. Standard Methods is revised about every five years.

Analysis of a typical American sewage is given below:

Total solids	600 mg/l
a) suspended solids	200 mg/l
1) settleable solids	120 mg/l
2) colloidal solids	80 mg/l
organic	60 mg/l
mineral	20 mg/l
b) Filterable solids	400 mg/l
1) colloidal solids	40 mg/l
organic	30 mg/l
mineral	10 mg/l
2) dissolved solids	360 mg/l
organic	100 mg/l
mineral	260 mg/l
c) BOD (5 days 20°C)	
1) suspended	42 gms/cap. × day
settleable	19 gms/cap. × day
non-settleable	23 gms/cap. × day
2) dissolved	12 gms/cap. × day
	Total 54 gms/cap. × day

Assuming an average daily flow of 100 gallons per capita, 54 gms/cap. × day = 0.12# BOD/cap. × day. This is a strictly domestic sewage. Per capita BOD values can vary from about 0.10#/day to above 0.25#/day. A commonly accepted value used for estimation is 0.17#/cap. × day.

In the above analysis, the determination of solids, mineral and organic, gives an indication of the loadings to be placed on the plant. Volatile solids give the organic loading and mineral solids are those which must be accommodated by sedimentation equipment.

BOD is a measure of the loadings placed on the oxygen resources of the receiving waters. Several methods have been used over the years for determination of the amount of oxygen necessary for stabilization of the waste. Permanganate and other oxidizing agents were formerly used. One method still finding application, but only as a general indication, is the putrescibility, or methylene blue, test. Methylene blue, a dye, decolorizes in the absence of dissolved oxygen. This test is only quasiquantitative, but finds use in day to day operation control of wastewater treatment plants. The method used today is an incubation and dilution method. Dissolved oxygen is determined at the beginning and end of a 5 day period by the Alsteberg azide modification of the Winkler method.

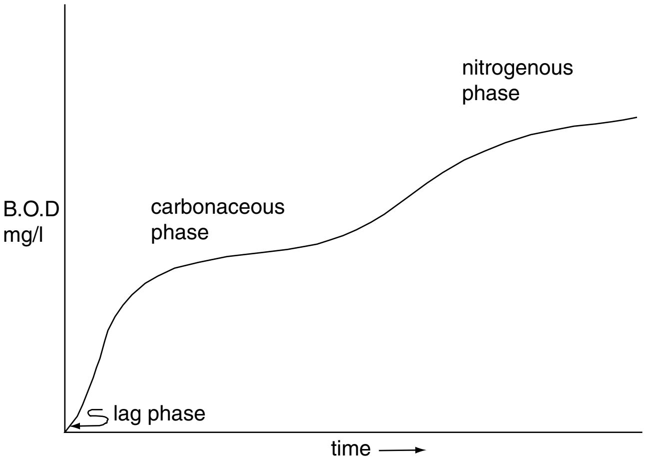


FIGURE 1

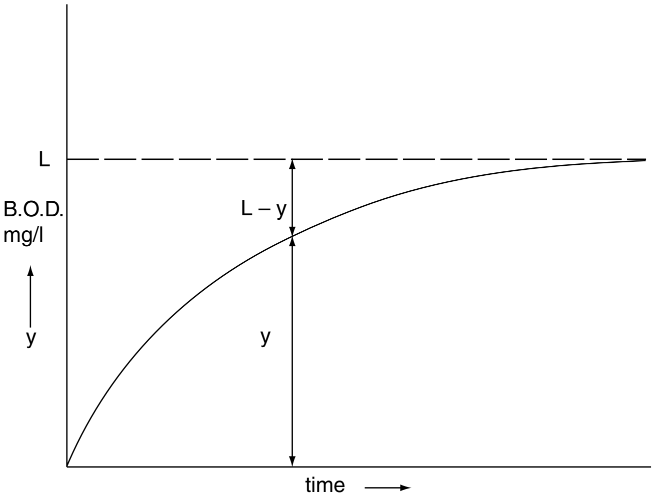


FIGURE 2

Water can hold about 10 mg/l dissolved oxygen. Since BOD values can vary from less than 1 mg/l for a clean stream to many thousands of milligrams per liter for some industrial wastes, it is necessary to dilute stronger wastes.

The dissolved oxygen concentration of a seeded nutrient dilution water is determined. A measured quantity of waste is then added to a known volume of dilution water. Total sample volume is usually about three hundred milliliters. The mixture is then incubated at 20°C for five days. At the end of this period the dissolved oxygen concentration of the mixture is again measured. BOD is then equal to the reduction in dissolved oxygen divided by the dilution factor.

Figure 1 shows the exertion of BOD over an extended period (more than 50 days). The trajectory displayed is one which might be followed when waste is added to a relatively clean stream and the watercourse acts as an incubation bottle. Such a complex reaction is of little utility for day to day control. In order to make the concept of BOD more useful, a simplified model is developed. The path followed in the idealized model is displayed in Figure 2.

The lag phase is eliminated by use of seeded dilution water. In this phase, microorganisms which utilize the waste material for food are becoming acclimatized and assuming dominance in the system. In the carbonaceous phase the dominant organisms are feeding upon the more easily degraded compounds. The nitrogenous phase, in which the character of the food and the microbial population change, does not usually begin for at least ten days. This portion is of interest in research on kinetics but can be ignored in the following. It is hypothesized that a simple monomolecular rate constant can describe the carbonaceous phase and the reaction is asymptotic to a limiting value L at infinite time. L is known as the ultimate BOD and is a mathematical artifact. The rate of reaction is assumed proportional to the BOD still remaining,

$$\frac{dy}{dt} = k_1'(L - y).$$

Integration and taking of proper limits gives

$$y = L(1 - e^{-k_1' t}) = L(1 - 10^{-k_1' t}).$$

Phelps' relation for decolorization of methylene blue with time

$$\frac{y}{L} = 1 - 0.794^t$$

is equivalent to the model equation when k_1 equals 0.10/day.

In practice, the 5 day 20°C BOD is used to describe the strength of a waste. Simple enumeration of the BOD value tells nothing about the path by which the number was reached since L and k_1 are not specified. Sewers and treatment plants are not isothermal entities but it is necessary that a common basis be established. One of the reasons advanced for the choice of the five day period is the fact that almost all rivers in England are within five days flow of the sea. Waste once discharged to the infinite ocean is no longer of interest. Times have changed and there is great concerns for effects of waste on the "infinite ocean."

In the literature reference to BOD means 5 day 20°C BOD (BOD_5) unless specifically stated otherwise. Recently, some popular writers have erroneously used the term "biological oxygen demand." There is no such parameter.

The reaction rate constant k_1 actually describes a series of complex microbiological reactions. In the initial stages of biological stabilization of sewage, carbonaceous material is oxidized to carbon dioxide and nitrogenous material is oxidized first to nitrite and then nitrate. One of the measures of degree of stabilization of organic matter in an effluent is the nitrate concentration. However, recent theories concerning eutrophication have raised serious questions concerning the desirability of high effluent nitrate concentrations.

Sewage contains the waste of a city and its inhabitants. It is possible to find almost any microorganism in sewage. In fact, in the sewer system can be found quite unexpected creatures. About 1960, sewer workers in New York City found a number of live and exceedingly unfriendly alligators in that city's sewers. These animals, gifts to city residents, had been disposed of through household toilets. The warm, wet environment of the sewers, rich in food, was excellent for rapid growth.

Traditionally, public health practice applied to water sanitation has placed major emphasis on prevention of typhoid fever, the causative organism of which is S. typhosa. This disease is spread by the pathway of anus to mouth. S. typhosa is a fairly delicate organism and is usually not found in high concentrations. It is felt that if pollution arising from human intestinal discharges is removed to a high degree, then S. typhosa will also be removed. As BOD is used as an indicator of pollutional loading on oxygen resources, so is the indicator organism E. coli used as an indication of pollutional loading due to human intestinal waste. E. coli is a normal inhabitant of the human intestinal tract. It is excreted in huge numbers and the presence of this organism in receiving waters may be evaluated quantitatively. The lactose broth-brilliant green bile test is commonly used to determine the presence of E. coli. A common soil organism, A. aerogenes, gives a false positive test and further confirmatory tests are necessary. Serial dilutions of the water being tested are prepared and, on the assumption that one organism is responsible for a positive test, the Most Probable Number (MPN) of organisms is determined. The MPN is based on statistical reasoning. Work by Kupchick, using the enzyme urease to determine the presence of uric acid, has shown a high degree of correlation between the concentration of this acid in wastewater and the MPN.

Most pathogenic organisms are not hardy and do not compete well for food. The use of E. coli as an indicator organism is in the way of a margin of safety. This is consistent with Phelps' concept of multiple barriers.

Microorganisms form the basis of secondary, or biological, wastewater treatment. Stated briefly, microorganisms establish themselves on trickling filter slime or in activated sludge liquor and feed on waste material in the sewage.

Large particles are removed in the primary settling portion of treatment. The larger particles, grease, etc. are in some ways not as objectionable as the colloidal or truly dissolved materials. The larger particles are, of course, objectionable from an aesthetic point of view, but the smaller particles place more of an immediate load upon the receiving waters. In almost all areas of environmental engineering surface area is one of the controlling parameters. Microorganisms, in carrying out their vital processes, utilize the waste as food and the smaller particles offer greater surface area per unit volume. Microbial activity is correspondingly greater and the oxygen required is also greater.

Microorganisms require a readily available source of oxygen. Aerobic conditions are said to exist when the oxygen is in the form of dissolved molecular oxygen or nitrates. At the point of exhaustion of nitrates and nitrites the system is in the anaerobic state and the oxygen sources are then sulfates, phosphates, borates, etc. Reduction of sulfates to give mercaptans (HS^-) carriers with it characteristic vile odors. Secondary treatment is an essentially aerobic process while conventional sludge digestion is an anaerobic process.

In the trickling filter an activated sludge processes microorganisms extract their food from the flowing wastewater. By means of extracellular enzymes large molecules are broken down so that they may pass through the microbial cell wall by diffusion. The food is further broken down for cell synthesis and energy by means of intracellular enzymes. End products are largely carbon dioxide and water. The waste material, now part of the cell mass, is removed in the final stage of treatment.

The primary reason for treating any waste is the need for avoiding nuisance and dangerous conditions in the environment. It is necessary to remove some of the waste so that the remaining can be discharged to the surroundings. This is, in effect, disposal by dilution. Discharge standards are usually based on concentration and total quantity.

Sewage purification works were formerly constructed for reasons based primarily on public health. More modern thinking has expanded the original rationale to include protection of oxygen resources of the receiving waters. If the second criterion is satisfied, the first will almost always be also satisfied. Demand for clean waters is increasing even more rapidly than population. Water will be used many times in its passage to the sea. Lakes are essentially a closed system. Leisure time is increasing and the question of water pollution has reached the point where aesthetics is now a significant parameter in planning. It is no longer possible to treat a watercourse as a separate entity. The approach now finding wider and wider application is that of basin management. This is the systems approach. An excellent example of this is the Ohio River Sanitation Commission. Here it has been possible to rise above local jealousies and self interest. The results are most gratifying and should serve as an example to other areas. It is unfortunate, but true, that one heavy pollution source can undo the efforts of many groups with social responsibility and foresight.

Because objectives in waste management can change drastically at political boundaries, it has occasionally been necessary for intervention by representatives of larger political entities when pollution problems effect, for example, several states. Federal agencies in the Unites States and other countries have come to play an increasingly larger role in waste management, particularly when problems do not respect political boundaries. The discussion presented here involves primarily technology of wastewater treatment and the underlying philosophy will not be extensively covered.

It has been mentioned that larger particles in wastewater can be removed by physical means. Removal of colloidal and dissolved components requires other methods and this must be accomplished economically. Treatment is classified as primary if it is without biological basis. Secondary treatment is generally accepted as biological treatment. A few regulatory authorities have endeavored to classify treatment degree on the basis of degree of removal. This relegates some biological treatment to primary status. Many experts do not agree with this change in definition. A widely accepted definition of tertiary treatment is the use of any process, in addition to conventional secondary treatment, for the purpose of further removals. From the foregoing it should be noted that chemical treatment, popular in the past, is a primary treatment process.

Water carriage of waste was practised in the Minoan civilization of Crete. It may be said that sanitation practices reflect the level of a civilization. Certainly, sanitary practices of the Middle Ages were of a rather low order. Methodical control of waterborne diseases was not attained until the 19th century. It is of historical interest to note that knowledge of the use of creosote for odor control at the Carlisle, England sewage purification works gave inspiration to Joseph Lister for the birth of asepsis in surgery.

The process by which wastewater is purified can best be understood by following the waste from its source through the collection system and treatment plant. Organic material discharged to a watercourse will eventually undergo stabilization. This is accomplished by natural processes (unit operations) and a wastewater treatment plant basically sets up, under controlled conditions, the processes which act in the river. Indeed, it has often been said that a sewage treatment plant is a river in miniature. In the river heavy particles settle out and lighter particles float to the surface. Biological decomposition takes place. Oxygen present is used by organisms that accomplish decomposition. Some of the settled material will be resuspended, increasing the organic loading. While oxygen is being withdrawn by BOD this resource is being replenished at a rate proportional to the deficit. The oxygen deficit is the difference between the amount that can be held at saturation (about 10 mg/l) and the amount actually present.

Stated mathematically, the oxygen concentration in a river as a function of time is

$$\frac{\mathrm{d}D}{\mathrm{d}t} = k_1' y - k_2' D.$$

This expression considers only the effects of BOD and atmospheric reaeration. A more complete equation can be written but effects of oxygen production by algae and oxygen reduction by benthal (bottom) deposits are not of major significance here. Solving the above gives

$$D_t = \frac{k_1 L}{k - k_1}\left\{10^{-k_1' t} - 10^{-k_1 t}\right\} + D_0 \times 10^{-k_2 t}.$$

This expression is commonly known as the oxygen sag equation and is displayed in Figure 3. It is of interest here because it illustrates the rationale underlying waste treatment requirements. Waste is treated so that undesirable conditions do not develop in the receiving waters. In effect, a limit has been placed on the allowable oxygen deficit.

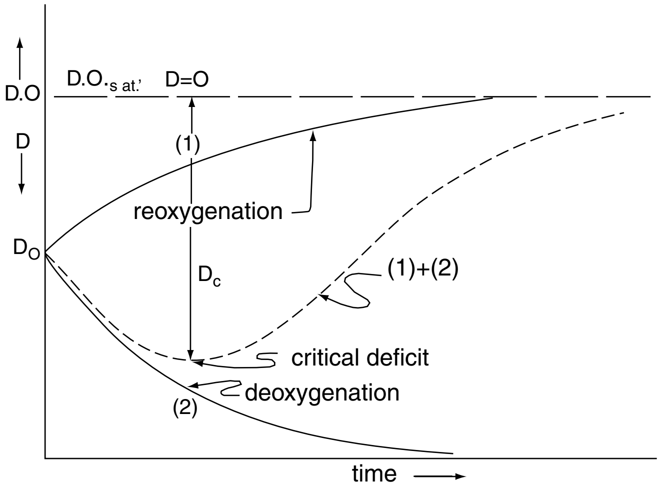


FIGURE 3

Regulatory authorities usually require that a minimum dissolved oxygen level be maintained. Normally, this will be stated as a percentage of dissolved oxygen saturation. This is the largest permissible critical deficit. Game fish may require a minimum of 5 mg/l D.O. while scavengers can survive in a much lower quality water.

The critical deficit is given by

$$D_c = \frac{k_1 L \times 10^{-k_1 t_c}}{k_2}.$$

Once the maximum deficit is specified, the BOD loading on the stream can be immediately estimated. The allowable BOD loading will be that impressed on the watercourse by the wastewater treatment plant effluent. It is of interest to observe the effect of the condition of the river at the point of discharge. A river in poor condition will have a large initial deficit, D_0. This can raise the treatment requirements. It can be seen that it is necessary to integrate the efforts on a basin wide basis. Parameters other than just dissolved oxygen must be controlled by the treatment processes. This is accomplished in some, or all, of the following steps.

Decomposition of the waste begins in the collection system. Ease, or difficulty, of treatment depends to a large degree on the condition of the sewage when it reaches the plant. Some substances are not permitted in the sewage system. Gasoline and other flammable substances, oil, hexavalent chromium are examples of prohibited substances. These can damage either the collection system or the treatment plant and processes. The legal vehicle by which such materials are excluded is known as a sewer ordinance.

It is most economical to collect sewage by gravity flow. If topography does not permit, pumping will be necessary in order to cross high points and to avoid excessively low flow velocities and deposition of waste material in the pipes.

A combined system will, at times, exceed the hydraulic capacity of the treatment plant and flow must be either bypasses or held in detention tanks until heavy flow has subsided.

Multiple treatment units are provided and treatment is not interrupted during periods of maintenance or repairs. Protection must be provided for pumps against large objects, such as floating pieces of wood. Coarse racks, with clear openings of more than 2 inches, may be placed at the entrance to the plant. Racks to be placed in advance of grit chambers and settling tanks will have clear openings of 1 to 2 inches. In smaller plants racks are cleaned by hand while larger plants have mechanically cleaned racks. Disposal is by burial, incineration or digestion with sewage sludge. Mechanically cleaned racks have smaller clear openings because head losses are lower with continuous cleaning. Racks with clear openings of 1 to 2 inches can be expected to give from 20 to 100 ft^3 of solids per 1000 people annually. Comminutors macerate floating material into sizes sufficiently small to be easily handled by centrifugal pumps. Racks and screens with very small openings have been almost completely replaced by comminutors.

Rate of flow into the plant will vary over a wide range during any 24 hour period with smaller plants exhibiting greater variation. Flow is measured by Parshall flume, Venturi meter or Sutro (keyhole) weir. The Parshall flume, sometimes called the open channel venturi, is most commonly used. The device operates on the principle of critical flow and measurement of water depth upstream of the flume throat. The governing equation is of the form $Q = cWH^{3/2}$, where Q is the discharge per unit time, W is the throat width, H is the water depth, and c is a constant. While c changes with throat width, it is closely constant for a constant throat width. Flume liners of reinforced fiberglass are all but replacing steel and concrete liners. Ease of fabrication, close tolerance, and corrosion resistance are advantages cited. The true Venturi operates on closed pipe (pressure) flow and is usually found in larger plants. The Sutro, or keyhole, weir is shaped as its name suggests. Its principal advantage is maintenance of a constant upstream velocity over a wide flow range, but it does have a high energy loss and metering is lost when the opening is submerged.

It is necessary to remove grit in order to protect pumps against excessive wear and to maintain capacity of sludge digesters. It has been found that digesters in plants serving low lying sandy areas can, if grit removal is not efficient, lose up to a third of capacity in but a few years. Grit chambers must operate in a fairly narrow velocity range of from 0.75 to 1.25 ft/sec. Above this range deposited material is scoured back into suspension and below the lower value organic material settles out. The resulting material, called detritus, is unsuitable for landfill uses due to its highly putrescible nature. Grit chambers are usually designed to remove particles with a specific gravity of 2.65 and a mean diameter of 0.02 cm. Because flow variation with depth follows a parabolic function, $Q = cWH \cdot H^{1/2}$, the grit chamber is often given an approximately parabolic shape and better velocity control is attained. The amount of grit collected per million gallons flow is found to vary from 1 to 12 cubic feet. Grit is removed manually in small plants and continuously by mechanical means in larger plants.

Settling tanks are provided for removal of larger, heavier organic particles, oil, and grease. Oil, grease, and other materials lighter than water are skimmed continuously from the surface and led to digestion. Both circular and rectangular surface configurations are used. Rectangular tanks of the flow through variety have length to width ratios of 4/1 to 6/1. Circular tank size is usually limited by structural requirements of trusses carrying skimming devices. Tank depths vary from 7 to 15 feet. Bottoms are sloped about 1% in rectangular tanks and about 8% in circular tanks to facilitate sludge removal.

Design is on the basis of hydraulic loading. A commonly used figure is 1000 gal/day × ft^2 surface area. It can be expected that a BOD removal of 30% will be achieved in a well operated primary sedimentation unit.

If treatment includes only screening, sedimentation, and chlorination of effluent, the treatment is classed as primary. Primary treatment, while inadequate for most areas, is better than no treatment. The adequacy of secondary treatment is now being seriously questioned. Nonetheless, it fits the economics of the situation. Only in the 1960's did wastewater treatment become of interest to any but a small number of people. Sanitary engineers were wont to say "It may be sewage to you but it is bread and butter to me." Theirs was not a profession to which much glamour was attached. Financing bodies were reluctant to invest adequate sums in waste treatment facilities. Hopefully, this has now changed.

There are two main processes utilized for biological (secondary) treatment. These are (1) the trickling filter and (2) activated sludge.

The trickling filter is not a true filter. It can best be described as a pile of stones, or other coarse material, over which sewage flows. This is the most widely used biological treatment process. Present day biological treatment technology is a logical development from sewage farms (irrigation areas) to intermittent sand filters to contact (fill and draw) beds to trickling filters and activated sludge units. Numerous modifications of the basic processes have evolved but the underlying principles remain unchanged. In biological treatment a suitable environment is provided so that microorganisms may thrive under controlled conditions. The suitable environment is one rich in food and maintained in the aerobic state. The zoogleal mass remains fixed on the filter media in the trickling filter while the sewage flows past. In the activated sludge process the sewage and organisms flow together. In both cases the microorganisms come from the sewage itself. Traditional secondary treatment plants operate in the declining growth phase.

Irrigation by sewage provides water return and some waste stabilization but this means of sewage disposal is in conflict with sound public health practice and ought not be used where there is a possibility that sewage can pass with little change into the groundwater table. Irrigation is best applied in arid regions. When it is utilized for food growing areas, care must be taken so that edible plants and fruit are not contaminated. Odors are a problem and removals decline markedly in cold weather.

Intermittent sand filters are much like the slow sand filters used for potable water production. The sewage is applied to a sandy area and allowed to percolate downward. Raw sewage may be applied at rates as high as 80,000 gal/acre × day and secondary effluents at rates as high as 800,000 gal/acre × day. Application of the secondary effluent would be considered tertiary treatment. Biological films that form on the sand grains undergo continuous stabilization. It is necessary to rest the bed between dosings so that objectional conditions do not develop. Surface accumulations of solids must be periodically removed. This method is not recommended for areas underlain by fissured limestone.

Fill and draw beds operate as the name indicates. A tank, packed with coarse material, is filled with sewage and allowed to stand full. It is then drained and allowed to rest. Air is drawn into the bed during filling and emptying. Loadings are about 200,000 gal/acre × ft × day. There is little application of this method today in treatment of municipal wastewater but it does find use in industrial waste treatment. Fill and draw beds are a batch operation and the trickling filter is a continuous operation.

Sewage is distributed over a trickling filter by slowly revolving arms equipped with nozzles and deflectors. Some earlier plants had fixed nozzles but this is no longer done. Revolving arms are driven by hydraulic head. Sewage discharged is allowed to flow slowly downward through the bed. Air is down into the bed by temperature differential, thus maintaining a supply of oxygen for the process. Filter media is usually stone. Sizes are in the range of 1 to 4 inches. Packing of this size permits air to be drawn into the bed and the bed is not clogged by biological slime. There appears to be a trend toward more use of plastic filter media. Filter depths range from 3 to 14 feet. A common depth is 6 feet. After passage through the filter the sewage is collected in tile underdrains. These underdrains serve two purposes: (1) collection of filter effluent and (2) circulation of air into the filter. The underdrains discharge to a main collection channel which, in turn, discharges to the final settling (humus) tank. The importance of the function of the final settling tank can be seen by an examination of what occurs in the filter itself.

A new filter is "broken in" by applying sewage as in normal operation. After a time the microbial (zoogleal) mass establishes itself on the filter media and carries on the work of waste stabilization. Waste material in the flowing sewage (food) is first absorbed into the zoogleal mass and then assimilated by the microorganisms. Much of the organic waste material has, at this point, been utilized for cell synthesis and energy. There must be continuous removal of filter slime or the process becomes sluggish due to a lower feeding rate of old organisms. Since waste material is now a part of the filter slime there must be a means provided for removal of sloughed off organisms or the waste material, now in different form, would still appear in the plant effluent and little constructive would have been accomplished. The required removal is carried out in the secondary settling tank. A rate of application lower than that of the primary tank is necessary here because of the different character of the material to be removed. Rates in this portion of the system are in the range of 600 gal/ft^2 × day.

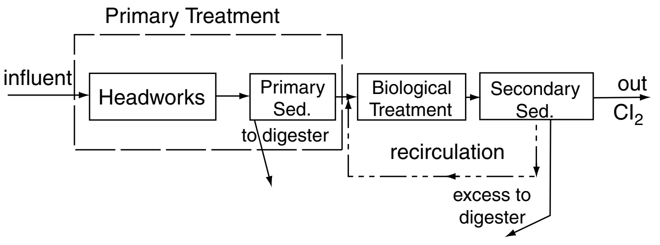


FIGURE 4

A portion of the effluent is recirculated, as shown in Figure 4. This is done in order to (1) smooth out flow, (2) keep the food concentration more constant, (3) lower the film thickness and, thus, control the psychoda fly, and (4) reseed the applied sewage with acclimatized organisms. The psychoda, or filter, fly is a very small insect which breeds in trickling filter slime. It does not bite but can be extremely bothersome because it does get into the nose and mouth. The range of flight is short but the creature can be carried great distances by the wind. Control of the fly in its developmental phase can be achieved by flooding the filter periodically or through chlorination of influent.

Trickling filters can be classified on the basis of (1) hydraulic loading per unit area and (2) applied pounds of BOD per 1000 ft^3 of filter volume.

The low rate trickling filter, with hydraulic loadings of 2 to 4 million gallons per acre per day (mgad) and 10 to 20 pounds BOD per 1000 ft^3, is usually found in use in smaller plants. With proper operation, BOD removals of 80 to 85% can be routinely expected. Raising the applied sewage to 10 mgad produced greater BOD removals per unit filter volume but the effluent organic concentration was found to be high. Influent organic concentration was reduced by greater effluent recirculation and lower effluent organic concentration was realized. Units that operate in the 10 to 40 mgad range are called high rate trickling filters. BOD loadings are up to 90 pounds per 1000 ft^3, but removals to be expected are in the range of 65 to 75%.

In the 4 to 10 mgad range operational difficulties were frequently encountered and this range was avoided for many years. It appears that, in this range, the hydraulic application was inadequate to keep the filter slime from attaining excessive thickness. Many plants had operated well in this range, but other plants had many problems. The solution seems to have been reached with use of relatively large, 2 to 4 inches, filter stones.

Experimental plants using plastics media have recently achieved very high removal efficiencies (97%) at hydraulic loading rates of 100 mgad. Much of the microbial mass is in the recirculated effluent and these plants are, in effect, modifications of the activated sludge process. Organic loadings are in the region of 100 pounds BOD per 1000 ft^3 filter volume.

Filter packing by plastic media is finding wider application. Design criteria, however, call for quite deep filters and this appears to be uneconomical in terms of power requirements.

Modifications of the trickling filter process over the years have dealt with improvements in media, air circulation, and loadings. One of great interest is that proposed by Ingram. Sewage is introduced at various levels in a very deep filter in an attempt to distribute the load more uniformly over the whole filter depth. Hydraulic loadings up to 500 mgad have been successfully achieved.

A new development in trickling filter technology is the Rotating Biological Contractor (RBC). A rotating drum is partially immersed in wastewater. A zoogleal mass develops on the drum surface, functioning in the same manner as trickling filter slime. Such installations may be completely enclosed in plant buildings, thus avoiding any effects of extremes in outside temperature. An excellent example of the application of this process for upgrading the municipal treatment works at North Bergen, New Jersey.

Activated sludge serves the same function as trickling filter slime. The major difference lies in the filter slime being fixed to the filter media while the activated sludge is carried along with the flowing wastewater. Development of the activated sludge process began with attempts to purify sewage by blowing air into it. It was observed that after prolonged aeration, flocs composed of voraciously feeding organisms developed. This floc settled after aeration was stopped. Addition of fresh sewage to tanks containing the settled sludge produced high purification in a practical time. The name activated sludge was assigned this means of waste treatment. At first, this was operated as a fill and draw system. Research showed that continuous operation could be practiced and this is the present means or operation. The process involves:

1) return of activated sludge to the aeration tank influent and discharge of excess sludge to digestion
2) aeration of the sludge-sewage mixture to maintain purification and
3) settling of the aeration tank effluent to remove floc before final discharge.

Step (3) is necessary for the same reason as the comparable portion of the trickling filter process-removal of waste material transferred to the microbial cell mass. Floc is formed in the tank through aerobic growth of unicellular and filamentous bacteria. Protozoa and other organisms will also be found in the floc. This is a strictly aerobic process and air requirements are high. Two aeration systems are used, (1) diffused air units and (2) mechanical aeration. Air diffusers are more commonly used in North America but mechanical aeration systems may be found in plants of less than 1 million gallons per day (mgd) capacity. Both methods of aeration perform three functions, (1) transfer of oxygen to the mixture and maintenance of aerobic conditions, (2) intimate mixing of floc and sewage, and (3) keeping the floc in suspension.

Aeration tanks are normally rectangular in cross section, 10 to 15 feet deep and 30 feet wide. Length to width should be greater than 5 to 1 in order to avoid short circuiting. Detention periods are from 4 to 6 hours. Air is introduced from diffusers in such a way as to set up a spiral flow pattern, thus aiding in mixing of floc and sewage and helping to prevent dead spaces in the tank. It was found that oxygen requirements decreased as the waste proceeded through the tank. The number of diffusers was, therefore, increased at the beginning of the unit and decreased at the effluent end. This is now the common practice and is known as tapered aeration. Mechanical aeration has the same function as air diffusers but is accomplished by rotating paddles or brushes. Peripheral velocity is about 2 ft/sec.

Floc returned to the aeration tank has the purpose as trickling filter slime but floc concentration can be varied as operation needs dictate. Returned sludge varies from about 10 to 30%. Mixed liquor suspended solids (MLSS) will vary from 600 mg/l to 4000 mg/l, on a dry weight basis. An important parameter in routine process control is the ratio of the volume of MLSS to the dry weight of MLSS. This is known as the sludge volume index (SVI) and is in the range of 50 to 100 in well operating plants. When the value approaches 200 operating difficulties can be expected. Factors which promote or inhibit microbiological growth are important and these include pH, temperature, and oxidation-reduction potential (ORP). Hydrogen ion potential, pH, will have a great effect on the dominant species of organisms. Bacteria predominate above pH of 6.5 while fungi assume greater importance below this value. There must be adequate buffering capacity if metabolic products are acidic.

Modifications of the basic activated sludge process have come about for solution to specific operating problems.

The municipal treatment plant at Peoria, Illinois received a waste high in carbohydrates. The resulting nitrogen deficiency caused a light and poorly settling activated sludge floc with attendant poor waste stabilization. Kraus, for whom the modification is named, aerated digester supernatant. This, added to the influent, gave a nitrifying activated sludge. The result was a readily settleable sludge with improved organic removals.

New York City has plants scattered throughout the five boroughs, treating more than 1 billion gallons of sewage per day. A major modification resulting from experimentation with plant operation has come from this city. In conventional plants sewage was added at one end of the aeration tank and allowed to flow through. This gave a high initial microbial food supply and correspondingly high oxygen requirement. The New York City modification involves introduction of sewage at intervals along the tank. This smooths out the food supply and lowers the oxygen requirements. The sewage is added at discrete steps along the unit and the name applied is step aeration.

A low mixed liquor suspended solids concentration of 200 to 500 mg/l is maintained in the high rate process. This gives a high food to microbial mass ratio. This keeps the floc in the active growth phase, but excess food will be discharged in the effluent. BOD removals are only 50 to 60%, but in some areas this is acceptable. New York City applied this method successfully because of a weak sewage and low temperatures. Philadelphia and Los Angeles met with indifferent success because of stronger sewage or higher temperatures.

A process originated simultaneously and independently by Smith and Eckenfelder made use of a phenomenon observed by many researchers but dismissed as experimental error. When activated sludge and raw sewage are mixed together in an aeration vessel there is a noted reduction in BOD, followed by a rise and then another reduction. The first decrease had been ignored by most research workers. Smith and Eckenfelder found that this was due to adsorption of waste material onto the activated sludge floc. This came from material desorbed from colloidal particles. Plants at Austin, Texas, and Bergen County, New Jersey were converted from overloaded to underloaded by changing to the biosorption process.

A recent activated sludge process modification is the Deep Shaft Process. As its name suggests, two concentric deep shafts (120–150 m) are sunk into the ground. Wastewater is injected into one of the concentric shafts and the effluent is withdrawn from the other. A constant ambient temperature is maintained due to the surrounding geological formations. Compressed air is injected at the bottom, giving high dissolved oxygen concentrations and provided intimate mixing. Waste sludge is removed in clarifiers as in conventional activated sludge plants.

Putrescible material collected from the primary settling tanks and excess sludge from humus tanks must be disposed of cheaply and efficiently. This material is highly unstable and a potential nuisance source. Because it is putrescible it can be stabilized by biological means, serving as food and energy sources for microorganisms naturally found in the sludge.

Raw sludge is about 95% water, but the water is not easily removed. As the sludge is broken down the water content is lessened, and the volume is markedly reduced. A rough rule is that sludge volume is reduced by half when water is lowered from 95 to 90%, and by two thirds when reduced from 95 to 85%. Fresh sludge has a gray color and can be easily pumped. Its odor is most disagreeable, being due principally to mercaptans. Digested sludge is black in color, granular and has a slight tarry odor.

Sludge digestion is carried out in order to reduce the volume of sludge to be handled, and reduce the number of pathogens. Sludge is usually withdrawn at regular intervals from primary and secondary tanks and led by gravity to a sludge well. It is then pumped to the digester.

Mixing is very important for efficient sludge digestion. Temperature is equally important.

Since destruction of sludge is carried on by microorganisms, kinetics of their life processes will be temperature dependent. It has been found that sludge temperature of about 95°F will give acceptably short detention times. Even shorter detention times for the same quality of digested sludge can be achieved with temperatures of about 125–130°F, but this temperature range is not widely used for reasons of economics. Above 95°F an increase in detention time is noted, up to 110°F, and then again a decrease. The reason for this is the changing character of the predominant organisms.

Heating of sludge for efficient digestion is carried out in one of two ways. The older installations have hot water coils in the periphery of the tank, and heat is transmitted to the digesting sludge. Mixing was felt to be adequately effected by turbulence due to gas generation. Mechanical mixers have been used. It was found, however, that mixing was not sufficient. In addition, heating of entire tank contents was not achieved due to "baking" of sludge in the vicinity of the heating coils. A second method of sludge heating and mixing was developed, involving the use of external heat exchangers. Sludge is pumped from the digestion tank through a heat exchanger and returned to the tank. Two objectives are accomplished (1) efficient mixing of sludge, thereby reducing the amount of inadequately digested sludge, (2) more uniform temperature throughout the tank, thus reducing digestion time. The use of external heat exchangers has almost completely supplanted heating coils and internal mixers in new plant design.

Sludge gas generated during digestion is approximately 72% methane and 28% carbon dioxide. Hydrogen and H_2S are present in trace amounts. Gas thus generated has a calorific content of about 600 BTU/ft^3. About 10 ft^3 of gas are produced per cubic foot of raw sludge digested. Generally, the amount of sludge gas produced is sufficient to provide heat used in maintaining digesting sludge at the required temperature, heating plant buildings, provide hot water and incineration of digested sludge, when practiced, and fuel and generators.

Volatile acids, reported as acetic acid, are perhaps the most important parameters in control of sludge digestion. Volatile acids below 1000 mg/l occur in a healthy digestion process. Volatile acids of 6000 mg/l indicate a malfunctioning process. pH values of 6.8 to 7.2 are optimum. Values less than 6.8 usually are due to excessive volatile acid production. In the past liming of malfunctioning tank contents was practiced in an effort to adjust pH to about 7.0. However, the change in volatile acids production was due to changing dominant process microorganisms. The lowered pH and high volatile acids concentrations were a sign of a sick process, rather than the cause.

Digested sludge is reasonably inert but it must be further dewatered and the question of final disposal of raw and digested sludge is one of the most pressing with which environmental engineers must deal today.

Sludge can be dewatered on open or covered drying beds. Open beds are exposed to the air and drying is accomplished by drainage and evaporation. Covered beds resemble a greenhouse. Temperatures are rather high and this aids evaporation. In both cases sludge is allowed to flow over sand beds and let stand for a suitable period. The dried sludge is then scraped from the beds.

Sludge can be dewatered by vacuum filtration. Filter drums rotate slowly, picking up wet sludge at the bottom. A slight vacuum is applied and the water drawn off is returned to the plant stream. At the end of the cycle the dewatered sludge is removed by a scraper.

Dried sludge can be incinerated, taken to a landfill disposal site, composed or subjected to superoxidation. Some coastal cities barged sludge to sea. This was to have ceased in 1981 but was permitted to continue for ten years. It is not allowed after 1991. Incineration of sludge has the potential for air pollution problems and often there is local opposition to installation of an incinerator. Incinerators are expensive to build and operate. Land disposal is expensive and disposal sites frequently are considerable distances from the generating wastewater treatment plants. Groundwater pollution can occur. Landfills which can accept digested sludge are in short supply. Transportation costs are a quite significant part of total disposal costs. Composting has been suggested as a possible ultimate solution. The requirement of relatively large land areas and odor production are problems.

A promising approach involves superoxidation. Here the sludge is treated with a strong oxidizing agent. Volume is reduced greatly and the end product is stable and inoffensive. Transportation costs are thus reduced.

It is common practice to chlorinate effluent for bacterial control. Regulations vary from state to state, but most regulations require chlorination to a specified residual. Requirements usually vary from season to season, the most stringent rules governing the swimming season.

A phenomena not yet fully understood is that of aftergrowth, wherein bacterial count is fairly low immediately after effluent discharge but then suddenly rises to a high figure.

In some plants chlorination of the influent is practiced for the purpose of odor control.

Chlorination of storm water overflow is commonly practiced. In some cases storm water overflow is subjected to simple sedimentation and/or screening, storage and chlorination, then discharged after cessation of the storm.

One commonly used definition of tertiary treatment is any treatment in addition to secondary (biological) treatment.

Tertiary treatment is practiced when an effluent of much higher quality is required than is attainable with conventional biological treatment. They type process used will depend on the final effluent quality necessary and the economics of the total process.

Commonly used tertiary treatment processes are listed below:

1) Sand filtration
2) Microstrainers
3) Oxidation ponds
4) Foam separation
5) Activated carbon adsorption
6) Chemical clarification and precipitation
7) Ion exchange

Disposal of human and kitchen wastes in areas not served by sewers and wastewater treatment plants presents unique problems.

Disposal must be in the immediate vicinity of the source of the wastes. In adequate controls are not exercised a closed system may results. An example is Suffolk County (Long Island), New York. Septic tanks are widely used, and there is strong local opposition to the considerable expense of installing sewers. Effluent from the septic tanks found its way into the ground water which is the supply for much of the county. Eventually, the problem was graphically pointed up by the appearance of foaming detergents in water issuing from the tap.

In more primitive societies waste disposal is a matter of convenience. A “cat hole” or communal straddle trench is utilized and covered when capacity is reached. This is the same as the practice with privies and cesspools.

Privies, as the name implies, are simply open pits with a structure to provide privacy. Human excreta is deposited into the pit and is slowly stabilized. Stabilization is slow, due principally to the presence of urine. Pits may be open earth or concrete vault. Drawbacks for both types are odors and fly problems. For the unlined pit there is the additional problem of ground water pollution. Older privy construction allowed access to flies around the edge of the pit. For proper protection against flies there must be a tight seal around the edge of the pit and adequate screening of openings in the privy structure itself.

In the 1930s a large number of the older privies were replaced by concrete vault types. Today, such methods of waste disposal are found only in the smaller rural communities where there is no municipal collection system.

Cesspools are simply pits into which waste is allowed to flow. The term leaching pit is sometimes used. Water seeps into the ground, leaving solid matter in the pit. Construction is of two types. A pit may be unlined, or it may be lined with sewer pipe laid on end. Almost nowhere in the United States are cesspools permitted by health authorities.

Septic tanks are widely used in smaller towns and outlying suburbs of larger cities. They are a combination sedimentation tank and anaerobic digester. Sanitary and kitchen wastes flow into the tank and grease and light material rise to the top. Heavier particles settle to the bottom where anaerobic stabilization occurs. Deflector plates are provided at inlet and outlet in order to minimize short circuiting. Effluent flows to a tile field where disposal is into the earth. The tile field is composed of perforated field tile fed by a manifold. The tile is underlain with granular material, usually gravel. Care must be taken that the earth does not become clogged by material carried over from the septic tank. Septic tanks are being replaced as more and more areas are served by municipal systems. Health authorities do not look with favor on septic tanks. Capacity will be a function basically of the number of persons or units served. Some experts feel that, in no case, should capacity be less than 1500 gallons. Lesser volumes are permitted in many codes and the thought that 1500 gallons ought to be the minimum permitted arouses home builders and land developers. Periodically it is necessary to employ a scavenger service for emptying the tank of accumulated solids. Solids thus collected may be discharged to a convenient treatment system or directly to a wastewater treatment plant.

The Imhoff Tank has not been discussed earlier because it is similar in many respects to the septic tank. Use is generally confined to small communities and isolated installations. Operation is a combination of sedimentation and anaerobic digestion. This tank was invented by Karl Imhoff, who first used them in the Essen District in Germany in 1907. The tank is composed of two chambers, one above the other. Surface configuration may be circular, square or rectangular. Depth is 25′ to 35′.

Sewage flows through the upper chamber, at a low velocity (about 1 fps). Solids settle out and slide through a slot into the bottom chamber. Detention period is about 2 hours.

Solids accumulating in the bottom, or digestion, chamber have an initial water content of 85 to 95%. After proper digestion of about 60 days the water content is reduced to about 50% and the volume is greatly reduced. Gases produced during digestion are vented to the atmosphere by gas vents located at the tank sides. Solids buoyed up by gas are prevented from escaping to the upper tank by deflector plates.

Attempts were made to hasten digestion by heating the lower compartment but were of limited success due to overturning of the tank contents.

Some rectangular tanks are arranged so that the direction of flow can be reversed, with outlets becoming inlets and vice versa.

In recent years manufacturers of waste treatment equipment have endeavoured to supply complete treatment plants for small communities or developments and isolated installations. Basically, these plants, called package plants, supply primary treatment and sometimes some biological treatment on a small scale without requiring extensive operating supervision. It is felt that such treatment is to be preferred to septic tanks or only primary treatment (Imhoff Tanks, for example), but such installations are not the ultimate solution.

REFERENCES

1. Anderson, E. and W.T. Lockett, *J. Soc. Chem. Ind. London*, **33**, 523, 1914.
2. Bewtra, J.K., Biological Treatment of Wastewater, *Encyclopedia of Environmental Science and Engineering*, 4th Ed., Vol. 1, Gordon and Breach, Inc., New York. 1998.
3. Bewtra, J.K. and H.I. Ali, Physical and Chemical Treatment of Wastewaters, *Encyclopedia of Environmental Science and Engineering*, 4th Ed., Vol. 2, Gordon and Breach, Inc., New York. 1998.
4. Bewtra, J.K., Recent Advances in Water-pollution-control Technology. Advances in *Environmental Science and Engineering*, **1**, Gordon and Breach, New York, 1979.
5. Camp, T.R., *Trans. ASCE*, **3**, p. 895, 1942.
6. Cecil, L.K., Water reuse, *Encyclopedia of Environmental Science and Engineering*, **2**, Gordon and Breach, Inc., New York, 1983.
7. Cecil, L.K., Water reuse and disposal, *Chemical Engineering*, May 5, 1969.
8. Disposal of Municipal Sewage. House Report No. 20 12th Report by Committee on Government Operation, House of Representatives, March 24, 1965.
9. Dobbins, W.E., *J. of the Sanitary Engineering Division, ASCE*, **90**, No. SA 3, Proc. Paper 3949, June 1964.
10. Eckenfelder, W.W. and D.J. O'Connor, *Biological Waste Treatment*, Pergamon Press, Ltd., London, 1961.
11. Fair, G.M., J.C. Geyer, and D.A. Okun, *Water and Waste Water Engineering*, **1** and **2**, John Wiley and Sons Inc., New York, N. Y., 1968.
12. Guttman, H.N., Microbiology, *Encyclopedia of Environmental Science and Engineering*, 4th. Ed., Vol., 2, Gordon and Breach, Inc., New York. 1998.
13. Imhoff, K., *Taschenbuch der Stadtentwässerung*, R. Oldenbourg Verlag, Munich, 1960.
14. Imholff, K. and G.M. Fair, *Sewage Treatment* 2nd Ed., John Wiley and Sons, Inc., New York, N.Y., 1956.
15. McKinney, R.E., *Microbiology for Sanitary Engineers*, McGraw-Hill Book Co., New York, N. Y., 1962.
16. Nolte, W.F., Effects of elevated pressure on secondary treatment of wastewater, *Advances in Environmental Science and Engineering*, **4**, Gordon and Breach, New York, 1981.
17. Pfafflin, J.R., Water: Sewage, *The Encyclopedia of Chemical Technology*, Interscience Publishers, Inc., New York, N.Y., 1981.
18. Pfafflin, J.R., Water: Reuse, *The Encyclopedia of Chemical Technology*, Interscience Publishers, Inc., New York, N.Y., 1981.
19. Phelps, E.B., *Stream Sanitation*, John Wiley and Sons, Inc., New York, N.Y., 1944.
20. Pollock, D.C. and M.A. Wilson, *Development of Deep Shaft Effluent Treatment Process in North America*, Second World Congress of Chemical Engineering, Montreal, 1981.
21. Qasim, S.R., Wastewater Treatment Plants—Planning, Design and Operation, Holt, Rinehart and Winston, New York, 1985.
22. Salvato, J., *Environmental Sanitation*, 2nd Ed., John Wiley and Sons, Inc., New York, N.Y., 1972.
23. Sawyer, C.N. and P.L. McCarty, *Chemistry for Sanitary Engineers*, 2nd Ed., McGraw-Hill Book Co., New York, N.Y., 1967.
24. *Standard Methods for the Examination of Water and Waste water*, 17th Ed., APHA, 1990.
25. Streeter, H.W. and E.B. Phelps. A study of the pollution and natural purification of the Ohio river. *Public Health Bulletin* 146, US Public Health Service, Washington, D.C., 1925.
26. UK Patent No. 2,128,980.
27. US Patent No. 4,464,257.
28. US Patent No. 4,500,428.

JAMES R. PFAFFLIN
Gillette, N.J.

CAMERON MACINNIS
University of Windsor.

MUNICIPAL WATER REUSE: see WATER REUSE
MUNICIPAL WATER SUPPLY: see WATER TREATMENT

N

NATURAL SYSTEMS FOR WASTEWATER TREATMENT

INTRODUCTION

In the continual search for a simple, reliable, and inexpensive wastewater treatment system, the natural systems are the latest discovery. They provide not only efficient methods of wastewater treatment, but also provide some indirect beneficial uses of the facility such as green space, wildlife habitat, and recreational areas.

In the natural environment, physical, chemical, and biological processes occur when water, soils, plants, microorganisms, and atmosphere interact. To utilize these processes, natural systems are designed. The processes involved in the natural systems include: sedimentation, filtration, gas transfer, adsorption, ion exchange, chemical precipitation, chemical oxidation and reduction, and biological conversions plus other unique processes such as photosynthesis, photooxidation, and aquatic plant uptake.[1] In natural systems, the processes occur at "natural" rates and tend to occur simultaneously in a single "ecosystem reactor" as opposed to mechanical systems in which processes occur sequentially in separate reactors or tanks in accelerated rates as a result of energy input. Generally, a natural system might typically include pumps and piping for wastewater conveyance but would not depend on external energy sources exclusively to maintain the major treatment responses.[2]

In this paper a general overview of natural systems for wastewater treatment is presented. The constructed wetlands are becoming a viable wastewater treatment alternative for small communities, individual homes, and rest areas. Therefore, in this paper, a great deal of information has been presented on site selection, design of physical facilities, performance expectations, hydraulic and organic loading rates, and cost of the constructed wetlands.

TYPES OF NATURAL SYSTEMS

Natural systems for effective wastewater treatment are divided into two major types: terrestrial, and aquatic systems. Both systems depend on physical and chemical responses as well as on the unique biological components.[2] Each of these systems are discussed below.

Terrestrial Treatment Systems

Land application is the sole terrestrial treatment system used to remove various constituents from the wastewater. It utilizes natural physical, chemical, and biological processes within the soil-plant-water matrix. The objectives of the land treatment system includes, irrigation, nutrient reuse, crop production, recharge of ground water and water reclamation for reuse. There are three basic methods of land application: (1) Slow-rate irrigation, (2) Rapid-infiltration–percolation, and (3) Overland flow. Each method can produce renovated water of different quality, can be adapted to different site conditions, and can satisfy different overall objectives.[3,4] Typical design features and performance expectations for the three basic terrestrial concepts are given in Table 1.

Slow-rate irrigation Irrigation is the most widely used form of land treatment systems. It requires presence of vegetation as a major treatment component. The wastewater is applied either by sprinkling or by surface technologies. In this process surface runoff is not allowed. A large portion of water is lost by evapotranspiration whereas some water may reach the groundwater table. Groundwater quality criteria may be a limiting factor for the selection of the system. Some factors that are given consideration in design and selection of irrigation method are (1) availability of suitable site, (2) type of wastewater and pretreatment, (3) climatic conditions and storage needed, (4) soil type, and organic and hydraulic loading rates, (5) crop production, (6) distribution methods, (7) application cycle, and (8) ground and surface water pollution.[2,3,5]

Rapid-infiltration–percolation In rapid-infiltration-percolation the wastewater percolates through the soil and treated effluent reaches the groundwater or underdrain systems. Plants are not used for evapotranspiration as in irrigation system. The objectives of rapid infiltration—percolation are (1) groundwater recharge, (2) natural treatment followed by withdrawal by pumping or under-drain systems for recovery of treated water, (3) natural treatment with groundwater moving vertically and laterally in the soil, and (4) recharging a surface water source.[5]

Overland flow In the overland flow system, the wastewater is applied over the upper reaches of the sloped terraces and allowed to flow overland and is collected at the toe of the slopes. The collected effluent can be either reused or discharged into a receiving water. Biochemical oxidation, sedimentation, filtration and chemical adsorption are the primary mechanisms for removal of the contaminants. Nitrogen removal is achieved through denitrification. Plant uptake of nitrogen and phosphorus are significant if crop harvesting is practiced.[3,5]

Aquatic Treatment Systems

Aquatic system utilize lagoons and ponds, and wetlands. The lagoons and ponds depend on microbial life, and lower plants and animals, while wetlands support the growth of rooted plants. Both pond systems and wetlands are discussed below.

Pond Systems

Within the aquatic systems, pond systems are the most widely accepted ones. Basically pond systems can be of three types based on oxygen requirement. These are aerobic, anaerobic and facultative pond systems. In all cases the major treatment responses are due to the biological components.[2,7] In most of the pond systems both performance and final water quality are dependent on the algae present in the system. Algae are functionally beneficial, providing oxygen to support other biological responses, and the algae-carbonate reactions are the basis for the effective nitrogen removal in the ponds. However, algae can be difficult to remove from the effluent. As a result, the stringent limits for suspended solids in the effluent can not be met. The design features and performance expectations for natural aquatic treatment units are summarized in Table 2.

Aerobic ponds Aerobic ponds, also called high rate aerobic ponds, maintain dissolved oxygen (DO) throughout their entire depth. They are usually 30 to 45 cm deep, allowing light to penetrate to the full depth.[3] Mixing is often provided to expose all algae to sunlight and to avoid deposition. As a result, formation of anaerobic sludge layer can be reduced. Oxygen is provided by photosynthesis and surface reaeration. Because of the presence of sufficient dissolved oxygen, aerobic bacteria can stabilize the waste. Detention time is short, usually 3 to 5 days. Aerobic ponds are limited to warm sunny climate and are used infrequently in the United States.[2,7]

TABLE 1
Design features and expected performance for terrestrial treatment units

				Typical criteria				
Concepts	Treatment goals	Climate needs	Vegetation	Area,[b] ha	Organic loading, kg BOD_5/ha.d	Hydraulic loading, m/year	Effluent characteristics, mg/L	
Slow rate	Secondary, or AWT[a]	Warmer season	Yes	23–280	—	0.5–6	BOD	<2
							TSS	<1
							TN	<3[c]
							TP	<0.1[c]
							FC	0[d]
Rapid infiltration	Secondary, or AWT, or groundwater recharge	None	No	3–23	45–180	6–125	BOD	5
							TSS	2
							TN	10
							TP	<1[e]
							FC	10
Overland flow	Secondary, nitrogen removal	Warmer season	Yes	6–40	<95	3–20	BOD	10
							TSS	10[f]
							TN	<10
							TP	<6

[a] Advanced wastewater treatment.
[b] For design flow of 3800 m^3/d.
[c] Total nitrogen removal depends on type of crop and management.
[d] FC = Fecal coliform, number per 100 ml.
[e] Measured in immediate vicinity of basin, increased removal with longer travel distance.
[f] Total suspended solids depends in part on type of wastewater applied.
Source: Adopted in part from Refs. 2, 3, 5, and 6.

TABLE 2
Design features and expected performance for pond treatment units

Parameter	Aerobic (High Rate)	Aerobic–anaerobic (Facultative)	Anaerobic
Detention time, days	5–20	10–30	20–50
Water depth, m	0.3–1	1–2	2.5–5
BOD_5 loading, kg/ha.d	40–120	15–120	200–500
Soluble BOD_5 removal, percent	90–97	85–95	80–95
Overall BOD_5 removal, percent	40–80	70–90	60–90
Algae concentration, mg/L	100–200	20–80	0–5
Effluent TSS, mg/L	100–250[a]	40–100[a]	70–120[a]

[a] TSS is high because of algae. Effluent quality can be improved significantly if algae is removed.

Source: Adopted in part from Refs. 3, 4 and 8.

Anaerobic ponds Anaerobic ponds receive such a heavy organic loading that there is no aerobic zone. They are usually 2.5 to 5 m in depth and have detention time of 20 to 50 days.[2,3] The principal biological reactions occurring are acid formation and methane production. Anaerobic ponds are usually used for treatment of strong industrial and agricultural wastes, or as a pretreatment step where an industry is a significant contributor to a municipal system. They do not have wide application to the treatment of municipal wastewater.

Facultative ponds It is the most common type of pond unit. These ponds are usually 1.2 to 2.5 m (4 to 8 ft) in depth with an aerobic layer overlaying an anaerobic layer, which often contains sludge deposits. The usual detention time is 5 to 30 days.[2] Anaerobic fermentation occurs in the lower layer, and aerobic stabilization occurs in the upper layer. The key to facultative operation is oxygen production by photosynthetic algae and surface reaeration. The algae cells are necessary for oxygen production, but their presence in the final effluent represents one of the most serious performance problems associated with the facultative ponds.[7]

Wetland Treatment Systems

The wetlands are inundated land areas in which water table is at or above the ground surface (usually 0.6 m or more). This water table stands long enough time each year to maintain saturated soil conditions and also to support the growth of related vegetation. The vegetation provides surface for attachment of bacteria films, and aids in the filtration and adsorption of wastewater constituents. Vegetation also translocate oxygen from leaves to the root systems to support a wide range of aerobic and facultative bacteria and controls the growth of algae by restricting the penetration of sunlight.[9] The unique ability of wetland plants to translocate oxygen to support a wide range of bacteria in the wetlands is shown in Figure 1. Wetlands can be of two types: natural wetlands and constructed wetlands. Both natural and constructed wetlands have been used for wastewater treatment, although the use of natural wetlands is generally limited to the polishing or further treatment of secondary or advanced wastewater treated effluent.[2]

Natural wetlands From a regulatory standpoint, natural wetlands are usually considered as part of the receiving waters. Consequently discharges to natural wetlands, in most cases, must meet applicable regulatory requirements, which typically stipulate secondary or advanced wastewater treatment.[10] Furthermore, the principal objective when discharging to natural wetlands should be enhancement of existing habitat. Modification of existing wetlands to improve treatment capability is often very disruptive to the natural ecosystem and, in general, should not be attempted.[10,11]

Constructed wetlands Constructed wetlands offer all of the treatment capabilities of the natural wetlands but without the constraints associated with discharging to a natural ecosystem. Additionally, the constructed wetland treatment units are not restricted to the special requirements on influent quality. They can ensure much more reliable control over the hydraulic regime in the system and therefore perform more reliably than the natural wetlands.[10] Two types of constructed wetlands have been developed for wastewater treatment (1) free water surface (FWS) systems, and (2) subsurface flow system (SF). The schematic flow diagrams of both types of system have been shown in Figure 2. A general description of both types of systems is provided below. The basic design considerations such as (a) site selection, (b) plants types, (c) physical facilities, (d) hydrologic factors, (e) organic loading factors, and (f) performance expectations, that are presented under a separate section entitled *Basic Design Considerations of Constructed Wetlands.*

Free water surface (FWS) wetlands The free water surface wetlands typically consist of a basin or channels with some type of barrier to prevent seepage, soil to support the root systems

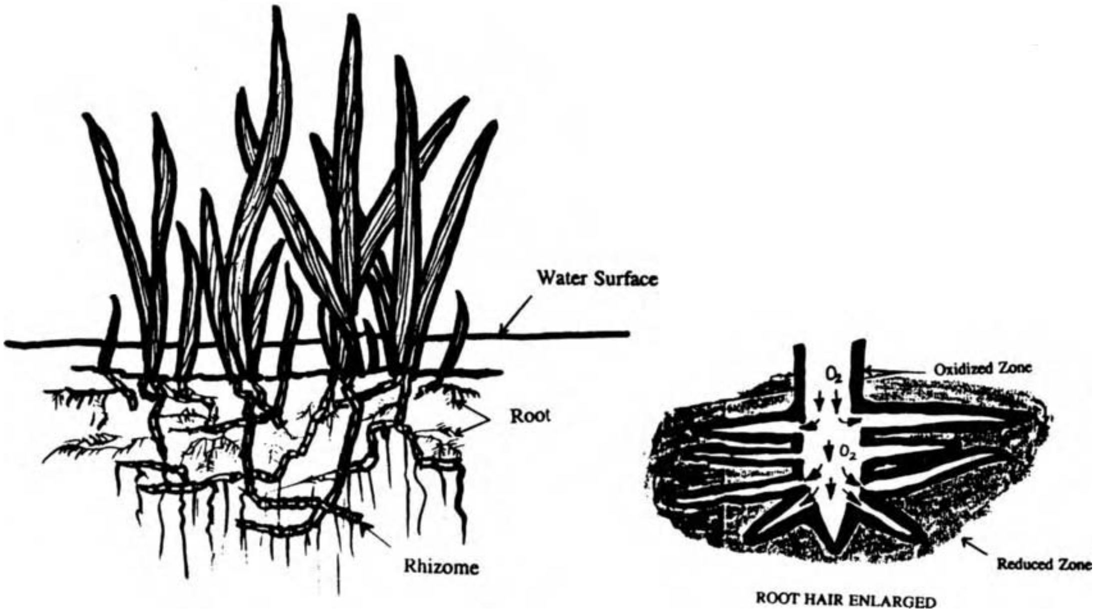


FIGURE 1 An enlarged root hair of wetland plants.

of the emergent vegetations, and water through the system at a relatively shallow depth.[2,9] The water surface in FWS wetlands is exposed to the atmosphere, and the intended flow path through the system is horizontal. Pretreated wastewater is applied continuously to such systems, and treatment occurs as the water flows slowly through the stems and roots of emergent vegetation. Free water surface systems may also be designed with the objective of creating new wildlife habitats or enhancing nearby existing natural wetlands. Figure 2(a) provides a schematic flow diagram of free water surface wetlands.

Subsurface flow (SF) wetlands The subsurface flow wetlands also consist of a basin or channel with a barrier to prevent seepage. The basin or channel is filled to a suitable depth with a porous media. Rock or gravel are the most commonly used media types. The media also supports the root systems of the emergent vegetation. The design of these systems assumes that the water level in the bed will remain below the top of the rock or gravel media. The flow path through the operational systems is usually horizontal.[12] The schematic flow diagram of submerged flow wetlands is provided in Figure 2(b).

Comparing the two types of constructed wetland systems, the SF type of wetlands offer several advantages over FWS type. If the water surface is maintained below the media surface there is little risk of odors, public exposure, or insect vectors. In addition, it is believed that the media provides larger available surface area for attached growth organisms. As a result, the treatment response may be faster, and smaller surface area may be needed for the same wastewater conditions.[2,10] Furthermore, the subsurface position of the water, and the accumulated plant debris on the surface of the SF bed, offer great thermal protection in cold climates as compared to the FWS type.[14] The reported disadvantage of the SF type system however, is clogging of the media possible overflow.

Basic Design Considerations of Constructed Wetlands

Constructed wetlands are relatively recent development, and are gaining popularity for treatment of wastewater from small communities, and residential and commercial areas. In this section the basic design information and economics of constructed wetlands are compared.

Site selection A constructed wetland can be constructed almost anywhere. In selecting a site for a free water surface wetland the underlying soil permeability must be considered. The most desirable soil permeability is 10E-6 to 10E-7 m/s.[10] Sandy clay and silty clay loams can be suitable when compacted. Sandy soils are too permeable to support wetland vegetation unless there is an impermeable restricting layer in the soil profile that result in a perched high ground water table. Highly permeable soils can be used for wastewater flows by forming narrow trenches and lining the trench walls and bottoms with clay or an artificial liner. In heavy clay soils, addition of peat moss or top soil will improve soil permeability and accelerate initial plant growth.

Plants Although natural wetlands typically contain a wide diversity of plant life, there is no need to attempt to reproduce the natural diversity in a constructed wetland. Such attempts in the past have shown that eventually cattails alone or in combination with either reeds or bulrushes will dominate

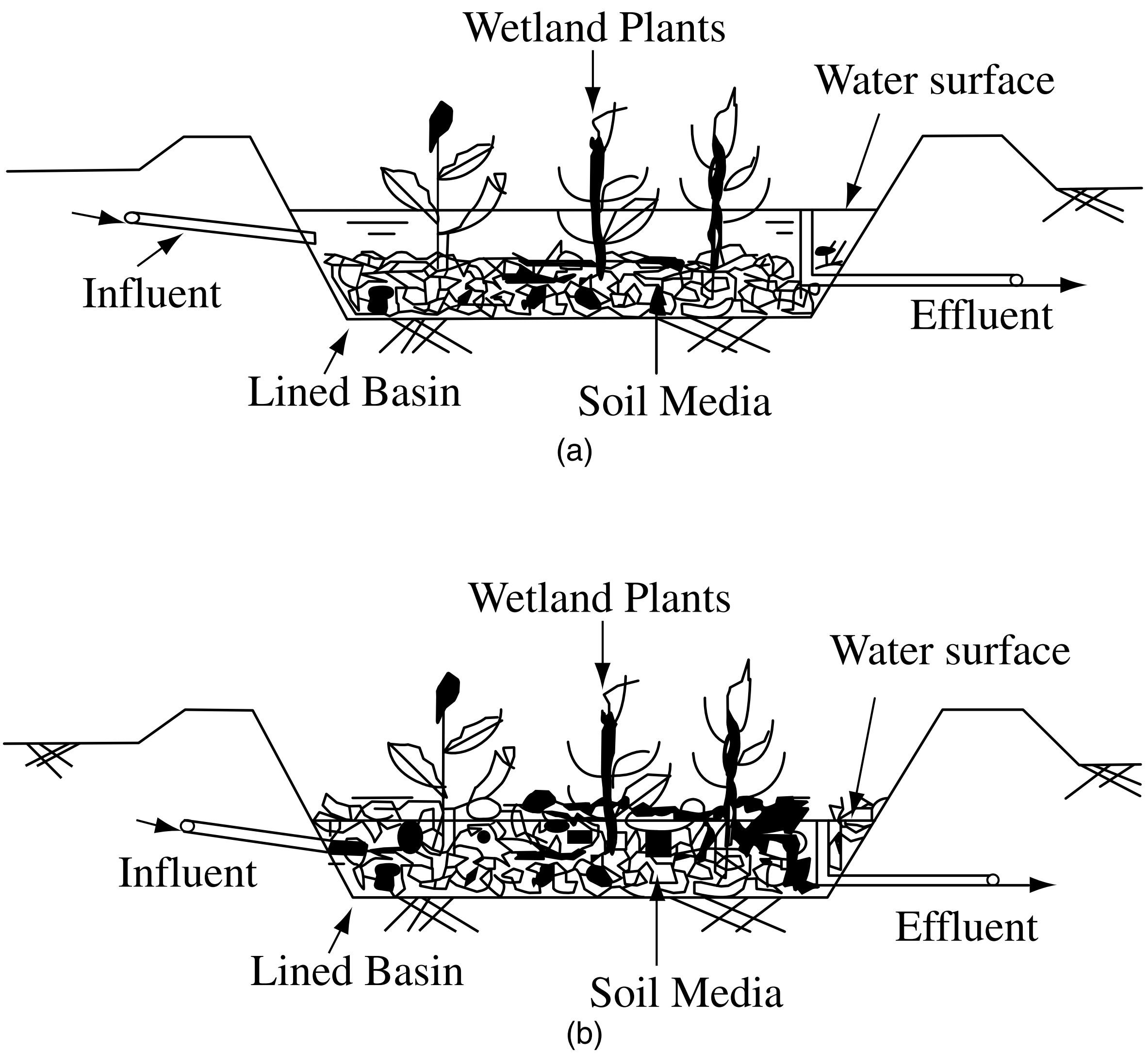


FIGURE 2 Schematic flow diagrams of constructed wetland systems, (a) Free water surface, (b) Subsurface flow.[13]

in a wastewater system owing to the high nutrient levels.[2] The emergent plants most frequently found in the wastewater wetlands include cattails, reeds, rushes, bulrushes and sedges. Some of the major environmental requirements of these plants are given in Table 3.

Physical facilities The constructed wetlands behave typically like a plug flow reactor. Constructed wetlands offer significant potential for optimizing performance through selection of proper system configuration, including aspect ratios, compartmentalization, and location of alternate and multiple discharge sites.

The aspect ratio, defined as the ratio of length to width, typically varies from 4:1 to 10:1. However, based on research data developed on experimental constructed wetlands, the aspect ratios approaching 1:1 may be acceptable.[13] Several alternative flow diagrams and configurations of constructed wetlands are provided in Figure 3.

Hydrologic factors The performance of the constructed wetlands system is dependent on the system hydrology as well as many other factors such as precipitation, infiltration, evapotranspiration, hydraulic loading rate, and water depth. These factors affect the removal of organics, nutrients, and trace elements not only by altering the detention time but also by either concentrating or diluting the wastewater.[9,10] For a constructed wetland, the water balance can be expressed by Eq. (1)

$$[dV/dt] = Q_i - Q_e + P - ET \tag{1}$$

where,

Q_i = influent wastewater flow, m^3/d
Q_e = effluent wastewater flow, m^3/d
P = precipitation, m^3/d
ET = evapotranspiration, m^3/d
$[dV/dt]$ = change in volume of water per unit time, m^3/d
t = time, d

TABLE 3
Emergent aquatic plants for constructed wetlands

Common name (*Scientific name*)	Temperature, °C		Maximum salinity tolerance, ppt[b]	Optimum pH
	Survival	Desirable[a]		
Cattail (*Typha spp.*)	10–30	12–24	30	4–10
Common reed (*Phragmites communis*)	12–33	10–30	45	2–8
Rush (*Juncus spp.*)	16–26	—	20	5–7.5
Bulrush (*Scirpus spp.*)	16–27	—	20	4–9
Sedge (*Carex spp.*)	14–32	—	—	5–7.5

[a] Temperature range for seed germination: roots and rhizomes can survive in frozen soils.
[b] ppt = parts per thousands.
Source: Adopted in part from Ref. 15.

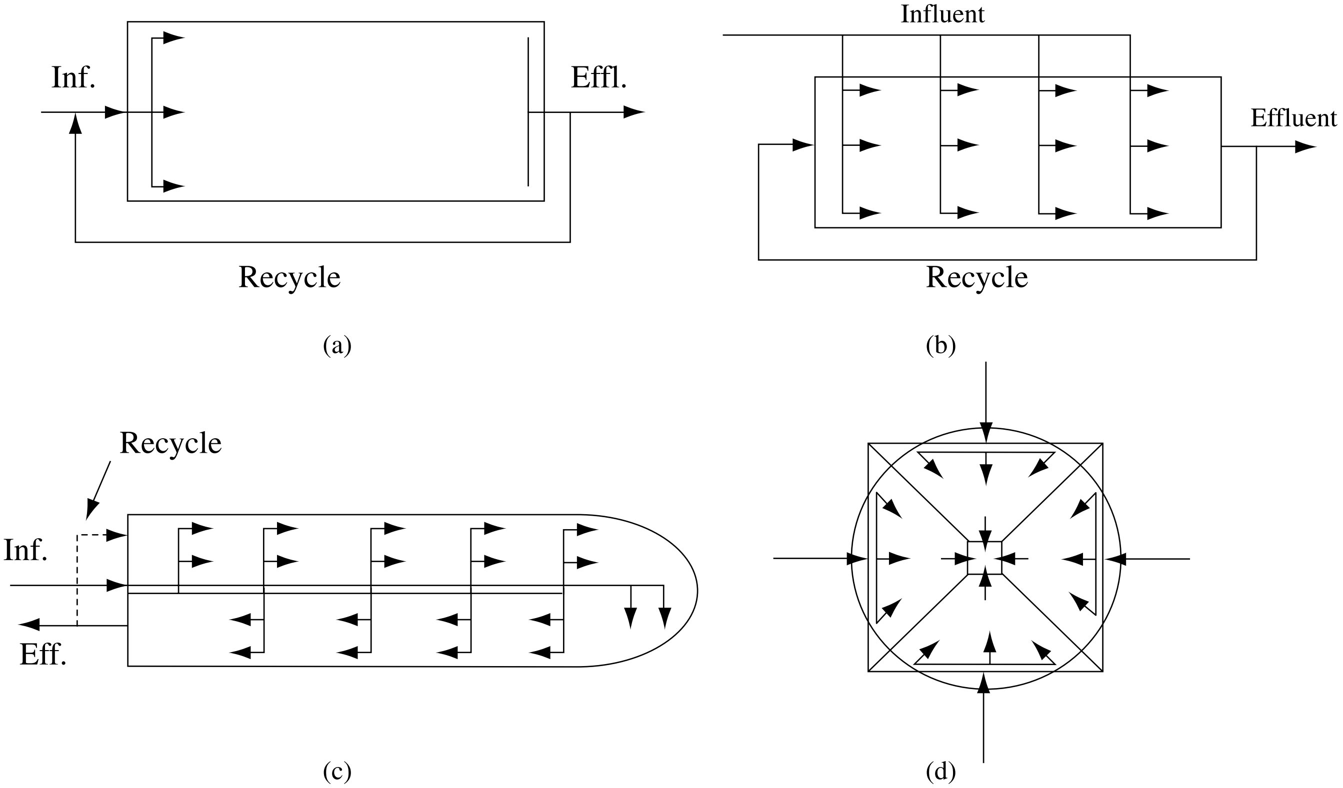


FIGURE 3 Alternate flow diagram and configurations of constructed wetlands: (a) Plug flow with recycle; (b) Step feed with recycle; (c) Step feed in wrap around pond; (d) Peripheral feed with center drawoff.

Hydraulic loading rate for FWS system is closely tied to the hydrologic factors and conditions specific to the site. Typical hydraulic loading rate of 198 m^3/d.ha (21,000 gpd/acre) is considered sufficient for optimum treatment efficiency.

BOD_5 loading rates There are two goals for organic load control in the constructed wetlands. The first is the provision of carbon source for denitrifying bacteria. The second goal is to prevent overloading of the oxygen transfer ability of the emergent plants. High organic loading, if not properly distributed, will cause anaerobic conditions, and plants die off. The maximum organic loading rate for both type of systems (FWS and SF) should not exceed 112 kg BOD_5/ha.d.[2,9,10]

Performance expectations Constructed wetland systems can significantly remove the biochemical oxygen demand (BOD), total suspended solids (TSS), nitrogen and phosphorus, as well as metals, trace organics and pathogens The basic treatment mechanisms include sedimentation, chemical precipitation, adsorption, microbial decomposition, as well as uptake by vegetation. Removal of BOD_5, TSS, nitrogen, phosphorus, heavy metals and toxic organics have been reported.[2,17] The performance of many well known constructed wetlands in terms of BOD_5, TSS, ammonia nitrogen and total phosphorus is summarized in Table 4.

Mosquito control and plant harvesting are the two operational considerations associated with constructed wetlands for wastewater treatment. Mosquito problems may occur when wetland treatment systems are overloaded organically and anaerobic conditions develop. Biological control agents such as mosquitofish (*Gambusia affins*) die either from oxygen starvation or hydrogen sulfide toxicity, allowing mosquito larvae to mature into adults. Strategies used to control mosquito populations include effective pretreatment to reduce total organic loading; stepfeeding of the influent wastewater stream with effective influent distribution and effluent recycle; vegetation management; natural controls, principally mosquitofish, in conjunction with the above techniques; and application of approved and environmentally safe chemical control agents.

The usefulness of plant harvesting in wetland treatment systems depends on several factors, including climate, plant species, and the specific wastewater objectives. Plant harvesting can affect treatment performance of wetlands by altering the effect that plants have on the aquatic environment. Further, because harvesting reduces congestion at the water surface, control of mosquito larvae using fish is enhanced. It has been reported in the literature that a total of 29,000 kg/ha dry weight of harvestable biomass of *Phragmites* shoots can be harvested for single harvesting in a year. Higher yield is achievable with multiple harvesting.

The BOD_5, TSS, nitrogen and phosphorus removal efficiencies of constructed wetlands are discussed below.

BOD_5 Removal in FWS wetlands In the free water surface constructed wetlands the soluble BOD_5 removal is due to microbial growth attached to plant root, stems, and leaf litter that has fallen into the water. BOD_5 removal is generally expressed by a first order reaction kinetic (Eq. 2).[2,3,10]

$$[C_e/C_o] = \exp(-K_T t) \tag{2}$$

where,
C_e = effluent BOD_5, mg/L
C_o = influent BOD_5, mg/L
K_T = reaction rate constant, d^{-1}
t = hydraulic retention time, d.

BOD_5 Removal in SF wetlands The major oxygen source for the subsurface components (soil, gravel, rock, and other media, in trenches or beds) is the oxygen transmitted by the vegetation to the root zone. In most cases, there is very little direct atmospheric reaeration as water surface remains below the surface of the media.[17] Removal of BOD_5 is expressed by Equation (3). This is also a first order equation and can be rearranged to calculate the area required for the subsurface flow system.

$$\log(C_e/C_o) = -[A_s K_t d e]/Q \tag{3}$$

where,
C_e = effluent BOD_5, mg/L
C_o = influent BOD_5, mg/L
K_t = reaction rate constant, d^{-1}
Q = flow rate through the system, m^3/d
d = depth of submergence, m
e = porosity of the bed,
A_s = surface area of the system, m^2.

Suspended solids removal Suspended solids removal is very effective in both types of constructed wetlands. Most of the removal occurs within few meters beyond the inlet. Control dispersion of the inlet flow will enhance removal near the inlet zone. Proper dispersion of solids can be achieved by low inlet velocities, even cross sectional loadings, and uniform flow without stagnation.[10]

Nitrogen removal Nitrogen removal is very effective in both the free water surface and submerged flow constructed wetlands, and the nitrification/denitrification is the major path of nitrogen removal. Total nitrogen removals of up to 79 percent are reported at nitrogen loading rates of (based on elemental N) up to 44 kg/(ha.day) [39 lb/(acre.day)], in a variety of constructed wetlands. If plant harvesting is practiced, a higher rate of nitrogen removal can be expected.[9,20]

Phosphorus removal Phosphorus removal in many wetlands is not very effective because of the limited contact opportunities between the wastewater and the soil. The exceptions are in the submerged bed design when proper soils are selected as the medium for the system.[21] A significant clay content and the presence of iron, and aluminum will enhance the potential for phosphorus removal.[9,20,21]

TABLE 4
Performances of some well known constructed wetlands

System name	Location	Type	Record (years)	Area (ha)	Flow (m^3/d)	BOD_5 (mg/L)		TSS (mg/L)		NH_3-N (mg/L)		TP (mg/L)	
						IN	OUT	IN	OUT	IN	OUT	IN	OUT
Lakeland	FL	FWS	2	498	26978	3	2.5	4	3.5	0.90	0.36	9.46	4.07
Reedy Creek	FL	FWS	11.2	35.24	12058	5.3	1.9	8.9	2.4	2.98	0.72	1.4	1.78
Reedy Creek	FL	FWS	11.2	5.87	2423	5.8	1.6	10.9	2.4	3.29	0.12	1.8	0.79
Ironbridge	FL	FWS	1	486	34254	4.8	2.1	10.5	65.9	3.99	0.94	0.53	0.11
Apalachicola	FL	FWS	3	63.7	3936	15.2	1.1	107	8	3.62	0.09	3	0.21
Fort Deposit	AL	FWS	0.67	6	628	29.9	5.4	78.7	10.4	13.59	1	—	—
West Jackson	MS	FWS	0.5	8.91	1953	21.6	10.5	65.9	24.5	2.69	0.19	5.13	3.56
Leaf River 1	MS	FWS	1.2	0.13	225	15.8	14	54.8	30.1	9.91	7.21	8.91	8.16
Leaf River 2	MS	FWS	1.2	0.13	254	15.8	15.7	54.8	34.9	9.91	6.33	8.91	8.19
Leaf River 3	MS	FWS	1.2	0.13	220	15.8	13.9	54.8	25.5	9.91	6.79	8.91	5.9
Cobalt	Canada	FWS	1	0.10	49	20.7	4.6	36.2	28	2.95	1.04	1.68	0.77
Pembroke	KY	FWS	0.75	1.48	188	67.4	9.4	91.9	8.2	13.8	3.35	6.03	3.16
Central Slough	SC	FWS	4	32	5372	16.3	6.5	27.7	14.8	7.49	1.38	4.09	1.46
Gustine 1A	CA	FWS	1	0.39	163	130	49.8	73.5	39.6	17.4	16.1	—	—
Gustin 2A	CA	FWS	1	0.39	174	151	44.8	99.8	33.8	18	23.2	—	—
Philips School	AL	SFS	2	0.20	58.7	15.3	1	63.7	2	11	1.7	6	0.3
Kingston	TN	SFS	0.7	0.93	76	56	9	83	3	22	16	3.4	2.1
Denham Sprg.	LA	SFS	1.5	2.10	2548	28.2	10.5	53	17	11	4.33	—	—
Monterey	VA	SFS	1.1	0.023	83	38	15	32	7	9.33	8.67	—	—

Source: Adopted in part from Ref. 13.

Cost Cost is often a significant factor in selecting the type of treatment system for a particular application. Unfortunately, the availability of reliable cost data for wetland treatment systems is limited. The cost of wetland treatment systems varies depending on wastewater characteristics, the type of wetland system, and the type of bottom preparation required. Subsurface flow systems are generally more expensive than free water surface systems. It has been reported that the Tennesee Valley Authority (TVA) wetland construction costs ranged from \$3.58/m^2 to \$32.03/m^2.[22] Estimates are that the construction, operation, and maintenance costs of constructed wetland systems are quite competitive with other wastewater treatment options.

REFERENCES

1. Mitsch, W.J. and J.G. Gosselink, *Wetlands,* Van Nostrand and Reinhold Company, New York, 1993.
2. Reed, S.C., E.J. Middlebrooks, and R.W. Crites, *Natural Systems for Waste Management and Treatment,* McGraw-Hill, Inc., New York, NY, 1988.
3. Qasim, S.R., *Wastewater Treatment Plants: Planning, Design, and Operation,* Technomic Publishing Company, Inc. PA, 1994.
4. Middlibrooks, E.J., C.H. Middlebrooks, J.H. Reynolds, G.Z. Watters, S.C. Reed, and D.B. George, *Wastewater Stabilization Lagoon Design, Performance and Upgrading,* Macmillan Book Co., New York, 1982.
5. US Environmental Protection Agency, *Process Design Manual for Land Treatment of Municipal Wastewater,* Office of Water Program Operation, EPA/COE/USDA, EPA 625/1-77-008, October 1977.
6. Sanks, R.L. and T. Asano (Eds.), *Land Treatment and Disposal of Municipal and Industrial Wastewater,* Ann Arbor Science Publisher, Ann Arbor, Mich. 1976.
7. Metcalf and Eddy, Inc., *Wastewater Engineering: Treatment, Disposal and Reuse,* McGraw-Hill, Inc., New York, NY, 1991.
8. Bastian, R.K. and S.C. Reed (Eds.), *Aquaculture Systems for Wastewater Treatment: Seminar Proceedings and Engineering Assessment,* US Environmental Protection Agency, Office of Water Program Operations, Municipal Construction Division, EPA 430/9-80-006, Washington, D.C. 20460, September 1979.
9. Hammer, D.A., *Constructed Wetlands for Wastewater Treatment,* Lewis Publishers, Ann Arbor, Michigan, 1989.
10. Water Pollution Control Federation, Natural Systems for Wastewater Treatment, Manual of Practice No. FD-16, *Water Pollution Control Federation,* 1990.
11. Dinges, R., *Natural Systems for Water Pollution Control,* Van Nostrand and Reinhold Company, New York, 1982.
12. Reed, S.C., *Constructed Wetlands for Industrial Wastewaters,* Presented at Purdue Industrial Waste Conference, May 10, 1993, Purdue University, West Lafayette, IN.
13. Moshiri, G.A., *Constructed Wetlands for Water Quality Improvement,* Lewis Publishers, Florida, 1993.
14. US Environmental Protection Agency, *Subsurface Flow Constructed Wetlands Conference,* Proceeding of a conference at University of Texas at El Paso, August 16 and 17, El Paso Texas, 1993.
15. Stephenson, M., G. Turner, P. Pope, J. Colt, A. Knight, and G. Tchobanoglous, *The Use and Potential of Aquatic Species for Wastewater Treatment,* Appendix A, The Environmental Requirements of Aquatic Plants, Publication No. 65, California State Water Resources Control Board, Sacramento, California, 1980.
16. Tchobanoglous, G., Aquatic Plant Systems for Wastewater Treatment: Engineering Considerations, in *Aquatic Plants for Water Treatment and Resource Recovery.* K.R. Reedy and W.H. Smiths, Eds. Magnolia Publishing, Orlando, FL, 1987, pp. 27–48.
17. US Environmental Protection Agency, *Subsurface Flow Constructed Wetlands for Wastewater Treatment—A Technology Assessment,* Office of Water, EPA 832R-93-001, July 1993.
18. Cueto, A.J., Development of Criteria for the Design and Construction of Engineered Aquatic Treatment Units in Texas, Chapter 9, *Constructed Wetlands for Water Quality Improvements,* G.A. Moshiri, Lewis Publishers, FL 1993, pp. 99–105.
19. Hammer, D.A., B.P. Pullin, and J.T. Watson, Constructed Wetlands for Livestock Waste Treatment, National Nonpoint Conference, St. Louis, MO, April 1989.
20. US Environmental Protection Agency, *Design Manual on Constructed Wetlands and Aquatic Plant Systems for Municipal Wastewater Treatment,* Office of Research and Development, Center for Environmental Research Information, Cincinnati, OH, September 1988.
21. Brix, H., Treatment of Wastewater in the Rhizosphere of Wetland Plants—The Root-Zone Method, *Water Science Technology,* Vol. 19, 1987, pp. 107–118.
22. Brodie, G.A., D.A. Hammer, and D.A. Tomljanovich, Constructed Wetlands for Acid Drainage Control in the Tennessee Valley, in *Mine Drainage and Surface Mine Reclamation,* Bureau of Mines Information Circular 9183, 1988, pp. 325–331.

MOHAMMED S. KAMAL

SYED R. QASIM
The University of Texas at Arlington

NITROGEN OXIDES REDUCTION

INTRODUCTION

Nitrogen oxides are one of the most persistent categories of globally emitted air pollutants because they are combustion products of both stationary and mobile sources. By far the highest concentration of the oxides formed during combustion are NO and NO_2. Given time, and in the presence of O_2 either in the flue gases or in the atmosphere most of the NO is converted to NO_2.

There are six commonly encountered oxides of nitrogen: (1) nitric oxide (NO); (2) nitrous oxide (N_2O); (3) nitrogen dioxide (NO_2); (4) nitrogen trioxide (N_2O_3); (5) nitrogen tetroxide (N_2O_4); (6) nitrogen pentoxide (N_2O_5). Typically, nitrogen oxides, or NO_x, refers to a collective name for nitric oxide, NO, and nitrogen dioxide NO_2, both of which are formed as a by-product during combustion. Nitric oxide is a colorless, odorless, toxic, nonflammable gas which is slightly soluble in water. Nitrogen dioxide, however, is a reddish-brown gas that is toxic and highly corrosive with a pungent odor. This gas can contribute to highly visible plumes. Emissions of nitrogen oxides are of concern due to their potential role in ozone formation, acid rain deposition, health effects, and formation of toxic air pollutants.

EFFECTS OF NO_X

Human and Animal Health Effects

Of the two compounds, nitrogen dioxide is the most toxic and dangerous to humans. A variety of studies has been performed to observe the effects of NO_2 on humans and animals. Most of these studies have been performed using pure NO_2. Effects of acute NO_2 exposure have been reported as nose and eye irritation, obliterative bronchiolitis, pulmonary congestion and edema, pneumonitis, and death. Most of these reactions, such as pulmonary edema and obliterative bronchiolitis, can occur at extremely high concentrations (150–500 ppm) for short periods of time minutes to an hour. It appears, however, that mixtures of oxides tend to lessen the discomfort and the potential to contract severe disorders. It has also been shown that chronic, intermittent exposure to NO_2 (10–40 ppm) can result in chronic pulmonary fibrosis and emphysema.[1] In animal studies, continuous exposure to NO_2 for 90 days at 5 ppm has resulted in the deaths of 18% of the rats, 13% of the mice and 66% of the rabbits.[1] On the other hand, intermittent exposure for 18 months at the same level (5 ppm) did not result in any deaths. In addition to the chronic and acute effects, it appears that daily exposure to NO_2 concentrations of 5 ppm can lead to slightly accelerated lung tumor formation, but not at a level of any statistical significance.

On the other hand, exposure strictly to nitric oxide has not been reported to result in human poisoning probably due to its relatively low toxicity and its conversion to NO_2. The relative toxicity can be seen in the exposure standards set by NIOSH (National Institute for Occupational Safety and Health) and OSHA (Occupational Safety and Health Administration). For nitric oxide, the threshold limit value (TLV) for an 8-hour time weighted average (TWA) exposure is 25 ppm.[2] For nitrogen dioxide, the TLV is 1 ppm for a short-term exposure limit (STEL). The short-term exposure limit is a 15-minute TWA exposure that should not be exceeded at any time during a workday. Likewise, the immediately dangerous to life or health (IDLH) concentration for nitric oxide is 100 ppm, whereas for nitrogen dioxide, the concentration is 50 ppm.[2] Toxicologists have reported that nitric oxide can be a mild nose, eye and throat irritant. At high concentrations, nitric oxide can lead to a progressive depression of the central nervous system.[1]

Environmental Effects

Not only does NO_x affect the human population, but it also has an adverse effect on the environment, in particular vegetation. Gaseous pollutants damage plants by entering through the stomata during the respiration cycle.[3] The pollutants can disrupt the photosynthesis process and can destroy plant chlorophyll. Experiments have shown that concentrations of NO_2 as low as 0.5 ppm can result in reduced plant growth by as much as 35%. In particular, it appears that plants are more susceptible to nitrogen dioxide effects at night than during the day. Scientists feel that NO_x has played a significant role in the deforestation of central Europe.

Not only does direct exposure to nitrogen oxides result in plant deterioration, but nitrogen oxides combine with certain hydrocarbons to form ozone and peroxyacyl nitrates (PAN's), two compounds that have been found be more toxic towards plants than nitrogen oxides alone. Exposure to these compounds has been shown to result in plant growth suppression, bleaching, glazing and silvering on the lower surface of the leaves.[3]

In addition to vegetative kill, nitrogen dioxide is an extremely corrosive gas that can have deleterious effects on a wide range of materials such as plastics, fabrics, rubber and metals. Studies on Nylon-6 and Kevlar have revealed that the ultimate tensile strengths of these materials are reduced when exposed to a NO_x environment in the range of 0.5–0.8% volume concentration.

Global Warming/Greenhouse Effect

The greenhouse effect occurs due to the buildup of gases which can absorb heat. The earth maintains a constant average temperature by radiating heat to space; thus, greenhouse gases absorb a portion of this heat and radiate it to the lower atmosphere. Scientists believe that the heat is trapped in the lower atmosphere, resulting in global warming, rising sea level and global climatological alterations. There is considerable debate over the global warming issue when viewed from the geologic time scale. In any case, there has been considerable research done in this field and it will continue to be a topic of concern for years to come.

Principally, carbon dioxide is the most important greenhouse gas, contributing to roughly half of the global warming that has been reported. The other half of the global warming has been attributed to approximately 20 other gases, most notably methane, chlorofluorocarbons (CFC-11 and CFC-12), ozone, nitrous oxide, and nitrogen dioxides. Some scientists have estimated that a 50% increase in the current concentration of N_2O will result in a mean global temperature increase of 0.2–0.5°C.[4] Of the nitrous oxide emissions, 25% has been attributed to fossil fuel combustion. If these predictions are accurate, then it becomes even more important to control the emissions of nitrogen oxides from combustion related processes.

Acid Rain Deposition

Acid rain forms when sulfur dioxide and nitrogen oxides mix with water vapor to form sulfuric and nitric acid. In particular, nitrogen oxides are transformed into nitric acid by the following reactions:[5]

$$2NO + O_2 \rightarrow 2NO_2$$

$$NO + O_3 \rightarrow NO_2 + O_2$$

$$2NO_2 + 2H_2O + O_2 \rightarrow 4HNO_3.$$

On an annual basis, nitric acid is responsible for approximately 30% of the acidity of rainfall. This percentage increases to around 50% during the winter. Acid rain is one of the most damaging effects of NO_x emissions. It leads to the destruction of ecosystems in lakes, deforestation, and the stripping of organic material in soils, creating erosion and potentially metal laden soils. In addition, NO_x emissions contribute substantially to the acid pulse in snowmelt, which in turn severely impacts the freshwater ecosystem. Therefore, acid rain is yet another reason for controlling and limiting NO_x emissions to the atmosphere.

Ozone Formation

In addition to being a cause of acid rain, nitrogen oxides are also considered one of two precursors to the formation of primary ozone, O_3. Ozone is thought to be formed from the complex reaction of certain hydrocarbons and nitrogen oxides.

The role of nitrogen oxides in ozone formation is significant because of the health effects associated with elevated levels of ozone. Ozone exposure can lead to coughing and chest discomfort, headaches, upper respiratory illness, reduced pulmonary function, eye irritation, and increased asthma attacks. For ozone, the NIOSH ceiling exposure limit is 0.1 ppm; the OSHA threshold limit value (TLV) for an 8-hour time weighted average (TWA) exposure is 0.1 ppm. The OSHA short-term exposure limit is 0.3 ppm.[2] The immediately dangerous to life or health (IDLH) concentration for ozone is 10 ppm, which is five times less than the IDLH for nitrogen dioxide and nitric oxide, respectively. In addition to human health effects, studies have shown that ozone damages agricultural crops and forest ecosystems.

NO_x REGULATIONS

Stationary Source Regulations

Because of the harmful effects associated with nitrogen oxides, governments around the world have established increasingly stricter regulations over the past few decades to curb and reduce the amount of NO_x emissions. In particular, the United States federal government enacted the Clean Air Act (CAA) to protect the nation's air quality. The CAA was first passed into legislation in 1970, with substantial amendments being added in 1977 and most recently in 1990. The amendments in 1990 greatly increased the scope of the existing CAA. The CAA is divided into seven primary Titles, I–VII. Some of he new major points in these Titles are as follows:[6, 7]

- **Title I**—A new system was established to determine if an area is classified as either an ozone attainment or an ozone nonattainment area. The CAA effectively divides the United States into attainment and nonattainment areas, based on the level of criteria pollutants in the area's ambient air. Because of the formation of ozone from volatile organic compounds (VOC's) and NO_x, Title I requires more stringent requirements for these two classes for compounds.
- **Title II**—Clean fuel requirements and increased restrictions on motor vehicle emissions were introduced in Title II. Under Title II, new vehicle tailpipe emissions of NO_x were to be reduced by 60%, starting with 40% of all new vehicles in the 1994 model year and increasing to 100% by the 1998 model year.
- **Title III**—Emission limitations for 189 hazardous air pollutants (HAP's) are to be set by the

United States Environmental Protection Agency (USEPA). The National Emissions Standards for Hazardous Air Pollutants (NESHAP's) set in earlier versions of the CAA will remain intact, but the USEPA will now be required to use the best demonstrated emissions control practices in a particular industry to regulate sources for that industry. Under standards set by the USEPA, major sources will be required to apply Maximum Available Control Technology (MACT). A major source is defined as one that emits 10 tons/year of a HAP or 25 tons/year of any combination of HAPs.

- **Title IV**—Under Title IV, the USEPA is required to establish a program to reduce the occurrence of acid rain. Because SO_2 and NO_x are the two main contributors to acid rain, facilities that fall under control of Title IV will be forced to meet certain standards and will be required to obtain an acid rain permit.
- **Title V**—A comprehensive operating permit program was established for all significant air emission sources.
- **Title VI**—Under this Title, a new national program was developed to phase out the use of chlorofluorocarbons (CFCs) and similar compounds to protect the stratospheric ozone layer.
- **Title VII**—The USEPA's enforcement ability was greatly enhanced with more criminal and civil powers.

Included in the Title I requirements are the National Ambient Air Quality Standards (NAAQS), which specify the maximum allowable concentrations for six criteria pollutants. These pollutants are: (1) carbon monoxide; (2) lead; (3) nitrogen oxides; (4) ozone; (5) particulates ($\leq$ 10 microns diameter); and (6) sulfur oxides. There are two types of NAAQS, which are defined in the USEPA 40 CFR Part 50 Regulations:

- primary—standards are designed to protect the public health.
- secondary—standards are designed to protect the public from a pollutant's effects on visibility, personal comfort, properly, etc.

If an airshed area exceeds the ambient air concentrations of one of these pollutants, then that area is considered to be in nonattainment. For nitrogen oxides, the primary and secondary National Ambient Air Quality Standards expressed as annual arithmetic mean concentration are 0.053 parts per million (100 $\mu g/m^3$).[8] This limit should not be exceeded, during any 12 consecutive month period, for the annual average of the 24-hour concentrations. Figure 1[9] shows the control technologies required by facilities to meet the NAAQS. The three main categories are BACT, LAER, and RACT. BACT or Best Available Control Technology, is defined by the CAA as " … an emission limitation based on the maximum degree of reduction of each pollutant … which the permitting authority … "[9] considers appropriate. BACT applies to new or modified sources of emissions in attainment areas. LAER or Lowest Achievable Emission Rate applies to new or modified sources in nonattainment areas. It refers to the most stringent emission limitation achieved by a similar facility or a particular source category.[9] RACT or Reasonably

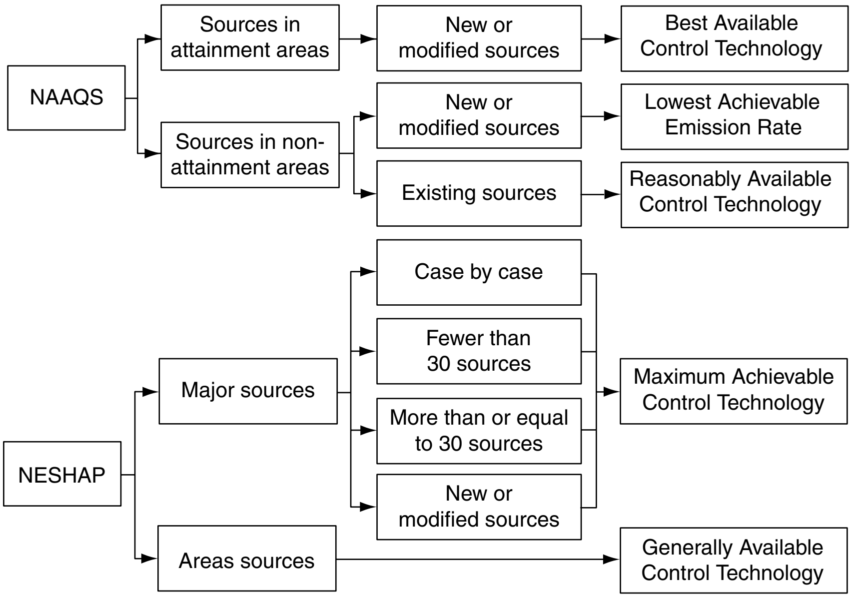


FIGURE 1 Relationships between control technologies and standards.[9]

Available Control Technology applies to existing sources in nonattainment areas. RACT has been defined by the USEPA as "the lowest emission limitation that a particular source is capable of meeting by the application of control technology that is reasonably available considering technological and economic feasibility".[9]

One of the ways the CAA provides for regulation of emissions is through the New Source Performance Standards (NSPS), which is also a component of Title I. The USEPA continually promulgates new standards, under the 40 CFR Part 60's, to regulate the emissions of criteria pollutants from new or substantially modified stationary sources. These new sources may also have to undergo a New Source Review (NSR) permitting process. Under this process, the regulating agency determines whether the facility can begin operation and under what conditions. The following is a list of new source performance standards that pertain to facilities that are regulated for NO_x emissions:[10]

- *Subpart D-Fossil Fuel Fired Steam Generators* (facilities that began construction after 8/17/71) (>250 MBtu/hr)[10]
- The NO_x emission limits are found in Table 1.
- For combinations of fuels, the following proration applies:

$$E_n = \frac{260w + 86x + 130y + 300z}{w + x + y + z}$$

where

E_n = NO_x limit (ng/J heat input)
w = % of total heat input derived from lignite
x = % of total heat input derived from gaseous fossil fuel
y = % of total heat input derived from liquid fossil fuel
z = % of total heat input derived from solid fossil fuel.

- *Subpart Da-Electric Utility Steam Generating Units* (facilities that began construction after 9/18/78) (>250 MBtu/hr)[10]
- The NO_x emission limits are found in Table 2.
- For combinations of fuels, the following proration applies:

$$E_n = \frac{86w + 130x + 210y + 260z + 340v}{100}$$

where

E_n = NO_x limit (ng/J heat input)
w = % of total heat input derived from combustion of fuels subject to the 86 ng/J heat input standard
x = % of total heat input derived from combustion of fuels subject to the 130 ng/J heat input standard
y = % of total heat input derived from combustion of fuels subject to the 210 ng/J heat input standard
z = % of total heat input derived from combustion of fuels subject to the 260 ng/J heat input standard
v = % of total heat input derived from combustion of fuels subject to the 340 ng/J heat input standard.

- *Subpart Db-Industrial-Institutional Steam Generating Units* (facilities that began construction after 6/19/84) (> 100 MBtu/hr)[10]
- The NO_x emission limits are found in Table 3.
- *Subpart Ea-Municipal Waste Combustors* (facilities that began construction after 12/2/89) (>250 tons/day)[10]
- 180 ppm by volume, corrected to 7% O_2 (dry basis).
- *Subpart G-Nitric Acid Plants* (facilities that began construction after 8/17/71)[10]
- 3.0 lb NO_x (as NO_2)/ton of acid produced expressed as 100% nitric acid).
- *Subpart GG-Stationary Gas Turbines* (facilities that began construction after 10/3/77)[10]
- The NO_x emission limits for stationary gas turbines are given by two equations. For units with a heat input at peak load of >107.2 GJ (100 MBtu) per hour, the allowable NO_x emission at 15% O_2, dry is given by the following equation:

$$E_n = 0.0075\,\frac{(14.4)}{Y} + F$$

For units with a heat input peak load of >10.7 GJ (10 MBtu) per hour but <107.2 GJ (100 MBtu) per hour and for units with a base load at ISO conditions of 30 MW or less, the

TABLE 1
USEPA 40 CFR Part 60 Subpart D-Fossil-fuel fired steam generators NO_x emission limits[10]

Fuel type	NO_x emission limits, ng/J(lb/MBtu) (expressed asNO_2) for heat input
(1) Gaseous fossil fuel	86 (0.20)
(2) Liquid fossil fuel	129 (0.30)
(3) Liquid fossil fuel + wood residue	129 (0.30)
(4) Gaseous fossil fuel + wood residue	129 (0.30)
(5) Solid fossil fuel[1]	300 (0.70)
(6) Solid fossil fuel + wood residue[1]	300 (0.70)
(7) Lignite, except (9)	260 (0.60)
(8) Lignite + wood residue, except (9)	260 (0.60)
(9) Lignite mined in North Dakota, South Dakota, or Montana, and which is burned in a cyclone-fired unit	340 (0.80)

[1] Except liginite or a solid fossil fuel containing 25%, by weight, or more coal refuse.

TABLE 2

USEPA 40 CFR Part 60 Subpart Da-electric utility steam generating units NO_x emission limits[10]

Fuel type	Emission limit for heat input ng/J	Emission limit for heat input (lb NO_x/MBtu)
Gaseous fuels		
Coal-derived fuels	210	0.50
All other fuels	86	0.20
Liquid fuels		
Coal-derived fuels	210	0.50
Shale oil	210	0.50
All other fuels	130	0.30
Solid fuels		
Coal-derived fuels	210	0.50
Any fuel containing more than 25%, by weight, coal refuse[1]	([1])	([1])
Any fuel containing more than 25%, by weight, lignite if the lignite is mined in North Dakota, South Dakota, or Montana, and is combusted in a slag tap furnace[2]	340	0.80
Any fuel containing more than 25%, by weight, lignite not subject to the 340 ng/J heat input emissions limit[2]	260	0.60
Subbituminous coal	210	0.50
Bituminous coal	260	0.60
Anthracite coal	260	0.60
All other fuels	260	0.60

[1] Exempt from NO_x standards and NO_x monitoring requirements.

[2] Any fuel containing less than 25%, by weight, lignite is not prorated but its % is added to the % of the predominant fuel.

allowable NO_x emission at 15% O_2, dry is given by the following equations:

$$E_n = 0.0150 \frac{(14.4)}{Y} + F.$$

where

E_n = NO_x limit (% by volume, dry)

Y = mfg.'s rated heat rate at mfg.'s rated peak load or the actual measured heat rate based on lower heating value of fuel as measured at actual peak load for the facility. The value of Y shall not exceed 14.4 kJ per watt hour.

F = defined according to the nitrogen content of the fuel as follows:

Fuel-bound nitrogen (% by weight)	F (NOx % by volume)
$N < 0.015$	0
$0.015 < N < 0.1$	0.04 (N)
$0.1 < N < 0.25$	0.004 + 0.0067 (N-0.1)
$N > 0.25$	0.005

In addition to NSPS and NAAQS, the USEPA is reducing NO_x emissions through Title IV, the Acid Rain Program. The regulations under this program were published as final in the April 13, 1995 *Federal Register* and became effective on May 23, 1995. The regulations are aimed directly at coal fired utility plants in which the combustion of coal on a BTU basis exceeds 50% of its annual heat input. The type of boilers used at these plants has been subdivided into Group 1 and Group 2 boilers. Group 1 boilers include tangentially fired boilers or dry bottom wall-fired boilers. Group 2 boilers include wet bottom wall-fired boilers, cyclone boilers, vertically fired boilers, arch-fired boilers, or utility boilers (i.e. fluidized bed or stoker boilers). In addition, the CAA has set standards based on Phase I and Phase II.

In the acid rain regulations, NO_x emission limits have been established for Phase I coal fired utility units with tangentially fired boilers (95 units) or with dry bottom wall-fired boilers (84 units), which were to be effective January 1, 1996. For tangentially fired boilers, the limit is 0.45 lb/MBtu of heat input on an annual average basis. For dry bottom wall fired boilers, the NO_x limit is 0.50 lb/MBtu of heat input expressed on an annual average basis. The facilities that cannot meet the requirements will be allowed to apply for a less restrictive emission standard or to join an "averaging pool," through which the overall emissions limit average is attained. Phase I standards for Group 2 boilers are scheduled to be set by January 1, 1997 with implementation by January 1, 2000. Phase II standards for Group 1 and 2 boilers are required to be established by 1997.[11]

Similar to the federal government under the Clean Air Act, State governments have the authorization to enact regulations that maintain ambient air quality and to set limits for sources of air pollution. In New York State, the agency charged with

TABLE 3
USEPA 40 CFR Part 60 Subpart
Db-industrial/commercial/institutional steam generating units
NO_x *emission limits*[10]

Fuel/steam generating unit type	NO_x emission limits, ng/J(lb/MBtu) (expressed as NO_2) for heat input
(1) Natural Gas and distillate oil, except (4)	
(i) Low heat release rate	43 (0.10)
(ii) High heat release rate	86 (0.20)
(2) Residual oil	
(i) Low heat release rate	130 (0.30)
(ii) High heat release rate	170 (0.40)
(3) Coal	
(i) Mass-feed stoker	210 (0.50)
(ii) Spreader stoker and fluidized bed combustion	260 (0.60)
(iii) Pulverized coal	300 (0.70)
(iv) Lignite, except (v)	260 (0.60)
(v) Lignite mined in North Dakota, South Dakota, or Montana, and is combusted in a slag tap furnace	340 (0.80)
(vi) Coal-derived synthetic fuels	210 (0.50)
(4) Duct burner used in a combined cycle system	
(i) Natural gas and distillate oil	86 (0.20)
(ii) Residual oil	170 (0.40)

enforcing these regulations is the New York State Department of Environmental Conservation (NYSDEC). The USEPA has delegated authority to the NYSDEC to issue permits to construct to sources for the modification or construction of any stationary source subject to federal NSPS, NESHAPS, or PSD (Prevention of Significant Deterioration) requirements, and to implement and enforce the federal standards and PSD requirements where they apply.[12] The NYSDEC requires that any new or modified air emission source acquire a permit to construct and a certificate to operate under the requirements of 6 NYCRR Part 201. Any air contamination source subject to the NSPS, NESHAPS, or PSD requirements must be in compliance with the federal standards as well as any applicable State standards, such as New York State Ambient Air Quality Standards (NYSAAQS) or New York State performance standards. The NYSAAQS for NO_2 is 0.053 ppm (100 $\mu g/m^3$)[13] with the same averaging period as that of the USEPA NAAQS. Recently, the NYSDEC made revisions to the 6 NYCRR Part 200, 201, and 227 air pollution regulations. In particular, these revisions included Subpart 227–2 entitled "Reasonably Available Control Technology (RACT) for Oxides of Nitrogen (NO_x)."[14] This subpart requires operators of existing major stationary sources of NO*x* to use RACT to limit NO_x emissions. A source is classified as a major source if it has the potential to emit 100 tons NO_x per year or if it has the potential to emit 25 tons NO_x per year in a "severe" ozone nonattainment area. The major stationary sources as defined in Subpart 227–2 are:[14]

- Very large boilers—$Q_{MAX} > 250$ MBtu/hr
- Large boilers—100 MBtu/hr $< Q_{MAX} \leq 250$ MBtu/hr
- Mid-size boilers—50 MBtu/hg $< Q_{MAX} \leq 100$ MBtu/hr
- Small boilers—20 MBtu/hr $< Q_{MAX} \leq 50$ MBtu/hr
- Combustion turbines—$Q_{MAX} > 10$ MBtu/hr
- Internal combustion engines—≥225 HP in severe nonattainment areas; >400 HP in all other areas
- Other combustion sources

where, Q_{MAX} is equal to the maximum heat input capacity or rate.

Tables 4–8[14] contain information regarding the NO_x emission limits for the above sources. These limits were to be met by the existing sources by May 31, 1995. Small boiler operators are only required to perform annual tune-ups and to maintain a log book containing process information. Other major combustion sources of NO_x not covered under one of the specific sources above were required to submit a RACT proposal detailing the technology and potential emissions to the NYSDEC by May 31, 1995.

Outside of state and federal regulations, certain cities promulgate their own air pollution control regulations.

The New York City Department of Environmental Protection (NYC DEP) revised their air pollution code in March, 1992. This document contains guidelines for obtaining local permits and certificates as well as standards for air pollutant emissions. In regards to NO_x emissions, the NYC DEP set a limit for boilers with a capacity of >500 MBtu/hr depending on whether the boiler was completed before (150 ppmv NO_x) or after (100 ppmv NO_x) August 20, 1971.[15]

The petrochemical and refinery sectors have been subjected to stringent regulatory requirements in recent years by state

TABLE 4
NYS DEC NO$_x$ RACT emission limits for very larger boilers (lbs NO$_x$/MBtu)[14]

	Boiler configuration			
Primary fuel type	Tangential	Wall	Cyclone	Stokers
Gas only	0.20	0.20	NA	NA
Gas/oil	0.25	0.25	0.43	NA
Coal wet bottom	1.00	1.00	0.60	NA
Coal dry bottom	0.42	0.45	NA	0.30[1]

[1] The limit is 0.33 lbs NOx/MBtu when 25% of other solid fuels, on a Btu basis, are utilized.

TABLE 5
NYS DEC NO$_x$ RACT emission limits for large boilers (lbs NO$_x$/MBtu)[14]

Primary fuel type	Emission limit
Gas only	0.20
Gas/oil	0.30
Pulverized coal	0.50
Coal (overfeed stoker)	0.30[1]

[1] The limit is 0.33 lbs NO$_x$/MBtu when 25% of other solid fuels, on a Btu basis, are utilized.

TABLE 6
NYS DEC NO$_x$ RACT emission limits for mid-size boilers (lbs NO$_x$/MBtu)[14]

Primary fuel type	Emission limit
Gas only	0.10
Distillate oil	0.12
Residual oil	0.30

TABLE 7
NYS DEC NO$_x$ RACT emission limits for combustion turbines (ppm, corrected to 15% O$_2$)[14]

Combustion turbine and fuel types	Emission limit
Simple cycle and regenerative	
Gas only	50
Multiple fuels	100
Combined cycle	
Gas only	42
Oil only	65

TABLE 8
NYS DEC NO$_x$ RACT emission limits for internal combustion engines (gms NO$_x$/brake HP-hr)[14]

Engine and fuel types	Emission limit
Rich burn	2.0
Lean burn	
Gas only	3.0
Other fuels	9.0

[1] Rich burn engine—O$_2$ by volume is less that 1.0% in the exhaust.

[2] Lean burn engine—O$_2$ by volume is greater than or equal to 1.0% in the exhaust.

and federal governments.[33,34,35] NO$_x$ control options have been applied to fired process heaters[34] whose uncontrolled emissions are in the range of 0.10 to 0.53 lb/10^6 Btu. The percentage reductions attained for key control technologies include,

(1) Low NOx burners (25–65%)
(2) Selective Catalytic Reductions (SCR) (65–90%) and
(3) Combined Low NOx and SCR (70–90%)

Additional factors involved in choosing a desired technology include capital, operating and maintenance costs.[35]

Mobile Source Regulations

In addition to stationary combustion, mobile sources, such as passenger cars and trucks, contribute almost 50% of the NO$_x$ produced in the United States. To reduce the emissions from these sources, federal and State governments have implemented standards on tailpipe emissions of NO$_x$. Typically, vehicles are divided into three main weight categories: (1) light duty, (2) medium duty, and (3) heavy duty. Vehicles are further subdivided in each main category; emission limits are applied to these subcategories. Emission limits are expressed two different units: (1) g/mi, grams per mile, or (2) g/bhp-hr, grams per brake horse-power-hour. In the federal and California regulations, standards are expressed in g/mi for light and medium duty vehicles, whereas g/bhp-hr is used for heavy duty vehicles. As a comparison of some of the emission limits, Table 9 shows the federal, California and New York[16,17] standards for passenger and light duty vehicles.

To further reduce air pollutant emissions from vehicles, the federal and State governments have implemented a number of other programs. One such program is the Clean-Fuel Vehicles Fleet Program which is designed to reduce emissions, in highly polluted regions, from vehicles belonging to a fleet. Tailpipe emissions generally account for 60% of the total vehicle emissions.[16] Thus, governments are beginning to focus on the other emission forms associated with vehicles. These include evaporative emissions, CO emissions at cold temperatures, air toxics emissions, emissions testing and procedures, and emissions control diagnostics systems. Regulating emissions is not the only method to curb NO production. The composition of the fuel has also been a target of regulations. Requirements have been promulgated to establish the use of reformulated gasoline, oxygenated gasoline, reduced volatility gasoline, and cleaner diesel fuel.

Thus, the federal and state governments have taken active role in the limiting of NO$_x$ emissions. Depending on the political climate in the United States, these gulations may become less or more stringent over the next few years.

TABLE 9

NO_x emission standards for passenger cars and light duty trucks (g/mi)[16,17]

		New York			California			Federal		
		PC	LDTI[2]	LDT2[3]	PC	LDTI[2]	LDT2[3]	PC	LDTI[2]	LDT2[3]
1993	5 yr/ 50,000 mi	0.4 (0.7)[1]	0.4 (0.7)[1]	1.0	0.4	0.4	1.0	1.0	1.2	1.7
	10 yr/ 100,000 mi	—	—	—	—	—	—	—	—	—
1994	5 yr/ 50,000 mi	0.4 (0.7)[1]	0.4 (0.7)[1]	1.0	0.4	0.4	0.7–1.0	0.4–1.0	0.4–1.2	0.7–1.7
	10 yr/ 100,000 mi	—	—	—	0.6	0.6	0.9	0.6	0.6	0.97
1995	5 yr/ 50,000 mi	0.4	0.4	0.7	0.4	0.4	0.7	0.4–1.0	0.4–1.2	0.7–1.7
	10 yr/ 100,000 mi	—	—	—	0.6	0.6	0.9	0.6	0.6	0.97

[1] The standard for in-use compliance for passenger cars and light duty trucks certifying to the 0.4 g/mi NO_x standard shall be 0.55 g/mi NO_x for 50,000 miles.

[2] LDTI refers to light duty trucks from 0–3, 750 pounds loaded vehicle weight.

[3] LDT2 refers to light duty trucks from 3,751–5750 pounds loaded vehicle weight, but less than 6,000 pounds gross vehicle weight.

Conversion of Emission Standards

As can be seen from the above regulations, NO_x standards come in various units. Some regulations state NO_x limits in ppm, on a dry volume basis, whereas other regulations may require the emissions in lbs NO_x/MBtu. Figures 2–4 have been developed for converting emission units at different oxygen percents in the flue gas for natural gas, oil, and coal fired units, respectively. Figure 2 is based on a typical gaseous fuel, a US/Texas gas (see Table 10[18,19]). Figures 3 and 4 were based on a number of different liquid and solid fuels (see Tables 11[18,20] and 12[18,21], respectively). Three Lotus spreadsheets were utilized to develop Figure 2. First, a spreadsheet was developed, Table 13, that converts the gas components, on a percent volume basis, to the elemental component mass fractions. In addition, the spreadsheet also calculates the lower and higher heating values of the natural gas based on the components in the natural gas. The lower heating value is compared for accuracy to experimental results. Next, the component mass fractions were read into the second spreadsheet, Table 14, to calculate the flue gas constituents and the pounds of flue gas produced per pound of fuel on a dry weight basis. Using the quantity of flue gas and the higher heating value of the natural gas, a "K" factor was developed for all three figures via:

$$K = \frac{HHV \times V_M}{FG \times MW}$$

where

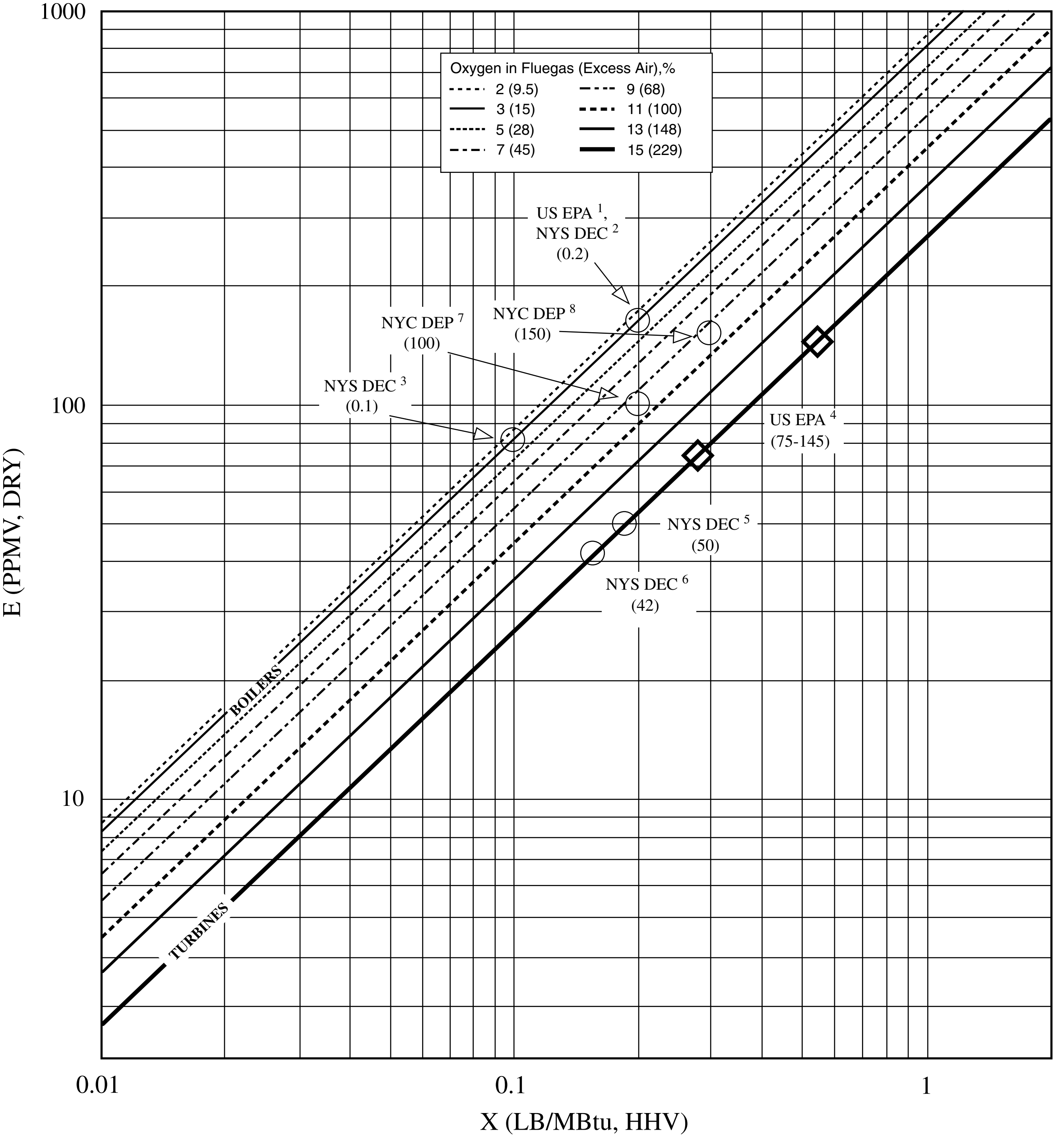
HHV = heating value of the natural gas, Btu/lb
FG = volume of flue gas per lb of fuel, dscf/lb
MW = molecular weight of NO_2, lb/lbmol
V_M = molar volume, dscf/lbmol

In this equation, V_M is equal to 385.1 dscf/lbmol at 68°F and 29.92 in. Hg. The "K" factors for a new selected fuels are found in Tables 10–12. As can be seen from the tables, the "K" factors do not vary widely for each type of fuel, which implies that Figures 2–4 could be readily applied to a particular fuel type. The exception to the rule are gaseous fuels that have a moderately high carbon dioxide content, such as the natural gas from Germany. Because the carbon dioxide does not play a role in the combustion process, a high carbon dioxide content gas will produce significantly lower flue gas per pound of fuel, thereby increasing the "K" factor. The "K" factor was then read into the third spreadsheet to be used in the following equation to develop the data points for each conversion graph:

$$X = \frac{K \times E \times (20.9 - \%O_2)}{20.9}$$

where

X = NO_x concentration, ppmvd
E = lb NO_x/Mbtu
$\%O_2$ = % oxygen in the flue gas.



[1] - New Source Performance Standards (NSPS)

[2] - > 100 MBtu/hr in size

[3] -50 - 100 MBtu/hr in size

[4] - Depends on heat rate (NSPS)

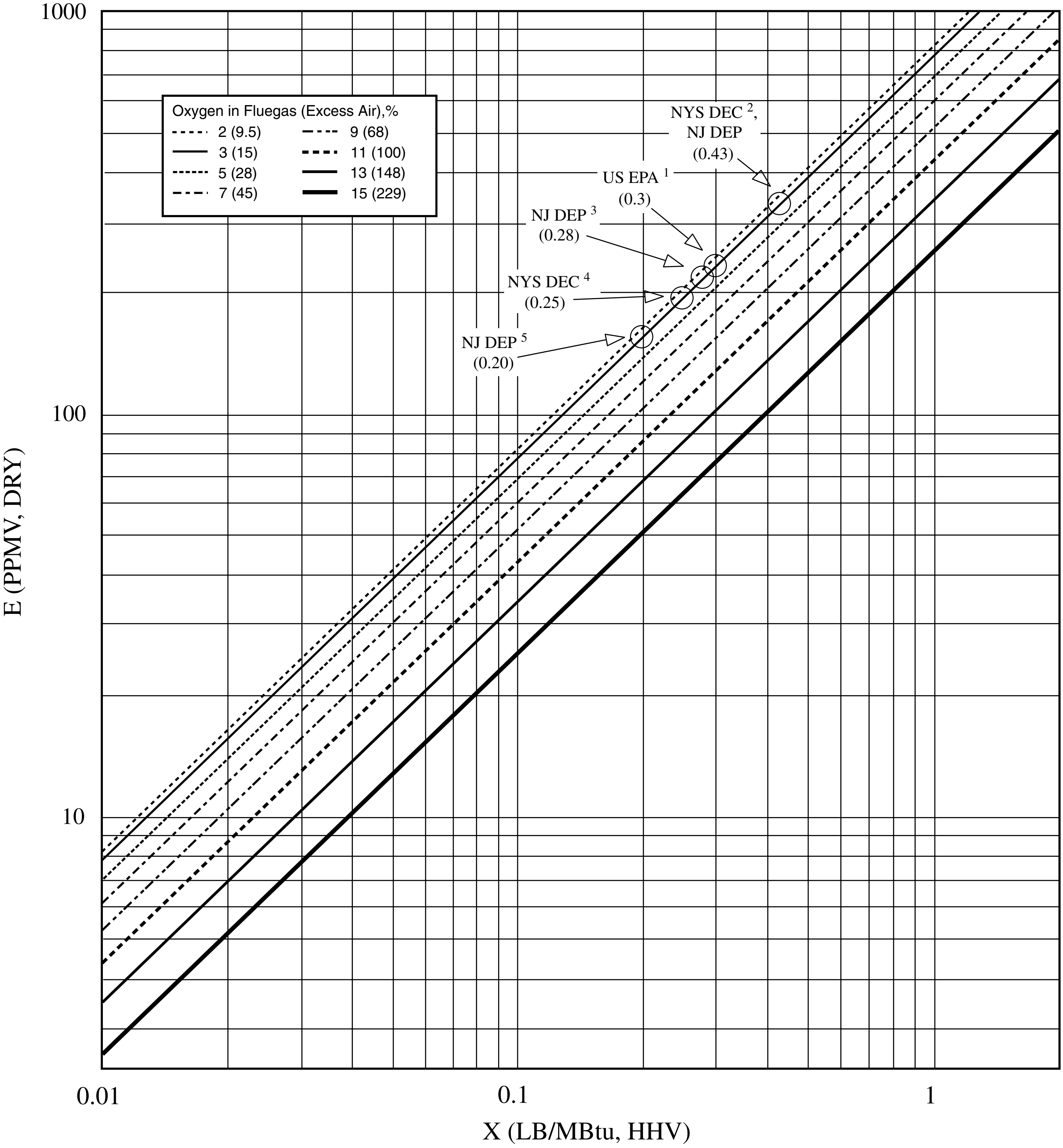
[5] - Simple cycle and regenerative combustion turbines

[6] - Combined cycle combustion turbines

[7] - > 500 MBtu/hr in size, after 8/20/71

[8] - > 500 MBtu/hr in size, before 8/20/71

FIGURE 2 Conversion of emission units and comparison of various standards for NO_x natural gas units.



[1] - NSPS-Subpart D- >250 MBtu/hr

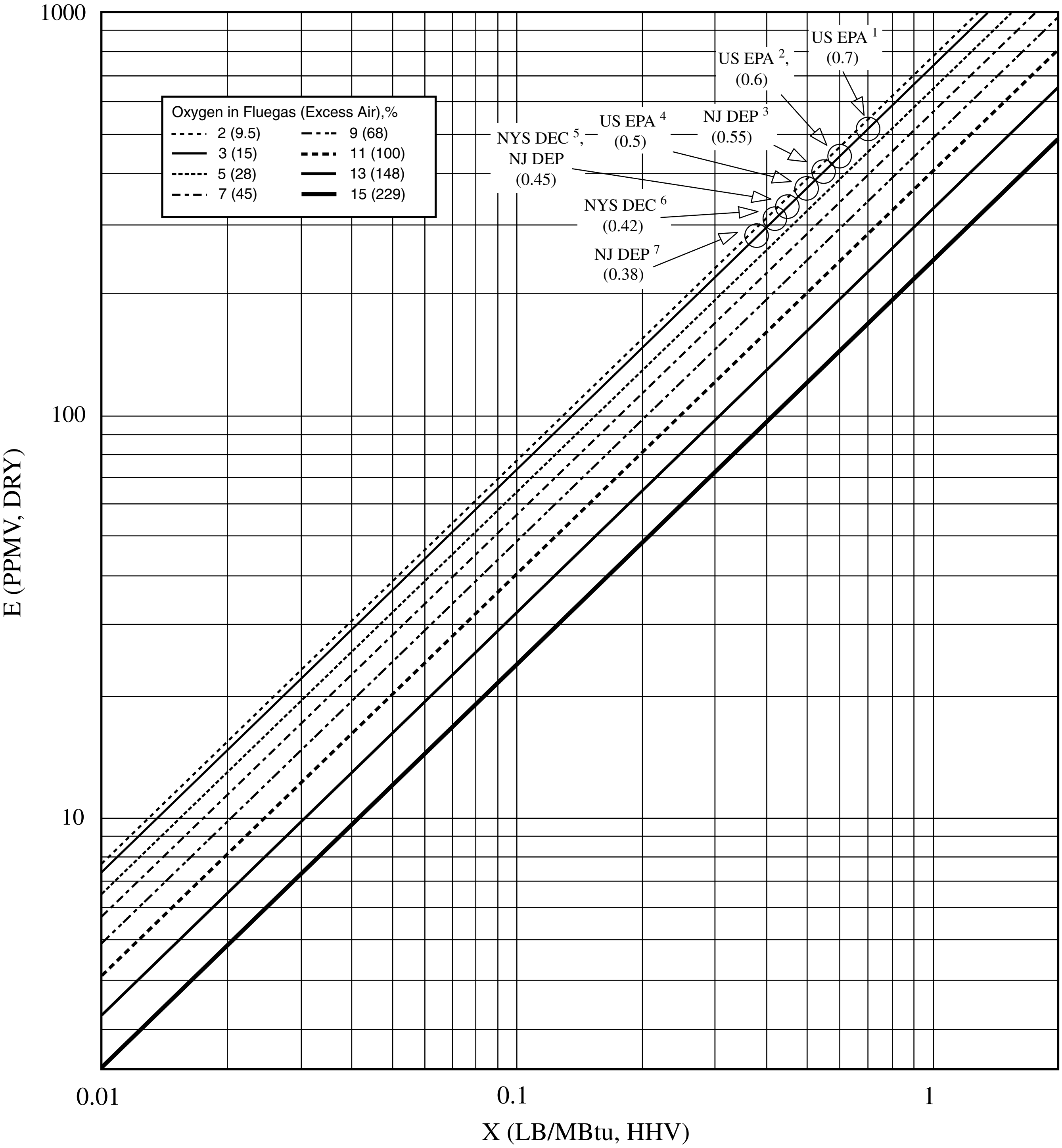
[2] - Dry bottom cyclone boiler
NJ DEP - utility boiler
NYS DEC - >250 MBtu/hr

[3] - Dry bottom wall utility boilers

[4] - Dry bottom tangential & wall boilers-> 250 MBtu/hr

[5] - Dry bottom tangential utility boilers

FIGURE 3 Conversion of emission units and comparison of various standards for NO_x oil-fired units.



1 - NSPS - Subpart D - > 250 MBtu/hr

2 - NSPS - Subpart Da - > 250 MBtu/hr (bituminous & anthracite coal)

3 - Dry bottom cyclone utility boilers

4 - NSPS - Subpart Da - > 250 MBtu/hr (subbituminous coal)

5 - Dry bottom wall utility boilers

6 - Dry bottom tangential boilers

7 - Dry bottom tangential utility boilers

FIGURE 4 Conversion of emission units and comparison of various standards for NO_x coal-fired units.

In addition to the conversion properties of the graph, some of the pertinent local, state and federal regulations have been included on the graph for quick comparison to the NO*x* levels.

CONTROL OF NO_x

Introduction

As noted earlier, there are numerous ways to control NO_x emissions from combustion systems. Control methods can be broken down into two categories: (1) technologies that reduce emissions during the combustion process; (2) technologies that reduce emissions after combustion has taken place. Technologies used during the combustion process typically utilize different stoichiometric balances or burner configurations to decrease NO_x formation. Post combustion technologies, such as SCR, utilize reactants, with or without catalysts, to break down the NO_x before it is exhausted to the atmosphere. Table 15[22,23] shows some of the control methods used in industry along with the NO_x reductions that can be expected. The following is a detailed description of each of the common processes used in industry today.

TABLE 10

Properties of gaseous fuels (% by weight)[18,19]

Fuel	CH_4	C_2H_6	C_2H_8	Other hydro-carbons	CO	H_2	H_2S	N_2	CO_2	"K" factor	HHV (MJ/m^3)
Natural Gas											
No. 1	77.7	5.6	2.4	1.8	—	—	7.0	—	—	972.34	40.4
No. 2[1]	88.8	6.4	2.7	2.0	—	—	0.0004	—	0.0	922.09	41.9
No. 3	59.2	12.9	—	—	—	—	—	0.7	26.2	1293.4	30.7
No. 4	99.2	—	—	—	—	—	—	0.6	0.2	892.52	36.3
US/Texas	80.9	6.8	2.7	1.6	—	—	—	7.9	0.1	952.93	41.1
North Sea	94.4	3.1	0.5	0.4	—	—	—	1.1	0.5	958.76	40.3
Denmark	91.2	4.8	1.7	1.1	—	—	—	0.5	0.7	956.19	42.6
Holland	81.3	2.9	0.4	0.2	—	—	—	14.3	0.9	973.06	35.2
Russia	98.0	0.6	0.2	0.1	—	—	—	1.0	0.1	969.24	39.2
Germany	74.7	0.1	—	—	—	—	—	7.2	18.0	1680.8	41.2
Refinery Gas											
No. 1	41.6	20.9	19.7	15.6	—	—	2.2	—	—	—	68.6
No. 2	4.3	82.7	13.0	—	—	—	—	—	—	—	67.1
No. 3	15.9	5.0	—	2.4	14.3	50.9	—	8.4	2.2	—	18.7
Coke oven gas	—	—	—	35.3	6.3	53.0	—	3.4	1.8	—	21.5
Blast furnace gas	—	—	—	—	26.2	3.2	—	57.6	13	—	3.4
Methane	100.0	—	—	—	—	—	—	—	—	—	38.9
Hydrogen	—	—		—	—	100.0	—	—	—	—	12.4

[1] "Sweetened"; H_2S removed.

TABLE 11

Properties of Liquid Fuels[18,20]

	Percent by weight								
Fuel	C	H	N	O	S	Ash	Specific gravity	"K" factor	HHV (MJ/kg)
Kerosene (No. 1) Fuel Oil	86.5	13.2	0.1	0.1	0.1	Trace	0.825	926.32	46.4
No. 2	86.4	12.7	0.1	0.1	0.4–0.7	Trace	0.865	916.47	45.5
No. 4	85.6	11.7	0.3	0.4	<2	0.05	0.953	895.69	43.4
No. 6	85.7	10.5	0.5	0.4	<2.8	0.08	0.986	896.05	42.5
No. 6 (Arthur Kill)	86.9	12.6	0.2	0.0	0.3	0.02	0.9059	903.35	44.9

Stationary Source Control

Low excess air operation (LEA) NO_x emissions are a function of the amount of available oxygen. Thus, one simple method of reducing NO_x emissions is by reducing the excess air level to the burners. Low excess air operation is effective in reducing fuel NO_x formation, but is limited in decreasing thermal NO_x emissions. Normally, this level is set to some constraint such as flame length, flame stability or carbon monoxide formation. As noted in Table 15, low excess air operation does not result in substantial NO_x reductions.

Off stoichiometric combustion (OS) In off stoichiometric combustion techniques, NO_x reduction is achieved by altering the fuel/air ratio in the primary combustion zone.

Burners-out-of-Service (BOOS) One such technique is known as burners-out-of-service. As the name implies, this operational control method involves taking one or more burners out of service, in a multiburner unit, by terminating fuel supply to the selected burners but leaving the air registers open. NO_x is reduced by lowering the peak flame temperature (PFT) in the remaining operating burners. As the temperature decreases in the combustion zone, the NO_x emissions will also decrease. The temperature decreases as a result of the remaining burners operating in a fuel rich environment, which corresponds to lower oxygen availability; thus, the peak flame temperature is lowered. In addition to the fuel/air ratio, the peak flame temperature is also dependent on the radiative heat effects in the boiler.[24] Each burner is in radiative exchange with adjacent burners; therefore, if the number of burners in-service are reduced, then the radiative effects are reduced along with the peak flame temperature.

Overfire Air (OA) One other method of changing the fuel/air ratio to the burners is by installing overfire air ports above the burner zone. A controlled portion of the combustion air, normally 10–20%, is redirected above; flames to the OFA ports.[25] Effective implementation this control method relies on a number of parameters, most notably adequate mixing of the overfire air with the primary combustion production. In addition, OFA is a function of the location and number of ports, ports spacing and geometry, and on the fan capacity and furnace dimensions. By itself, OFA can yield 15–30% reductions in NO_x emissions. However, there are certain advantages of OFA which have been noted in several cases. Because NO_x reduction requires a separation of the OFA ports from the primary combustion zone, poor temperature distribution in the convective zone and deterioration in carbon burnout has been observed.

One variation of the overfire air control method is called lance air. This method involves the installation of air tubes around the periphery of each burner to supply staged air.[22]

Flue gas recirculation (FGR) One of the most effective methods of reducing NO_x emissions for gas fired units is

TABLE 12

Properties of selected solid fuels[18,21]

	Percent by weight										
	Proximate analysis				Ultimate analysis						
	Carbon	Volatile matter	Moisture	Ash	C	H	N	O	S	HHV (MJ/kg)	"K" factor
Meta-anthracite (RI)	65.3	2.5	13.3	18.9	64.2	0.4	0.2	2.7	0.3	21.7	790.60
Anthracite (PA)	77.1	3.8	5.4	13.7	76.1	1.8	0.6	1.8	0.6	27.8	815.63
Semianthracite (PA)	78.9	8.4	3.0	9.7	80.2	3.3	1.1	2.0	0.7	31.3	838.32
Bituminous (PA)	70.0	20.5	3.3	6.2	80.7	4.5	1.1	2.4	1.8	33.3	856.67
High-volatile bituminous											
(PA)	58.3	30.3	2.6	9.1	76.6	4.9	1.6	3.9	1.3	31.7	850.68
(CO)	54.3	32.6	1.4	11.7	73.4	5.1	1.3	6.5	0.6	30.7	861.10
(KY)	45.3	37.7	7.5	9.5	66.9	4.8	1.4	6.4	3.5	28.1	850.04
(IL)	39.1	40.2	12.1	8.6	62.8	4.6	1.0	6.6	4.3	26.7	855.34
Subbituminous (CO)	45.9	30.5	19.6	4.0	58.8	3.8	1.3	12.2	0.3	23.6	862.60
Lignite (ND)	30.8	28.2	34.8	6.2	42.4	2.8	0.7	12.4	0.7	16.8	—
Brown coal (Australia)	15.3	17.7	66.3	0.7	—	—	—	—	0.1	8.6	—
Wood (Douglas fir)	17.2	82.0	35.9	0.8	52.3	6.3	0.1	40.5	0.0	21.0	—
August Victoria (Germany)	NA	33.5	NA	5.5	85.5	5.2	1.5	—	1.1	—	—
Prosper (Germany)	NA	34.3	NA	7.5	88.7	2.8	1.6	—	1.2	—	—
Göttleborn (GB-Germany)	NA	36.5	NA	10.2	79.5	4.9	1.5	—	1.0	—	—
Emil Mayrisch (EM-Germany)	NA	14.3	NA	8.5	89.3	4.2	1.5	—	0.9	—	—
Blends											
40% GB-60%EM	NA	23.6	NA	9.1	85.3	4.5	1.5	—	1.0	—	—
70% GB-30%EM	NA	30.1	NA	9.5	83.4	4.6	1.5	—	1.0	—	—

TABLE 13

Calculation of component mass percents

Type of gas USA/Texas Calculated density of gas 0.86379

Density of gas 0.868 Mass fraction ash 0 Mass fraction of water 0

	Formula					Density	Molecular weight	Volume fraction	Lower heat value	[1] Grams A/ total moles	Mass fraction	Component mass fraction				
	C	H	N	O	S							C	H	N	O	S
Methane	1	4	0	0	0	0.71	16.04	0.8090	35.3300	12.9780	0.6665	0.49901	0.16753	0.00000	0.00000	0.00000
Ethane	2	6	0	0	0	1.34	30.07	0.0680	63.5000	2.0446	0.1050	0.08389	0.02112	0.00000	0.00000	0.00000
Propane	3	8	0	0	0	1.99	44.09	0.0270	90.0000	1.1905	0.0611	0.04996	0.01118	0.00000	0.00000	0.00000
Butane	4	10	0	0	0	2.63	58.12	0.0110	117.0000	0.6393	0.0328	0.02714	0.00569	0.00000	0.00000	0.00000
Pentane	5	12	0	0	0	3.30	72.15	0.0050	144.0000	0.3607	0.0185	0.01542	0.00311	0.00000	0.00000	0.00000
Hexane	6	14	0	0	0	4.44	86.17	0.0000	165.0000	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
Heptane	7	16	0	0	0	5.15	100.20	0.0000	190.0000	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
Octane	8	18	0	0	0	5.86	114.22	0.0000	216.0000	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
Hydrogen	0	2	0	0	0	0.09	2.02	0.0000	10.6200	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
Carbon dioxide	1	0	0	2	0	1.95	44.01	0.0010	0.0000	0.0440	0.0023	0.00062	0.00000	0.00000	0.00164	0.00000
Carbon Monoxide	1	0	0	1	0	1.23	28.01	0.0000	12.4800	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
Oxygen	0	0	0	2	0	1.41	32.00	0.0000	0.0000	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
Nitrogen	0	0	2	0	0	1.23	28.02	0.0790	0.0000	2.2136	0.1137	0.00000	0.00000	0.11369	0.00000	0.00000
Hydrogen Sulfide	0	2	0	0	1	1.52	34.08	0.0000	23.1000	0.0000	0.0000	0.00000	0.00000	0.00000	0.00000	0.00000
TOTALS	—	—	—	—	—	—	—	1.0000	37.3370	19.4708	1.0000	0.67542	0.20863	0.11369	0	0

Notes: 1. Grams A/total moles—parameter was calculated by multiplying the molecular weight of the component by its volume fraction.

TABLE 14
Combustion-Emission Calculations

Type of Gas USA\Texas Molecular weight of gas to be converted 45,9929

Higher Heat Value (MJ/m3n) 41.1

Lower Heat Value (MJ/3n) 37.8

Calc. Heat Value (MJ/nm3) 37.3

			Dulong	Measured	Calculated		
Dulong	(Btu/lb) =	14,544*C + 62,028*(H − O/8) + 4050*S =	22,764	18,720	18,491	lbmol/MBtu	22.82
	(Kj/g) =	33.81*C + 144.18*(H − O/8) + 9.41*S =	53	44	43	“K” Factor	952.93

(1) Fuel Consistent per lb fuel		(2) lb Oxygen required/ lb of fuel constituent	(3) = (1)* (2)Theoretical Amt. of Oxygen Required (lb)	(4) lb Fuel Gas Produced/lb of fuel	(5) Flue Gas Constituents				
					CO_2	H_2O	SO_2	O_2	N_2
C	0.67542	2.664	1.799	3.664	2.477				
H	0.20863	7.937	1.656	8.937		1.865			
S	0	0.998	0	1.998			0.000		
O	0		0	See below					
N	0.11369								0.114
H_2O	0					0.000			
Ash	0								
Total	0.998								
Oxygen Excess Factor For Burner Air & Leakage – OEF% =		OEF 0	3.455	0				0	
Total Oxygen Supplied				3.455					
Nitrogen Supplied = Oxygen * 3.321				11.474					11.474
Dry Combustion Air Supplied				14.929					
H_2O in air @ 0.013 lb H_20/lb BDA ~				0.194		0.194			
Wet Combustion Air Supplied				15.123					
Flue Gas Constituents (lb/lb fuel)					2.47726031	2.059	0	0	11.474
Specific Volumes * (scf/lb gas)					8.750	21.383	6.011	12.034	13.714
Flue Gas Constituents * (scf/lg fuel), vi					21.676	44.028	0.000	0.000	157.354
Vwet - Total Flue Gas (scf/lb fuel) (wet basis) Constituents Vol. % (wet basis) = (vi/vwet) * 100%				223.058	9.718	19.738	0.000	0.000	70.544
Vdry - Total Flue Gas (scf/lb fuel) (dry basis) Constituents Vol. % (dry basis) = (vi/vdry) * 100%				179.030	12.107		0.000	0.000	87.893
Emissions (based on lower heat value):									
E (lb/MBtu)									
E (grains/dscf))									

* Reference conditions 68 deg.F, 29.92 in Hg. Specific volumes based on international gas constant of 8.31439 Joules/mol-deg.K or 1545.4 ft-lb/lb mol deg.R and thus 385.1 scf/lbmol. The specific volume for nitrogen is based on composition of dry air (20.9% O_2, 79.1% N_2) where N_2 is the “atmospheric” nitrogen containing about 1% Ar, 0.03% CO_2, and trace amounts of Kr, Ne, Xe; molecular weight of atmospheric nitrogen is 28.08.

~ Humidity at 60% R.H. and 80 deg.F often used as standard.

TABLE 15
NO_x Control methods and reduction[22,23]

Control method	NO_x reduction (%)
Low excess air	1–15
Off-stoichiometric combustion	30–60
Low-NO_x burner	
Staged-air burner	25–35
Staged-fuel burner	40–50
Low excess air burner	20–25
Burner w/external FGR	50–60
Burner w/internal FGR	40–50
Air or fuel-gas staging w/internal FGR	55–75
Air or fuel-gas staging w/external FGR	60–80
Flue gas recirculation	40–80
Waster/stream injection	40–70
Selective catalytic reduction	70–90
Selective noncatalytic reduction—Urea	25–50
Selective noncatalytic reduction—Ammonia	25–50

through internal or external flue gas recirculation (FGR). In FGR, a portion of the hot inert flue gas (300°F–500°F) is recirculated from the stack back to the burner box. The recirculation rate is normally limited to 15–25%. Figure 5[22] shows the NO_x reductions that can be expected from different amounts of flue gas recirculation. In external FGR, flue gas can be mixed with combustion air or can be injected into the primary combustion zone through the use of a recirculation fan. Internal FGR utilizes the pressure energy of the combustion air, fuel gas or steam or recirculate the flue gas. Although internal FGR eliminates the need for fans and controls, it does necessitate larger burner sizes.

Flue gas recirculation reduces NO_x because the flue gas acts as a diluent. The additional flue gas reduces the peak flame temperature, cutting thermal NO_x formation. In addition, the partial pressure of oxygen is reduced, also limiting NO_x formation. In industrial applications, FGR is typically used in forced draft gas fired units which burn low nitrogen fuels. As noted above, external FGR requires substantial equipment for installation: (1) recirculation

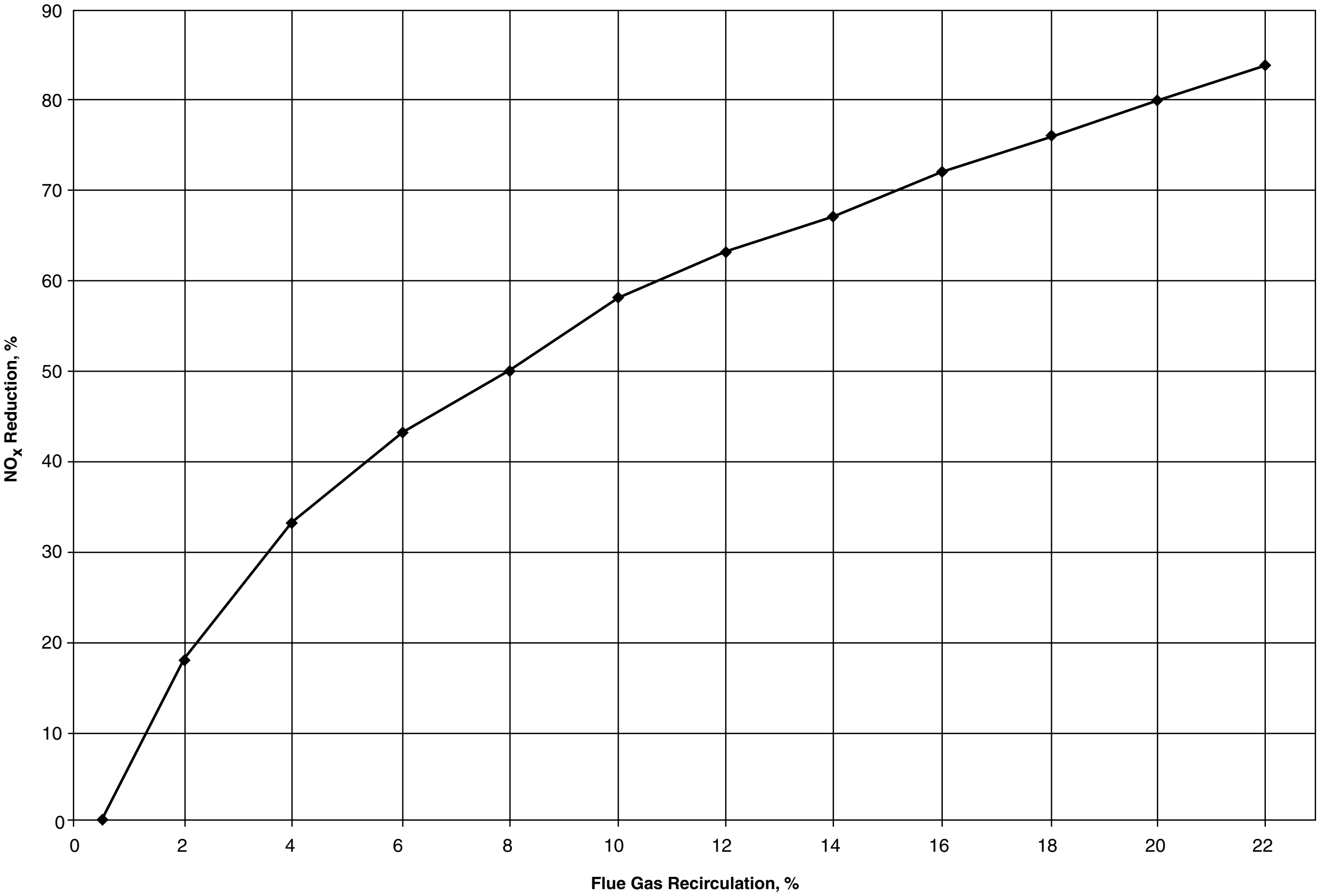
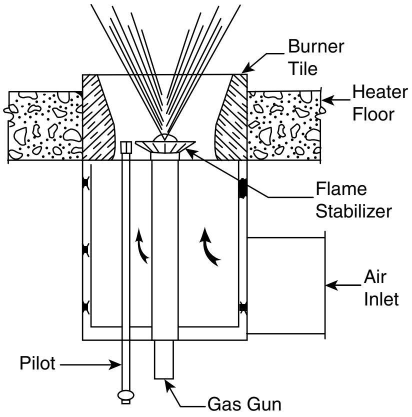


FIGURE 5 NO_x reduction as a function of % flue gas recirculation.[22]

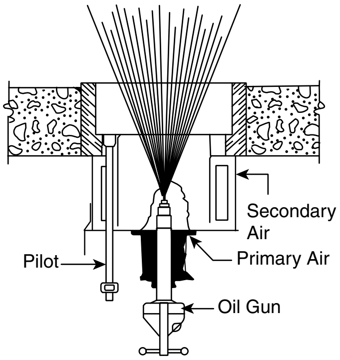
fan, (2) additional ductwork, (3) flow control dampers, and (4) carbon monoxide and oxygen analyzers.[26] The disadvantages of FGR include flame instability and a decrease in net thermal output at high recirculation rates; thus, the reasoning for limiting the recirculation rate.

Low NO_x Burners (LNB) Another way to achieve the staging effects similar to that of OS combustion is to internally design the effect into the burner. A variety of low NO_x burners have been developed to achieve the desired fuel/air ratio required for NO_x reduction in different applications. Figure 6[23] displays some of the burner types used in industry. Table 15 lists some of the burner types and the NO_x reductions that can be expected. Retrofitting a facility with low NO_x burners offers many advantages and include the following: (1) requires no added instrumentation for controlling the fuel/air ratio, (2) needs no additional ductwork or flue gas recirculation fans, as discussed in the previous section, (3) minimal long-term capital investment and maintenance requirements, and (4) cost effective compliance with the regulatory standards.

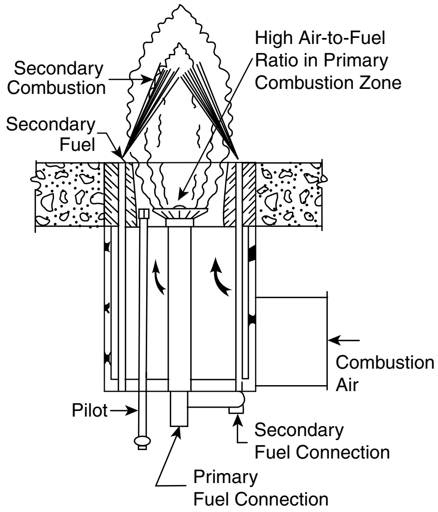
Low Excess Air Burner The first burner type listed in Table 15 is the low excess air burner, which reduces NO_x emissions by combusting the fuel with the lowest amount of excess air possible, usually 5–8%. Typically, boilers operate at 18–30% excess air at full load depending on the fuel. Case studies have shown that a reduction of excess air from 30% to 10% can cut NO_x emissions by 30%. The NO_x is reduced by the same principles as noted in LEA operations. LEA burners are most commonly of a forced-draft design and obviously have the ability to operate at low levels of excess air. Furnaces can operate in two modes: (1) forced draft; (2) natural draft.[27] Forced draft operation involves the pressurization of the air by a fan into the furnace, whereas the natural draft mode requires the stack to draft the combustion



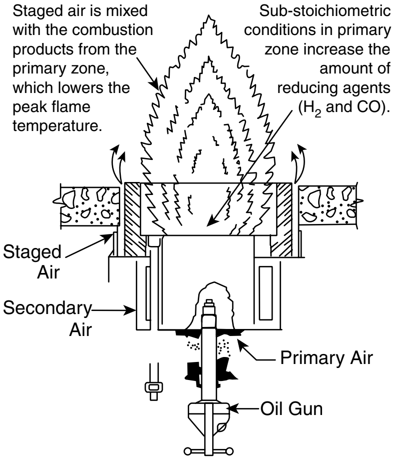
Standard Gas Burner



Standard Oil Burner



Staged Fuel Gas Burner



Staged Air Burner

FIGURE 6 Burner configurations.[23]

air into the furnace. For multiple burner applications, it is imperative that all burners receive equal amounts of air. One burner that receives more excess air will produce more NO$_x$, whereas the resulting burner that receives less excess air will generate more carbon monoxide and unburned carbon.

Staged Air Burner Another type of burner is the staged air burner, which is primarily used in forced draft liquid fuel fired applications, although natural draft applications do exist. The fuels used in this burner are normally such fuels as butane, pentane, diesel, and No. 6 fuel oil. This burner design reduces NO$_x$ by lifting the combustion air into a primary and a secondary zone. Fuel is injected into the throat of the burner and mixed with the primary air. This zone is fuel rich and produces partial combustion. NO$_x$ is minimized in this zone because the nitrogen is converted into reducing agents, which are subsequently oxidized to elemental nitrogen. In addition, because the generated heat in the primary zone is rapidly dissipated, the peak flame temperature is reduced and NO$_x$ formation is also lowered. The fuel lean or secondary zone completes combustion by injecting air through refractory ports, which also stabilizes the flame profile. Although this method can lead to NO$_x$ reductions of 20–35%, staged air burners can lead to long flame profiles, which must be closely monitored and controlled.

Staged Fuel Burner In gas fired applications, staged fuel burners are typically used to lower NO$_x$ emissions. Because liquid fuels can create fouling problems with the secondary gas nozzles, this burner normally is used for firing natural gas and other light fuels. Similar to the staged air burner design, fuel supply is divided into primary and secondary zones. The primary zone involves the mixing of the combustion air with a portion of the fuel gas, resulting in a fuel lean environment. This fuel lean combustion zone reduces peak flame temperature and lowers the formation of thermal NO$_x$. Combustion is completed in the secondary zone where nozzles inject the remaining fuel to create fuel rich conditions. Part of the NO$_x$ formed in the first stage is reduced by hydrogen and carbon monoxide in the secondary zone. Staged fuel burners normally produce a flame that is 50% longer than that of normal standard gas burners.

Ultra Low NO$_x$ Burners The last type of burner design is the ultra low NO$_x$ burner. This burner combines the staged air or staged fuel with internal or external flue gas recirculation to reduce NO$_x$. Typically, internal FGR is utilized. In a staged air internal FGR burner, fuel is mixed with part of the combustion air to create a fuel rich zone. The recirculated flue gas is developed by high pressure atomization of the liquid or gaseous fuel. Combustion is completed by piping the secondary air to the burner block. In a staged fuel internal FGR burner, flue gas is recirculated by the pressure of the fuel gas. The fuel gas creates a fuel lean zone and reduces the partial pressure of the oxygen, thereby reducing NO$_x$ emissions.

Water/steam injection One of the most seldom used methods in controlling NO$_x$ is water/steam injection. Normally, these techniques are applied to gas turbines. These methods have not been used extensively because of the lower thermal efficiency resulting from the absorption of usable energy. Both of these processes accomplish NO$_x$ reduction by lowering the peak flame temperature. The PFT is reduced by directly abstracting heat from the burner flame and by diluting the oxygen concentration near the burner front. Figure 7[22] shows the approximate NO$_x$ reductions that can be expected for different water injection rates.

Although substantial NO$_x$ reductions can be expected from these processes, a number of operational consequences must be considered. At low loads, carbon monoxide and unburned hydrocarbon emissions increase. As noted earlier, usable energy is lost, particularly due to the heat of vaporization of water, which results in an increase in fuel consumption of up to 5%. Further considerations include the additional wear and tear on turbine parts and the additional complexity in controlling and monitoring the process.[28]

Selective Catalytic Reduction (SCR) In lieu of or in addition to changing the air/fuel ratio in the combustion zone, some facilities utilize post combustion techniques to control NO$_x$. There are two basic post combustion control technologies of varying types on the market today: selective catalytic reduction (SCR) and selective noncatalytic reduction (SNCR). These methods have been used extensively on an international scale and have become a common feature on gas-turbine cogeneration and combined cycle systems in the United States. These systems can provide NO$_x$ reductions of up to 90%.

One of the most popular post combustion techniques is selective catalytic reduction. SCR works on the premise of reacting NO$_x$ with ammonia to produce water and elemental nitrogen. The reactions involved in the SCR process are as follows:[29]

$$
\begin{gathered}
4NO + 4NH_3 + O_3 \rightarrow 4N_2 + 6H_2O \\
6NO + 4NH_3 \rightarrow 5N_2 + 6H_2O \\
2NO_2 + 4NH_3 + O_2 \rightarrow 3N_2 + 6H_2O \\
6NO_2 + 8NH_3 \rightarrow 7N_2 + 12H_2O \\
NO + NO_2 + 2NH_3 \rightarrow 2N_2 + 3H_2O.
\end{gathered}
$$

Performance tests indicate that the first reaction is the dominant reaction. An SCR system consists mainly of an ammonia injection grid, catalyst reactor and ductwork.

The first part of the SCR, the ammonia injection grid, involves the mixing of ammonia with the flue gas stream. Two types of ammonia are used in the process: anhydrous and aqueous. A typical anhydrous ammonia injection system involves the following:[29]

- Storage of the anhydrous ammonia in a pressurized tank
- Piping of the anhydrous ammonia to a liquid vaporizer
- Mixing of the ammonia vapor with a predetermined amount of ambient air
- Distribution of the ammonia-air mixture to the grid for injection.

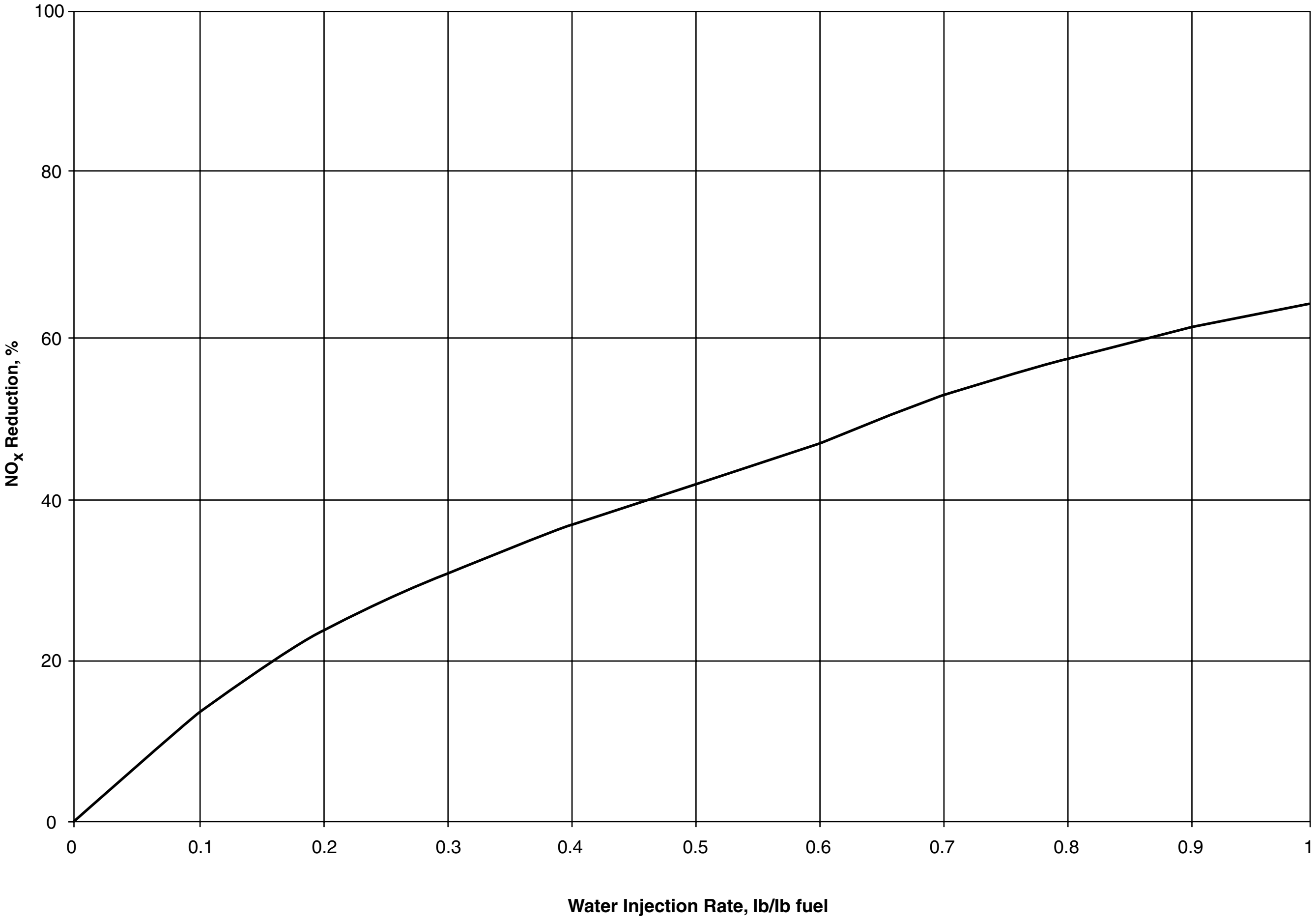


FIGURE 7 NO_x reduction as a function of water injection rate.[22]

One of the drawbacks of using anhydrous ammonia is that the chemical is hazardous and dangerous to handle and transport. Secondly, it requires a double lined steel tank due to its high vapor pressure. The other type of ammonia used for injection is aqueous ammonia, NO_3 H_2O, which typically comes in an industrial grade of 27% NH_3, 73% H_2O by weight. An injection system for aqueous ammonia contains the following:[29]

- Storage of the aqueous ammonia in an ordinary tank
- Vaporizing the ammonia through atomizing nozzles
- Mixing of the ammonia with a predetermined amount of heated ambient air in the vaporizer
- Injection of the ammonia-air mixture into the flue gas stream via a parallel network of pipes containing several orifices.

The advantages of this type of system are that the aqueous ammonia is less hazardous, can be transported on U.S. highways, and can be kept in an ordinary storage tank.

The second part of the SCR system, the catalyst reactor, is the most important. There are a variety of catalysts available on the market today, such as aluminum oxide (Al_2O_3), vanadium pentoxide (V_2O_5), titanium dioxide (TiO_2), tungsten trioxide (WO_3), molybdenum trioxide (MoO_3), and iron oxides.[29,30] In addition to metal catalysts, many SCR systems use zeolites or activated carbon. The use of zeolites has been found to increase catalyst activity, reduce the amount of metals for disposal, and reduce the size and capital cost of the SCR system.

As noted above, SCR's can provide 70–90% reductions in NO_x emissions. The performance of a SCR, however, depends on many parameters, including:[29]

- sulfur content in the fuel
- type and age of catalyst
- residence time of flue gas in the reactor
- oxygen and water vapor concentration in the flue gas
- flue gas temperature
- effective mixing and distribution of ammonia and air in the flue gas stream

- amount of ammonia
- ammonia slip
- catalyst surface area
- dust loading.

Ammonia slip and sulfur content of the fuel are two parameters that are interrelated and must be closely monitored. Theoretically, according to the balanced chemical reactions, only one mole of ammonia is needed to break down one mole of NO_x. Because mixing of the ammonia-air mixture and distribution of the mixture with the flue gas is not ideal, ammonia is normally injected at higher quantities than is theoretically required. This overinjection can lead to residual ammonia escaping into the flue gas, which is called ammonia slip. Although higher values of ammonia slip lead to increased NO_x reductions (to an asymptotic limit), it is environmentally detrimental to allow excess ammonia to the atmosphere. Furthermore, because the sulfur in the fuel leads to SO_2 and SO_3 production, excess ammonia can result in the formation of ammonium sulfate and ammonium bisulfate by the following reactions:[29]

$$SO_3 + 2NH_3 + H_2O \rightarrow (NH_4)_2SO_4$$
$$SO_3 + NH_3 + H_2O \rightarrow NH_4HSO_4.$$

The formation of sulfates can lead to higher particulate concentrations in the flue gas and can lead to catalyst poisoning and equipment deterioration. The formation of sulfates can be mitigated by keeping the temperature of the gas relatively high (>600°F) and by limiting ammonia slip to <10 ppm for low sulfur fuels and to 3–5 ppm for high sulfur fuels.

Oxygen and water vapor concentrations have opposing effects on NO_x reduction. Typically, higher oxygen concentrations result in better catalyst performance leading to improved NO_x reductions, to a certain limit. On the other hand, water vapor decreases the performance of the catalyst as its concentration increases.

Obviously, the type of catalyst is extremely important to the performance of the SCR. As noted earlier, catalysts are available in a variety of materials. In addition, the catalyst shape can vary, including rings, pellets, honeycomb plates and parallel-ridged plates. Although it depends on the particular catalyst, flue gas temperature is critical to the performance of the catalyst. Typically 600°F to 700°F is the optimum range of operation. Below these temperatures, formation of ammonium sulfates becomes more substantial leading to further complications. Figures 8 and 9[30] show the temperature dependence of ammonia slip and NO_x reductions on two catalysts, VNX and ZNX. The VNX catalyst is a low temperature catalyst and is comprised of V_2O_5 and TiO_2. The ZNX catalyst is a high temperature zeolite catalyst. In addition to temperature, the age of the catalyst affects the catalyst performance. Usually, the rate of deterioration is large at the onset of operation, but levels out after the initial settlement. Another important parameter is the specific surface area of the catalyst. The larger the area of the catalyst exposed to the flue gas, the better the SCR performance.

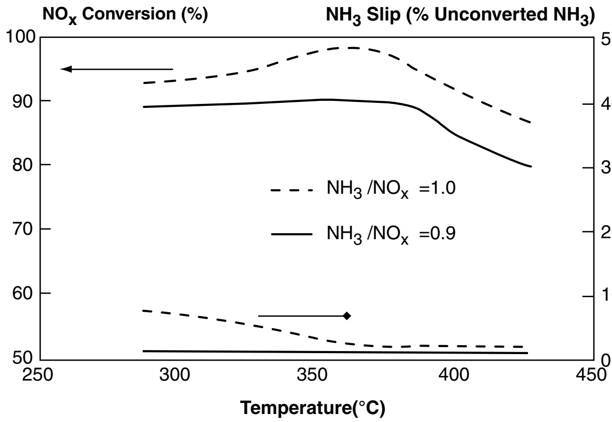


FIGURE 8 NO_x Conversion and NH_3 Slip for VNX™ SCR Catalyst.[30]

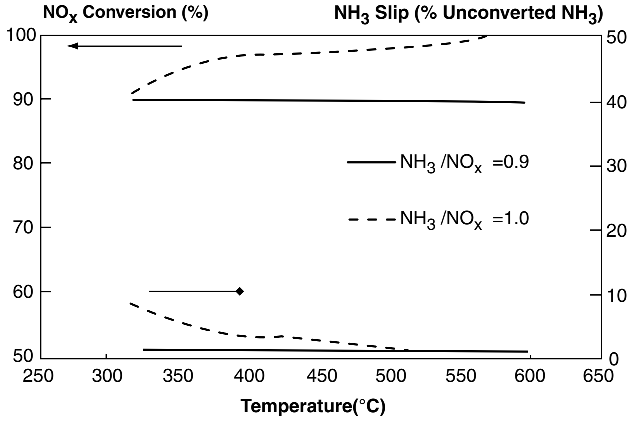


FIGURE 9 NO_x Conversion and NH_3 Slip for ZNX™ SCR Catalyst.[30]

In addition to the surface area of the catalyst, the residence time of the flue gas in the reactor is critical to good performance. Related to the residence time is the space velocity, which is a parameter that is normally used for SCR reactor design. The efficiency of the SCR can be expressed in terms of the space velocity through the following equation:[29]

$$\eta = m(1 - e^{-k/SV})$$

where

η = NO_x removal efficiency
m = molar ratio of ammonia to NO_x
SV = space velocity
k an activity constant that takes into account most of the operating parameters listed earlier.

The molar ratio of ammonia to NO_x is determined by:[29]

$$m = \eta + S/NO_{x,in}$$

where

m = molar ratio of ammino to NO_x
η = NO_x removal efficiency
S = limit of ammonia slip, ppmdv
$NO_{x,in}$ = inlet concentration of NO_x, ppmdv

With the values of m and η know, one can then calculate the space velocity. The volume of catalyst is then determined by dividing the volumetric gas flow by the space velocity.

The last parameter that affects SCR performance is dust loading. This parameter is particularly important when considering coal-fired plants. The dust from the coal pulverization process can accumulate on the SCR catalyst, creating plugging and fouling problems and decreasing performance. Thus, in pulverized coal-fired applications, soot blowers are sometimes installed upstream of the catalyst layers to "dust off" the catalyst surface.

Selective noncatalytic reduction (SNCR) The other major type of post combustion control is selective noncatalytic reduction. SNCR reduces NO_x through the same reactions as SCR. However, SNCR differs from selective catalytic reduction in a number of ways, most notably the lack of catalyst for reaction. Because of the lack of catalyst, NO_x reduction in a SNCR must occur at higher temperatures. Typically, the temperature of the process is in the range of 1,600–2,000°F, with an optimum temperature of approximately 1,750°F.[26] In a SNCR system, a urea based compound can also be used besides ammonia as a reagent. Urea based compounds break down to ammonia and carbon dioxide after injection into the flue gas stream. SNCR does not typically provide the same reductions as SCR, but some studies and tests have shown NO_x reductions using SNCR to be in the range of 60–80%.

As noted, one of the most important operating parameters for SNCR performance is the temperature. Typically, for ammonia injection, the temperatures should be maintained from 1,600°F to 1,750°F. For urea based compounds, 1,000°F to 1,900°F is the optimum temperature range. If the temperature increases above the optimal range, then the reagent will react more with oxygen than with the NO_x, leading to increased NO_x formation. On the other hand, if the temperature falls below the optimal range, then less ammonia will react, creating more ammonia slip.

The disadvantages of using SNCR are the same as that of SCR. Both ammonium sulfate and ammonium bisulfate can form in a SNCR system, creating the fouling and corrosion problems. Ammonium chloride, NH_4Cl can also be formed, resulting in plumes that are more visible. Furthermore, SNCR requires strict control of the temperature and does not provide the same reductions in NO_x as SCR.

The advantages of SNCR over SCR include lower capital and operating costs and safer handling of chemicals when urea based reagents are used. A comparative study of the two processes on a 200 ton/hr oil and gas fired boiler revealed that the SNCR would require a capital cost of \$0.4 million as compared to a capital cost of \$2 million for SCR. The NOx reductions in this study for the SNCR were predicted to be 65–70%.[31]

Fluidized bed combustors or boilers For new power or cogeneration plants, one method of meeting the increasingly stringent air pollutant emission standards is by selecting fluidized bed combustion as the technology for combusting the fuel. There are a number of different fluidized bed technologies available: (1) atmospheric fluidized bed combustion (AFBC); (2) circulating fluidized bed combustion (CFBC); and (3) pressurized fluidized bed combustion (PFBC). Based on the technology selected and on process variations, the FBC technology has an efficiency range of 37–45%. Figure 10[32] is a schematic of a typical PFBC.

The basic principle behind all FBC's is that air is injected upwards through a bed of inert particles (i.e. sand) at a sufficient rate so as to give the bed fluid-like characteristics. The particle bed is normally heated to just above the ignition temperature of the fuel, normally 850°C. Most conventional boilers are operated at 1600°C. Recall that NO_x emissions increase with increasing flame temperature. Thus, one can see that NO_x emissions will be reduced substantially in a FBC because of the reduced combustion temperature. As a matter of fact, almost all NO_x emissions from an FBC can be attributed to fuel NO_x. Combined with other NO_x control techniques, such as staged combustion or SNCR, FBC provides a reasonable alternative to conventional fired boilers for reducing pollutant emissions without sacrificing efficiency.

Mobile Source Control

Spark ignition engines There are a number of methods used to reduce pollutant emissions from spark ignition engines. These methods are based on two different approaches: (1) combustion based; (2) exhaust gas treatment.

Combustion based emission controls for NO_x include spark retard and exhaust gas recirculation (EGR).[18] In spark retard, combustion initiation is delayed to reduce the peak pressure and peak temperature. As noted earlier, reducing the peak temperature of combustion results in lower NO_x emissions. The other type of method, EGR, involves the recirculation of cool combustion products for mixing with the pre-combusted air and fuel mixture. This method reduces the peak temperature through dilution and by increasing the specific heat of the gas due to the addition of water carbon dioxide. Although effective in reducing NO_x emissions, spark retard and EGR must be somewhat limited due to the engine power loss, increased fuel consumption and increased carbon monoxide emissions associated with these control measures.

There are two basic types of exhaust gas treatment: (1) noncatalytic thermal reactors; and (2) catalytic converters[18]. Noncatalytic thermal reactors can reduce hydrocarbon and CO emissions substantially, but is not as efficient in alleviating the NO_x problem. In such a system, NO_x is limited by ammonia reduction. However, the ammonia-NO_x reaction, without catalyst, requires a high and narrow range of temperatures, thereby limiting the application of this technology. On the other hand, catalytic converters have become the standard on most automobiles in the United States. The basic

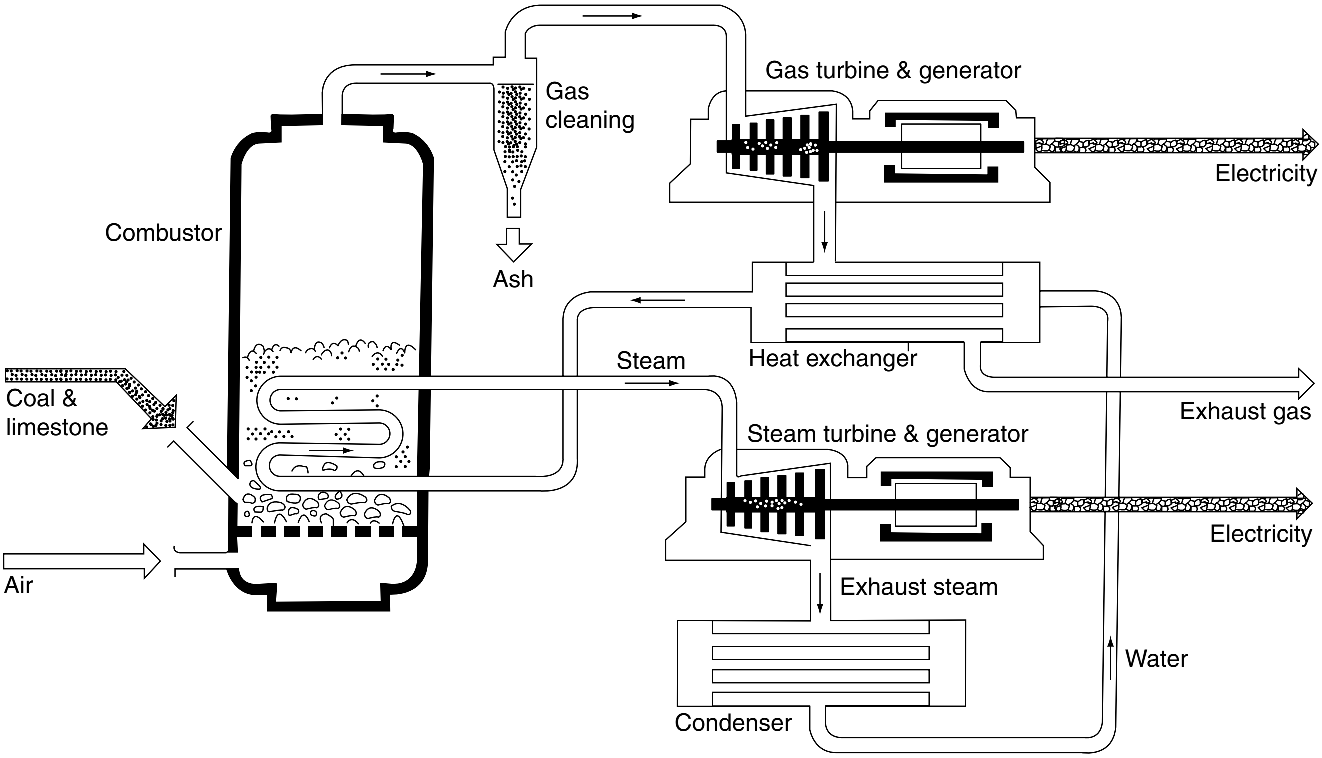


FIGURE 10 Schematic of pressurized fluid bed combustor.[32]

principle behind the catalytic converter is that the exhaust gases are passed through a catalyst bed which effectively lowers the temperatures at which CO and hydrocarbons are oxidized and at which NO_x is reduced.[18] Typical catalysts include alumina, platinum, or metal oxides.

Diesel engines Control measures for diesel engines are similar to those for spark ignition engines. However, because diesel engines generally operate at fuel lean conditions, catalytic reduction of NO_x is not feasible. Therefore, NO_x emissions are reduced from diesel engines through combustion based control measures. These methods include exhaust gas recirculation and injection timing retardation. As with the spark ignition engine, EGR reduces NO_x emissions by lowering the peak combustion temperature through dilution with cool combustion gases. Likewise, retarding the injection timing results in a delay of the heat release in the cycle, thereby lowering the peak flame temperature. As noted above, both of these techniques result in reduced NO_x emissions, but at the potential cost of fuel economy and engine performance.

CONCLUSION

As our knowledge of NO_x advances and as political climates change, NO_x emission regulations will inevitably change, prompting new or modified control technologies to be developed. Although sometimes viewed as cumbersome and costly, regulations are the key to providing a healthy environment for generations to come.

REFERENCES

1. Stern, A.C., *Air Pollution, Volume 1, Air Pollution and Its Effects,* 2nd Ed. Academic Press, New York (1968).
2. US Department of Health and Human Services, Centers for Disease Control, and National Institute for Occupational Safety and Health. “NIOSH Pocket Guide to Chemical Hazards” (1990).
3. Ross, R.D., *Air Pollution and Industry,* Van Nostrand Reinhold Company, New York (1972).
4. Abrahamson, D.E., *The Challenge of Global Warming,* Island Press, Washington, D.C. (1989).
5. Howard, R. and Perley, M., *Acid Rain, The North American Forecast,* Anansi Press, Toronto (1980).
6. Gallagher, John, and Foxen, Robert J., “Clean Air: Stricter Standards, New Technology and Better Planning,” *Engineering News Record,* **232,** E71–E73 (1994).
7. Clean Air Act, 42 U.S.C.A., pp 7401–7671q.
8. USEPA. “National Primary and Secondary Ambient Air Quality Standards,” 40 CFR Part 50.11 (1992).
9. Santire, Stanley, “Regs Serve Up Alphabet Soup,” *Pollution Engineering,* 25, 86–87 (1993).
10. USEPA. “New Source Performance Standards,” 40 CFR Part 60 (1995).
11. Kuehn, Steven E., “Retrofit Control Technology Reducing NO_x Emissions,” *Power Engineering,* 23–31 (1994).
12. NYSDEC. “New York Air Pollution Control Regulations,” NYCRR, Title 6, Part 200 (1993).
13. NYSDEC. “New York Ambient Air Quality Standards,” NYCRR, Title 6, Part 256 and 257 (1989).

14. NYSDEC. “Reasonably Available Control Technology for Oxides of Nitrogen (NO_xRACT),” NYCRR, Title 6, Parts 200, 201 and 227 (1994).
15. NYCDEP. “Air Pollution Code” (1992).
16. Renna, Steven P. and Rubin, Laurence M., *Present and Future Vehicle Clean Air Regulations—Federal and California Requirements,* Brooklyn Union Gas Company, New York (1991).
17. NYSDEC. “New York Air Pollution Control Regulations,” NYCRR, Title 6, Part 218 (1992).
18. Flagan, R.C. and Seinfeld, J.H., *Fundamentals of Air Pollution Engineering,* Prentice Hall, New Jersey (1988).
19. Wartsila Diesel. “Pure Energy Plant Project Guide,” 2–1 thru 2–2.
20. Con Edison Laboratory Analysis of Arthur Kill No. 6 Fuel Oil.
21. Maier, H., Spliethoff, H., Kicherer, A., Fingerle, A., and Hein, K.R.G., “Effect of Coal Blending and Particle Size on NO_x Emission and Burnout,” *Fuel,* **73,** 1447–1452 (1994).
22. Wood, S.C., “Select the Right NO_x Control Technology,” *Chemical Engineering Progress,* **90,** 32–38 (1994).
23. Garg, A., “Specify Better Low-NO_x Burners For Furnaces,” *Chemical Engineering Progress,* **90,** 46–49 (1994).
24. Colannino, J., “Low-cost Techniques Reduce Boiler NO_x,” *Chemical Engineering,* **100,** 100–106 (1993).
25. Jones, C., “NO_x Emissions Control: Small Boilers Pose Great Challenges,” *Power,* **138,** 34–41 (1994).
26. Garg, A., “Trimming NO_x From Furnaces,” *Chemical Engineering,* **99,** 122–129 (1992).
27. Straitz, J.F. and Mendoza, V.A., “Combat NO_x with Better Burner Design,” *Chemical Engineering,***101,** EE-4–EE-8 (1994).
28. Makansi, J., “Reducing No*x* Emissions,” *Power,* **132,** S1–S13 (1988).
29. Cho, S.M., Properly Apply Selective Catalytic Reduction for No_x Removal,” *Chemical Engineering Progress,* **90,** 39–45 (1994).
30. Heck, R.M., Chen, J.M. and Speronello, B.K., “Operating Characteristics and Commercial Operating Experience with High Temperature SCR NO_x Catalyst,” *Environmental Progress,* **13,** 221–225, (1994).
31. McInnes, R., Wormer, M.B.V., “Cleaning Up NO*x* Emissions,” *Chemical Engineering,* **90,** 130–135 (1990).
32. Redman, J., “Fluidised bed combustion, So_x and $NO_{x,}$” *The Chemical Engineer,* 32–38 (1989).
33. NJ Department of Environmental Protection—State implementation plan for the attainment and maintenance of the ozone and carbon monoxide national ambient air quality standards. (1999).
34. Baukal, C.F., Hayes, R., Grant, M., Singh P., and Foote, D., “Nitrogen Oxides Emissions Reduction Technologies in the Petrochemical and Refining Industries”. Environmental Progress 23, No. 1 19–28 (2004).
35. Mussatti, D.C. (Ed.) EPA Air Pollution Control Cost Manual, 6th Edition, Report EPA/452/B-02–001, U.S. EPA, Wash, D.C. (2002).

W. MICHAEL SUTTON
New York City Department of
Environmental Protection

EDWARD N. ZIEGLER
Polytechnic University

NOISE

Noise and **sound** refer to audible pressure fluctuations in air. Both are characterized by sound level in decibels and frequency content in hertz. Although sound is vital for communication, noise is one of our greatest problems. Intentionally generated acoustic signals including speech and music are usually referred to as *sound. Noise* is a term used to identify unwanted sound, including sound generated as a byproduct of other activities such as transportation and industrial operations. Intrusive sound, including speech and music unwelcome to the hearer, are also considered noise. Thus, the distinction between noise and sound is subjective, and the two terms are often used interchangeably.

When a body moves through a medium or vibrates, some energy is transferred to that surrounding medium in the form of sound waves. Sound is also produced by turbulence in air and other fluids, and by fluids moving past stationary bodies. In general, gases, solids and liquids transmit sound.

Well-documented effects of noise include hearing damage, interference with communication, masking of warning signals, sleep interruption, and annoyance. Noise detracts from the quality of life and the environment; it contributes to anger and frustration and has been implicated as a contributor to psychological and physiological problems.

The National Institute for Occupational Safety and Health (NIOSH) named hearing loss as a priority research area, noting that noise-induced hearing loss is 100% preventable, but once acquired, it is permanent and irreversible. The Occupational Safety and Health Administration (OSHA) noted that hearing loss can result in a serious disability, and put employees at risk of being injured on the job. The World Health Organization (WHO) notes that noise-induced hearing impairment is the most prevalent irreversible occupational hazard, and estimates that 120 million people worldwide have disabling hearing difficulties. In developing countries, not only occupational noise but also environmental noise is an increasing risk factor for hearing impairment.

The European Union (EU) identified environmental noise caused by traffic, industrial and recreational activities as one of the main local environmental problems in Europe and the source of an increasing number of complaints. It is estimated that 20% of the EU population suffer from noise levels that both scientists and health experts consider unacceptable. An additional 43% of the population live in ‘gray areas’ where noise levels cause serious daytime annoyance. Estimates of the cost of noise to society range from 0.2% to 2% of gross domestic product.

Noise control involves reduction of noise at the source, control of noise transmission paths, and protection of the receiver. Source control is preferred. For example, design of transportation systems and machinery for lower noise output may be the most effective means of noise control. But, after trying all feasible noise source reduction, airborne noise and/or solid-borne noise may still be objectionable. Interruption of noise transmission paths by means of vibration isolation, source enclosures, sound absorbing materials, or noise barriers is then considered.

In some industrial situations, excessive noise is still present after all attempts to control noise sources and transmission paths. Administrative controls—the assignment of employees so that noise exposure in reduced—should then be considered. As a last resort, employees may be required to use personal hearing protection devices (muff-type and insert-type hearing protectors). Communities often resort to ordinances that limit noise levels and restrict hours of operation of noise-producing equipment and activities. Community noise control methods also include zoning designed to separate noise sources from residential and other sensitive land uses.

FREQUENCY, WAVELENGTH AND PROPAGATION SPEED

Frequency. Audible sound consists of pressure waves with frequencies ranging from about 20 hertz (Hz) to 20,000 Hz, where 1 Hz = 1 cycle per second. Sound consisting essentially of a single-frequency sinusoidal pressure wave is called a **pure tone.** In most cases, noise consists of sound waves arriving simultaneously from a number of sources, and having a wide range of frequencies. A sound wave which has a frequency below the audible range is called **infrasound** and sound of frequency above the audible range is called **ultrasound.**

Propagation speed. The propagation speed of airborne sound is temperature dependent. It is given by:

$$c = 20.04[T' + 273.16]^{1/2} \qquad (1.1)$$

where c = propagation speed, i.e. the speed of sound, (m/s) T' = air temperature (°C).

At an air temperature of $T' = 20°\text{C}$ (68°F), the propagation speed is $c = 343$ m/s (approx). Sound waves propagate at a different speed in solids and liquids. The propagation speed for axial waves in a steel rod is about 5140 m/s. Note that

c is a wave a propagation speed; it does not represent particle velocity within the medium.

Wavelength. If a pure-tone pressure wave could be observed at a given instant, the length of one cycle of the wave in the propagation directly could be identified as the wavelength. Thus,

$$\lambda = c/f \tag{1.2}$$

where λ = wavelength (m), c = propagation speed (m/s) and f = frequency (Hz). The effectiveness of noise barriers and sound-absorbing materials is dependent on the sound wavelength (thus, effectiveness is frequency-dependent).

SOUND PRESSURE AND SOUND PRESSURE LEVEL

One standard atmosphere is defined as a pressure of 1.01325×10^5 Pa (about 14.7 psi). Typical sound pressure waves represent very small disturbances in ambient pressure. Sound pressure level is defined by

$$L_\mathrm{p} = 10\ 1_\mathrm{g}\left[p_\mathrm{rms}^2/p_\mathrm{ref}^2\right] = 20\ \mathrm{lg}\left[p_\mathrm{rms}/p_\mathrm{ref}\right] \tag{2.1}$$

where L_p = sound pressure level in decibels (dB), lg = common (base-ten) logarithm, p_rms = root-mean-square sound pressure (Pa) and p_ref = reference pressure = 20×10^{-6} Pa.

Sound pressure represents the difference between instantaneous absolute pressure and ambient pressure. For a pure-tone sound wave of amplitude P,

$$p_\mathrm{rms} = P/2^{1/2}.$$

The reference pressure is the nominal threshold of hearing, corresponding to zero dB. Sound pressure may be determined from sound pressure level by the following relationship:

$$p_\mathrm{rms} = p_\mathrm{ref} \times 10^{Lp/20} = 2 \times 10^{[L_\mathrm{P}-100]/20.} \tag{2.2}$$

A-WEIGHTING

Human hearing is frequency-dependent. At low sound levels, sounds with frequencies in the range from about 1 kHz to 5 kHz are perceived as louder than sounds of the same sound pressure, but with frequencies outside of that range. A-, B- and C-weighting schemes were developed to compensate for the frequency-dependence of human hearing at low, moderate and high sound levels. Other weightings are also used, including SI-weighting which relates to speech interference. A-weighting has gained the greatest acceptance; many standards and codes are based on sound levels in A-weighted decibels (dBA). When noise is measured in frequency bands, the weighting adjustment may be added to each measured value. Sound level meters incorporate weighting networks so that weighted sound level is displayed directly. A-weighting adjustments are shown in Table 1.

TABLE 1
A-weighting

Frequency Hz	Adjustment dB
20	−50.5
25	−44.7
31.5	−39.4
40	−34.6
50	−30.2
63	−26.2
80	−22.5
100	−19.1
125	−16.1
160	−13.4
200	−10.9
250	−8.6
315	−6.6
400	−4.8
500	−3.2
630	−1.9
800	−0.8
1,000	0
1,250	+0.6
1,600	+1.0
2,000	+1.2
2,500	+1.3
3,150	+1.2
4,000	+1.0
5,000	+0.5
6,300	−0.1
8,000	−1.1
10,000	−2.5
12,500	−4.3
16,000	−6.6
20,000	−9.3

Some representative sound levels are given in Table 2. Most values are approximate; actual noise sources produce a wide range of sound levels.

EQUIVALENT SOUND LEVEL

Sound energy is proportional to mean-square sound pressure. Equivalent sound level is the energy-average A-weighted sound level over a specified time period. Thus,

$$L_\mathrm{eq} = 10\ \mathrm{lg}\left[\left(1/T\right)\int_0^T 10^{L/10}\,\mathrm{d}t\right] \tag{4.1}$$

where L_eq = equivalent sound level (dBA), L = instantaneous sound level (dBA) and T = averaging time, often 1 hour, 8

TABLE 2
A-weighted sound levels

Approximate sound level	Noise source or criterion
140	Threshold of pain
122	Supersonic aircraft^{A}
120	Threshold of discomfort
112	Stage I aircraft^{A}
110	Leaf blower at operator
105	OSHA 1 hr/da limit^{B}
99	EEC 1 hr/da limit^{B}
90	OSHA and EEC 8 hr/da limit^{B}
70	EPA criterion for hearing conservation^{C}
67	DOT worst hour limit^{D}
65	Daytime limit, typical community ordinance
45	Noise limit for virtually 100% indoor speech intelligibility
35	Acceptable for sleep
0	Threshold of hearing

Notes:
A: Aircraft measurements 500 ft beyond end of runway, 250 ft to side.
Stage 3 aircraft in current use are quieter.
B: Criteria for worker exposure (US Occupational Safety and Health Administration and European Economic Community).
C: Environmental Protection Agency identified 24-hr equivalent sound level.
D: Department of Transportation design noise level for residential use.

hours, 24 hours, etc. The time period may be identified by the subscript, e.g. L_{eq24} for a 24 hour averaging time.

Integrating sound level meters compute and display equivalent sound level directly. If equivalent sound level is to be determined from a number of representative instantaneous measurements or predictions, the above equation may be rewritten as follows:

$$L_{\mathrm{eq}} = 10\ \mathrm{lg}\left[(1/N)\sum_{i=1}^{N} 10^{L_i/10}\right]. \tag{4.2}$$

If a large number of readings are involved, it is convenient to incorporate the above equation into a computer program. If the base-10 logarithm is not available on the computer it may be obtained from

$$\mathrm{lg}(x) = \ln(x)/\ln(10) \tag{4.3}$$

where ln is the natural (base-e) logarithm.

Example Problem: Equivalent Sound Level

Considering four consecutive 15-minute intervals, during which representative sound levels are 55, 58, 56 and 70 dBA respectively. Determine equivalent sound level for that hour. Solution:

$$L_{\mathrm{eq}} = 10\ \mathrm{lg}(1/4)\left(10^{55/10} + 10^{58/10} + 10^{56/10} + 10^{70/10}\right)$$
$$= 64.5\ \mathrm{dBA}.$$

It can be seen that higher sound levels tend to dominate when determining L_{eq}. Note that the mean average sound level (55 + 58 + 56 + 70)/4 = 59.8 has no significance.

DAY–NIGHT SOUND LEVEL

Day–night sound level takes into account the importance of quiet during nighttime hours by adding a 10 dBA weighting to noise during the period from 10 pm to 7 am. It is given by

$$L_{\mathrm{DN}} = 10\ \mathrm{lg}\left\{[1/24]\left[\int_{7\mathrm{am}}^{10\mathrm{pm}} 10^{L/10}\,\mathrm{d}t + \int_{10\mathrm{pm}}^{7\mathrm{am}} 10^{(L+10)/10}\,\mathrm{d}t\right]\right\} \tag{5.1}$$

where L_{DN} = day–night sound level and t = time (hours).

COMBINING NOISE FROM SEVERAL SOURCES

Correlated sound waves. Sound waves with a precise time and frequency relationship may be considered correlated. A sound wave arriving directly from a source may have a precise phase relationship with a reflected sound wave from the same source. The sound level resulting from combining two correlated sound waves of the same frequency depends on the phase relationship between the waves. Reactive mufflers and silencers are designed to produce reflections that cancel the progressive sound wave.

Active noise control is accomplished by generating sound waves out-of-phase with the noise which is to be cancelled. Active noise control systems employ continuous measurement, signal processing, and sound generation.

Uncorrelated noise sources. Most noise sources are not correlated with one another. The combined effect of two or more uncorrelated sources is obtained by combining the energy from each at the receiver. To do this, we may add mean-square sound pressures. In terms of sound levels, the result is

$$L_{\mathrm{T}} = 10\ \mathrm{lg}\sum_{i=1}^{N} 10^{L_i/10} \tag{6.1}$$

where L_T = total sound level due to N contributions L_i (dBA as measured or predicted at the receiver).

For two contributions, the total sound level is

$$\begin{aligned} L_{\mathrm{T}} &= 10\ \mathrm{lg}\left[10^{L1/10} + 10^{L2/10}\right] \\ &= L_1 + 10\ \mathrm{lg}\left[1 + 10^{-\mathrm{DIF}/10}\right] \end{aligned} \tag{6.2}$$

where L_1 = the greater sound level and DIF = $L_1 - L_2$, the difference between the two sound levels.

The last term in equation 6.2 may be identified as L(add), the quantity to be added to L_1 to obtain total sound level L_T. L(add) is tabulated against DIF, the difference in levels, in Table 3. Values are given to the nearest one-tenth decibel. Although measured and predicted sound levels are often reported to the nearest whole decibel, fractional values are often retained for comparison purposes, and to insure accuracy of intermediate calculations.

Note that addition of the contributions of two equal but uncorrelated sources produces a total sound level 3 decibels higher than the contribution of one source alone.

If the difference between contributions is 10 or more decibels, then the smaller contribution increases total sound level by less than one-half decibel. If the difference is 20 or more decibels, the smaller contribution has no significant effect; for DIF ≥ 20, $L(\text{add}) < 1/20$. This is an important consideration when evaluating noise control efforts. If several individual contributions to overall sound level can be identified, the sources producing the highest sound levels should be considered first. Figure 1 is a graph showing the effect of combining noise levels.

Example Problem: Combining Noise Contributions

The individual contributions of five machines are as follows when measured at a given location: 85, 88, 80, 70 and 95 dBA. Find the sound level when all five are operating together.
Solution:

Using equation 6.1, the result is $L_T = 10\ \lg[10^{85/10} + 10^{88/10} + 10^{80/10} + 10^{70/10} + 10^{95/10}] = 96.25$ dBA. We could use Table 3 instead. Combining the levels in ascending order, the result is

$$70 + 80 = 80.4 \text{ and}$$
$$80.4 + 85 = 86.3 \text{ and}$$
$$86.3 + 88 = 90.3 \text{ and}$$
$$90.3 + 95 = 96.3 \text{ dBA}$$

Fractional parts of one dBA are only retained for purposes or illustration.

For several sources which contribute equally to sound level at the receiver, total sound level is given by

$$L_T = L_1 + 10 \lg n \tag{6.3}$$

where L_1 = sound level contribution at the receiver due to a single source and n = the number of sources. Table 3 and Figure 2 show the effect of combining n equal (but uncorrelated) contributions.

SOUND FIELDS

The region within one or two wavelengths of a noise source or within one or two typical source dimensions is called the **near field.** The region where reflected sound waves have a significant effect on total sound level is called the **reverberant field.** Consider an ideal nondirectional noise source which generates a spherical wave. For regions between the near field and the reverberant field, sound intensity is given by

$$I = W/\left[4\pi r^2\right] \tag{7.1}$$

TABLE 3
Combining noise from two uncorrelated sources

DIF	L(add)	DIF	L(add)
0.0	3.0	5.0	1.2
0.2	2.9	5.5	1.1
0.4	2.8	6.0	1.0
0.6	2.7	6.5	0.9
0.8	2.6	7.0	0.8
1.0	2.5	7.5	0.7
1.2	2.5	8.0	0.6
1.4	2.4	8.5	0.6
1.6	2.3	9.0	0.5
1.8	2.2	9.5	0.5
2.0	2.1	10.0	0.4
2.2	2.0	10.5	0.4
2.4	2.0	11.0	0.3
2.6	1.9	11.5	0.3
2.8	1.8	12.0	0.3
3.0	1.8	12.5	0.2
3.2	1.7	13.0	0.2
3.4	1.6	13.5	0.2
3.6	1.6	14.0	0.2
3.8	1.5	14.5	0.2
4.0	1.5	15.0	0.1
4.2	1.4	15.5	0.1
4.4	1.3	16.0	0.1
4.6	1.3	16.5	0.1
4.8	1.2	17.0	0.1
—	—	17.5	0.1
—	—	18.0	0.1
—	—	18.5	0.1
—	—	19.0	0.1
—	—	19.5	0.0

L1 = greater sound level, L2 = lower sound level. DIF = L1 − L2, Combined sound level LT = L1 + L(add).

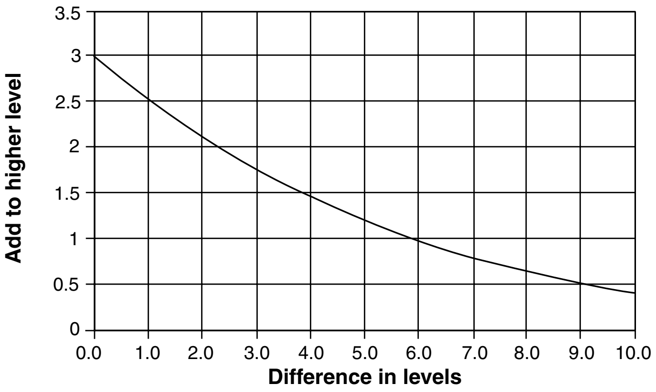


FIGURE 1 Combining noise levels.

where I = sound intensity (W/m^2), where sound pressure and particle velocity are in-phase, W = sound power of the source (W) and r = distance from the source (m). The above equation is called the inverse square law.

Scalar sound intensity level is given by

$$L_I = 10\lg[I/I_{ref}] \tag{7.2}$$

where L_I = scalar sound intensity level and $I_{ref} = 10^{-12}$ W/m^2. For airborne sound under typical conditions, sound pressure level and scalar sound intensity level are approximately equal, from which

$$L_p \approx L_I = 10\lg W - 20\lg r - 109 \tag{7.3}$$

for the spherical wave where L_p and L_I are expressed in dB. If sound power has been A-weighted, L_p and L_I are in dBA.

When the inverse-square law applies, then sound levels decrease with distance at the rate: 20 lg r. Thus, if sound level is known at one location, it may be estimated at another location. Table 4 and Figure 3 show the distance adjustment to be added to sound level at distance r_1 from the source to obtain the sound level at distance r_2.

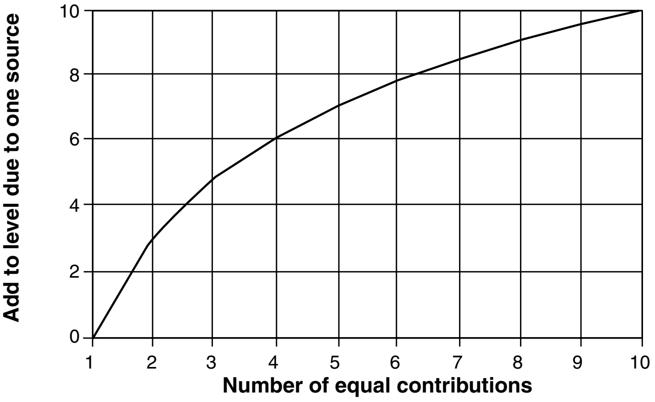


FIGURE 2 Combining n equal contributions.

FIGURE 3 Distance adjustment based on the inverse-square law.

TABLE 4
Spherical wave attenuation

r2/r1	ADJ
0.5	6.0
0.6	4.4
0.7	3.1
0.8	1.9
0.9	0.9
1.0	0.0
1.1	−0.8
1.2	−1.6
1.3	−2.3
1.4	−2.9
1.5	−3.5
1.6	−4.1
1.7	−4.6
1.8	−5.1
1.9	−5.6
2.0	−6.0
2.1	−6.4
2.2	−6.8
2.3	−7.2
2.4	−7.6
2.5	−8.0
2.6	−8.3
2.7	−8.6
2.8	−8.9
2.9	−9.2
3.0	−9.5

Distance adjustment based on the inverse-square law. L(r2) = L(r1) + ADJ.

MEASUREMENT AND INSTRUMENTATION

Sound level meters. The sound level meter is the basic tool for making noise surveys. A typical sound level meter is a hand-held battery-powered instrument consisting of a microphone, amplifiers, weighting networks, a rootmean-square rectifier, and a digital or analog sound level display. The A-weighting network is most commonly used. This network electronically adjusts the signal in accordance with Table 1, so that sound level is displayed in dBA. When measuring out-of-doors, a windscreen is used to reduce measurement error due to wind impinging on the microphone. Integrating sound level meters automatically calculate equivalent sound level. If a standard sound level meter is used, equivalent sound level may be calculated from representative measurements, using the procedure described later.

Frequency analysis. The cause of a noise problem may sometimes be detected by analyzing noise in frequency bands. An octave band is a frequency range for which the upper frequency limit is (approximately) twice the lower limit. An octave band is identified by its center frequency defined as follows:

$$f_c = \left[f_L f_u\right]^{1/2} \tag{8.1}$$

where f_c = the center frequency, f_L = the lower band limit, and f_u = the upper band limit, all in Hz. The center frequencies of the preferred octave bands in the audible range are 31.5, 63, 125, 250 and 500 Hz and 1, 2, 4, 8 and 16 kHz. The center frequencies of the preferred one-third-octave bands are those listed in the first column of Table 1 (Section 3).

Real-time analyzers and Fast-Fourier-Transform (FFT) analyzers examine a signal in all of the selected frequency bands simultaneously. The signal is then displayed as a bargraph, showing the sound level contribution of each selected frequency band.

Sound intensity measurement. Vector sound intensity is the net rate or flow of sound energy. Vector sound intensity measurements are useful in determining noise source power in the presence of background noise and for location of noise sources. Sound intensity measurement systems utilize a two-microphone probe to measure sound pressure at two locations simultaneously.

The signals are processed to determine the particle velocity and its phase relationship to sound pressure.

Calibration. Acoustic calibrators produce a sound level of known strength. Before a series of measurements, sound measurement instrumentation should be adjusted to the calibrator level. Calibration should be checked at the end of each measurement session. If a significant change has occurred, the measured data should be discarded. Calibration data should be recorded on a data sheet, along with instrumentation settings and all relevant information about the measurement site and environmental conditions.

Background noise. When measuring the noise contribution of a given source, all other contributions to total noise are identified as background noise. Let the sound level be measured with the given source operating, and then let background noise alone be measured. The correction for background noise is given by

$$\mathrm{COR} = 10\ \lg\left[1 - 10^{-\mathrm{DIF}/10}\right] \tag{8.2}$$

where DIF = Total noise level – background noise level, and the noise level contribution of the source in question is given by:

$$L_{\mathrm{SOURCE}} = \text{Total noise level} + \mathrm{COR}.$$

Background noise corrections are tabulated in Table 5 and plotted in Figure 4. Whenever possible measurements should be made under conditions where background noise is negligible. When total noise level exceeds background noise by at least 20 dB, then the correction is less than 1/20 dB. Such ideal conditions are not always possible. Truck noise, for example, must sometimes be measured on a highway with other moving vehicles nearby. If the difference between total noise level and background noise is less than 5 dB, then the contribution of the source in question cannot be accurately determined.

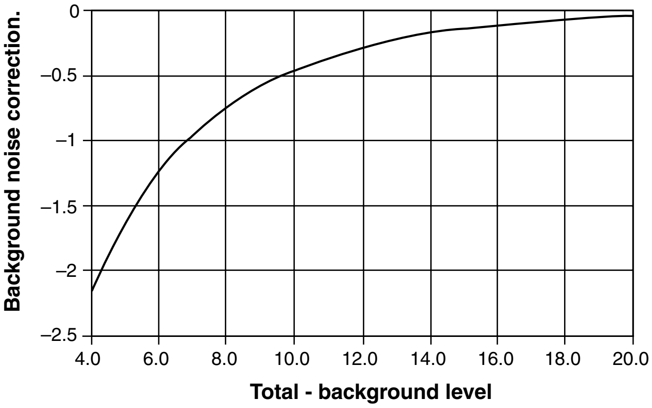


FIGURE 4 Background noise correction.

HEARING DAMAGE RISK

The frequency range of human hearing extends from about 20 Hz to 20 kHz. Under ideal conditions, a sound pressure level of 0 dB at 1 kHz can be detected. Human hearing is less sensitive to low frequencies and very high frequencies.

Hearing threshold. A standard for human hearing has been established on the basis of audiometric measurements at a series of frequencies. An individual's hearing threshold represents the deviation from the standard or audiometric-zero levels. A hearing threshold of 25 dB at 4 kHz, for example, indicates that an individual has "lost" 25 dB in ability to hear sounds at a frequency of 4 kHz (assuming the individual had "normal" hearing at one time).

A temporary threshold shift (TTS) is a hearing threshold change determined from audiometric evaluation before, and immediately after exposure to loud noise. A measurable permanent threshold shift (PTS) usually occurs as a result of long-term noise exposure. The post-exposure audiometric measurements to establish PTS are made after the subject has been free of loud noise exposure for several hours. A compound threshold shift (CTS) combines a PTS and TTS. There is substantial evidence that repeated TTS's translate into a measurable PTS. Miller (1974) assembled data relating TTS, CTS and PTS resulting from exposure to high noise levels.

Occupational Safety and Health Administration (OSHA criteria. OSHA (1981, 1983) and the Noise Control Act (1972) set standards for industrial noise exposure and guidelines for hearing protection. OSHA criteria have resulted in reduced noise levels in many industries and reduced the incidence of hearing loss to workers. However, retrospective studies have shown that some hearing loss will occur with long-term exposure a OSHA-permitted sound levels.

The basic OSHA **criterion level** (CL) is a 90 dBA sound exposure level for an 8 hour day. An **exchange rate** (ER) of 5 dBA is specified, indicating that the permissible daily

TABLE 5
Background noise correction

DIF	COR	DIF	COR
0.2	−13.5	—	—
0.4	−10.6	—	—
0.6	−8.9	—	—
0.8	−7.7	6.5	−1.1
1.0	−6.9	7.0	−1.0
1.2	−6.2	7.5	−0.9
1.4	−5.6	8.0	−0.7
1.6	−5.1	8.5	−0.7
1.8	−4.7	9.0	−0.6
2.0	−4.3	9.5	−0.5
2.2	−4.0	10.0	−0.5
2.4	−3.7	10.5	−0.4
2.6	−3.5	11.0	−0.4
2.8	−3.2	11.5	−0.3
3.0	−3.0	12.0	−0.3
3.2	−2.8	12.5	−0.3
3.4	−2.7	13.0	−0.2
3.6	−2.5	13.5	−0.2
3.8	−2.3	14.0	−0.2
4.0	−2.2	14.5	−0.2
4.2	−2.1	15.0	−0.1
4.4	−2.0	15.5	−0.1
4.6	−1.8	16.0	−0.1
4.8	−1.7	16.5	−0.1
5.0	−1.7	17.0	−0.1
5.2	−1.6	17.5	−0.1
5.4	−1.5	18.0	−0.1
5.6	−1.4	18.5	−0.1
5.8	−1.3	19.0	−0.1
6.0	−1.3	19.5	−0.0

DIF = total noise level–background noise level. Sound level due to source = total noise level + COR.

TABLE 6
Allowable exposure times

Time *T* hr/da	Sound level *L* dBA
32*	80
16	85
8	90
4	95
2	100
1	105
1/2	110
1/4 or less	115

*The 32 hr exposure time is used in evaluating noise dose when sound levels vary.

exposure time is halved with each 5 dBA sound level increase. The **threshold level,** the sound level below which no contribution is made to daily noise dose, is 80 dBA (threshold level is not to be confused with hearing threshold). When noise exposure exceeds the **action level** (85 dBA) a hearing conservation program is to be implemented. A hearing conservation program should include noise exposure monitoring audiometric testing, employee training, hearing protection and record-keeping.

According to OSHA standards continuous noise exposure (measurable on the slow-response scale of a sound level meter) is not to exceed 1151 dBA. For sound levels L where $80 \leq L \leq 115$ dBA allowable exposure time is given by

$$T = 8/\left[2^{(L-\mathrm{CL})/\mathrm{ER}}\right] = 8/\left[2^{(\mathrm{L}-90)/5}\right] \qquad (9.1)$$

where T = allowable exposure time (hours/day). The result is shown in Table 6.

Noise dose. When sound levels vary during the day, noise dose is used as an exposure criterion. Noise dose is given by

$$D_{\%} = 100\sum_{i=1}^{N} C_{\mathrm{i}}/T_{\mathrm{i}}$$

where C = actual exposure of an individual at a given sound level (hr), T = allowable exposure time at that level and N = the number of different exposure levels during one day. Noise dose $D_{\%}$ should not exceed 100%. As an alternative to monitoring and calculations, workers may wear dosimeters which automatically measure and calculate daily dose.

An exchange rate of 3 dB is used in occupational noise exposure criteria by some European countries. This exchange rate is equivalent to basing noise exposure on L_{eq}.

Environmental Protection Agency (EPA) identified levels. Using a 4 kHz threshold shift criterion, protective noise levels are substantially lower than the OSHA criteria. EPA (1974, 1978) in its "Levels" document identified the equivalent sound level of intermittent noise:

$$L_{\mathrm{eq24}} = 70\ \mathrm{dBA}$$

as the "(at ear) exposure level that would produce no more than 5 dB noise-induced hearing damage over a 40 year period". This value is based on a predicted hearing loss smaller than 5 dB at 4 kHz for 96% of the people exposed to 73 dBA noise for 8 hr/da × 250 da/yr × 40 yr. With the following corrections, the 73 dBA level is adjusted to L_{eq24} = 70 dBA the protective noise level:

−1.6 dBA to account for 365 da/yr exposure,
−4.8 dBA to correct for 24 hr/day averaging,
+5 dBA assuming intermittent exposure and
−1.6 dBA for a margin of safety.

NON-AUDITORY EFFECTS OF NOISE

The relationship between long-term exposure to industrial noise and the probability of noise-induced hearing loss is well-documented. And, we can estimate the effect of intrusive noise on speech intelligibility and masking of warning signals. Equivalent sound levels and day-night sound levels based on hearing protection, activity interference, and annoyance are given in Table 9. The United States Census Bureau identified noise as the top complaint about neighborhoods, and the major reason for wanting to move. In a typical city, about 70% of citizen complaints relate to noise. The most common complaints are aircraft noise, highway noise, machinery and equipment, and amplified music.

Noise tolerance varies widely among individuals. It is difficult to relate noise levels to psychological and non-auditory physiological problems. But there is anecdotal evidence that violent behavior can be triggered by noise. In a New York case, one man cut off another man's hand in a dispute over noise. In another noise-related incident, a New Jersey man operated his motorcycle engine inside his apartment, leading a neighbor to shoot him.

Chronic noise exposure has been related to children's health and cognitive performance. In a study of British schools, Stansfield and Haines (2000) compared reading skills of students at four schools with 16-hour equivalent sound levels less than 57 dBA and four schools with levels greater than 66 dBA. After adjustment for socioeconomic factors, lower average reading scores were found at the noisier schools. The difference was equivalent to six months of learning over four years.

A study by Zimmer et al. (2001) examined aircraft noise exposure and student proficiency test results at three grade levels. Communities with comparable socioeconomic status were selected for the study. Noise-impacted communities with a day-night sound level greater than 60 dBA and communities with a level of less than 45 dBA were compared. If proficiency test results are extrapolated to educational attainment and salary level, one could predict a 3% salary level disadvantage for students from the impacted communities.

COMMUNITY NOISE

Contributors to community noise include aircraft, highway vehicles, off-road vehicles, powered garden equipment, construction activities, commercial and industrial activities, public address systems and loud radios and television sets. The major effects of community noise include sleep interference, speech interference, and annoyance.

Highway noise. Noise levels due to highway vehicles may be estimated from the Federal Highway Administration (FHWA) model summarized by the sound level vs. speed relationships in Table 7. These values make it possible to predict the impact of a proposed highway or highway improvement on a community.

The contribution that a given class of vehicles makes to hourly equivalent sound level is given by

$$L_{eqH} = L_o + 10\,\lg\left[D_o V/S\right] + A_B + A_D + A_F + A_G + A_S - 25 \tag{10.1}$$

TABLE 7
Energy mean emission levels for vehicles

Vehicle class	Sound level L_0dBA	Speed S km/hr
Autos	$31.8 \lg S - 2.4$	$\geqslant 50$
Autos	62	<50
Med. trucks	$33.9 \lg S + 16.4$	$\geqslant 50$
Med. trucks	74	<50
Heavy trucks	$24.6 \lg S + 38.5$	$\geqslant 50$
Heavy trucks	87	<50

Sound levels at 15 meters. Source: Barry and Reagan (1978).

where D_o = 15 m, V = volume (vehicles/hr), S = speed (km/hr). $A_{B, D, F, G}$ and $_S$ are adjustments for barriers, distance, finite highway segments, grade and shielding due to buildings, respectively. Each term is applied to a given class of vehicles and traffic lane. For acoustically absorptive sites, the distance adjustment is

$$A_D = 15\lg\left[D_o/D\right] \tag{10.2}$$

where D = distance from the traffic lane (m). Hourly equivalent sound level at any location is predicted by combining the contributions from all vehicle classes and traffic lanes. The result is

$$L_{eqH(COMBINED)} = 10\lg\sum_{i=1}^{N} 10^{L_{eqHi}/10}. \tag{10.3}$$

Design noise levels for highways. Design noise levels specified by the Federal Highway Administration (1976) are summarized in Table 8. Noise predictions based on projected traffic on proposed highways are compared with the design levels. These data aid in selecting a highway design and routing alternative including the "no-build" alternative.

Aircraft noise. Noise contour maps are available for most major airports. These enable one to make rough predictions of the impact of aircraft noise on nearby communities. Federal Aviation Administration publications (1985a and b) outline aircraft noise certification procedures and aircraft noise compatibility planning. Many of the existing airport noise contour maps are based on the descriptor **Noise exposure forecast (NEF).** An approximate conversion from NEF to L_{DN} is given by

$$L_{DN} \approx \text{NEF} + 35 \tag{10.4}$$

where L_{DN} = day-night sound level (± about 3 dBA).

Community noise criteria. There are thousands of different community noise ordinances, with a wide range of permitted sound levels. Their effectiveness depends largely on the degree of enforcement in a particular community. The Environmental Protection Agency has identified the noise

TABLE 8
Design noise levels

Sound level L_{eqH}dBA	Measurement location	Land use category
57	Exterior	Tracts of land in which serenity and quiet are of extraordinary significance.
67	Exterior	Residences, schools, churches, libraries, hospitals, etc.
72	Exterior	Commercial and other activities.
52	Interior	Residences, schools, churches, libraries, hospitals, etc.

TABLE 9
Protective noise levels

Effect	Level (dBA)	Area
Hearing protection	$L_{eq24} \leqslant 70$	All areas. See Section 9.
Outdoor activity	$L_{DN} \leqslant 55$	Outdoors in residential areas.
Interference and annoyance	$L_{eq24} \leqslant 55$	Outdoor areas where people spent limited amounts of time.
Indoor activity	$L_{DN} \leqslant 45$	Indoor residential areas.
Interference and annoyance	$L_{eq24} \leqslant 45$	Other indoor areas with human activities such as schools, etc.

Source: EPA (1974, 1979).

levels in Table 9 as protective of public health and welfare. All are based on an average 24 hour day.

Control of community noise. Environmental noise problems are particularly difficult to solve due to problems of shared responsibility and jurisdiction. In many cases, Federal laws preempt community regulations. Highway noise and aircraft noise, often the most significant contributors to community noise levels, are largely exempt from local control. In spite of the difficulties encountered, however, the importance of protecting the quality of life makes environmental noise control efforts worthwhile. Depending on the circumstances, some of the following courses of action may be considered.

a) Review the applicable noise ordinance. Compare it with a model noise ordinance. Check to see if specific limits are set in terms dBA. Determine whether or not sound level meters are available and whether or not the ordinance is actually enforced.
b) Meet with representatives of the local governing body or environmental commission. Make them aware of noise related problems in the community.
c) Initiate a campaign for public awareness with regard to the environment including the noise environment. Make use of the local papers.
d) Consider a ban or limitation on all-terrain-vehicles (ATV's). Determine whether muffler requirements are actually enforced.
e) Encourage planning and zoning boards to require an environmental impact statement (EIS), including a noise report, before major projects are approved.
f) Support noise labeling for lawn mowers and other power equipment.
g) Attend and participate in hearings involving plans for airports, heliports, and highways. Consider noise impact when evaluating the cost/benefit ratio for proposed facilities.
h) Evaluate the feasibility of noise barriers on existing and proposed highways in sensitive areas.
i) Support legislation to reduce truck noise emission limits.
j) Support legislation enabling airport curfews.

REFERENCES

Barry, T.M. and Reagan, J.A. *FHW A highway traffic noise prediction model* FHWA-RD-77-108, 1978.

Environmental Protection Agency, *Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety,* EPA 550/9-74-004, 1974.

Environmental Protection Agency, *Model community noise control ordinance,* EPA 550/9-76-003, 1975.

Environmental Protection Agency, *Protective noise levels,* EPA 559/979-100, 1978.

Federal Aviation Administration, *Noise standards: aircraft type and airworthiness certification,* FAR part 36, 1985(a).

Federation Aviation Administration, *Airport noise compatibility planning,* FAR part 150, 1985(b).

Federal Highway Administration, *Procedures for abatement of highway traffic noise and construction noise,* FHPM 7-7-3, 1976.

Federal Register, *Code of federal regulations,* **29,** parts 1900 to 1910, 1985.

Miller, J.D., "Effects of noise on people," *J. Acoust Soc. Am.* **56,** no. 3, pp. 729–764, 1974.

Noise control act of 1972, PL 92-574, HR 11021, Oct. 27, 1972.

Occupational Safety and Health Administration, "Occupational noise exposure hearing conservation amendment" *Federal Register,* **46**(11), 4078–4181 and **46**(162), 42622–42639, 1981.

Occupational Safety and Health Administration, "Guidelines for noise enforcement", *OSHA Instruction,* CPL2-2.35,29 CFR1910.95(6) (1), 1983.

Peterson, A.P.G., *Handbook of noise measurement,* GenRad, Concord, MA, 9th ed., 1980.

Stansfield, S. and M. Haines, "Chronic aircraft noise exposure and children's cognitive performance and health: the Heathrow studies", FICA symposium, San Diego CA, 2000.

Wilson, C., *Noise Control,* Krieger, Malabar FL, 1994.

Zimmer, I.B., R. Dresnack, and C. Wilson, Modeling the impact of aircraft noise on student proficiency", NOISE-CON Portland ME, 2001.

The following Internet resources may contain current information of interest:

www.faa.gov	Federal Aviation Administration
www.icao.int	International Civil Aviation Administration
www.fhwa.dot.gov. environment/noise	Federal Highway Administration
www.osha.gov	Occupational Safety and Health Administration

www.cdc.gov/niosh	National Institute for Occupational Safety and Health
europe.osha.eu.int	European Agency for Safety and Health at Work
europa.eu.int.en/record/green	European Union
www.epa.gov	Environmental Protection Agency
www.who.int/environmental_information/noise	World Health Organization

CHARLES E. WILSON
New Jersey Institute of Technology

NON-IONIZING RADIATIONS

Lasers, microwave ovens, radar for pleasure boats, infrared inspection equipment and high intensity light sources generate so-called "non-ionizing" radiation.

Electromagnetic radiations which do not cause ionization in biological systems may be presumed to have photon energies less than 10–12 eV and may be termed "non-ionizing."

Because of the proliferation of such electronic products as well as a renewed interest in electromagnetic radiation hazards, the Congress enacted Public Law 90-602, the "Radiation Control for Health and Safety Act." This Act has as its declared purpose the establishment of a national electronic product radiation control program which includes the development and administration of performance standards to control the emission of electronic product radiation. The most outstanding feature of the Act is its omnibus coverage of all types of electromagnetic radiation emanating from electronic products, that is, gamma, X-rays, ultraviolet, visible, infrared, radio frequencies (RF) and microwaves. Performance standards have already been issued under the Act for TV sets, microwave ovens and lasers. In similar fashion, the recent enactment of the federal Occupational Safety and Health Act gives attention to the potential hazards of non-ionizing radiations in industrial establishments.

For the purposes of this chapter more formal treatment is given to ultraviolet (UV) radiation, lasers, and microwave radiation than the visible and infrared (IR) radiations. However the information on visible and IR radiation presented in the section on Laser Radiation is generally applicable to non-coherent sources. It should become obvious in reading the material which follows that the eye is the primary organ at risk to all of the non-ionizing radiations.

NATURE OF ELECTROMAGNETIC ENERGY

The electromagnetic spectrum extends over a broad range of wavelengths, e.g. from $<10^{-12}$ to $>10^{10}$ cm. The shortest wavelengths are generated by cosmic and X-rays, the longer wavelengths are associated with microwave and electrical power generation. Ultraviolet, visible and IR radiations occupy an intermediate position. Radiation frequency waves may range from 10 kHz to 10^{12} Hz, IR rays from 10^{12} ; 4×10^{13} Hz (0.72 μm), the visible spectrum from approximately 0.7–0.4 μm, UV from approximately 0.4–0.1 μm and γ- and X-radiation, below 0.1 μm. The photon energies of electromagnetic radiations are proportional to the frequency of the radiation and inversely proportional to wavelength, hence the higher energies (e.g.10^{8} eV) are associated with X- and γ-radiations, the lower energies (e.g.10^{-6} eV) with RF and microwave radiations.

Whereas the thermal energy associated with molecules at room temperature is approximately 1/30 eV, the binding energy of chemical bonds is roughly equivalent to a range of <1–15 eV, the nuclear binding energies of protons may be equivalent to 10^{6} eV and greater. Since the photon energy necessary to ionize atomic oxygen and hydrogen is of the order of 10–12 eV it seems in order to adopt a value of approximately 10 eV as a lower limit in which ionization is produced in biological material.

An extremely important qualification however is that non-ionizing radiations may be absorbed by biological systems and cause changes in the vibrational and rotational energies of the tissue molecules, thus leading to possible dissociation of the molecules or, more often, dissipation of energy in the form of fluorescence or heat.

In conducting research into the bioeffects of the non-ionizing radiations the investigator has had to use several units of measurement in expressing the results of his studies. For this reason Appendix A, containing definitions of many useful radiometric terms has been included. Appendix B provides a simple means for expressing radiant exposure and irradiance units in a number of equivalent terms.

ULTRAVIOLET RADIATION

Physical Characteristics of Ultraviolet Radiation

For the purpose of assessing the biological effects of UV radiation the wavelength range of interest can be restricted to 0.1–0.4 μm. This range extends from the vacuum UV (0.1 μm) to the near UV (0.4 μm). A useful breakdown of the UV region is as follows:

UV region	γ-range (±m)	(eV)
Vacuum	<0.60	>7.7
Far	0.16–0.28	7.7 4.4
Middle	0.28–0.32	4.4 3.9
Near	0.32 0.4	3.9 3.1

The photon energy range for wavelengths between 0.1 and 0.4 μm is 12.4–3.1 eV, respectively.

Representative Sources of Ultraviolet Radiation

The manor source of UV radiation is the sun, although absorption by the ozone layer permits only wavelengths greater than 0.29 μm to reach the surface of the earth. Low and high pressure mercury discharge lamps constitute significant manmade sources. In low pressure mercury vapor discharge lamps over 85% of the radiation is usually emitted at 0.2537 μm, viz. at germicidal wavelengths. At the lower pressures (fractions of an atmosphere) the characteristic mercury lines predominate whereas at higher pressures (up to 100 atmos.) the lines broaden to produce a radiation continuum. In typical quartz lamps the amount of energy at wavelengths below 0.38 μm may be 50% greater than the radiated visible energy, depending on the mercury pressure. Other manmade sources include xenon discharge lamps, lasers, and relatively new types of fluorescent tubes, which emit radiation at wavelengths above 0.315 μm reportedly at an irradiance less than that measured outdoors on a sunny day.

Biological Effects of Ultraviolet Radiation

The biological action spectrum for erythema (reddening) produced by UV radiation of the skin has been the subject of investigation for many years. The most recent data show that a maximum erythemal effect is produced at 0.260 μm with the secondary peak at approximately 0.290 μm. Erythemal response to wavelengths above 0.32 μm is predictably poor. The greatly increased air absorption of wavelengths below 0.25 μm and difficulty in obtaining monochromatic radiations in this region probably account for the lack of definitive bioeffects data. This may change with the increase in the number of UV lasers.

Wavelengths between 0.28 and 0.32 μm penetrate appreciably into the corium of the epidermis; those between 0.32 and 0.38 μm are absorbed in the epidermis, while those below 0.28 μm appear to be absorbed almost completely in the stratum corneum of the epidermis.

Depending on the total UV dose, the latent periods for erythema may range from 2 to several hours; the severity may vary from simple erythema to blistering and desquamation with severe secondary effects. A migration of melanin granules from the basal cells to the maphigian cell layers of the epidermis may cause a thickening of the horny layers of the skin. The possible long-term effects of the repeated process of melanin migration is not completely understood. The available data seem to support the contention that some regions of the UV may produce or initiate carcinogenesis in the human skin. The experiments which have supported this contention indicate that the biological action spectrum for carcinogenesis is the same as that for erythema.

Cases of skin cancer have been reported in workers whose occupation requires them to be exposed to sunlight for long periods of time. The reportedly high incidence of skin cancer in outdoor workers who are simultaneously exposed to chemicals such as coal tar derivatives, benzpyrene, methyl cholanthrene, and other anthracene compounds raises the question as to the role played by UV radiation in these cases. It is a matter of common knowledge that significant numbers of workers who routinely expose themselves to coal tar products while working outdoors experience a photosensitization of the skin.

Abiotic effects from exposure to UV radiation occurs in the spectral range of 0.24–0.31 μm. In this part of the spectrum, most of the incident energy is absorbed by the corneal epithelium at the surface of the eye. Hence, although the lens is capable of absorbing 99% of the energy below 0.35 μm only a small portion of the radiation reaches the anterior lenticular surface.

Photon-energies of about 3.5 eV (0.36 μm) may excite the lens of the eye or cause the aqueous or vitreous humor to fluoresce thus producing a diffuse haziness inside the eye that can interfere with visual acuity or produce eye fatigue. The phenomenon of fluorescence in the ocular media is not of concern from the bioeffects standpoint; the condition is strictly temporary and without detrimental effect.

The development of photokeratitis usually has a latency period varying from 30 min to as long as 24 hrs depending on the severity of the exposure. A sensation of "sand in the eyes" accompanied by varying degrees of photophobia, lacrimination and blepharospasm is the usual result. Blepharospasm is a reflex protective mechanisms characterized by an involuntary tight closing of the lids, usually over a damaged cornea.

Exposure Criteria

The biological action spectrum for keratitis peaks at 0.28 μm. At this wavelength, the threshold for injury has been determined to be approximately 0.15×10^6 ergs. It has been suggested that the corneal reaction in due primarily to selective absorption of UV by specific cell constituents, for example, globulin.

Verhoeff and Bell (1916) gave the first quantitative measurement of the UV energy necessary for threshold damage as 2×10^6 ergs/cm^2 for the whole UV spectrum. More recent data by Pitts, using 10 nm bands of radiation produced a threshold of approximately 0.5×10^5 ergs/cm^2 in rabbit eyes.

The exposure criteria adopted by the American Medical Association based on erythemal thresholds at 0.2537 μm radiation are as follows: 0.5×10^{-6} W/cm^2 for exposure up to 7 hr; 0.1×10^{-6} W/cm^2 for exposure periods up to and exceeding 24 hr. Although these criteria are generally thought to be very conservative, i.e. stringent, they are nevertheless in common use.

The American Conference of Governmental Industrial Hygienists (1982) recommend threshold limit values (TLV) for UV irradiation of unprotected skin and eyes for active wavelengths between 0.2 and 0.315 μm (200 and 315 nm)[37]. Typical values are: for 200 nm, a TLV of 100 mJ/cm^2; for 240 nm, a TLV of 10 mJ/cm^2; for 280 nm, a TLV of 3.4 mJ/cm^2; and for 315 nm, a TLV of 1 J/cm^2.

Measurement of Ultraviolet Radiation

Various devices have been used to measure UV radiation, e.g. photoelectric cells, photoconductive cells, photovoltaic cells, and photochemical detectors. It is common practice to employ the use of selective filters in front of the detecting device in order to isolate that portion of the UV spectrum of interest to the investigator.

A commonly used detector is the barrier or photovoltaic cell. Certain semiconductors such as selenium or copper oxide deposited on a selected metal develop a potential barrier between the layer and the metal. Light falling upon the surface of the cell causes the flow of electrons from the semiconductor to the metal. A sensitive meter placed in such a circuit will record the intensity of radiation falling on the cell.

Ultraviolet photocells take advantage of the fact that certain metals have quantitative photoelectric responses to specific bands in the UV spectrum. Therefore a photocell may be equipped with metal cathode surfaces which are sensitive to certain UV wavelengths of interest. One of the drawbacks of photocells is solarization or deterioration of the envelope, especially with long usage or following measurement of high intensity UV radiation. This condition requires frequent recalibration of the cell. The readings obtained with these instruments are valid only when measuring monochromic radiation, or when the relationship between the response of the instrument and the spectral distribution of the source is known.

A desirable design characteristic of UV detectors is to have the spectral response of the instrument closely approximate that of the biological action spectrum under consideration. However, such an instrument is unavailable at this time. Since available photocells and filter combinations do not closely approximate the UV biological action spectra it is necessary to standardize (calibrate) each photocell and meter. Such calibrations are generally made at a great enough distance from a standard source that the measuring device is in the "far field" of the course. Special care must be taken to control the temperature of so-called standard mercury lamps because the spectral distribution of the radiation from the lamps is dependent upon the pressure of the vaporized mercury.

A particularly useful device for measuring UV is the thermopile. Coatings on the receiver elements of the thermopile are generally lamp black or gold black to simulate black body radiation devices. Appropriate thermopile window material should be selected to minimize the effects of air convection, the more common windows being crystal quartz, lithium chloride, calcium fluoride, sodium chloride, and potassium bromide.

Low intensity calibration may be made by exposing the thermopile to a secondary standard (carbon filament) furnished by the National Bureau of Standards.

Other UV detection devices include (1) photodiodes, e.g. silver, gallium arsenide, silver zinc sulfide, and gold zinc sulfide. Peak sensitivity of these diodes is at wavelengths below 0.36 μm; the peak efficiency or responsivity is of the order of 50–70%; (2) thermocouples, e.g. Chromel-Alumel; (3) Golay cells; (4) superconducting bolometers, and (5) zinc sulfide Schottky barrier detectors.

Care must be taken to use detection devices having the proper rise time characteristics (some devices respond much too slowly to obtain meaningful measurements). Also, when measurements are being made special attention should be given to the possibility of UV absorption by many materials in the environment, e.g. ozone or mercury vapor, thus adversely affecting the readings. The possibility of photochemical reactions between UV radiation and a variety of chemicals also exists in the industrial environment.

Control of Exposure

Because UV radiations are so easily absorbed by a wide variety of materials appropriate attenuation is accomplished in a straightforward manner.

In the case of UV lasers no firm bioeffects criteria are available. However the data of Pitts may be used because of the narrow band UV source used in his experiments to determine thresholds of injury to rabbit eyes.

LASER RADIATION

Sources and Uses of Laser Radiation

The rate of development and manufacture of devices and systems based on stimulated emission of radiation has been truly phenomenal. Lasers are now being used for a wide variety of purposes including micromachining, welding, cutting, sealing, holography, optical alignment, interferometry, spectroscopy, surgery and as communications media. Generally speaking lasing action has been obtained in gases, crystalline materials, semiconductors and liquids. Stimulated emission in gaseous systems was first reported in a helium-neon mixture in 1961. Since that time lasing action has been reported at hundreds of wavelengths from the UV to the far IR (several hundred micrometers). Helium–neon (He–Ne) lasers are typical of gas systems where stable single frequency operation is important. He–Ne systems can operate in a pulsed mode or continuous wave (CW) at wavelengths of 0.6328, 1.15, or 3.39 μm depending upon resonator design. Typical power for He–Ne systems is of the order of 1–500 mW. The carbon dioxide gas laser system operates at a wavelength of 10.6 μm in either the continuous wave, pulsed, or Q-switched modes. The power output of CO_2–N_2 systems may range from several watts to greater than 10 kW. The CO_2 laser is attractive for terrestrial and extra-terrestrial communications because of the low absorption window in the atmosphere between 8 and 14 qm. Of major significance from the personal hazard standpoint is the fact that enormous power may be radiated at wavelength which is invisible to the human eye. The argon ion gas system operates predominantly at wavelengths of 0.488 and 0.515 μm in either a continuous wave or pulsed mode. Power generation is greatest at 0.488 μm, typically at less than 10 W.

Of the many ions in which laser action has been produced in solid state crystalline materials, perhaps neodymium (Nd^{3+}) in garnet or glass and chromium (Cr^{3+}) in aluminum oxide are most noteworthy. Garnet (yttrium aluminum garnet) or YAG is an attractive host for the trivalent neodymium ion because the 1.06 μm laser transition line is sharper than in other host crystals. Frequency doubling to 0.530 μm using lithium niobate crystals may produce power approaching that available in the fundamental mode at 1.06 μm. also through the use of electro-optic materials such as KDP, barium–sodium niobate or lithium tantalite, "tuning" or scanning of laser frequencies over wide ranges may be accomplished. The ability to scan rapidly through wide frequency ranges requires special consideration in the design of protective measures.

Perhaps the best known example of a semi-conductor laser is the gallium arsenide types operating at 0.840 μm; however, semiconductor materials have already operated over a range of approximately 0.4–5.1 μm. Generally speaking, the semiconductor laser is a moderately low-powered (milliwatts to several watts) CW device having relatively broad beam divergence thus tending to reduce its hazard potential. On the other hand, certain semiconductor lasers may be pumped by multi-kV electron beams thus introducing a potential ionizing radiation hazard.

Through the use of carefully selected dyes, it is possible to tune through broad wavelength ranges.

Biological Effects of Laser Radiation

The body organ most susceptible to laser radiation appears to be the eye; the skin is also susceptible but of lesser importance. The degree of risk to the eye depends upon the type of laser beams used, notably the wavelength, output power, beam divergence, and pulse repetition frequency. The ability of the eye to refract long UV, visible, and near IR wavelengths is an additional factor to be considered in assessing the potential radiation hazard.

In the UV case of UV wavelengths (0.2–0.4 μm) produced by lasers the expected response is similar to that produced by non-coherent sources, e.g. photophobia accompanied by erythema, exfoliation of surface tissues and possible stromal haze. Absorption of UV takes place at or near the surface of tissues. The damage to epithelium results from the photochemical denaturization of proteins.

In the case of IR laser radiation damage results exclusively from surface heating of the cornea subsequently to absorption of the incident energy by tissue water in the cornea. Simple heat flow models appear to be sufficiently accurate to explain the surface absorption and damage to tissue.

In the case of the visible laser wavelengths (0.4–0.75 μm) the organ at risk is the retina and more particularly the pigment epithelium of the retina. The cornea and lens of the eye focus the incident radiant energy so that the radiant exposure at the retina is at least several orders of magnitude greater than that received by the cornea. Radiant exposures which are markedly above the threshold for producing minimal visions on the retina may cause physical disruption of retinal tissue by steam formation or by projectile-like motion of the pigment granules. In the case of short transient pulses such as those produced by Q-switched systems, acoustical phenomena may also be present.

There are two transition zones in the electromagnetic spectrum where bio-effects may change from one of a corneal hazard to one of a retinal hazard. These are located at the interface of the UV-visible region and the visible–near IR region. It is possible that both corneal and retinal damage as well as damage to intermediate structures such as the lens and iris could be caused by devices emitting radiation in these transitional regions. Several investigators noticed irreversible changes in electroretinograms with attendant degeneration of visual cells and pigment epithelium, when albino and pigmented rats were exposed to high illumination environments.

The chronic and long term effects of laser radiation have not been fully explored.

The biological significance of irradiating the skin with lasers is considered to be less than that caused by exposure of the eye since skin damage is usually repairable or reversible. The most common effects on the skin range from erythema to blistering and charring dependent upon the wavelength, power, and time of exposure to the radiation. Depigmentation of the skin and damage to underlying organs may occur from exposure to extremely high powered laser radiation, particularly Q-switched pulses. In order that the relative eye-skin hazard potential be kept in perspective, one must not overlook possible photosensitization of the skin caused by injection of drugs or use of cosmetic materials. In such cases the maximum permissible exposure (MPE) levels for skin might be considerably below currently recommended values.

The thresholds for producing retinal lesions at all visible wavelengths were considered to be approximately the same i.e., 5 to 10 W/cm^2, until more recent investigations discovered a much greater sensitivity of the eye to blue wavelengths. The mechanism for this enhanced sensitivity is explained on the basis of photochemical, rather than thermal effects.

Exposure Criteria

Permissible levels of laser radiation impinging upon the eye have been derived from short term exposure and an examination of damage to eye structures as observed through an ophthalmoscope. Some investigators have observed irreversible visual performance changes at exposure levels as low as 10% of the threshold determined by observation through an ophthalmoscope. McNeer and Jones found that at 50% of the ophthalmoscopically determined threshold the ERG B wave amplitude was irreversibly reduced. Mautner has reported severe changes in the visually evoked cortical potential at 25% of the ophthalmoscopically determined threshold. Since most, if not all, of the so-called laser criteria have been based on ophthalmoscopically-determined lesions on the retina, the findings of irreversible functional changes at lower levels causes one to ponder the exact magnitude of an appropriate safety factor which should be applied to the ophthalmoscope data in order to derive a reasonable exposure criterion.

There is unanimous agreement that any proposed maximum permissible exposure (MPE) or threshold limits value (TLV) does not sharply divide what is hazardous from what is safe. Usually any proposed values take on firm meaning only after years of practical use. However, it has become general practice in defining laser exposure criteria to:

1) Measure the radiant exposure (J/cm^2) or irradiance (W/cm^2) in the plane of the cornea rather than making an attempt to calculate the values at the retina. This simplifies the measurements and calculations for the industrial hygienists and radiation protection officers.
2) Use a 7 mm dia. limiting aperture (pupil) in the calculations. This assumes that the largest amount of laser radiation may enter the eye.
3) Make a distinction between the viewing of colimated sources, for example lasers and extended sources, such as fluorescent tubes or incandescent lamps. The MPE for extended source viewing takes into account the solid angle subtended at the eyes in viewing the light source; therefore the unit is W/cm^2·sr (Watts per square centimeter and steradian).
4) Derive permissible levels on the basis of the wavelength of the laser radiation, e.g. the MPE for neodymium wavelength (1.06 μm) should be increased, i.e. made less stringent by a factor of approximately five than the MPE for visible wavelengths.
5) Urge caution in the use of laser systems that emit multiple pulses. A conservative approach would be to limit the power of energy in any single pulse in the train to the MPE specified for direct irradiation at the cornea. Similarly the average power for a pulse train could be limited to the MPE of a single pulse of the same duration as the pulse train. More research is needed to precisely define the MPE for multiple pulses.

Typical exposure criteria for the eye proposed by several organizations are shown in Wilkening (1978). These data do not apply to permissible levels at UV wavelengths or to the skin. A few supplementary comments on these factors are in order: There appears to be general agreement on maximum permissible exposure levels of radiation for the skin, e.g. the MPE values are approximately as follows for exposure times greater than 1 sec, an MPE of 0.1 W/cm^2; exposure times $10^{-1}-1$ sec, 1.0 W/cm^2; for $10^{-4}-10$ sec, 0.1 J/cm^2, and for exposure times less than 10^{-4} sec, 0.01 J/cm^2. The MPE values apply to visible and IR wavelengths. For UV radiations the more conservative approach is to use the standards established by the American Medical Association. These exposure limits (for germicidal wavelengths viz. 0.2537 μm) should not exceed 0.1×10^{-6} W/cm^2 for continuous exposure. If an estimate is to be made of UV laser thresholds then it suggested that the more recent work of Pitts be consulted.

Major works to be consulted on hazard evaluation and classification, control measures, measurement, safety and training programs, medical surveillance and criteria for exposure of the eye and skin to laser radiation are the American National Standards Institute (ANSI) and Bureau of Radiological Health (BRH) documents. Also see the ACGIH document for additional laser, microwave and ultraviolet exposure criteria. A major work on laser safety, soon to be released, is the laser radiation standard of the International Electrotechnical Commission (IEC).

Measurement of Laser Radiation

The complexity of radiometric measurement techniques, the relatively high cost of available detectors and the fact that calculations of radiant exposure levels based on manufacturers' specifications of laser performance have been found to be sufficiently accurate for protection purposes, have all combined to minimize the number of measurements needed in a protective program. In the author's experience, the output power of commonly used laser systems, as specified by the manufacturers, has never been at variance with precision calibration data by more than a factor of two.

All measurement systems are equipped with detection and readout devices. A general description of several devices and their application to laser measurements follow.

Because laser radiation is monochromatic, certain simplifications can be made in equipment design. For example, it may be possible to use narrow band filters with an appropriate type of detector thereby reducing sources of error. On the other hand, special care must be taken with high powered beams to prevent detector saturation or damage. Extremely short Q-switched pulses require the use of ultrafast detectors and short time-constant instrumentation to measure instantaneously power. Photoelectric detectors and radiation thermopiles are designed to measure instantaneous power, but they can also be used to measure total energy in a pulse by integration, provided the instrumental timeconstants are much shorter than the pulse lengths of the laser radiation. High current vacuum photo-diodes are useful for measuring the output of Q-switched systems and can operate with a linear response over a wide range.

Average power measurements of cw lasers systems are usually made with a conventional thermopile or photovoltaic cells. A typical thermopile will detect signals in the power range from 10 μW to about 100 mW. Because thermopiles are composed of many junctions the response of these instruments may be non-uniform. The correct measure of average power is therefore not obtained unless the entire surface of the thermopile is exposed to the laser beam. Measurements of the cw power output of gas lasers may also be made with semiconductor photocells.

The effective aperture or aperture stop of any measurement device used for determining the radiant expose (J/cm^2) or irradiance (W/cm^2) should closely approximate, if not be identical to, the papillary aperture. For purposes of safety the diameter should correspond to that of the normal dark-adapted eye, i.e. 7 mm. The response time of measurement system should be such that the accuracy of the measurement is not affected especially when measuring short pulse durations or instantaneous peak power.

Many calorimeters and virtually all photographic methods measure total energy, but they can also be used for measuring power if the time history of the radiation is known. Care should be taken to insure that photographic processes are used within the linear portion of the film density vs. log radiant exposure (gamma) curve.

Microammeters and voltmeters may be used as read out devices for cw systems; microvoltmeters or electrometers coupled to oscilloscopes may be used for pulsed laser systems. These devices may be connected in turn to panel displays or recorders, as required.

Calibration is required for all wavelengths at which the instrument is to be used. It should be noted that Tungsten Ribbon filament lamps are available from the National Bureau of Standards as secondary standards of spectral radiance over the wavelength region from approximately 0.2–2.6 μm. The calibration procedures using these devices permit comparisons within about 1% in the near UV and about 0.5% in the visible. All radiometric standards are based on the Stefan–Boltzmann and Planck laws of blackbody radiation.

The spectral response of measurement devices should always be specified since the ultimate use of the measurements is a correlation with the spectral response of the biological tissue receiving the radiation insult.

Control of Exposure

In defining a laser hazard control program, some attempt should be made to classify the lasers or laser system according to their potential hazard. For example, one may wish to classify the lasers in terms of their potential for exceeding the Maximum Permissible Exposure (MPE) level or Threshold Limit Values (TLV). This could mean that a classification of "low powered," "exempt" or special "protected" lasers could evolve. "Exempt" may apply to lasers and laser systems which cannot, because of inherent design parameters, emit radiation levels in excess of the MPE; "low powered" could refer to systems emitting levels greater than the MPE for direct exposure to collimated beams but less than the MPE for extended sources; "high powered" could refer to systems that emit levels greater than the MPE for direct exposure to collimated laser beams as well as the MPE for extended sources; a "protected" laser system could be one where by virtue of appropriate engineering controls the emitted levels of radiation are less than any MPE value. Other variations are possible. Once a classification scheme has been established it is possible to devise engineering measures and operating procedures to maintain all radiation at or below the desired levels, the stringency of the controls being directly related to the degree of risk to personnel in each category.

It stands to reason that certain basic control principles apply to many laser systems: the need to inform appropriate persons as to the potential hazard, particularly with the discharge of capacitor banks associated with solid state Q-switched systems, the need to rely primarily on engineering controls rather than procedures, e.g. enclosures, beam stops, beam enlarging systems, shutters, interlocks and isolation of laser systems, rather than sole reliance on memory or safety goggles. The "exempt" laser system is an exception to these measures. In all cases, particular attention must be given to the safety of unsuspecting visitors or spectators in laser areas.

"High powered" systems deserve the ultimate in protective design: enclosures should be equipped with interlocks. Care should be taken to prevent accidental firing of the system and where possible, the system should be fired from a remote position. Controls on the high powered systems should go beyond the usual warning labels by installing an integral warning system such as a "power on" audible signal or flashing light which is visible through protective eye wear.

Infrared laser systems should be shielded with fireproof materials having an appropriate optical density (O.D.) to reduce the irradiance below MPE values. The main hazard of these systems is absorption of excessive amounts of IR energy by human tissue or by flammable or explosive chemicals.

Before protective eye wear is chosen, one must determine as a minimum the radiant exposure or irradiance levels produced by the laser at the distance where the beam or reflected beam is to be viewed, one must know the appropriate MPE value for the laser wavelength and finally one must determine the proper O.D. of protective eyewear in order to reduce levels below the MPE. Likewise, the visible light transmission characteristics should be known because sufficient transmission is necessary for the person using the device to be able to detect ordinary objects in the immediate field of vision.

MICROWAVE RADIATION

Physical Characteristics of Microwave Radiation

Microwave wavelengths vary from about 10 m to about 1 mm; the respective frequencies range from 30 MHz–300 GHz. Certain reference documents, however, define the microwave frequency range as 10 MHz–100 GHz. The region between 10 MHz and the IR is generally referred to as the RF or radiofrequency region.

Certain bands of microwave frequencies have been assigned letter designations by industry; others, notably the ISM (Industrial, Scientific, Medical) frequencies have been assigned by the Federal Communications Commission for industrial, scientific and medical applications.

Source of Microwave Radiation

Microwave radiation is no longer of special interest only to those in communications and navigational technology. Because of the growing number of commercial applications of microwaves, e.g. microwave ovens, diathermy, materials drying equipment, there is widespread interest in the possible new applications as well as an increased awareness of potential hazards. Typical sources of microwave energy are klystrons, magnetrons, backward wave oscillators and semiconductor transmit time devices (impatt diodes). Such sources may operate continuously as in the case of some communications systems or intermittently, e.g. in microwave ovens, induction heating equipment and diathermy equipment or in the pulsed mode in radar systems. Natural sources of RF and microwave energy also exist. For example, peak field intensities of over 100 V/m are produced at ground level by the movement of cold fronts. Solar radiation intensities range from 10^{-18} to 10^{-17} watts per square meter per Hz ($\mathrm{Wm^{-2}Hz^{-1}}$) however, the integrated intensity at the earth's surface for the frequency range of 0.2–10 GHz is approximately 10^{-8} mW/cm^2. This value is to be compared with an average of 10^2 mW/cm^2 on the earth's surface attributable to the entire (UV, visible IR and microwave) solar spectrum.

Biological Effects of Microwave Radiation

The photon energy in RF and microwave radiation is considered to be too low to produce photochemical reactions in biological matter. However, microwave radiation is absorbed in biological systems and ultimately dissipated in tissue as heat. Irradiation of the human body with a power density of 10 mW/cm^2 will result in the absorption of approximately 58 W with a resultant body temperature elevation of 1°C, a value which is considered acceptable from a personal hazard standpoint. By way of comparison, the human basal metabolic rate is approximately 80 W for a person at rest; 290 for a person engaged in moderate work.

Microwave wavelengths less than 3 cm are absorbed in the outer skin surface, 3–10 cm wavelengths penetrate more deeply (1 mm–1 cm) into the skin and at wavelengths from 25–200 cm penetration is greatest with the potential of causing damage to internal body organs. The human body is thought to be essentially transparent to wavelengths greater than about 200 cm. Above 300 MHz the depth of penetration changes rapidly with frequency, declining to millimeter depths at frequencies above 3000 MHz. Above 10 GHz the surface absorption of energy begins to approach that of the IR radiation.

The observed effects of radiofrequency radiation on biological systems seem to depend more on a differential rate of energy deposition than in the case with ionizing radiation where biological effects seem to be related more to energy and integral (time independent) quantities, such as absorbed dose.

The National Council on Radiation Protection and Measurements (NCRP) has attempted to consolidate the many quantities and units used to describe absorption of radio frequency electromagnetic energy by introducing the term "specific absorption rate" (SAR). The specific absorption rate is the rate at which electromagnetic energy is absorbed at a point in a medium per unit mass of the medium, and is expressed in W/kg. Energy absorption is a continuous and differentiable function of space and time and one may speak of its gradient and its rate, hence the time derivative of the incremental energy (dW) absorbed in an incremental mass (dm) contained in a volume element (dV) of a given density (ρ) may be expressed:

$$\mathrm{SAR} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}W}{\mathrm{d}m}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}w}{\rho\,\mathrm{d}V}\right).$$

Carpenter and Van Ummersen (1968) investigated the effects of microwave radiation on the production of cataracts in rabbit eyes. Exposures to 2.45 GHz radiation were made at power densities ranging from 80–400 mW/cm^2 for different exposure times. They found that repeated doses of 67 J/cm^2 spaced a day, a week, or 2 weeks apart produced lens opacities even though the single threshold exposure dose at that power density (280 mW/cm^2) was 84 J/cm^2. When the single exposure dose was reduced to 50 J/cm^2 opacities were produced when the doses were administered 1 or 4 days apart, but when the interval between exposures was increased to 7 days no opacification was noted even after 5 such weekly exposures. At the low power density of 80 mW/cm^2 (dose of 29 J/cm^2) no effect developed but when administered daily for 10 or 15 days cataracts did develop. The conclusion is that microwaves may exert a cumulative effect on the lens of the eye if the exposures are repeated sufficiently often. The interval between exposures is an important factor in that a repair mechanism seems to act to limit lens damage if adequate time has elapsed between exposures.

Certain other biological effects of microwave radiation have been noted in literature. One of these is the so-called "pearl chain effect" where particles align themselves in chains when subjected to an electric field. There is considerable disagreement as to the significance of the pearl chain effect.

Investigators at the Johns Hopkins University have suggested a possible relationship between mongolism (Down's Syndrome) in offspring and previous exposure of the male parent to radar. This suggested relationship was based on the finding that of 216 cases of mongolism, 8.7% of the fathers having mongol offspring vs. 3.3% of the control fathers (no mongol offspring) had contact with radar while in military service. This possible association must be regarded with extreme caution because of many unknown factors including the probability of a variety of exposures to environmental agents (including ionizing radiation) while in military service.

Soviet investigators claim that microwave radiation produces a variety of effects on the central nervous system and without a temperature rise in the organism. Claims are also made for biochemical changes, specifically a decrease in cholinesterase and changes in RNA at power density levels of approximately 10 mW/cm^2. The reported microwave effects on the central nervous system usually describe initial excitatory action, e.g. high blood pressure followed by inhibitory action and low blood pressure over the long term. Electroencephalographic data have been interpreted as indicating the presence of epileptiform patterns in exposed subjects. Other reported effects ranged from disturbances of the menstrual cycle to changes in isolated nerve preparations.

Field interactions with brain tissue in cats have been assessed by effects on calcium ion fluxes. Increases in calcium efflux of the order of 20% have been reported under conditions of direct stimulation of synaptic terminals. Moreover, exposure of intact animals (cats) to a 450 MHz 0.375 mW/cm^2 field, amplitude modulated at 16 Hz produced a sharp rise in calcium efflux, with a response curve identical to that obtained by direct electrical stimulation of brain tissue at the same intensity.

In addition, power and frequency "windows" have been reported, that is enhanced biological responses have been elicited within narrow bands of incident power and radiation frequency.

What is often overlooked in any description of the biological effects of microwave radiation is that such radiations have produced beneficial effects. Controlled or judicious exposure of humans to diathermy or microthermy is widely practiced. The localized exposure level in diathermy may be as high as 100 mW/cm^2.

Exposure Criteria

Schwan in 1953 examined the threshold for thermal damage to tissue, notably cataractogenesis. The power density necessary for producing such changes was approximately 100 mW/cm^2 to which he applied a safety factor of 10 to obtain a maximum permissible exposure level of 10 mW/cm^2. This number has been subsequently incorporated into many official standards. The current American National Standards Institute C95 standard requires a limiting power density of 10 mW/cm^2 for exposure periods of 0.1 hr or more; also an energy density of 1 milliwatt-hour per square centimeter (1 mWh/cm^2) during any 0.1 hr period is permitted. The latter criterion allows for intermittency of exposure at levels above 10 mW/cm^2, on the basis that such intermittency does not produce a temperature rise in human tissue greater than 1°C. More recently, Schwan has suggested that the permissible exposure levels be expressed in terms of current density, especially when dealing with measurements in the near or reactive field where the concept of power density loses its meaning. He suggests that a permissible current density of approximately 3 mA/cm^2 be accepted since this value is comparable to a far field value of 10 mW/cm^2. At frequencies below 10 100 KHz this value should be somewhat lower and for frequencies above 1 GHz it can be somewhat higher.

The most recent proposal of the American National Standards Institute (ANSI) specifies a frequency dependent criterion, with a minimal elvel of 1 mW/cm^2 in the so-called resonant frequency range of the human body (approximately tens of MHz to several hundred MHz) and higher permissible levels at lower and higher frequencies.

The performance standard for microwave oven specifies a level of 1 mW/cm^2 at any point 5 cm or more from the external oven surfaces at the time the oven is fabricated by manufacturer. 5 mW is permitted throughout the useful life of the oven.

Because Soviet investigators believe that effects on the central nervous system are more appropriate measure of the possibly detrimental effects of microwave radiation than are thermally induced responses, their studies have reported "thresholds" which are lower than those reported in Western countries. Soviet permissible exposure levels are several orders of magnitude below those in Western countries.

The Soviet Standards for whole body radiation are as follows: 0.1 mW/cm^2 for 2 hr exposure per day and 1 mW/cm^2 for a 15–20 min exposure provided protective goggles are used. These standards apply to frequencies above 300 MHz. Recent reports indicate that the Soviet Union has raised the above mentioned value of 0.01 mW/cm^2 to 0.025 mW/cm^2; also, the Soviet value of 0.001 mW/cm^2 for continuous exposure of the general population has been raised to 0.005 mW/cm^2.

There appears to be no serious controversy about the power density levels necessary to produce thermal effects in biological tissue. The nonthermal CNS effects reported by the Soviets are not so much controversial as they are a reflection of the fact that Western investigators have not used the conditioned reflex as an end point in their investigations.

Measurement of Microwave Radiation

Perhaps the most important factor underlying some of the controversy over biological effects is the lack of standardization of the measurement techniques used to quantify results. To date, unfortunately, there seems to be little promise that such standardization will be realized in the near future.

The basic vector components in any electromagnetic wave are the electric field (*E*) and the magnetic field (*H*). The simplest type of microwave propagation consists of a plane wave moving in an unbounded isotropic medium, where the electric and magnetic field vectors are mutually perpendicular to each other and both are perpendicular to the direction of wave propagation. Unfortunately the simple proportionality between the *E* and *H* fields is valid only in free space, or in the so-called "far field" of the radiating device. The far field is the region which is sufficiently removed from the source to eliminate any interaction between the propagated wave and the source. The energy or power density in the far field is inversely proportional to the square of the distance from the source and in this particular case the measurement of either *E* of *H* suffices for their determination.

Plane-wave detection in the far field is well understood and easily obtained with equipment which has been calibrated for use in the frequency range of interest. Most hazard survey instruments have been calibrated in the far field to read in power density (mW/cm^2) units. The simplest type of device uses a horn antenna of appropriate size coupled to a power meter.

To estimate the power density levels in the near field of large aperture circular antennas one can use the following simplified relationship

$$W = \frac{16P}{\pi D^2} = \frac{4P}{A}\left(\text{near field}\right),$$

where *P* is the average power output, *D* is the diameter of the antenna, *A* is the effective area of the antenna and *W* is power density. If this computation reveals a power density which is less than a specified limit, e.g. 10 mW/cm^2, then no further calculation is necessary because the equation give the maximum power density on the microwave beam axis. If the computed value exceeds the exposure criterion then one assume that the calculated power density exists through-out the near field. The far field power densities are then computed from the Friis free space transmission formula

$$W = \frac{GP}{4\pi r^2} = \frac{AP}{\lambda^2 r^2}\left(\text{far field}\right),$$

where λ is the wavelength, r is the distance from the antenna and G is the far field antenna gain.

The distance from the antenna to the intersection of the near and far fields is given by

$$r_1 = \frac{\pi D^2}{8\lambda} = \frac{A}{2\lambda}.$$

These simplified equations do not account for reflections from ground structures or surfaces; the power density may be four times greater than the free space value under such circumstances.

Special note should be made of the fact that microwave hazard assessments are made on the basis of average, not peak power of the radiation. In the case of radar generators, however, the ratio of peak to average power may be as high as 10^5.

Most microwave measuring devices are based on bolometry, calometry, voltage and resistance changes in detectors and the measurement of radiation pressure on a reflecting surface. The latter three methods are self-explanatory. Bolometry measurements are based upon the absorption of power in a temperature sensitive resistive element, usually a thermistor, the change in resistance being proportional to absorbed power. This method is one of the most widely used in commercially available power meters. Low frequency radiation of less than 300 MHz may be measured with loop or short ship antenna. Because of the larger wavelengths in the low frequency region, the field strength in volts per meter (V/m) is usually determined rather than power density.

One troublesome fact in the measurement of microwave radiation is that the near field (reactive field) of many sources may produce unpredictable radiative patterns. Energy density rather than power density may be a more appropriate means of expressing hazard potential in the near field. In the measurement of the near field of microwave ovens it is desirable that the instrument have certain characteristics, e.g. the antenna probe should be electrically small to minimize perturbation of the field, the impedance should be matched so that there is no backscatter from the probe to the source, the antenna probe should behave as an isotropic receiver, the probe should be sensitive to all polarizations, the response time should be adequate for handling the peak to average power of the radiation and the response of the instrument should be flat over a broad band of frequencies.

In terms of desirable broad band characteristics of instruments it is interesting to note that one manufacturer has set target specifications for the development of a microwave measurement and monitoring device as follows: frequency range 20 KHz–12.4 GHz and a power density range of 0.02–200 mW/cm^2 ± 1 dB. Reportedly two models of this device will be available: one a hand held version complete with meter readout, the other a lapel model equipped with audible warning signals if excessive power density levels develop.

APPENDIX A

Useful radiometric and related units

Term	Symbol	Description	Unit and abbreviation
Radiant energy	O	Capacity of electromagnetic wages to perform work	Joule (J)
Radiant power	P	Time rate at which energy is emitted	Watt (W)
Irradiance or radiant flux density (dose rate in photobiology)	E	Radiant flux density	Watt per square meter ($W \cdot m^{-2}$)
Radiant intensity	I`	Radiant flux of power emitted per solid angle (steradian)	Watt per steradian ($W \cdot sr^{-1}$)
Radiant exposure (dose in photobiology)	H	Total energy incident on unit area in a given time interval	Joule per square meter ($J \cdot m^{-2}$)
Beam divergence	ϕ	Unit of angular measure. One radian ≈ 57.3° 2π radians = 360°	Radian

APPENDIX B

Conversion factors A−radiant energy units

	erg	joule	W/sec	± W/sec	g-cal
erg =	1	10^{-7}	10^{-7}	0.1	2.39×10^{-8}
10 joule =	1	1	10^{6}	0.239	
W sec =	107	1	1	10^{4}	0.239
±W sec =	10	10^{-6}	10^{-6}	1	2.39×10^{-7}
g-cal =	4.19×10^{7}	4.19	4.19	4.19×10^{6}	1

B−radiant exposure (dose) units

	erg/cm^2	joule/cm^2	W/sec cm^2	±W/sec cm^2	g-cal/cm^2
erg cm^2 =	10^{-7}	10^{-7}	0.1	2.39×10^{-8}	
joule cm^2 =	10^{7}	1	1	10^{6}	0.239
W sec cm^2 =	10^{7}	1	1	10^{6}	0.239
±W sec cm^2 =	10	10^{-6}	10^{-6}	2.39×10^{-7}	
g−cal cm^2 =	4.19×10^{7}	4.19	4.19	4.19×10^{6}	1

C-irradiance (dose rate) units

	erg/cm^2 · sec	joule/cm^2 · sec	W/cm^2	±W/cm^2	g−cal/cm^2 · sec
erg/cm^2 · sec =	1	10^{-7}	10^{-7}	0.1	2.39×10^{-6}
joule cm^2· sec =	10^{7}	1	1	10^{6}	0.239
W/cm^2 =	10^{7}	1	1	10^{6}	0.239
±W/cm^2 =	10	10^{-6}	10^{-6}	1	2.39×10^{-7}
g-cal/cm^2 · sec =	4.19×10^{7}	4.19	4.19	4.19×10^{6}	1

A tabular summary of typical characteristics of instrumentation used for electromagnetic field measurements is available in an NCRP report.

Control Measures

The control of excessive exposures to microwave radiation is basically an engineering matter. The engineering measures may range from the restriction of azimuth and elevation settings on radar antennas to complete enclosures of magnetrons in microwave ovens. The use of personnel protective devices have their place but are of much lower priority importance to engineering controls. Various types of microwave protective suits, goggles and mesh have been used for special problems.

It has been shown that cardiac pacemakers, particularly those of the demand type, may have their function compromised by microwave radiation. Furthermore, the radiation levels which cause interference with the pacemaker may be orders of magnitude below levels which cause detrimental biological effects. The most effective method of reducing the susceptibility of these devices to microwave interference seems to be improved shielding. Manufacturers of cardiac pacemakers "… have successfully redesigned and shielded the units so that they can adequately withstand power densities of at least 10 mW/cm^2 without interference with their function."

PREFERRED READING

1. Clarke, A.M. (1970), "Ocular Hazards from Lazers and other Optical Sources," *CRC Critical Reviews in Environmental Control,* **1**, 307.
2. Cleary, S.F. (1970), "The Biological Effects of Microwave and Radiofrequency," *CRC Critical Reviews in Environmental Control,* **1**, 257.

REFERENCES

1. Matelsky, I., The non-ionizing radiations, *Industrial Hygiene Highlights* **1**, Indus, Hygiene Foundation of America Inc. Pittsburgh, Pa., 1968.
2. *Ibid.* p. 145.
3. *Ibid.* p. 149.
4. Cogan, D.G. and V.E. Kinsey (1946), Action spectrum of keratitis produced by ultraviolet radiation, *Arch. Ophthal.*, **35**, 670.
5. Verhoeffr, F.H. and L. Bell (1916), Pathological Effects of Radiant Energy on the Eye, *Proc. Amer. Acad. Arts and Sci.*, **51**, 630.
6. Pitts, D.G., J.E. Prince, W.I. Butcher, K.R. Kay, R.W. Bowman, H.W. Casey, D.G. Richey, L.H. Mori, J.E. Strong, and T.J. Tredici, The effects of ultraviolet radiation on the eye, *Report SAM-TR*-69-10, USAF School of Aerospace Medicine, Brooks AFB, Texas, Feb., 1969.
7. Pitts, D.G. and K.R. Kay (1969), The photophthalmic threshold for the rabbit, *Amer. J. Optom.*, **46**, 561.
8. Permissible limit for continuous ultraviolet exposure, *Council on Physical Therapy, Amer. Med. Assn.,* Chicago, 1948.

9. Bulletin No. 3, The Eppley Laboratory Inc., Newport, Rhode Island, 1963.
10. Richardson, J.R. and R.D. Baertsch (1969), Zinc sulfide schottky barrier ultraviolet detectors, *Solid State Electronics* **12**, 393.
11. Javan, A., W.R. Bennett, and D.R. Herriott (1961), Population inversion and continuous optical maser oscillation in a gas discharge containing a He-Ne mixture, *Phys. Rev. Lett.,* **6**, 106.
12. Miller, R.C. and W.A. Nordland (1967), Tunable Lithium Niobate Optical Oscillator with external mirrors, *Appl. Phys. Lett.,* **10**, 53.
13. Ham, W.T., R.C. Williams, H.A. Muller, D. Guerry, A.M. Clarke, and W.J. Geeraets (1965), Effects of laser radiation on the mammalian eye, *Trans. N.Y. Acad. Sci.,* **28**, 517.
14. Clarke, A.M., W.T. Ham, W.J. Geeraets, R.C. Williams, and H.A. Mueller (1969), Laser Effects on the eye, *Arch. Environ. Health,* **18,** 424.
15. Noell, W.K., V.S. Walker, B.S. Kang, and S. Berman (1966), Retinal damage by light in rats, *Invest. Ophthal.,* **5**, 450.
16. Kotiaho, A., I. Resnick, J. Newton, and H. Schwell (1966), Temperatures rise and photocoagulation of rabbit retinas exposed to the CW Laser, *Amer. J. Ophthal.,* **62**, 644.
17. Davis, T.P., and W.J. Mautner (1969), Helium–neon laser effects on the eye, *Annual Report Contract No. DADA* 17–69-C-9013, US Army Medical Research and Development Command, Wash., DC.
18. McNeer, K.W., M. Ghosh, W.J. Geeraets, and D. Guerry (1963), Erg after light coagulation, *Acta. Ophthal. Suppl.* **76**, 94.
19. Jons, A.E., D.D. Fairchild, and P. Spyropoulos (1968), Laser radiation effects on the morphology and function of ocular tissue, *Second Annual Report, Contr. No. DADA*-17–67-C-0019, US Army Medical Research and Development Command, Wash., DC.
20. Safety level of microwave radiation with respect to personnel, committee C95–1 USA Stds Inst. (Now Amer. Natl. Stds. Inst.) New York, N.Y., 1966.
21. Mumford, W.W. (1969), Heat stress due to R.F. radiation, *Proceedings of IEEE,* **57,** 171.
22. Carpenter, R.L. and C.A. Van Ummersen (1968), *J. Microwave Power,* **3**, 3.
23. Sigler, A.T., A.M. Lillienfeld, B.H. Cohen, and J.E. Westlake (1965), Radiation exposure in parents of children with mongolism (Down's Syndrome), *Johns Hopkins Hosp. Bull.,* **117**, 374.
24. Tolgskaya, M.S., and Z.V. Gordon (1960), *Trans. Inst. of Labor Hygiene and Occupational Diseases of the Acad. of Med. Sci.,* **99**.
25. Orlova, A.A. (1959), *Proc. on Labor Hygiene and the Biol. Effects of Electromagnetic Radio Frequency Waves* 25.
27. Kholodov, Yu. A. (1962), *Proc. on Problems of the Biological Effects of Superhigh Frequency Fields,* **58**.
28. Schwan, H.P. and K. Li (1953), *Proc. IRE,* **41**, 1735.
29. Schwan, H.P. (1970), *Biol. Effects and Health Implications of Microwave Radiation,* US Govt. Printing Office.
30. USA Standards Institute (Now American Technical Standards Institute) C95.3, *Specifications for Techniques and Instrumentation for Evaluating Radio Frequency Hazards to Personnel,* New York, NY., 1968.
31. Wacker, P. (1970), *Biol. Effects and Health Implication of Microwave Radiation,* US Govt. Printing Office.
32. Bowman, R., *Ibid.*
33. King, G.R., A.C. Hamburger, F. Parsa, S.J. Heller, and R.A. Carleton (1970), Effect of microwave oven on implanted cardiac pacemaker, *Jama,* **312**, 1213.
34. Wilkening, G.M., The Potential Hazards of Laser Radiation, *Proceedings of Symposium on Ergonomics and Physical Environmental Factors,* Rome, Italy, 16–21, September 1968, International Labor Office, Geneva.
35. Sliney, D.H., F.C. Bason, and B.C. Freasier (1971), Instrumentation and measurement of ultraviolet, visible and infrared radiation, *Amer. Indus. Hygiene Assn. J.,* **32**, No. 7, July.
36. Cleary, S.F. (1970), The biological effects of microwave and radiofrequency radiation, *CRC Critical Review in Environmental Control* **1**, 257.
37. Threshold Limit Values for Physical Agents (1982), American Conference of Governmental Industrial Hygienists, P.O. Box 1937, Cincinnati, Ohio 45201.
38. Wilkening, G.M. (1978), Nonionizing Radiation Chapter 11, Patty's Industrial Hygiene and Toxicology, **1**, General Principles, John Wiley and Sons, 3rd Rev. Ed., New York, NY.
39. Charschan, S. *et al.,* American National Standards for the Safe Use of Lasers Z136.1 (1980), American National Standards Institute, Inc. 1430 Broadway, New York, NY 10018.
40. Laser Products Performance Standard (1975), US Department of Health and Human Services, Food and Drug Administration, Bureau of Radiological Health, Rockville, Maryland.
41. Wilkening, G.M. *et al.*, Laser Radiation Standard (1982), Technical Committee on Laser Products, International Electrotechnical Commission (IEC), Geneva, Switzerland.
42. Wilkening, G.M. *et al.* (1981), Radiofrequency Electromagnetic Fields, Properties, Quantities and Units, Biophysical Interaction and Measurements National Council on Radiation Protection and Measurements, Report No. 67, Washington, DC, 20014.
43. Adey, W.R. (1979), Neurophysiologic Effects on Radiofrequency and Microwave Radiation, Bulletin of the New York Academy of Medicine **55**, 1079.
44. Guy, A.W. *et al.* (1982), "Safety Level of Electromagnetic Radiation with Respect to Personnel" American National Standards Institute Inc., 1430 Broadway NY, NY 10018.
45. Nonionizing Radiation Protection, 2nd Ed. WHO, Copenhagen, 1989.

GEORGE M. WILKENING (DECEASED)
Bell Telephone Laboratories

O

OCEANOGRAPHY

INTRODUCTION

Before beginning a discussion of the origin and structure of the world's oceans, one should first reflect on the reasons for our interest in these bodies of water, which together comprise approximately seventy percent of the earth's surface area. Perhaps one of the more obvious reasons is the enormous wealth of natural resources hidden in the oceans' depths, including vast reserves of food and energy sources. This abundance has of course led to an increased human presence in and along much of the world's oceans, prompting considerable concern for the short and long term consequences of man's activities in this fragile environment.

An additional focus of recent oceanographic research has been the role of the sea in global climate variations, spurred by the recognition that these variations are driven largely by fluctuations in the air-sea momentum and heat exchange. This role was dramatically illustrated in the 1982–1983 El Nino event, which sparked such wide-ranging weather aberrations as flooding in southern California and droughts in Australia, and is widely believed to have had an oceanic origin (Rienecker and Mooers, 1986).

In addition to these varied and multi-disciplinary research activities, the increasing strategic importance of the world's oceans has been one of the prime motivations behind recent advances in underwater communication, depth measurement, and perhaps most significantly, the remote sensing of ocean physical processes. In the political forum, this strategic importance, along with the economic benefits of the natural resources mentioned earlier, continues to provoke international debate over such issues as fisheries management, ocean dumping, and the demarcation of territorial ("exclusive zone") boundaries.

Clearly, therefore, the oceans mean many different things to many different people. The common thread is the need and desire to learn more about these fascinating and fundamentally important bodies of water. In the following, we shall attempt to provide some insight into the makeup and behavior of the sea, from its origins to its various modes of motion. Realizing the short-comings of an overview treatment of such complex subject matter, the interested reader is urged to pursue in more detail any of the specific topics addressed in this article.

ORIGIN AND STRUCTURE OF THE OCEANS

It is generally believed that most of the water now present on the surface of the earth originated in the earth's interior. The prevailing scientific opinion holds that the earth was formed through the gradual compaction of an accumulation of particles, primarily silicon compounds, and iron and magnesium oxides. The heat released as a result of this compaction and radioactive decay was sufficient to produce a molten mass, the heaviest material sinking under the action of gravity to form the earth's core and the lightest "floating" to the surface to form the crust. This formative period was characterized by intense volcanic activity. It is widely accepted that the molten material introduced to the earth's surface through this activity served as the conduit for the release of hydrogen and oxygen into the atmosphere (primarily in the gaseous form) with some eventually condensing to form liquid water at the surface. Sediments discovered in Australia in 1980 appear to be of marine origin and have been dated at approximately 3.5 billion years old (Ross, 1982), indicating that surface water has been present over a significant portion of the earth's history.

The horizontal movements that produced the present-day sea-floor are, in geologic terms, a very recent occurrence. In fact, measurements indicate that the most ancient regions of the present deep ocean floor are no more than 225 million years old (Gross, 1977). The oldest dated sediments from the Atlantic Ocean indicate an age of approximately 165 million years (Sclater and Tapscott, 1979), while those of the Pacific reveal a relatively "young" water body at approximately 120 million years old (Heezen and MacGregor, 1973).

The mechanism for the creation and growth of the ocean basins is commonly referred to as continental drift, or in more general terms, as plate tectonics. The basic concept can be explained as follows. The outer shell of the earth (the lithosphere) is actually composed of a number of rigid plates.

These plates are in constant motion relative to each other. As one would expect, the borders between adjoining plates are regions of intense seismic activity.

The visible manifestation of this activity varies according to its strength, and ranges from seismically inactive mountain belts to active volcanic chains. Along areas where plates are diverging ("ridge axes"), there is a constant formation of new surface area, as volcanic material from the earth's interior rises to fill the gap caused by the plate divergence. Quite logically then, and in light of the fact that the earth is undergoing little or no expansion, there must be zones where plates are converging. These areas are called subduction zones and are characterized by the sliding of one plate beneath the edge of the adjoining plate. The theory of seafloor spreading states that the ocean basins were formed, and are continuing to change at a rate of order 2 centimeters per year (EOS, 1988) as a result of the divergence of plates along axes called oceanic ridges. As an example of this phenomenon, as recently as 165 million years ago, the continents bordering what is now the Atlantic Ocean were very much more closely separated, possibly constituting one very large land mass. The plates on which these continents rest then began diverging along an axis known as the Mid-Atlantic ridge, creating the ocean basin that presently exists. This ridge runs along the approximate centerline of the Atlantic Ocean, from Iceland south to approximately 1800 kilometers north of Antarctica.

For the purpose of categorization, we can divide the world's oceans into three bodies: the Atlantic Ocean (including the Arctic, Baltic, and Mediterranean Seas), bordered by the Americas, Africa, Europe, and the Arctic land mass; the Pacific Ocean, bordered by the Americas, Asia, Australia, and Antarctica; and the Indian Ocean, bordered by Africa, Asia, Australia, and Antarctica.

Table 1 indicates the area and average depth of the three ocean basins (including adjacent seas). Note that the Pacific Ocean encompasses the largest area and has the largest average depth of the three. We should caution that this separation of the oceans into three, distinct bodies of water is somewhat misleading, since the adjacent seas (e.g., the Mediterranean and Arctic) are often quite different from the major ocean basins in physical, chemical and biological characteristics. As partial evidence of this point, we also list in Table 1 the area and depth of the three ocean basins, excluding the adjacent seas. Note that in comparison with the Pacific and Indian Oceans, a much larger percentage of the Atlantic Ocean's assigned area consists of smaller seas. The neglect of these seas results in a more uniform average depth among the three major oceans, although the Pacific remains the largest and deepest on average.

In general, if one were to move seaward from the boundary between continent and ocean, one would first encounter a continental shelf, characterized by relatively shallow water with depth gradually increasing in the seaward direction to a maximum of the order to 200 meters. The continental slope would then be encountered, representing the transition structure from shallow to deep ocean, and characterized by an abrupt increase in water depth to the order of 3000 meters. One typically finds that the water depth will continue to increase away from the continent (although at a slower rate) to maximum depths of between 5000 and 6000 meters until reaching the spreading centers, or mid-ocean ridges discussed earlier. The water depth generally decreases toward the ridges to the order of 2000 meters.

This pattern of increasing depth away from the continents, followed by decreasing depth as the ridge axis is approached appears at first glance to be a contradiction to the theory of ocean basin formation explained earlier. If we are to accept the notion of a mid-ocean ridge, or zone of divergence, at which new ocean crust is continually being uplifted and forced out on either side of the ridge axis, we would expect to find shallow water depths along these axes, and symmetrically increasing depths as one approaches the continental land masses on either side. The reason for this apparent contradiction is the process of sedimentation, the deposition of land-derived sediments along the ocean floor. As one would expect, ocean regions most closely bordering the continents experience the highest rates of sedimentation, both through wind-driven atmospheric input as well as the more dominant mechanism of water-borne inputs at the coast. Rates of sediment accumulation in these regions vary considerably, but have typical magnitudes of several hundred meters (vertically) per million years on most continental shelves. The rate of sedimentation decreases by an order of magnitude on the continental slope, to tens of meters per million years. Finally, there is in general very little or negligible deposition in the deep ocean. Figure 1 illustrates a typical cross-section of the ocean floor as it exists today; the product of the combined actions of seafloor spreading and sedimentation.

SEA WATER PROPERTIES

As a forcing mechanism for both deep water and coastal motions, a tracer for the identification of water movements, and a critical parameter in the acoustic characteristics of ocean waters, density, ρ, is perhaps the most important seawater property to ocean engineers and scientists. In fact, many oceanographic calculations require a knowledge of the water density to an accuracy of 6 significant digits! In practice, the

TABLE 1
Area and average depth of major ocean basins

Ocean Basin	Area (10^6 km^2)	Average Depth (meters)
	Including Adjacent Seas	
Indian	74.917	3897
Atlantic	106.463	3332
Pacific	179.679	4028
	Excluding Adjacent Seas	
Indian	73.443	3963
Atlantic	82.441	3926
Pacific	165.246	4282

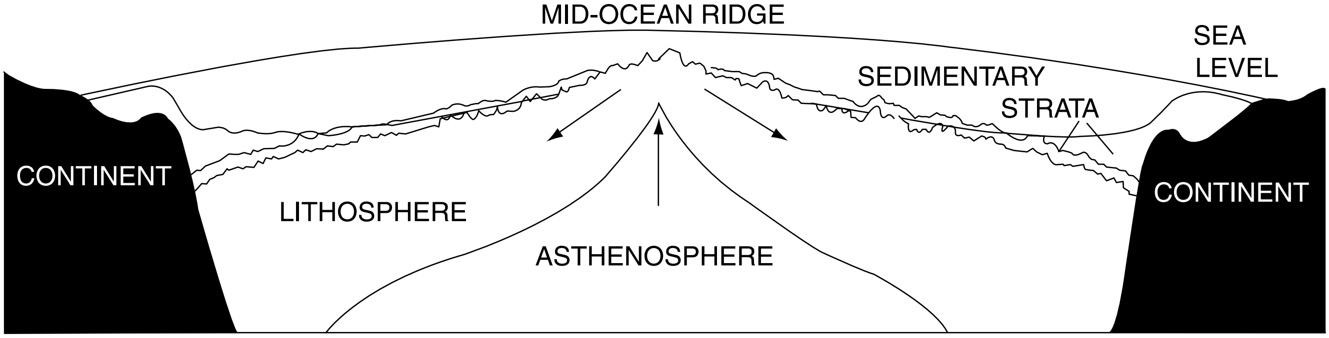


FIGURE 1 Typical seafloor cross-section (Sclater and Tapscott, 1979).

density is often related to the water salinity, S, and temperature, T, through an equation of state (e.g., Fofonoff, 1962).

In many oceanographic applications, the water density is expressed in terms of a sigma-t value, σ_t, which is a measure of the density at atmospheric pressure, $p = 0$. Furthermore, since very slight variations in water density can have significant physical, chemical and biological implications, the sigma-t values are tabulated in the form: $\sigma_t, = (\rho - 1) \times 10^3$. As an example, the density of sea water at 20°C and salinity 35% (35 parts per thousand) at atmospheric pressure is not written as $\rho = 1.024785$ g/cm^3, but rather as $\sigma_t = 24.785$.

The temperature distribution in the ocean, although highly variable in space and time, is generally characterized by a layer of relatively warm water at the surface, underlain by dense, colder water. Variations in deep ocean temperature, that is, excluding shallow coastal areas, can exceed 25°C, with a typical range of 20°C at the surface to 2°C near the bottom. These values exclude isolated aberrations such as the very saline, 56°C water found at the bottom of the Red Sea, and the hot springs ("black smokers") located along the mid-ocean ridges of the eastern Pacific, where fluid having temperatures as high as 350°C seeps through cracks in the diverging ridge axis (Edmond, 1986).

The dense bottom water in most cases originates in the polar regions, as we shall discuss later. Typically, the temperature of the surface layer is strongly dependent on surface atmospheric conditions (i.e., air temperature, solar intensity and wind speed). In the absence of strong vertical mixing, the temperature in this region decreases with depth in a region commonly referred to as the thermocline. Below the thermocline, the temperature is usually significantly more uniform with depth. Thermocline development varies in depth and duration with latitude. In the temperate regions, the thermocline is seasonal, developing only during the warmer months of the year, and achieving an average depth on the order of 100 meters. In the tropical regions, the thermocline is a relatively constant feature, with depths often exceeding 1000 meters. In both regions, periods of strong surface winds can induce vertical turbulent mixing sufficient to break down the thermocline and create a surface layer of uniform density exceeding several hundred meters in depth.

The salinity of the world's oceans is, in general, much less variable than the temperature, typically ranging from 33 to 37 parts per thousand. Exceptions include coastal regions, where rainfall and river runoff can add appreciable amounts of fresh water to the nearshore waters and reduce the salinity. In addition, tropical ocean regions tend toward higher salinity as a result of the high evaporation (which removes only fresh water) and low precipitation (which adds fresh water) rates relative to temperate ocean areas. Large variations in salinity can also occur in the polar regions. Ice formation, through its extraction of very nearly fresh water from the underlying ocean surface, results in the deposition of brine at the surface and the creation of a layer of very cold, saline water. Conversely, during ice melting, large amounts of fresh water are introduced to the surface layer, considerably reducing the local salinity.

OCEAN CIRCULATION

For the purposes of this discussion, we shall divide the currents that together comprise the oceanic circulation into two components, each distinguished by the forcing mechanism responsible for the water motion: 1) Wind-driven surface currents, and 2) Density-driven deep ocean currents. We shall follow these discussions with a brief examination of shallow water coastal circulation.

Wind-Driven Currents

The surface currents that describe much of the wind-driven ocean circulation are familiar to most readers. On the east coast of North America, the Gulf Stream carries warm water from off the coast of Florida northward to Nova Scotia, then running northeast and becoming the North Atlantic Current. In the Pacific Ocean, the Kuroshio carries warm water from a region south of Japan in a northeasterly direction, becoming the North Pacific Current.

Before discussing these and other features of the wind-driven ocean circulation, we should first examine the characteristics of the atmospheric circulation. Of course, this motion varies considerably in space and time, making any prediction of global synoptic wind patterns an extremely difficult, if not impossible, task. We can, however, derive some useful information by averaging the wind field observed at each latitude circle over many years. Figure 2 illustrates the resulting, long-term averaged wind characteristics. The following patterns are apparent:

1) Between latitudes 30°N and 30°S, there is a zone of easterly (from the east) winds, with an equatorward component, commonly referred to as the "Trade Winds".
2) Between latitudes 30°N and 60°N, and 30°S and 60°S, there exists zones of westerly winds, each with a poleward component.
3) Near each pole, there exists a zone of easterly winds, each with an equatorward component.

We note that the Trade Winds are a fairly consistent feature of the overall wind field, whereas the winds in the northern and southern zones are highly variable and often associated with large, transient storm systems. For this reason, the zones of westerly winds north and south of the Trade Wind zone are often referred to as the "Roaring Forties". Along the equator, at the convergence of the two components of the Trade Winds, lies a region of relatively light winds, commonly referred to as the "Doldrums".

Note that at the wind divergence located at 30°N and the wind convergence at 60°S, we can expect regions of anticyclonic (clockwise) wind gyres, whereas in the vicinity of 60°N and 30°S, we expect cyclonic (anticlockwise) gyres, as illustrated in Figure 2.

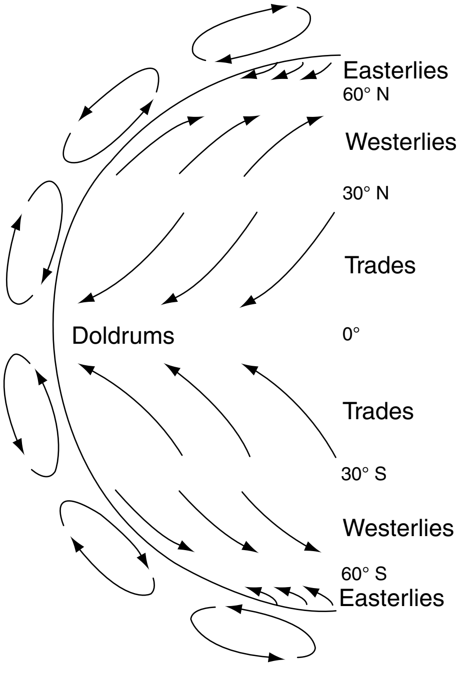


FIGURE 2 Long-term averaged wind pattern (Ross, 1982).

The long-term averaged surface circulation of the world's oceans bears a strong resemblance to the above described pattern of atmospheric motion, with some modifications due to the influence of the continental landmasses. The main features of this circulation are illustrated in Figure 3. Note especially the presence of wind-induced surface current gyres, as expected from our description of the atmospheric circulation.

We here depart from what has been up to now a purely descriptive treatment of the wind-driven motion, to examine the equations governing this air-sea interaction. For simplicity, and in the interest of generality, we shall here confine ourselves to the problem of a steady (constant in time), uniform (constant in space) surface wind blowing over an infinitely wide, infinitely deep, constant density ocean. For a much more detailed treatment of the complexities of upper ocean wind-driven dynamics, the reader is referred to Price *et al.* (1987).

In physical terms, our problem specification corresponds to a very large, relatively stationary storm system acting on the deep ocean (that is, excluding shallow, continental shelf areas).

The primary forcings acting on our water column, or vertical slice of the water body, are: 1) the surface wind stress; 2) the internal, turbulent shear stresses; and 3) the Coriolis "force".

The surface wind stress will be represented as a steady, uniform shear stress acting in the x-direction: τ^S. The internal turbulent shear stresses, which dominate over the viscous shear stresses, can be defined in a manner analogous to the viscous stresses, so that:

$$\tau_x = \rho A_v \frac{du}{dz},$$

$$\tau_y = \rho A_v \frac{dv}{dz},$$

where our coordinate system is defined in Figure 4. The quantities τ_x, and τ_y represent the horizontal (x and y) components of the turbulent shear stress. The quantities u and v are the x and y components, respectively, of the water velocity. Note that u and v are functions only of the vertical coordinate, z, because of our assumption of a steady, horizontally uniform wind stress. The constant, A_v, represents the turbulent eddy viscosity coefficient. Strictly speaking, A_v is a function of the turbulent flow field and should be modelled accordingly as another unknown in our system of equations (see, e.g., Blumberg and Mellor, 1983). In the interest of obtaining an analytic solution to the governing equations and so obtain useful information about the gross features of the wind-driven water motion, we shall here set A_v as constant. The quantity, ρ, is the (constant) water density.

The Coriolis "force" is not, strictly speaking, a force, but is rather the result of applying Newton's second law to the earth's rotating reference frame. In physical terms, this "force" causes a deflection of motion to the right (left) in the northern (southern) hemisphere.

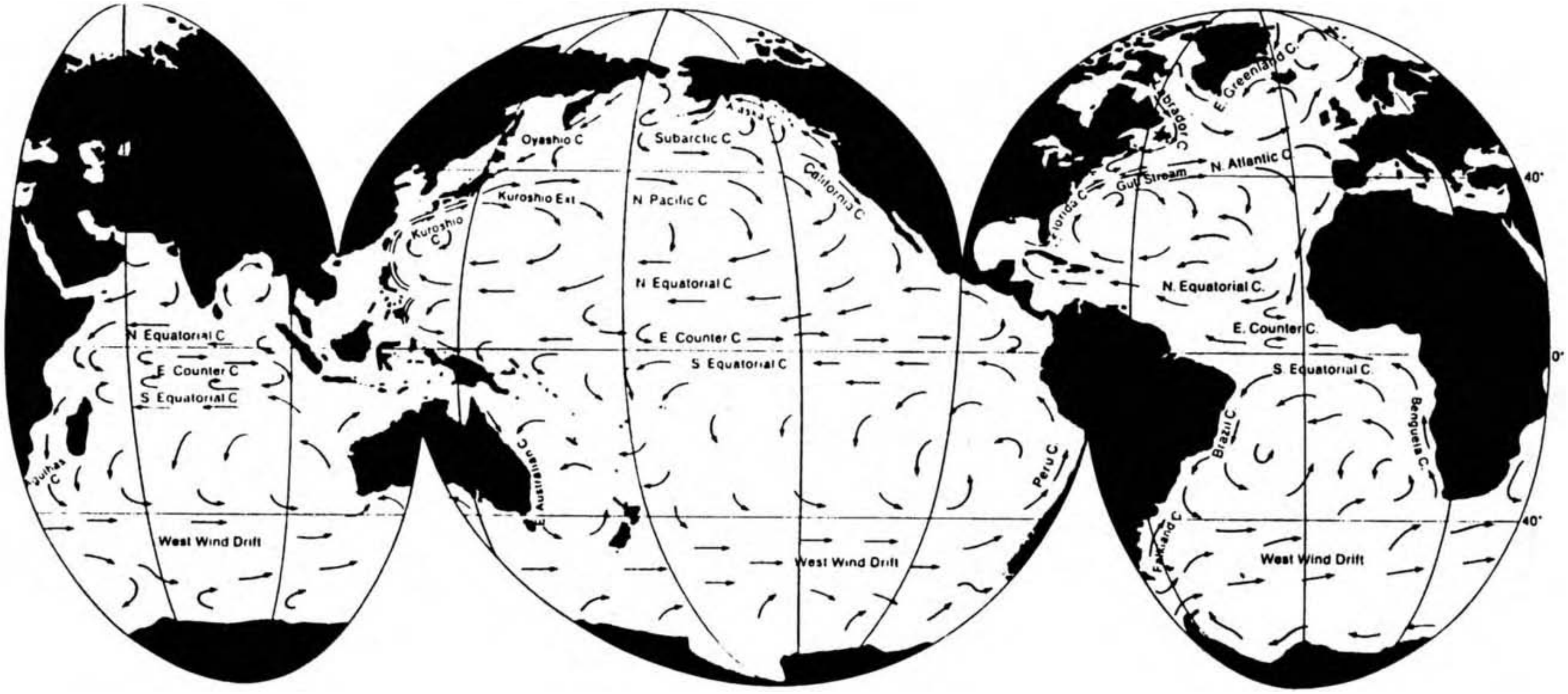


FIGURE 3 Long-term averaged ocean surface circulation (Ross, 1982).

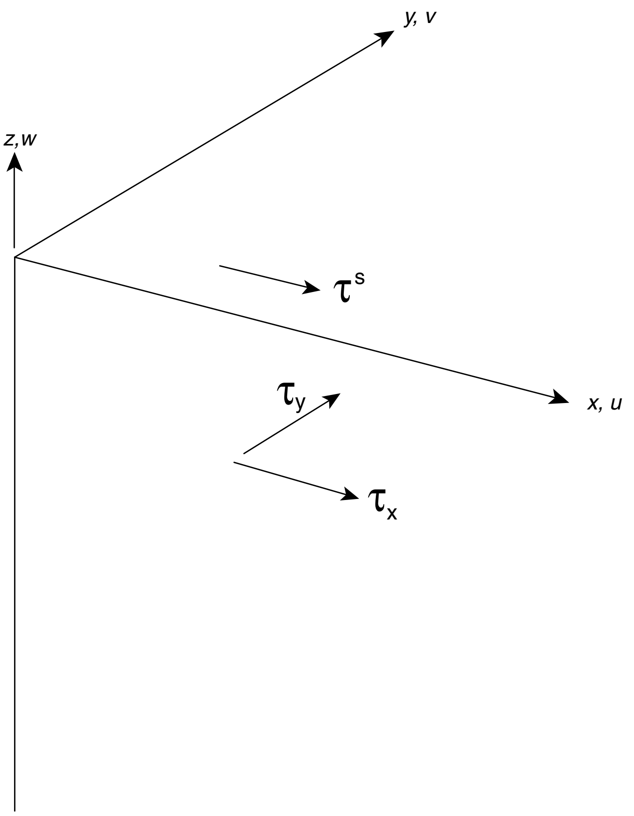


FIGURE 4 Coordinate system.

With the simplifications and assumptions described above, the Navier-Stokes equations governing the motion of a homogeneous ocean in the northern hemisphere become:

$$\rho fu = A_v \frac{d^2v}{dz^2},$$

$$-\rho fu = A_v \frac{d^2u}{dz^2},$$

where f is the Coriolis parameter, and is equal to $2\omega \sin\phi$, where ω is the angular velocity of the earth and ϕ is the latitude of interest.

The previous two equations can be combined to give two fourth-order differential equations, one for u and one for v. Solution of these equations requires knowledge of the boundary conditions for u and v. We specify that u and v must decrease to zero at infinite depth (i.e., as z goes to negative infinity). At the water surface, $z = 0$, we specify that the turbulent shear stress must equal the applied surface wind stress. With these boundary conditions, the solution of our equations becomes:

$$u = \frac{\tau^s}{\rho(A_v f)^{1/2}} \cos((f/2A_v)^{1/2} z - 45)\exp((f/2A_v)^{1/2} z),$$

$$v = \frac{\tau^s}{\rho(A_v f)^{1/2}} \sin((f/2A_v)^{1/2} z - 45)\exp((f/2A_v)^{1/2} z).$$

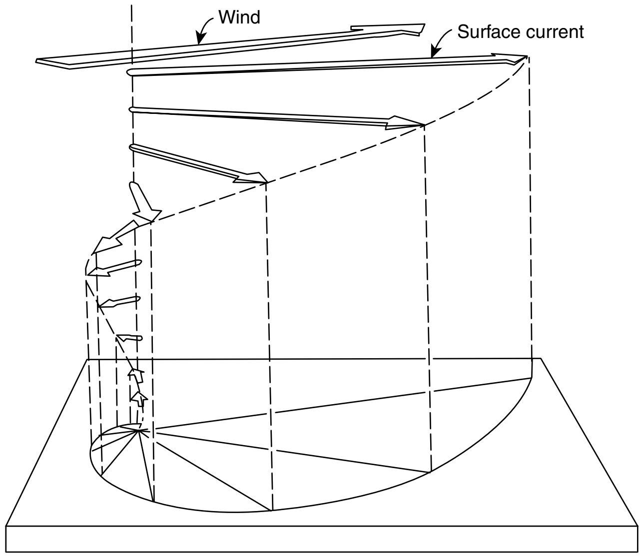


FIGURE 5 Vertical variation of wind-driven water motion (Gross, 1977).

Examination of our result illustrates that the water velocity decreases exponentially with depth from a maximum at the surface, while at the same time experiencing a rotation in direction. This vertical spiral, illustrated in Figure 5, is commonly referred to as the "Ekman Spiral", after V.W. Ekman, who first investigated the problem (Ekman, 1905). Note that the surface water velocity (i.e., at $z = 0$) is directed at an angle of 45° to the right of the wind direction (in the northern hemisphere). This theoretical result, obtained with several simplifications and assumptions, is not far removed from observations which indicate surface water motions directed at approximately 10° to 40° to the right of the wind direction. For a more complete theoretical treatment of this problem, including the effects of unsteadiness, and a varying eddy viscosity, the reader is referred to Madsen (1977).

A further interesting conclusion can be obtained by vertically integrating our expressions for u and v from the surface to the "bottom", z = negative infinity. The quantities thus obtained, U and V, represent the total volume flux of water in the x and y direction, respectively. Performing the integration, we obtain $U = 0.0$ and $V = -\tau^S(f\rho)$. This result indicates that the total water flux in the wind-driven water column (termed the "Ekman Transport") is directed 90° to the right (left) of the wind stress in the northern (southern) hemisphere!

Our solution for the Ekman Transport offers an explanation for the coastal phenomena known as "upwelling" and "downwelling". Clearly, if a wind acts along a coastline with the coast to the left (in the northern hemisphere), the total Ekman Transport will be directed in the offshore direction, with most of the flux occurring in the surface region of the water column (remember, the velocity decreases exponentially from the surface). Conservation of water mass dictates that this volume of water must be replaced, and this "replacement" water can only come from the offshore, bottom region. This process, known as upwelling, is actually much more complex than the general explanation given here, with the onset and subsidence dependent on many factors in addition to the surface wind magnitude and direction, including temperature variations and turbulent mixing (Send *et al.,* 1987). The phenomenon of downwelling is induced by the opposite forcing scenario, with the wind blowing along a coast with the shoreline to the right (left) in the northern (southern) hemisphere. Downwelling is characterized by a transport of nearshore surface waters downward and in the offshore direction.

Upwelling is of great concern to ocean scientists because of its potentially significant impact on climate and water quality. The offshore bottom water brought to the nearshore region is often nutrient-rich. This input of nutrients can have either beneficial or detrimental effects, providing sufficient food sources to support a large fisheries population (e.g., the Peruvian coastline) or promoting algal growth and deoxygenation in the coastal waters, as seen in the 1976 shellfish die-off along the coast of New Jersey (Swanson and Sindermann, 1979). The transport of relatively cold bottom water to the shallow nearshore region can also alter the heat exchange between the atmosphere and the ocean, with potentially significant consequences for the regional climate.

In our introduction, we alluded to the 1982–1983 El Nino event as evidence for the ocean's influence on climate variations. Generally speaking, an El Nino event is characterized by a warming of the coastal waters off the coasts of Peru and Ecuador. Under normal circumstances, these waters are held relatively cool by the upwelling activity typical of this region of the eastern Pacific. The atmospheric circulation of this area is characterized by what is commonly referred to as the Southern Osciollation: the seasonal shift in atmospheric surface pressure between the Australian Indian Ocean region and the southeastern Pacific. The Southern Oscillation has been shown to influence surface pressure, temperature and rainfall variations over much of the earth (Barnett, 1985). It has been postulated by many scientists that the disastrous effects of severe El Nino events (e.g., abnormal rainfall variations on a global scale) are due to the alternation of the Southern Oscillation by the nearshore surface warming along Peru and Ecuador. Although evidence of a direct connection between El Nino and alternations in the Southern Oscillation is by no means complete (Deser and Wallace, 1987), there is little doubt that the warming and cooling of ocean surface waters has a significant impact on atmospheric circulation, potentially on a global scale.

Density-Driven Deep Ocean Currents

The wind-driven ocean circulation, because of its forcing from an applied shear stress at the water surface, does not persist at great depth, as shown in our solution indicating an exponential decrease in velocity with depth. Using this solution, we can define an "Ekman Depth", or depth of influence of the surface wind stress as: $D_e = \pi(2A_v/f)^{1/2}$.

Depending on the magnitude of the turbulent eddy viscosity, A_v, and the latitude of interest, the Ekman Depth can vary from order of 100 meters to 1000 meters. Since the average depth of the world's oceans is approximately 4000 meters, one may be led to believe that the waters of the deep ocean remain motionless. However, quite the contrary is true. Measurements indicate that the oxygen content of the ocean's deep waters is much higher than would be expected for a motionless water volume, thereby indicating that motion does indeed occur at great depth. As alluded to earlier, water masses can be characterized by their temperature and salinity. This T/S structure can be employed to track specific water volumes from point of origin to final destination. Using this methodology, scientists have, over the last 200 years, identified several large scale, deep ocean flows.

The initial question in addressing the deep ocean transport of water masses must necessarily concern the identification of point of origin of specific water types. The two dominant water masses associated with deep ocean transport are North Atlantic Deep Water and Antarctic Bottom Water.

The first water type we will consider is Antarctic Bottom Water. This water is formed primarily on the south-west continental shelf of Antarctica (i.e., the Weddell Sea). During ice formation at the water surface, brine is expelled from the sea water, creating a layer of cold, highly saline water immediately below the ice. Because of its significantly higher density, this water sinks at the shelf break (the steep slope connecting the continental shelf with the deep ocean), and mixes with the relatively warm, saline deep water transported by the North Atlantic Deep Water current (discussed later). This mixture is more dense than either of the constituent water types, and therefore sinks toward the ocean bottom, becoming Antarctic Bottom Water. This water then moves in two directions, northward to the North Atlantic, and eastward around the Antarctic continent.

The second water type, North Atlantic Deep Water, is formed in the Norwegian Sea, which is separated from the Atlantic Ocean by a submarine ridge running from Greenland to Europe. We know that at the lower latitudes near the equator, the surface waters undergo considerably more heating and evaporation than those of the mid and high latitudes, where surface cooling and precipitation dominate. As one would expect, therefore, the surface waters of the lower latitudes are considerably warmer and more saline than those of the higher latitudes. The wind-driven Gulf Stream carries a large volume of this warm, saline water from the lower latitudes northward, some of it eventually being transported into the Norwegian Sea. This transport is sufficiently fast that the water remains highly saline relative to the surrounding North Atlantic water. As this saline water is cooled at the surface, it becomes increasingly dense, finally sinking to the bottom layers of the Norwegian Sea. This water then flows back into the North Atlantic over the submarine ridge mentioned earlier. Since the water depth at the ridge is only 800 meters at its deepest location, much shallower than the equilibrium depth of the very dense bottom water, this water flows over the ridge and down toward the bottom of the North Atlantic, falling thousands of meters much like an underwater waterfall. The water mixes with the surrounding water during its vertical flow, resulting in a very dense water mass, although not as dense as the Antarctic Bottom Water. The North Atlantic Deep Water therefore flows in a southerly direction, on top of the northward flowing Antarctic Bottom

Water, eventually reaching the South Atlantic and flowing eastward around Antarctica.

The travel time for the North Atlantic Deep Water to reach the Antarctic continent has been estimated as being on the order of 1000 years. Of course, we do not expect this water to remain unchanged during such a journey. Mixing processes alter the density of the water by mixing it with the less dense water overlying it. It is this gradual mixing, in fact, that creates the relatively warm, saline deep water which arrives at the Weddel Sea to be mixed with the cool, saline shelf water and form Antarctic Bottom Water.

Current speeds of the North Atlantic Deep Water and the Antarctic Bottom Water vary greatly. Maximum speeds as high as 20 cm/sec have been measured near the ocean bottom beneath the Gulf Stream (Pierce and Joyce, 1988). However, as indicated by the estimated travel time of the North Atlantic Deep Water, the average deep water current velocity is much lower, most likely on the order of 1 cm/sec.

Coastal Ocean Circulation

We have already discussed the generation of surface (wind-driven) and bottom (density-driven) currents, identifying the dominant, or best known, of each. Our examples, however, were confined to the "interior", or deep ocean regions, far removed from the shallow water areas that border most of the continents. As mentioned earlier, the coastal oceans are typically characterized by a shallow continental shelf, gently sloping seaward to a continental slope, which drops relatively abruptly to the deep ocean. Most seismically inactive areas, such as the eastern coast of the United States, possess a relatively wide (100–200 km) continental shelf, whereas seismically active areas, such as the Pacific coast of the United States, typically possess a very narrow shelf-slope region.

Coastal ocean regions differ from the deep ocean both in forcing and response. As one would expect, many of these differences arise from the relatively shallow water depths encountered along a continental shelf. Clearly, the surface wind stress is responsible for much of the coastal ocean circulation, especially the short term, highly variable components of the flow. However, the response of the water column differs from that of the deep ocean for the simple reason that the water depth is often considerably less than the "depth of penetration" of the wind stress, the Ekman Depth alluded to earlier. Since the wind-induced motion does not therefore decrease to zero before the ocean bottom is reached, the flow experiences a resistance due to skin friction and form drag over roughness elements on the bottom. For ocean circulation modelers, the estimation-parameterization of this bottom flow resistance is a formidable task. A knowledge is required, not only of the roughness characteristics of the bottom, but also the variation of these characteristics in both time and space as the sediments are influenced by near-bottom motions and bottom-dwelling marine organisms. For an excellent treatment of the complexities of the flow within the continental shelf "bottom boundary layer", the reader is referred to Grant and Madsen (1986).

Of course, the proximity of coastal ocean regions to land also influences the water circulation. In the vicinity of freshwater inflows, the nearshore circulation is altered by the presence of density-driven motions. Typically, the fresh water moves on top of the saline ocean water, eventually becoming mixed in the vertical and horizontal directions, primarily through wind-induced turbulent mixing. Before this mixing is complete, however, the horizontal and vertical density gradients will induce water motions, often quite different from those anticipated from the local (i.e., wind) forcing. As an example, the mean flow along the Middle Atlantic Bight (the continental shelf region running from Cape Cod south to Cape Hatteras) is toward the southwest, opposed to the mean eastward wind stress experienced in the area. Although still a topic of active research, the explanation for this flow appears to lie in the presence of a mean along shelf pressure gradient, quite possibly originating from freshwater input north of the region (Chapman *et al.,* 1986). The land boundary itself is also responsible for phenomena unique to the coastal ocean. Topographic variations, both in the horizontal plane and along the ocean bottom can induce secondary motions by virtue of the variation in flow resistance along the boundaries. These motions, as in the case of the density-driven flows mentioned previously, can often run counter to the local, primary forcing.

One of the most significant differences between the two types of coastal boundaries mentioned earlier (narrowshelf and wide-shelf) is the influence of deep ocean water motions on the nearshore circulation. As one might expect, narrow shelf areas are more prone to deep ocean forcing. For this reason, fluctuations in the location and magnitude of deep ocean surface currents (e.g., the California Current along the Pacific coast of the United States) can greatly influence the nearshore water motion, as can wind-driven upwelling events. Regions possessing a wide continental shelf, however, are characterized by quite different flow regimes. Recent evidence (Chapman and Brink, 1987; Chapman *et al.*, 1986) indicates that shallow continental shelves remain effectively isolated from deep ocean forcing. The circulation on these shelves is dominated by wind-driven motions, with tidal forcing and freshwater inflows (hence, density-driven currents) important, especially in nearshore areas.

We should here note that in regions of energetic deep water motions, the shelf and slope are occasionally influenced by these motions. As a case in point, continental shelf motions along the east coast of the United States can be affected by the northward flowing Gulf Stream. This influence takes the form of circulation cells, or eddies (commonly referred to as Gulf Stream Rings) shed from the Gulf Stream and driven toward the shelf region. As these short-lived eddies (typically 1–3 weeks, Lee and Atkinson (1983)) approach the shallow shelf-slope boundary, their forward motion is impeded by the bottom topography, thereby limiting their influence on inner shelf motions. However, the rotational motion of the eddies themselves can entrain extremely large volumes of shelf-slope water, removing this water out to the deep ocean and replacing it with Gulf Stream water (Garfield and Evans, 1987; Evans *et al.*, 1985).

WIND WAVES

In this section, we will treat the generation and characteristics of wind-induced surface gravity waves. These are waves formed at the air–sea interface by the action of a surface wind stress. The term "gravity wave" infers that the restoring force, or that force which seeks to restore the water surface to its equilibrium position, is the force due to gravity. An example of a different type of restoring force is the force due to surface tension. This force is actually the result of molecular attraction among water molecules, which, at the air—water interface, creates a net, adhesive force retaining the water molecules at the water surface. As one would expect, this force is only important in cases of very small scale motions, such as capillary waves, and is therefore unimportant for our purposes.

Early theories of wind-wave generation include the work of Kelvin (1871) and Helmholtz (1888), who examined the onset and growth of wave forms arising from instabilities at the interface of two fluids, each with different density and moving at different velocities. Although providing the basis for much of our present knowledge of flow instabilities and turbulence, the Kelvin–Helmholtz mechanism proved inadequate as a description of the onset of wind wave generation. The theory predicts wave generation only at surface wind speeds exceeding approximately 14 miles/hour, far greater than the minimum wind speeds required for wave generation observed in nature.

Jeffreys (1924, 1925) introduced what is commonly referred to as the "sheltering hypothesis" of wind wave generation. He proposed that the airflow over a previously rough water surface separates on the downwind sides of crests much the same as the flow separation observed in the lee of a cylinder. The resulting asymmetry in wind velocity leads to an asymmetric pressure distribution along the water surface, giving rise to a resultant force in the direction of the airflow. Theoretically, this pressure forcing will continue to add energy to the waveform as long as the velocity of the airflow exceeds the phase velocity of the wave.

Although providing valuable insights into the growth of wind waves, the sheltering hypothesis failed to address the critical question of the onset of wave generation because of its assumption of a previously rough water surface. This problem of wave generation on an initially smooth water surface was examined by Phillips (1957), who proposed that the initiation of wave formation is due to the presence of turbulence in the airflow overlying the water surface. Associated with this turbulence are random fluctuations in velocity, and hence, pressure. One can imagine that in areas of high pressure, the water surface will be depressed, whereas in regions of low pressure, the interface will, relatively speaking, rise, thereby creating a waveform at the air–water interface. Phillips proposed that the pressure fluctuations will continue to act on those waveforms having a phase speed equal to the speed of the turbulent pressure fluctuations, so that selective growth of only certain wave frequency components occurs.

Phillips' mechanisms was found to provide an excellent description of the initial stages of wind-wave formation. The problem of continued wave growth was addressed by Miles (1957). Using a model of shear flow instability similar to the Kelvin–Helmholtz theory mentioned earlier, Miles illustrated that a coupling exists between the airflow and the wave motion. In addition to the airflow doing work on the water surface, the perturbations in the water surface (i.e., the waveforms) can induce instabilities in the airflow. The energy transfer due to these instabilities in what is essentially a two-fluid shear flow is responsible for continued wave growth.

Although the Phillips–Miles mechanism for the initiation and growth of surface wind waves was proven quite accurate, measurements of the long-term evolution of wave fields indicated the importance of a third factor, the interaction among the individual wave components themselves. Recall from our previous discussion that selective generation and growth of different wave frequencies occurs, depending on the relative speeds of the wave forms and the turbulent pressure fluctuations overhead. The resulting water surface is therefore not characterized by a single, uniform wave form, but is rather composed of a multitude of wave components, each with a different frequency, amplitude, and wavelength. For this reason, wave data are typically analysed with the use of a wave spectrum (analogous to the color spectrum of optics), separating the observed sea into its various frequency or wavelength components. Phillips (1960) and Hasselmann (1962) illustrated that these individual components interact nonlinearly, resulting in a transfer of energy from the central frequencies of the spectrum to the high and low frequencies. As measurement and modeling techniques have improved over the last 25 years, functional relations for all of the forcing mechanisms mentioned here have been developed, and have led to the creation of quite accurate computer algorithms for the prediction of surface wind waves.

In seeking to describe the characteristics of a specific wave field, perhaps the most important relationship is the "dispersion relation", which defines the relationship between the wave frequency, ω, and the wavenumber, k:

$$\omega^2 = gk \tanh(kh),$$

where g is the acceleration due to gravity, h is the water depth, $\omega = 2\pi/T$, T is the wave period (time of travel of one waveform), $k = 2\pi/L$, and L is the wavelength. Figure 6 illustrates the most commonly used surface wave parameters.

The term, "dispersive," signifies that waves of different frequencies move at different phase speeds, $C = \omega/k$. Clearly, therefore, longer waves have higher phase speeds. Note also the dependence on water depth, h, with a given wave in deep water having a higher phase speed than the corresponding were in shallow water. Simply speaking, this depth dependence is responsible for the refraction of wave as they approach at an angle to the shoreline, with the portion of the wavetrain in deeper water leading that portion located in shallow water.

Several interesting observations can be obtained by examining the dispersion relation in the two limits of very deep and very shallow water depths. Note that "deep" and "shallow" water waves are defined by the water depth relative to the wavelength, as illustrated in the dependence on (kh). For this reason, a deep water wave is often termed a "short" wave and a shallow water wave, a long wave. In the case of short waves, we have a very large value for (kh) so that the quantity, tanh(kh) goes to unity and the dispersion relation

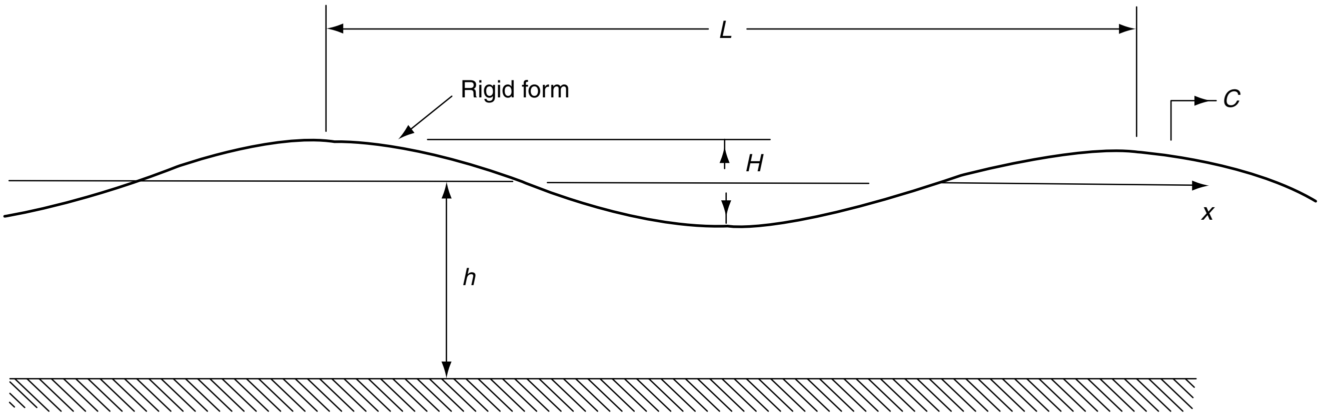


FIGURE 6 Definition sketch-surface water characteristics.

becomes: $\omega^2 = gk$. The wave characteristics (e.g., wavelength, phase speed, etc.) are therefore independent of depth.

For very shallow (long) waves, we have (kh) much less than one, so that tanh(kh) goes to (kh) and the dispersion relation becomes: $\omega^2 = gk^2h$. The wave characteristics are now strongly dependent on water depth. A further observation is that shallow water waves all move at the same phase speed, $C = (gh)^{1/2}$.

The two-dimensional fluid motion beneath a surface wave can be described by the following relations:

$$U = \frac{agk \cosh k(z + h)}{\omega \cosh(kh)} \cos(kx - \omega t),$$

$$W = \frac{agk \sinh k(z + h)}{\omega \cosh(kh)} \sin(kx - \omega t),$$

where U and W are the horizontal and vertical components, respectively, of the water motion; z is the vertical position of interest, decreasing from $z = 0$ at the surface to $z = -h$ at the bottom; and a is the wave amplitude, measured from the equilibrium, or still water level, to the wave crest.

Note that the wave-induced water velocities, U and W, are 90° out of phase, indicating that the fluid particle trajectories beneath surface waves are elliptical in shape. Since the value of $\sinh k(z + h)$ decreases to zero while $\cosh k(z + h)$ decreases to unity as one approaches the bottom ($z = -h$), the fluid motion becomes more and more horizontal with depth. Theoretically, at the bottom the wave-induced motion is purely horizontal.

Once again, if we examine the limits of shallow water and deep water waves, we can make some interesting observations. In the case of deep water (short) waves, we have very large (kh), and, noting that $\sinh(x) = \cosh(x)$ for large x, the particle trajectories are circular in shape. For shallow water, or long waves, we have very small (kh) so that the vertical velocity, W, is much smaller than the horizontal velocity, U, and the particle motions are almost exclusively horizontal.

We should caution that the equations noted here are derived from the linear form of the Navier-Stokes equations. That is, the non-linear, or convective acceleration terms have been assumed very small. This assumption is satisfied when the ratio of the wave height, $H = 2a$, to the wavelength, L, is very small: $H/L \ll 1$, or in physical terms, when the wave steepness is very small. For this reason, the equations listed here constitute a description of "linear," or "small amplitude" water waves. It follows that in regions where the wave steepness is relatively large (in the breaking wave zone, for example) our theory loses some of its accuracy, and one should be careful in its application. A classic treatment of the derivation of the governing equations for both linear and non-linear water waves can be found in Whitham (1974).

TIDES

Ocean tides can be described as the periodic rise and fall of the ocean surface. The forcing responsible for this motion is the gravitational attraction of the moon, sun and (to a much lesser extent) the other planets. Because of its close proximity to the earth, however, the moon plays the dominant role in tidal forcing.

A simple explanation for the generation of tides can be obtained from an examination of the earth-moon system. We know that the attractive force due to gravity is inversely proportional to the square of the separation distance between two bodies: $F \propto 1/d^2$. For this reason, the ocean surface closest to, or facing the moon experiences a much higher attractive force than does the ocean surface on the opposite side of the earth. The water surface on the opposite side of the earth. The water surface facing the moon is therefore "pulled" toward the moon. At the same time, the attractive force of the moon "pulls" the entire earth from under the ocean surface on the opposite side of the earth. The net result, therefore, is an increase in water surface elevation on both the side of the earth facing the moon and the side opposite. Two peak tides are therefore created at the same time on opposite sides of the earth. Since the earth rotates once in 24 hours, we expect two high tides and two low tides each day. However, the cycle, or lunar day is actually approximately 24 hours, 50 minutes, so that each day the high and low tides occur approximately 50 minutes later.

Of course, the above description of the generation of ocean tides is rather simplistic. In actuality, we have the additional influences of the sun and planets, sometimes enhancing and sometimes reducing the moon-induced tide. Perhaps more importantly, variations in ocean water depth greatly influence the movement of the tidal "wave", this influence becoming very apparent in the continental shelf-slope region. In spite of the these complexities, we have at the present time an excellent understanding of the physics of tide generation and tidal wave propagation. Tide prediction algorithms have been developed and proven surprisingly accurate among much of the world's coastlines. Of course, accuracy is a relative word and should be measured against the requirements of ocean scientists. As these requirements dictate still more accurate tidal elevation predictions, new algorithms must be developed. An example of such a development is the recent work in spectral modelling of shallow water tidal wave propagation, with accuracy in sea surface elevation of the order 2 centimeters (Vincent and Le Provost, 1988).

Recent Advances in Oceanographic Research The global impacts of El Nino events have become more clearly defined in the interval since this article first appeared. The need to develop models for prediction of El Nino occurrences became obvious following the severe event in 1982/83 and the even stronger one in 1996/97. There has been a significant effort to integrate advances in observations from satellites (with coverage of the global ocean) with *in situ* measurements from ships and buoys and to incorporate these into coupled numerical models of the climatology of the ocean and atmosphere. The evidence seems compelling that the occurrence of El Nino is closely associated with the oscillation in sea level pressure between the western and central regions of the tropical Pacific Ocean. A non-dimensional index of this oscillation is found by subtracting the measured pressure in Darwin, Australia from that measured in Tahiti and then dividing by the standard deviation of this pressure difference obtained from the long term records at these locations.

Under conditions without an El Nino this index is generally slightly positive but can undergo reversals of sign. During significant El Nino events this index becomes strongly negative, indicating that the typical low pressure on the western side of the tropical Pacific has changed dramatically.

Conceptually, during normal years the strong easterly trade winds move warm surface water into the western portion of the Pacific with much cooler water occurring along the west coast of South America and extending into the Eastern Equatorial Pacific. The elevation of the sea surface is about 0.5m higher near Indonesia than off Ecuador. Prior to the onset of El Nino the east to west pressure gradient along the equator diminishes as the Southern Oscillation Index becomes negative. As a result the trade winds diminish (even reversing direction) and the warmer surface water in the western tropical Pacific moves eastward eventually warming the near surface waters along the equator and the west coast of South America. The effects of the occurrence of an El Nino on weather are significant and on a global scale. Following an El Nino there can be an over recovery with stronger trade winds and more intense upwelling of colder deep water along the Peruvian Coast. This La Nina is associated with a positive Southern Oscillation index and has important global effects as well. The connection of El Nino and the Southern Oscillation is sufficiently strong to give rise to the acronym ENSO to denote these two phenomena and their interconnections.

Another large scale oscillation has been detected for the North Pacific Region. Termed the Pacific Decadal Oscillation, PDO, it is chiefly characterized by an alternation of a warm phase with a cool phase, with time scales of 2–3 decades. During the warm phase, sea surface temperatures in the central North Pacific are lower than average with warmer than average temperatures occurring along the west coasts of North and South America. For the cool phase these anomalies reverse with cooler water occurring along the west coast and the central region sea surface temperatures being warmer than average. Although there is no persistent connection between the PDO and ENSO the most severe El Nino events have all coincided with the warm phase of the PDO.

Observations of the oceanographic and meteorological conditions at the sea surface on a global scale from satellites have become increasingly useful. An example of this is provided by the 12 year success of the measurement of the dynamic topography of the sea surface using a high frequency radar altimeter and sea surface temperature using a microwave radiometer onboard the Topex Poseidon satellite. The technical difficulties in the accurate determination of sea surface topography are formidable. The position of the satellite relative to the Earth must be accurately known, the earth's geoid (the level surface most closely aligned with mean sea level) must be well defined and the effect of open ocean tides must be removed from the observations. The Topex Poseidon satellite repeats observations over the same oceanic region once every 10 days and during that period it collects data from 95% of the ice free global ocean. The reported accuracy of the departures of the mean elevation of the sea surface from the geoid is about 0.04m. Thus, the change in elevation of the sea surface across the Gulf Stream of the order of 1.0 meters is easily discerned. Given the spatial distribution of sea surface elevation, $\varsigma(x, y)$ then the surface geostrophic current components, u and v, in the x and y directions respectively are found from the following equations expressing the near balance between the Coriolis force and the component of gravity along the sloping isobaric sea surface:

$$fu = -g\,\partial\,\varsigma(x, y)/\partial y\;,\quad fv = g\,\partial\,\varsigma(x, y)/\,\partial y\;,$$

where g is the acceleration of gravity and f is, as before, the Coriolis parameter. The observations from the Topex Poseidon satellite provide estimates of the geostrophic currents at very high spatial resolution and with a temporal resolution of 10 days. These data are useful since surface currents measured directly are in substantial agreement with the calculated geostrophic currents. The 12+ year mission of this satellite (1992–2005) allowed for detailed observations of development and time history of the very intense ENSO event in 1997/98 and the subsequent La Nina. In addition to radar altimeter observations other satellite borne instruments have been used to map the distribution and evolution of sea surface temperatures using infrared radiometers, to estimate surface wind stress over the sea and ocean surface wave characteristics from scattering of microwave radars.

More conventional observational techniques are required to monitor subsurface distributions of temperature, salinity, and currents. A consequence of the intense ENSO event in 1982/83 was the recognition that atmospheric and oceanographic processes in the tropical Pacific were important precursors of such events. One result of this was the initiation of the ten year international Tropical Ocean Global Atmosphere, TOGA, program and the deployment over this period of an array of instrumented buoys called the Tropical Atmosphere Ocean, TAO, array and presently the TAO/TRITON array. The present array of 67 buoys are deployed along eleven longitudinal transects from 135°E to 95°W with coverage from 8°N to 8°S. Measurements include air temperature, relative humidity, barometric pressure, long and short wave radiation, wind velocity, together with sea surface and subsurface temperatures to a depth of 500m. Buoy data are transmitted in near real time via satellite.

An important feature of extensive ocean observation programs such as TOPEX POSEIDON and TAO/TRITON is the widespread and timely dissemination of data products via the Internet. The Pacific Marine Environmental Laboratory of the National Atmospheric and Oceanic Agency manages the TAO array and is responsible for data quality control and availability. TAO data may be obtained from the Internet site http://www.pmel.noaa.gov/tao/jsdisplay/. The Physical Oceanography Distributed Active Archive Center, PODAAC, of the NASA Jet Propulsion Laboratory at the California Institute of Technology provides data management for a variety of satellite observations of the ocean and atmosphere. These data may be accessed at the Internet site http://podaac.jpl.nasa.gov/. The development of numerical models of the ocean and of the atmosphere began in the 1950s. Increases in computer power and in global observational programs have led to a steady evolution of such models into useful predictive tools. One significant recent development has been the creation of fully coupled atmosphere and ocean climatological models. One example of this model development is the Community Climate System Model, CCSM, developed by the National Center for Atmospheric Research. It includes the atmosphere, the ocean, the land areas, and regions of sea-ice. All are coupled by computing fluxes between them. This model is an example of a recent trend in oceanography (and other sciences) to make models readily available to the "community" of potential users. In the case of CCSM it is available online from its web site together with user manuals and other aids for its successful implementation. The CCSM has demonstrated its predictive skill through comparison with observed climatic variability. A particular example is its ability to hindcast recent El Nino events with reasonable accuracy.

REFERENCES

Barnett, T.P., Variations in near-global sea level pressure, *J. Atmos Sci.,* **42**, 478–501, 1985.

Blumberg, A.F. and G.L. Mellor, Diagnostic and prognostic numerical circulation studies of the South Atlantic Bight, *J. Geophys. Res.,* **88**, 4579–4592, 1983.

Chapman, D.C. and K.H. Brink, Shelf and slope circulation induced by fluctuating offshore forcing, *J. Geophys. Res.,* **92**, 11, 741–760. 1987.

Chapman, D.C., Barth, J.A., Beardsley, R.C. and R.G. Fairbanks, On the continuity of mean flow between the Scotian Shelf and the Middle Atlantic Bight, *J. Phys. Oceanogr.,* **16**, 758–772, 1986.

Deser, C. and J.M. Wallace, El Nino events and their relation to the southern oscillation: 1925–1986, *J. Geophys. Res.,* **92**, 14, 189–196. 1987.

Edmond, J.M., The geochemistry of ridge crest hot springs, *Oceanus.* **27**, 15–20, 1984.

Ekman, V.W., On the influence of earth's rotation on ocean currents. *Ark. Math. Astron.* Fys., **2**, 1–53, 1905.

Evans, R.H., Baker, K.S., Brown, O.B. and R.C. Smith, Chronology of warm-core ring 82B, *J. Geophys. Res.,* **90**, 8803–8812, 1985.

EOS, News Item, 69, 49, 1988.

Fofonoff, N.P., Physical properties of sea water, in *The Sea,* **1,** 3–30, New York: Interscience, 1962.

Garfield, N. and D.L. Evans, Shelf water entrainment by Gulf Stream warm-core rings, *J. Geophys. Res.,* **92**, 13, 3–12, 1987.

Grant, W.D. and O.S. Madsen, The continental shelf bottom boundary layer, *Ann. Rev. Fluid Mech.,* **18**, 265–305, 1986.

Gross, M.G., Oceanography a View of the Earth. New Jersey: Prentice-Hall, 1977.

Hasselmann, K., On the non-linear energy transfer in a gravity wave spectrum, 1 general theory, *J. Fluid Mech.,* **12**, 481–500, 1962.

Heezen, B.C. and I.D. Mac Gregor, The evolution of the Pacific, *Scientific American,* **229**, 102–112, 1973.

Helmholtz, H., Uber atmospharische bewegungen, S. Ber. Preuss. Akad. Wiss., Berlin, Mathem.-Physik, 1888.

Jeffreys, H., On the formation of waves by wind, *Proc. Roy. Soc. Lond., Ser. A,* Vol. 110, 341–347, 1925.

Jeffreys, H., On the formation of waves by wind, *Proc. Roy. Soc. Lond., Ser. A,* Vol. 107, 189–206, 1924.

Kelvin, On the waves produced by a single impulse in water of any depth, or in a dispersive medium. Math. And Phys. Papers, Vol. IV, London, Cambridge University Press, 1887.

Lee, T. N. and L.P. Atkinson, Low-frequency current and temperature variability from Gulf Stream frontal eddies and atmospheric forcing along the southeast U.S. outer continental shelf, *J. Geophys. Res.,* **88**, 4541–4567, 1983.

Madsen, O.S., A realistic model of the wind-induced Ekman boundary layer, *J. Phys. Oceanogr.,* **7**, 248–255, 1977.

Miles, J.W., On the generation of surface waves by shear flows, *J. Fluid Mech.,* **3**, 185–204, 1957.

Phillips, O.M., On the dynamics of unsteady gravity waves of finite amplitude, Part 1, *J. Fluid Mech.,* **9**, 193–217, 1960.

Phillips, O.M., On the generation of waves by turbulent wind, *J. Fluid Mech.,* **2**, 417–445, 1957.

Price, J.F., Terray, E.A. and R.A. Weller, Upper ocean dynamics, Contributions in Oceanography, 193–203, *American Geophys.* Union, 1987.

Reid, J.L., Deep ocean circulation, in Oceanography, the Last Frontier, 203–217. New York: Basic Books. 1973.

Rienecker, M.M. and C.N.K. Mooers, The 1982–1983 El Nino signal off northern California, *J. Geophys. Res.,* **91**, 6597–6608, 1986.

Ross, D.A., Introduction to Oceanography. New Jersey: Prentice-Hall, 1982.

Sclater, J.G. and C. Tapscott, The history of the Atlantic, *Scientific American,* **240**, 156–174, 1979.

Send, U., Beardsley, R.C. and C.D. Winant Relaxation from upwelling in the Coastal Dynamics Experiment, *J. Geophys. Res.,* **92**, 1683–1698, 1987.

Swanson, R.L. and C.J. Sindermann (eds.), Oxygen Depletion and Associated Benthic Mortalities in New York Bight, 1976. NOAA Professional Paper 11, U.S. Govt. Printing Office, 1979.

Vincent, P. and C. Le Provost, Semidiurnal tides in the northeast Atlantic from a finite element numerical model, *J. Geophys. Res.,* **93**, 543–555, 1988.

Whitham, G.B., Linear and Nonlinear Waves. New York: Wiley, 1974.

MICHAEL BRUNO
Stevens Institute of Technology

RICHARD HIRES
Stevens Institute of Technology

OIL SPILLAGE INTO WATER—TREATMENT

The general subject of treatment of oil spillage represents a relatively new area of technology that is unique in that it encompasses chemical, mechanical, and biological disciplines. There are three major aspects to the problem of oil spills.

1) PREVENTION OF THE SPILL;
2) CONTAINMENT AND RECOVERY OF THE SPILL;
3) TREATMENT OF THE SURFACE OIL.

Although this discussion is mainly directed toward the treatment of the spilled oil, the related areas will also be considered in order to put the overall subject in the correct perspective.

PREVENTION OF THE SPILL

The Prevention of Oil Spillages is the Primary Consideration

It should be emphasized that prevention is the first consideration and, of course, the most complete solution. In the industrial and governmental communities, the major effort has been directed toward this area. There is extensive ongoing research for example, ranging from operational areas such as collision avoidance techniques and training to more novel approaches such as the gellation of crude oil. In this latter approach several chemical systems have been developed to gel the oil cargo. This in-situ solidification thereby prevents the release of oil from a damaged cargo compartment that may be in danger of failure.

Details of this gellation system can be found in US Patent 3,634,050.[1] Other details, such as the effects of mixing, crude oil type, chemical concentration, and so on, on the strength of the gel have been outlined by Corino.[2] Gellation is a novel approach to prevent the release of oil. However the fact that there have been no commercial uses of this method since its conception twenty five years ago raises questions regarding its practicality.

Finally, the removal of the oil cargo from a grounded tanker is another area where the threat of the release of a fluid and mobile oil cargo to the marine environment has been mitigated by advances in salvage techniques. The offloading of the grounded SS General Colocotronis on a reef off Eleuthera Island in March–April 1969 and the well documented recovery of Bunker C oil from the sunken tanker SS Arrow in Chedabucto Bay, Nova Scotia during the winter of 1970[3] are two outstanding examples of this prevention technique.

This latter incident represented a singular achievement in light of the weather conditions encountered during early March in Nova Scotia. Over 6000 tons of viscous Bunker C oil were recovered from the sunken wreck. The salvage team used a hot tap technique to penetrate the tanker cargo tanks and then used a steam traced pumping system to transfer the oil to a barge at the surface.

A more recent and massive removal of oil was the EXXON VALDEZ in March 1989 after its grounding on a reef in Prince William Sound. Although approx 250,000 Bbls of North Slope Crude oil was spilled from the grounded vessel, 80 percent of its cargo was still in the tanker. This offloading was a significant marine engineering feat since care must be taken to offload such a large vessel in the correct sequence since otherwise hull stresses could cause the vessel to break up.

CONTAINMENT, RECOVERY OR REMOVAL OF THE SPILLED OIL

If a spill has occurred, it is universally agreed that the recommended procedure is to contain and physically recover it with or without the use of adsorbents. It is obviously the most direct solution to spill incident, if conditions permit its execution. This approach may entail three processes:

1) Confinement of the spill by spill booms.
2) Recovery of the spill by sorbing agents. In this area, more recent advancements have been solidifying agents (Solidifiers).
3) Physical removal of the contained oil by oil pickup devices.
4) Controlled burning of spilled oil.

These aspects of the recovery approach are interrelated as will be appreciated by the following discussion.

Confinement of the Spill by Spill Booms There are many oil spill booms commercially available today. Unfortunately they are significantly limited by the velocity of the surface current and wave height. Although there are variations in the materials of construction, strength, geometry, etc., of these various boom designs, as evidenced by the number available and the range of costs, their general forms are quite similar. Almost any type of floating barrier will hold back and contain some amount of oil under quiescent

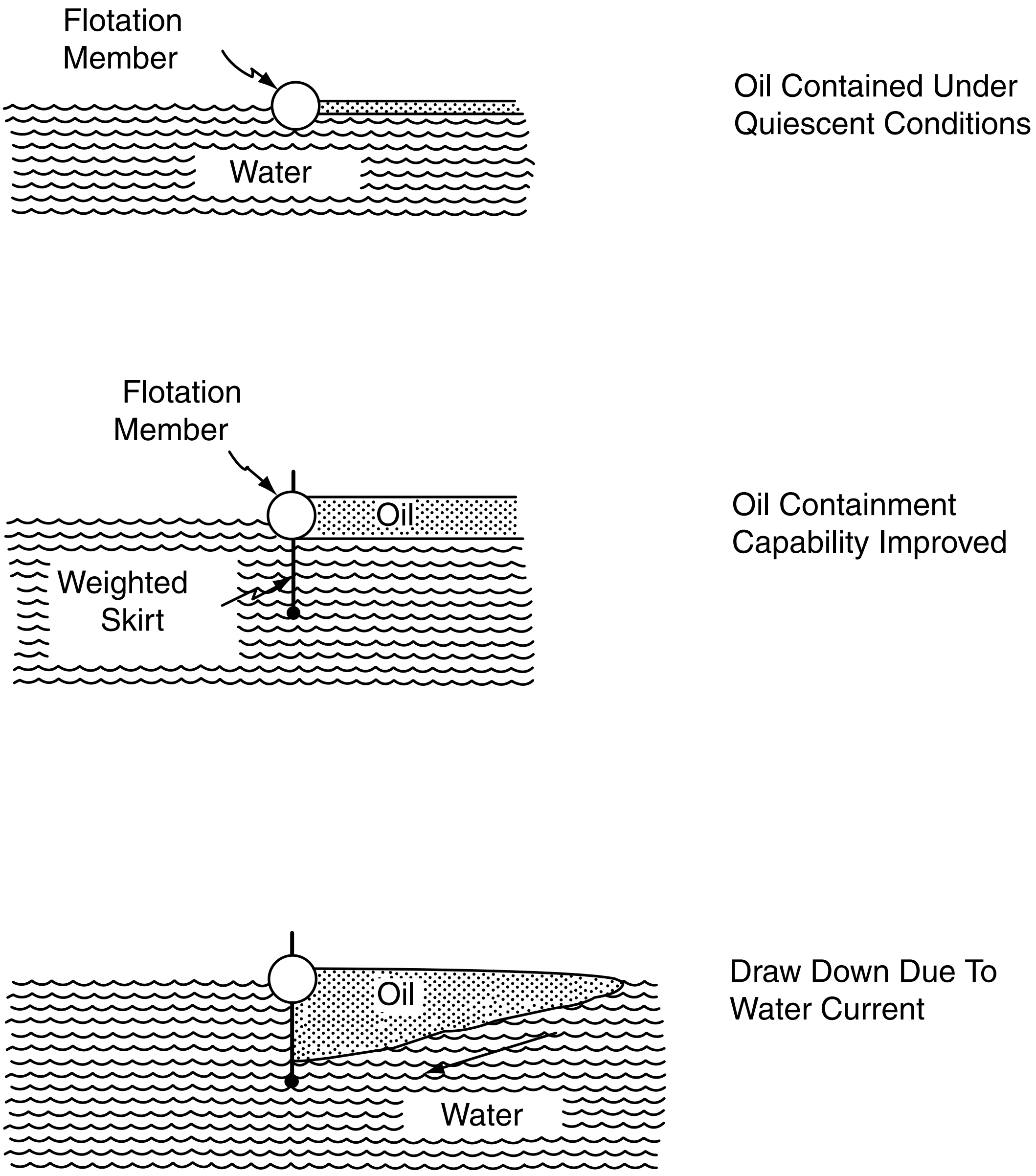


FIGURE 1 Mechanical boom principle.

conditions. Indeed, telephone poles have been employed in more than one spill instance as a jury rig emergency measure. To improve the capacity of such a floating barrier, a weighted skirt is hung from the flotating member illustrated in Figure 1. Design requirements for spill booms have been published by Lehr and Scherer[4] and Hoult,[5] among others. By a rather cursory inspection of Figure 1, we may now appreciate some of these requirements such as:

Sufficient freeboard to prevent overtopping by waves; Adequate skirt length below water surface to confined a sufficient quantity of oil;

Adequate flexibility to permit the boom to bend under wave action and maintain its retention of the oil spill;

Sufficient mechanical strength to withstand the forces imposed by the environment.

Some of the difficulties of oil retention against the action of a steady current are illustrated in Figure 1. A discussion of draw down phenomena by Hoult[6] outlines that a gradient in oil thickness, h, is established by the stress imposed by the current flow. There is, based on the fluid dynamics of the contained volume of oil in the presence of a water current, a limiting water velocity above which oil droplets are entrained and flow underneath the barrier.

The deployment of the above described mechanical booms is also an important consideration. In the event of a spill, the speed of response is, of course, most critical. Hence,

TABLE 1
Boom Classification

	Hs Maximum		Freeboard		Draft	
Environment	ft	meters	inches	centimeters	inches	centimeters
Calm Water	1	0.3	4–10	10–25	6–12	15–30
Harbor	3	0.9	10–18	25–46	12–24	30–61
Offshore	6	1.8	>18	>46	>24	>61

easily deployed lightweight booms are desirable. However, these desired properties are not necessarily consistent with making booms stronger and more capable of withstanding severe sea conditions.

These are the criteria and mechanism of operation of oil spill booms. It is beyond the scope of this chapter to present the many commercial and changing commercial products. In the world Catalog of Oil Spill Response Products, booms have been divided into three categories based on maximum operating significant wave height (Hs). Table 1 shows the ranges of freeboard and draft corresponding to the expected maximum waves. A boom size can thus be selected based on the expected environment.

In the *World Catalog of Oil Spill Response Products,* booms have been divided into three categories based on maximum operating significant wave height (Hs). Table 1 shows the ranges of freeboard and draft corresponding to the expected maximum waves. A boom size can thus be selected based on the expected environment.

Boom Selection Matrix The selection of a boom depends on how rapidly it is needed and how readily it can be utilized. Deployment speed and ease relate to the number of people, the amount of time, and any special equipment (Winches, etc.—even wrenches) necessary to move the required amount of boom from storage to the launch site, to deploy it, and to position it around the spill. For example, self-inflatable booms can be deployed very rapidly either from reels or bundles. Experience has shown, however, that this rapid response boom should be replaced by a more rugged boom if extended deployment is required. Thus, deployment ease must often be traded off against ruggedness and durability.

The matrix shown in Table 2 can be used to select the optimum boom for a specific application since it indicates:

- Generic types of boom that are most suitable in a given environment
- Selected booms that have the most needed performance characteristics
- Choices with the most desirable convenience features.

Excess or reserve buoyancy is the surplus of flotation over boom weight as deployed, and is a measure of resistance to boom submergence. Wave response is a measure of conformance to the water surface and is usually improved by increasing boom water-plane area and flexibility. Other characteristics should be evident from the headings.

To use the matrix correctly, follow these steps:

1. Identify the most probable environmental conditions in which the boom will be used. Note those types of booms with an acceptable rating (1 or 2).
2. Identify the most needed performance characteristics for the intended application. From the booms chosen above, select the ones that have an acceptable rating (1 or 2) in the most important performance characteristics.
3. Identify the most desirable convenience features. With booms from steps 1 and 2 above, select the boom with the best rating in the convenience features of interest.

These data (Table 1 and 2) were extracted and used with the permission of EXXON from a very informative OIL SPILL RESPONSE FIELD MANUAL by Exxon Production Research Company published in 1992.

Recovery of the Spill by Sorbing Agents A most direct manner of physically removing the spilled oil is by use of sorbents. These materials are buoyant, and preferentially wetted by and adsorb oil. In essence, they permit this sorbed oil to be physically "picked up" from the water. In addition to making the collection of oil an easier task, the oil is prevented from spreading and remains as a more congealed mass.

Materials that have been found useful for this service vary from simple, naturally occurring materials such as straw, sawdust, and peat to synthetic agents, such as polyurethane foam and polystyrene powder. The oil pickup capability varies greatly. For example, values of oil pickup, i.e., weight of oil sorbed per weight of adsorption material, have been reported by Struzeski and Dewling[7] for straw as 3 to 5, although higher values have been reported. Polyurethane foam, by comparison, is capable of oil pick up values of 80. A complete investigation of sorbents for oil spill removal has been published by Schatzberg and Nagy.[8] Of interest is the variation in the oil pickup capability of a given sorbent based on the type of spilled oil. For example, in Schatzberg's controlled tests, oil pickup by straw was 6.4 for heavy crude oil and 2.4 for light crude oil. For urea formaldehyde foam, however, oil pickup was 52.4 for heavy crude and 50.3 for light crude. Also some

TABLE 2
Boom Selection Matrix

		Type of Boom				
	Legend 1—Good 2—Fair 3—Poor	Internal Foam Flotation	Self-Inflatable	Pressure-Inflatable	External Tension Member	Fence
Environmental	Offshore					
Conditions	Hs > 3 ft;	2	2	1	1	3
	V > 1 kt	—	—	—	—	—
	Harbor	—	—	—	—	—
	Hs < 3 ft;	1	1	1	2	2
	V < 1 kt	—	—	—	—	—
	Calm Water	—	—	—	—	—
	Hs < 1 ft;	1	1	1	2	1
	V < .5 kt	—				—
	High Currents	—	—	—	—	—
	V > 1 kt	2*	3	2	1	3
	Shallow Water	—				—
	(Depth < 1 ft)	1	2	2	3	3
Performance	Operation	1	3	2	3	2
Characteristics	in Debris	—	—	—	—	—
	Excess	2	1	1	2	3
	Buoyancy	—	—	—	—	—
	Wave	2	2	1	1	3
	Response	—	—	—	—	—
	Strength	2	3	1	1	1
Convenience	Ease of	2	1	2	3	2
Characteristics	Handling	—	—	—	—	—
	Ease of	1	1	1	3	1
	Cleaning	—	—	—	—	—
	Compactability	3	1	1	2	3
	Cost/Ft	—	—	—	—	—
	1—Low	—	—	—	—	—
	2—Medium	1	3	2	3	2
	3—High	—	—	—	—	—

Notes:

* Hs = Significant Wave Height.

* V = Velocity of Surface Current.

Not *all* the booms of a particular generic type have the rating shown in the matrix. But at least one or more commercially available booms of the generic type in question have the rating shown.

* Specially-designed high-current models may be available (river boom).

sorbents are much less effective for oil adsorption if contacted by water prior to application to the spill.

Although highly effective sorbents are available as noted above, techniques for harvesting (recovering) the oil soaked sorbent have been limiting. For example, there have been prior instances of oil soaked straw recovery by manual pickup with pitchforks. However, there is development work underway to mechanize this step as well as the application procedure. In this regard, some very practical observations on the use of sorbents have been made by an IMCO subcommittee on Marine Pollution. This guidance manual outlined that "the use[9] of absorbents involves six basic operations, the supply, storage, and transportation of the material and then the application, harvesting and disposal of the contaminated absorbent." The manual further observes that some of the early applications of sorbents such as the Torrey Canyon and Santa Barbara suffered because of the lack of effective and efficient harvesting techniques.

More recently, since the early 1990s a new approach to oil pickup was conceived by the use of SOLIDIFIERS.

Solidifiers are products which, when mixed with oil, turn the oil into a coherent mass. They are usually available in dry granular form. Unlike sorbents that physically soak up liquid, solidifiers bond the liquid into a solid carpet-like mass with minimal volume increase, and retain the liquid for easy removal. The bonded material also eliminates dripping-sponge effect by not allowing the material to be squeezed out, minimizing residue or contamination. Some polymers, in sufficient quantity or of high molecular weight, can actually convert the oil to a rubber-like substance.

Solidifiers are most commonly used during very small oil spills on land or restricted waterways to immobilize the oil and enhance manual recovery. There has been little documented use of solidifiers on large spills or open water. However, the possibility that they may reduce the spread of waterborne oil by solidifying it and increase recovery and removal rates is a concept with significant potential benefit.

The effectiveness of a solidifier is based on the amount of product and time it takes to "fix" a given volume of oil. The less effective products require larger amounts to solidify oil. Fingas *et al.* (1994) presented results from effectiveness tests on various solidifiers and found that generally between 13–44 percent by weight of the product to oil was required to solidify Alberta Sweet Crude over a 30-minute period.

The entire treatment of solidifiers as an aid to oil spill response is well covered in an MSRC publication.[10]

Physical Removal of the Contained Oil by Oil Pickup Devices Since oil containment booms have a fixed capacity for oil spill containment, it is important to consider means to physically remove the contained oil from the surface. The use of sorbents has been discussed. An alternate approach is to remove the fluid oil by means of skimming devices.

Oil skimmers have been divided into five categories:[11]

- Oleophilic surfaces (belts, disc, ropes, and brushes, either acting independently, mounted on a vessel or used in combination with a boom)
- Weirs (simple, self-leveling, vortex assisted, auger assisted, vessel-mounted, and weir/boom systems)
- Vacuum units (portable units and truck-mounted units)
- Hydrodynamic devices (hydrocyclone and water jet types)
- Other methods (including paddle belt and net trawl).

The selection of the optimum skimmer for a particular spill is based on site conditions such as the sea state and characteristics of the spilled oil e.g. viscosity and emulsion-forming tendency.

There are over 100 commercially available skimmers on the market that fall within the generic types previously mentioned. These are summarized in publications such as the WORLD CATALOG OF OIL SPILL RESPONSE PRODUCTS.

For example, for the principle of oleophilic surfaces, these can comprise either a sorbent belt, an oleophilic rope or a solid oleophilic disc that rotates through the surface oil film. In heavy sea conditions this type would be more effective than a wier type that is more suited to protected in-shore areas.

Controlled Burning of the Spilled Oil Burning represents a surface treatment of an oil spill that is attractive in that the oil is essentially removed from the water. However, some of the negative aspects of this approach that have hampered its widespread acceptance and use may be summarized as follows:

In many spill instances, there is an obvious concern regarding the combustion of the oil for safety reasons. Spills near harbors, tankers, offshore platforms would create an obvious hazard if set afire.

A minimum thickness of oil is required to establish combustion.

Air pollution is a concern in some instances. There is continuing evaluation and development burning of agents.

As reported by Alan Allen,[12] there are fire retardant booms and ignition methods available to burn the oil under proper conditions e.g. oil film thickness and amount of emulsified water in the oil. An effective burn after the EXXON VALDEZ spill on Sat. March 25, 1989 was reported by Allen in this publication.

The very encouraging burn rate statistics suggest that only 2% of the original relatively fresh oil remained as residue.

In this regard, it is relevant to quote the author of this publication in its entirety because of its concise and sufficient analysis of this technique by one well recognized in this method.

"It should be recognized that the elimination of spilled oil using in-situ burning must be considered in light of the full range of potential impacts (safety, air quality, etc.) associated with the burning of oil an water. The mechanical removal of spilled oil is by far the preferred cleanup technique whenever possible. Burning, on the other hand, may provide a safe, efficient and logistically simple method for eliminating oil under certain conditions. As a backup for mechanical cleanup techniques, in-situ burning can provide a useful means of eliminating large quantities of oil quickly, while avoiding the need for recovered oil storage containers. Anyone considering the use of burning should be sure that all regulatory controls have been satisfied, that the ignition and burning operations can be carried out safety, and that the temporary reductions in local air quality represent the lower of all other environmental impacts should the spilled oil not be burned."

TREATMENT OF THE SURFACE OIL

Chemical Treatment of Surface Oil Should Be Considered as an Alternate Solution

It is generally agreed, as indicated above, that situations can arise where the spill cannot be contained and recovered because sea conditions, weather state, and so on, are beyond the current operating capability of containment devices. There are also instances wherein the logistics of containment and recovery equipment, that is, containment boom availability and/or deployment time and effort, could indicate chemical treatment as the most practical and expedient handling technique. When physical recovery of the oil pollutant is impractical, there are, in effect, two courses of action possible. In one case, the oil may be permitted to remain as intact cohesive slick on the surface of the water and possibly reach shore. The alternate course is to "treat" this surface oil—such treatment essentially directed toward the removal of the oil from the water surface and the enhancement of its ultimate removal from the environment. This many be accomplished by chemical dispersion.

The Ecological and Economic Damage Caused by an Untreated Oil Spill Can Be Extensive

The damage resulting from an untreated oil spill is both visually apparent and extensive. It encompasses both biological as well as property damage. The potential damage may be summarized as follows:

Marine fowl, particularly diving birds, are particularly vulnerable to an oil spill. As reported by Nelson-Smith,[13] sea birds are most obvious victims of an oil spill due to "mechanical damage." The oil penetrates and clogs the plumage which the bird depends upon for waterproofing and heat insulation. For example, a duck with oil-impregnated plumage is under the same stress at a moderate temperature of +59°F as a normal bird would be under a more severe temperature condition at –4°F. Some statistics regarding bird damage have been cited by McCaull.[14] More than 25,000 birds, mostly guillemots and razorbills, were killed after the Torrey Canyon grounding. The guillemot casualties equaled the entire breeding stock between the Isle of Wright and Cardigan Bay. Bird losses in the Santa Barbara spill, according to the state Department of Fish and Game, totaled 3500.

Shore contamination by beached oil represents biological, as well as property damage. The tendency of oil to cling to shore surfaces, such as beach sand, sea walls, and the resultant property damage, are well established. This is perhaps the most apparent and widely publicized damaging aspect as attested by lawsuits on the part of tourist interests, property owners, etc. There is also, in a biological sense, a physical smothering effect on some attached, intertidal organisms such as mussels and barnacles. The effects of untreated oil coming ashore is well illustrated by Blumer *et al.*[15] regarding a No. 2 diesel fuel spill from the barge Florida in Buzzards Bay, Massachusetts in September 1969. Oil was incorporated into the bottom sediment to at least 10 meters of water depth, testifying to the wetting effect of untreated oil in this instance, the oil was physically dispersed by the heavy seas but retained its adhesive characteristics. Therefore, it is deduced that the oil droplets probably came into contact with and wetted and upswept, suspended particulates which later settled again to the bottom. Other spill instances depicting the importance of this aspect that of the incorporation of oil into the sediment-have been reported by Murphy.[16] In the Buzzards Bay and several other spill incidents of distillate fuels cited by Murphy, there has been a significant kill of all marine life in the area since these highly aromatic products are known to be much more toxic than whole crude oil.

Persistent tarry agglomerates are formed as the spilled oil weathers at sea. There has been increasing attention directed to the presence of tar-like globules ranging up to 10 cm in diameter in the open sea. As reported by Baker[17] during the voyage of Thor Heyerdahl's papyrus boat, Ra, during five separate days, they sailed through masses of these agglomerates whose age could be substantiated by the growth of goose barnacles adhering to them. There have been other incidents reported recently by the International Oceanographic Foundation[18] and a well documented survey was made by the research craft, R.V. Atlantis, as reported by Horn *et al.*[19] In this latter investigation, tarry agglomerates were present in 75% of over 700 hauls with a surface skimming (neuston) net in the Mediterranean Sea and eastern North Atlantic. The amount of tar in some areas was estimated at 0.5 milliliter in volume per square meter of sea surface.

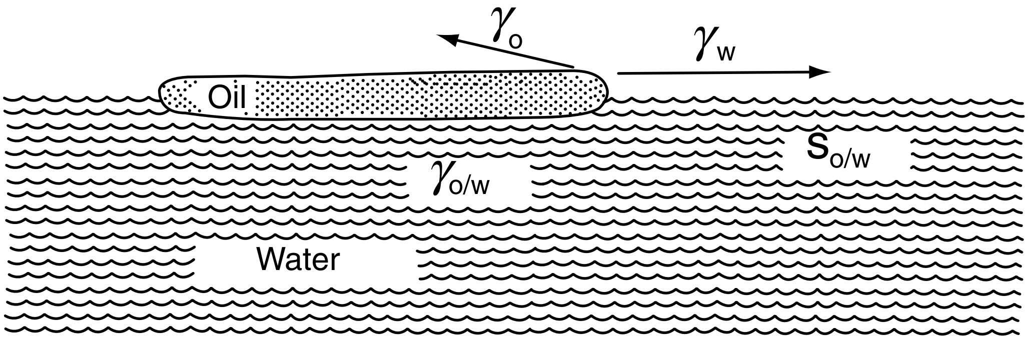
The Behavior of Spilled Oil at Sea

Before consideration of the mechanism of dispersing oil and its associated effects, an understanding of the behavior of spilled oil at sea will be useful. When a volume of oil is spilled onto the surface of water, the oil has a driving force to film out or spread-in essence, a spreading pressure usually expressed as a Spreading Coefficient. This Spreading $S_{o/w}$, is readily quantified and is determined by a balance of the surface tension forces as follows.

$$S_{o/w} = \gamma_w - \gamma_{o/w} - \gamma_{o,} \tag{1}$$

wherein:

$S_{o/w}$ is the spreading coefficient for oil on water ergs/cm^2 or dynes/cm



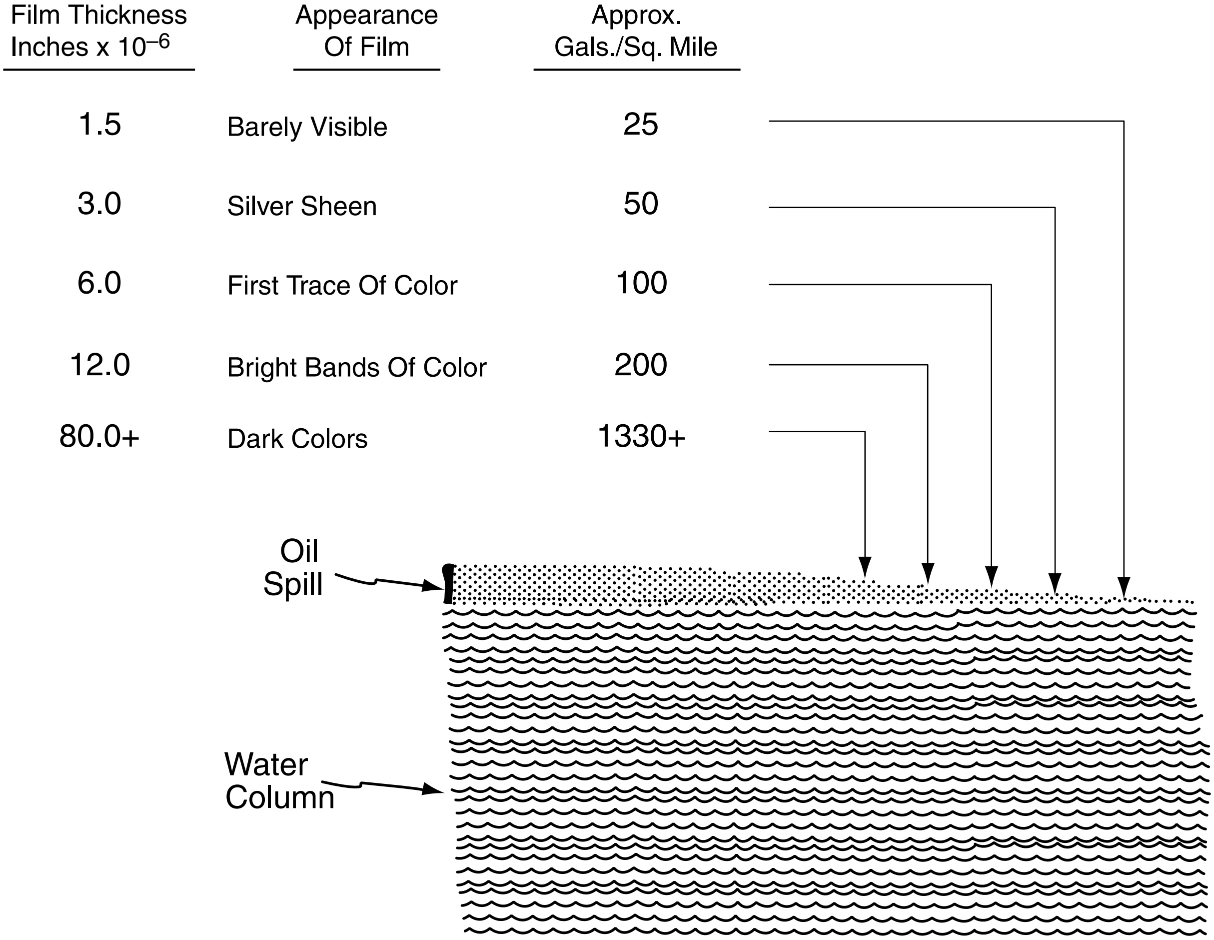
$S_{o/w}$, Spreading Coefficient For Oil On Water,

$$= \gamma_w - \gamma_o - \gamma_{o/w}$$

Measured Value For Kuwait Crude Oil On Sea Water

$$S_{o/w} = 61 - 28 - 22$$
$$= 11 \text{ Dynes/cm}$$

FIGURE 2 The spreading behavior of spilled oil.



Film Thickness Inches x 10^{-6}	Appearance Of Film	Approx. Gals./Sq. Mile
1.5	Barely Visible	25
3.0	Silver Sheen	50
6.0	First Trace Of Color	100
12.0	Bright Bands Of Color	200
80.0+	Dark Colors	1330+

FIGURE 3 Oil slick appearance during spreading.

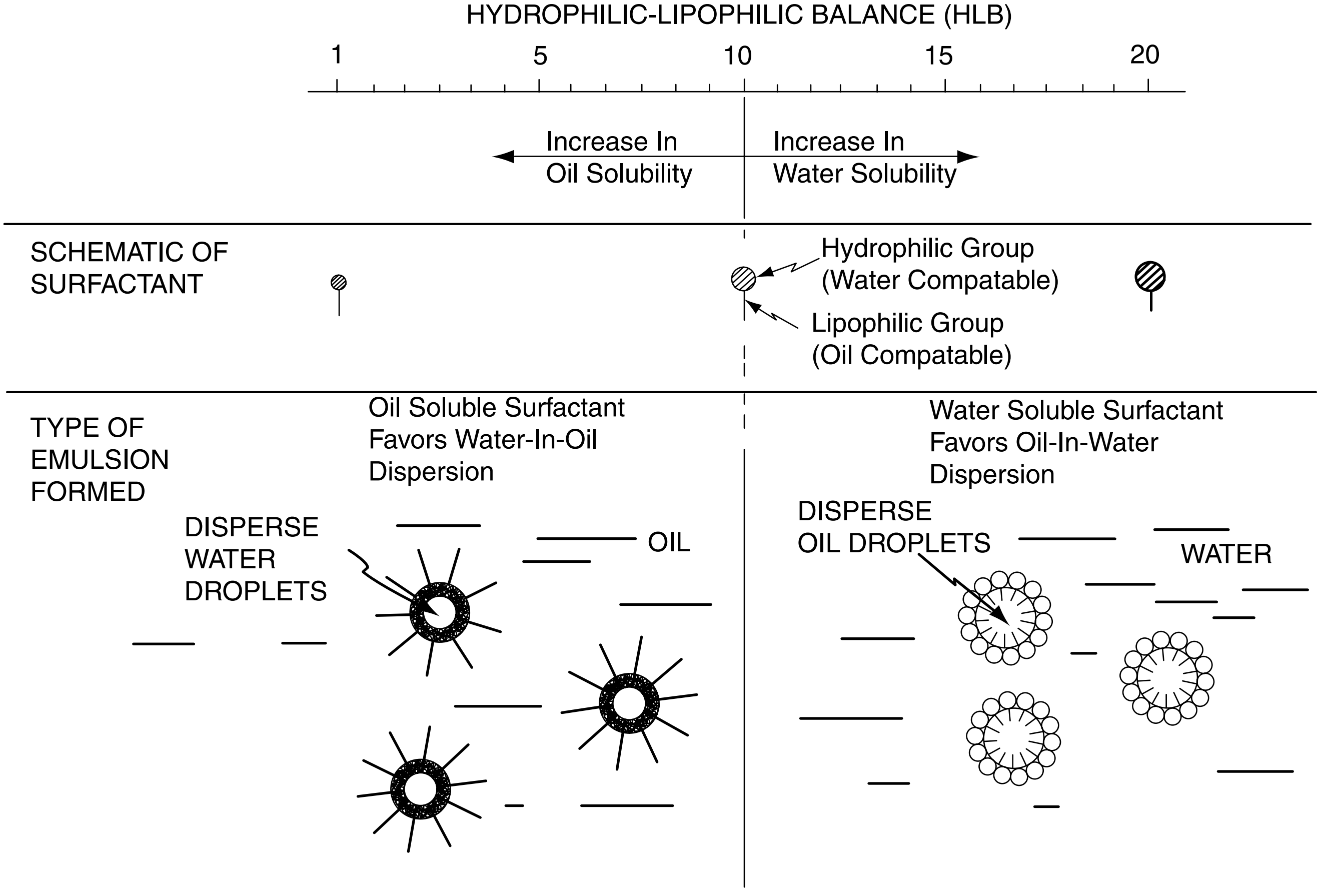


FIGURE 4 Influences of surfactant structure on type of dispersion.

γ_w is surface tension of water, dynes/cm
γ_o is surface tension of oil, dynes/cm
$\gamma_{o/w}$ interfacial tension of oil and water, dynes/cm.

By an examination of the force balance shown in Figure 2 it can be seen that if $S_{o/w}$—the resultant spreading force is positive, the oil will spread on the water; if negative, it will not spread but remain a "lens" of Liquid. For example, spreading coefficient values for Kuwait Crude on sea water, reported by Canevari[20] are positive and confirm that for this system the oil readily spreads on the water phase. Garrett[21] has summarized spreading pressures of various oils on sea water that vary from 25 to 33 dynes/cm. Cochran[22] has also published values that generally agree with these level on sea water. As one can see from Figure 3, for positive spreading coefficients, the oil is capable of filming out to very thin films. A film thickness of only 3.0×10^{-6} inches representing a spill of 50 gallons of oil distributed over a surface area of one mile will be quite visible as a "flat" silver sheen on the surface of the water.

However, the initial spreading rate of a large volume of spilled oil is based on the volume and density of the oil in essence, sort of static head that overcomes other factors such as interfacial tension.

The Mechanism of Dispersing Surface Oil Slicks by Chemical Dispersants

The dispersion of surface oil films as fine oil droplets into the water column is promoted by the use of a chemical dispersant. This oil spill dispersant consists primarily of a surface active agent (surfactant) and a solvent. The solvent is added as a diluent or vehicle for the surfactant. It also reduces the viscosity and aids in the uniform distribution of the surfactant to the oil film.

A surfactant is a compound that actually contains both water compatible (hydrophilic) and oil compatible (lipophilic) groups. Due to this amphiphatic nature, a surfactant locates and arranges itself at an oil–water interface as schematically shown in Figure 4. The surfactant's molecular structure, e.g. ratio of hydrophilic to lipophilic portion, determines the type of dispersion (oil droplets dispersed in water phase or water droplets dispersed in oil phase), as well as stability of the dispersion. In essence, a surfactant that is principally water soluble disperses oil-in-water and established water as the continuous phase; a surfactant that is principally oil soluble, the converse. This is Bancroft's Law,[23] which has been tested and proven empirically true over the years. A convenient classification for surfactants therefore, is based on the ratio or balance of the water compatible portion to the oil compatible portion-sometimes referred to as HLB (Hydrophilic–Lipophilic Balance).[24]

This relationship between the molecular structure of the surfactant and the emulsion type is also shown in Figure 4 and the physical concept behind Bancroft's Law may be appreciated. For example, it can be visualized that for a more water compatible surfactant, the physical location of the larger hydrophilic group on the outside of the dispersed oil droplets results in a more effective "fender" to parry droplet collisions and prevent droplet coalescence. The converse, location and the larger portion of the surfactant in the dispersed rather than the continuous phase, would be geometrically awkward and unstable.[25] The mechanism of oil slick dispersion by the application of chemical dispersants has been covered in some detail by Poliakoff[26] and Canevari,[27,28,29] among others. From the above discussion, one can see that the chemical dispersant (surfactant) will locate at the oil–water interfaced reduce interfacial tension. This will then act to increase the spreading tendency of the oil film as shown by Eq. (1). More important, it promotes fine droplet formation which can be expressed as:

$$W_k = A_{o/w}\gamma_{o/w,}, \qquad (2)$$

where:

Wk mixing energy, ergs
$A_{o/w}$ interfacial area, cm^2
$\gamma_{o/w}$ interfacial tension, dynes/cm.

Thus, for the same amount of mixing energy, a reduction of $\gamma_{o/w}$ will result in a corresponding increase in $A_{o/w}$.

It is important to emphasize that, as can be realized from the above discussion, the chemically dispersed oil does not sink. Rather, the surfactant merely enhances small droplet formation for a given amount of mixing energy. Smaller diameter oil droplets have a much lower rise velocity per the familiar Stokes Law. Hence, once the oil is chemically treated, and placed 3 to 5 feet below the surface of the water by the mixing process, it does not rise to the surface as readily, as illustrated by Figure 5.

There are many surfactants that will aid the formation of fine droplets in the above manner. It has already been noted that the surfactant structure (Hydrophilic–Lipophilic Balance) influences the efficiency of the emulsifier. However, a more subtle and less tractable requirement for an effective dispersant is the prevention of droplet coalescence once the fine oil droplets are formed. This is illustrated by Figure 6 wherein a volume of oil has been dispersed by a chemical surfactant and maintained in suspension by gentle bubbling of air. After 24 hours, there has been no coalescence or separation of these fine oil droplets. In the control sample, with similar volume of oil and mixing energy, the oil separated almost immediately and reformed an intact, cohesive film of oil.

In essence then, an effective dispersant must parry droplet collisions physically. For example, dispersed oil may separate in a sample bottle but even though there may be a "creaming" effect, i.e. oil droplets concentrate near the surface, the droplets should not coalesce to reform an intact slick. It is this same "fendering" action that reduces the tendency of the droplets to stick to a solid surface.

The Physical and Environmental Incentives for Dispersing Oil Slicks

Consideration of the previous summary of the potential damaging aspects of an untreated and unrecoverable oil spill indicates that the removal of the intact, cohesive mass of oil from the surface of the water yields more than a cosmetic effect as is often claimed. For this alternate approach when conditions do not permit the recovery of the spilled oil, the removal of oil from the surface by dispersing it into fine droplets yields established benefits that can be summarized by the following discussion:

1) Oil properly dispersed with a chemical dispersant will not stick to a solid surface. As previously outlined, the physical fending action of a properly selected surface-active agent prevents the oil droplets from coalescing after dispersion. This same property also inhibits the oil from wetting out on a solid surface. This has become a controversial point and it has actually been claimed that the converse is true. For example, in the First Report of the President's Panel on Oil Spills,[30] it has been stated that such agents cause the oil to "spread into the sand-surfaces which untreated oil would not wet." A laboratory experiment was conducted to evaluate this aspect. A mixture of 256 cc of sea water, 95 cc of beach sand (New Jersey shore area), and 20 cc Kuwait Crude, were placed in a graduate. This represented a vertical cross section of the marine environment after an oil spill. The mixture was then agitated to simulate the possible contact of sediment by the oil when turbulent conditions existed. After mixing, the sample was settled to separate the oil–sand water phases. In a body of water, either the oil may be driven down into contact with the sandy bottom or the sand may be suspended in the body of water by wave action, such as deduced from the previously cited Buzzards Bay spill. The graduate was then purged with clean water to simulate the return of the environment to a non-contaminated condition.

   The experiment was then repeated using 20 cc of Kuwait Crude Oil and 4 cc of a chemical dispersant (5 parts oil/1 part dispersant).

   Virtually no "treated" oil impregnated the sand. For the experiment with the untreated crude oil, an analysis of the oil content of the sand bed indicated that 11.20 cc of oil remained of the initial 20 cc.
2) Oil removed from surface water prevents bird damage. The aforementioned hazard to marine fowl that is presented by the surface oil film is

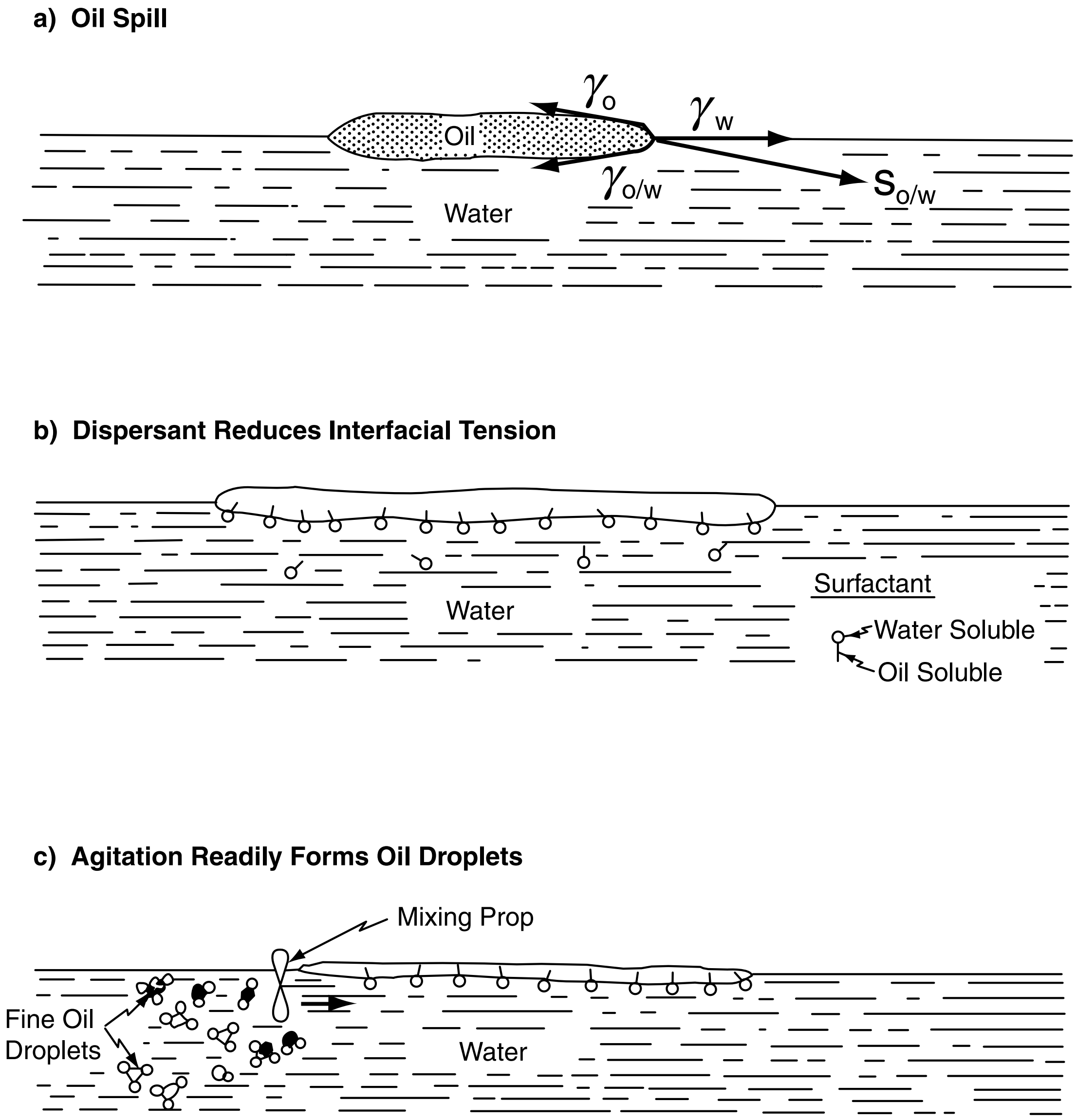


FIGURE 5 Dispersant enhances droplet formation.

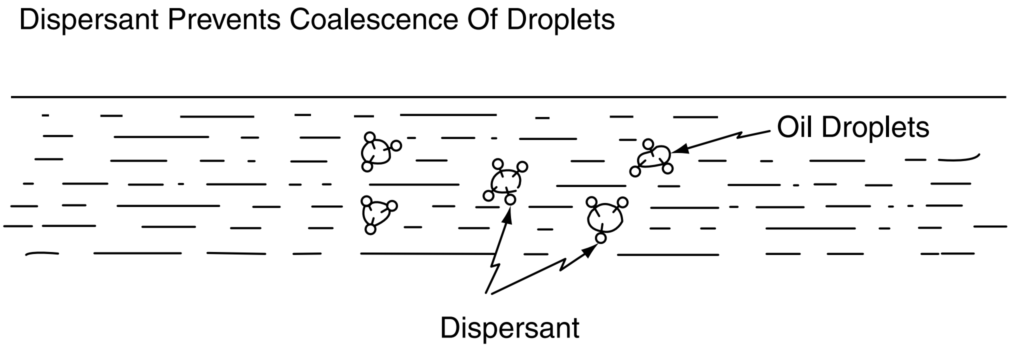


FIGURE 6 Dispersant maintains oil droplets in suspension with mild agitation..

clearly eliminated when the oil is dispersed as fine droplets into the water column. These dispersed droplets are placed several feet below the water surface by the mixing process.

3) The fire hazard from the spilled oil is reduced by dispersion of the oil several feet into the water column. The removal of this combustible material from the water's surface and from contact with the atmosphere prevent possible combustion of the spilled oil. This is perhaps the most accepted benefit accruing from the use of dispersants. It has provided the motivation for many past instances of dispersant applications.
4) The rate of biodegradation of the oil is enhanced. This is the historical basis for the dispersion of oil. It is perhaps the most significant contribution of dispersants. The order of magnitude increases in interfacial area that are generated by the dispersant greatly increases the rate of biodegradation of the oil. ZoBell[31] has reported biodegradation rates that are one or two orders of magnitude higher in laboratory experiments in which the oil is emulsified. Not only is the physical state of the oil, that is, small droplets, more conducive to bacterial action, but it is also made available to a much larger population of microbial organisms. This particular reference has been one of the most complete treatments of the subject of oil biodegradation to date.

   A study by Robichaux and Myrick[32] presented the results of a study of the effects of chemical dispersing agents on the rate of microbial destruction of crude oil in aqueous environments. Increased destruction rates of up to 15 times the rate of untreated oil/water mixtures were reported.
5) The formation of persistent tar lumps from an untreated oil spill is prevented. The tarry agglomerates (up to 10 cc dia.) found on the ocean surface, as mentioned previously represent a small percent residue of the crude oil. If the crude oil had been dispersed into 10 μ to 1 mm diameter droplets, these large residue agglomerates would not have formed and their persistence in the marine environment would have been greatly reduced.

The Concern Regarding the Chemical Dispersion of Oil Spills

Clearly then, from a consideration of the foregoing, the removal of oil from the surface of the sea has merits in mitigating the damage resulting from a spill. It is more than a cosmetic, hide-it-from view effect. What then are the negative aspects to their use? What is the ecological price for introducing the chemical dispersant and dispersed oil into the water column?

The major concern regarding the use of dispersants are twofold as covered in the following discussion.

The Toxic Effects of the Chemical Dispersants

This has been an area of great concern since their use has become significant. There is a basis for this concern. It was highlighted by the investigation by Smith *et al.*,[33] after the Torrey Canyon that indicated that in some areas, particularly in the intertidal zone, the chemicals used were more toxic to the marine life than the oil itself.

During this period (1967–1968), the chemical formulations available to disperse spilled oil were derived mainly from cleaning agents, hence the term "detergent" was used quite commonly. To permit these agents to dissolve tar-like residues and perform their cleaning function, an aromatic solvent, such as heavy aromatic naphtha, was generally employed. The short term acute toxicity of aromatic hydrocarbon to marine life is well-known. Blumer[34] states that low boiling aromatics are toxic to man as well as all other organisms and that it was the great tragedy of the Torrey Canyon that the detergents used were dissolved low boiling aromatics. The Toxicity of these aromatic solvent constituents were extensively studied by the Marine Biological Laboratory of the UK. Their acute toxicity was evident since 5 ppm of kerosene extract solvent killed 50% of the Elminius nauplius larvae in 21 minutes. Their analyses of the more common detergents (dispersants) used during the Torrey Canyon indicated that they contained some portion of aromatics.

In addition to these toxic aromatic solvents, the surfactants were typically selected from the class compounds formed by the reaction of hydroxy-containing compounds (e.g. phenol or alcohol) with ethylene oxide. A typical surfactant might be ethoxylated nonylphenol. The number of ethylene oxide groups added to the nonylphenol hydro-phobe may be controlled to any desired extent to adjust the degree of water solubility of the material. These types of surfactants, although effective emulsifiers, were quite detrimental to marine life.

However, there has been research directed toward formulating dispersants that would have little effect on marine life. For example, water is now used as the solvent in products where it is compatible with the particular surfactants. High boiling saturated hydrocarbons which are similar to the type of hydrocarbon that occur naturally in the marine environment have a low order of toxicity and are also employed as solvents in some of the more recent dispersants. This modification of the solvent, and the selection of generic types of surface-active agents that are not considered to be chemically toxic, have resulted in the development of dispersants that have greatly reduced toxicity. This can be illustrated by the study of J.E. Portmann,[35] summarized in Table 3. For example, three dispersant products used during the Torrey Canyon spill and identified as Torrey Canyon Dispersants A, B, C, have 48 hr LC_{50} values of 8.8, 5.8, and 6.6 ppm, respectively. These concentrations represent the amount of the specific agent to kill 50% of the test species (Crangon crangon) in 48 hours. The toxicity of a typical Torrey Canyon surfactant, ethoxylated nonlphenol is shown at 89.5 ppm. By

TABLE 3
Development of low toxicity dispersants illustrated by Portmann Study

Chemical	48 hour LC_{50}, ppm brown shrimp (Crangon Crangon)
Torrey Canyon Dispersant "A"	8.8
Torrey Canyon Dispersant "B"	5.8
Torrey Canyon Dispersant "C"	6.6
Post Torrey Canyon Dispersant "D"	7,500–10,000
Post Torrey Canyon Dispersant "E"	3,300–10,000
Post Torrey Canyon Dispersant "F"	3,300
Nonyl Phenol-Ethylene Oxide	89.5

TABLE 4
Summary of Canadian Fish. Res. Bd. evaluation of 10 dispersants

Classification	Numbers of dispersants	48 hour LC50, ppm Salmon (Salmo Salar L)
Toxic	8	1–100
Moderately toxic	1	100–1000
Slightly toxic	0	1000–10,000
Practically non toxic	1	>10,000

TABLE 5
Toxicity of dispersants with and without crude oil

Chemical	96 hour TLM, ppm Fathead minnow (Pimephales promelas)
Dispersant A	5.6
Dispersant A + oil	14.0
Dispersant B	14.0
Dispersant B + oil	27.0
Dispersant C	25.0
Dispersant C + oil	42.0
Dispersant D	32.0
Dispersant D + oil	44.0
Dispersant E	56.0
Dispersant E + oil	75.0
Dispersant F	3200^{+}
Dispersant F + oil	1800^{+}

comparison, the Toxicity levels of three dispersant products developed since the Torrey Canyon, identified as Post Torrey Canyon, Dispersants D, E, F, are 7500–10,000; 3300–10,000; and >3300 ppm, respectively. These concentrations are orders of magnitude greater than the level applied by conventional application in the field.

Other agencies have confirmed this finding. Table 4 illustrates results of a recent study by the Fisheries Research Board of Canada entitled, "Toxicity Tests with Oil Dispersants in Connection with Oil Spill at Chedabucto Bay N.S."[36] Again, the large difference in toxicity due to the surfactant-solvent recipe can be noted in the summary of results (Table 4). These values represent 4 day LC_{50} values in fresh water to Salmon (Salmo salar L) and vary from "Toxic" (1–100 ppm) to "Practically non-toxic" (>10,000 ppm). Over 25 research institutions are known to have conducted studies on these lower toxicity chemicals. Testing by Dr. Molly Spooner,[37,38] among others, has encompassed juvenile species, planktonic life and other very sensitive forms of marine life.

Clearly then, the concern and conclusion that all chemical dispersants are in themselves inherently toxic is incorrect. Some of the most effective emulsifiers/dispersants available are those derived from and found in the natural environment.

The Toxic Effects of the Dispersed Oil

When the surface film of oil is dispersed several feet or more into the water column, it is unfortunately made available to other forms of marine life in addition to the hydrocarbonoxidizing bacteria. Necton and other filer feeder many now come into contact with dispersed oil droplets that they otherwise may have escaped as surface oil. This is, effect, the "ecological price" for the cited benefits of dispersing oil. There are published data on the acute toxicity levels of dispersed oil such as that from the State of Michigan[39] presented as Table 5. This does indicate an approximate tolerance level of a thousand ppm or more for dispersed oil. It can also be noted that the toxicity of the chemical is reflected in the toxicity level of 1000 ppm or so for dispersed oil, however it should be noted that (1) it is unlikely that fish would remain in this inhospitable environment for 96 hours and (2) the dispersed oil has a driving force to dilute itself. Of greater concern than these short term acute effects is the possibility that the finely dispersed oil droplets represent a more subtle contaminant and may cause long-range detrimental effects. However, it should also be noted that crude oil is a natural rather than man-synthesized material. Wheeler North[40] reported after extensive research into several spill incidents, "Unlike many of the products man liberates into the environment, crude oil is a naturally occurring substance. From time to time it appears on the earth's crust by natural processes of exudation."

More Recent Dispersant Research Has Involved Improvement in Effectiveness

The previous discussion regarding the dispersion mechanism cited the need for mixing energy, W_k. This is normally supplied by means of a work boat applying the chemical. However, consider the rate by which this work is accomplished by the boat's wake and propeller. A typical work boat may apply energy to swath 50 ft wide at a speed of 5 knots thereby only mixing 35 acres per hour of ocean.

Therefore, in recent years, research has been directed at eliminating the need for the tedious, time consuming mixing process.

In essence, a "self-mix" dispersant formulation has been developed that requires essentially no energy to be applied to the oil-water interface in order to generate a dispersion of fine oil droplets. This has greatly enhanced the scope and potential of chemical dispersion particularly for large spills. For example, since mixing is no longer needed, aerial application alone would be feasible. Some aircraft uniquely adapted for this service, such as the canadiar CL-215, carries 1500 gallons of dispersant and covers 3000 acres per hour based on a 150 knot speed and treated swath width of 150 feet. Extensive use has already been made of commercial DC-4's and DC-6's for this purpose. A very novel development of a load on tank and spray system for even larger aircraft is now in place.

The Mechanism of More Recently Developed Self-Mix Dispersants

The mechanism of the self-mix chemical dispersants goes beyond the simple thesis represented by Eq. (2). In an ideal no-mixing system true spontaneous emulsification (or "self-mixing") is postulated to occur in the following manner. The chemical surfactant formulation is made compatible with the bulk oil. However, when the oil phase comes into contact with a water boundary rather than air, part of the surfactant

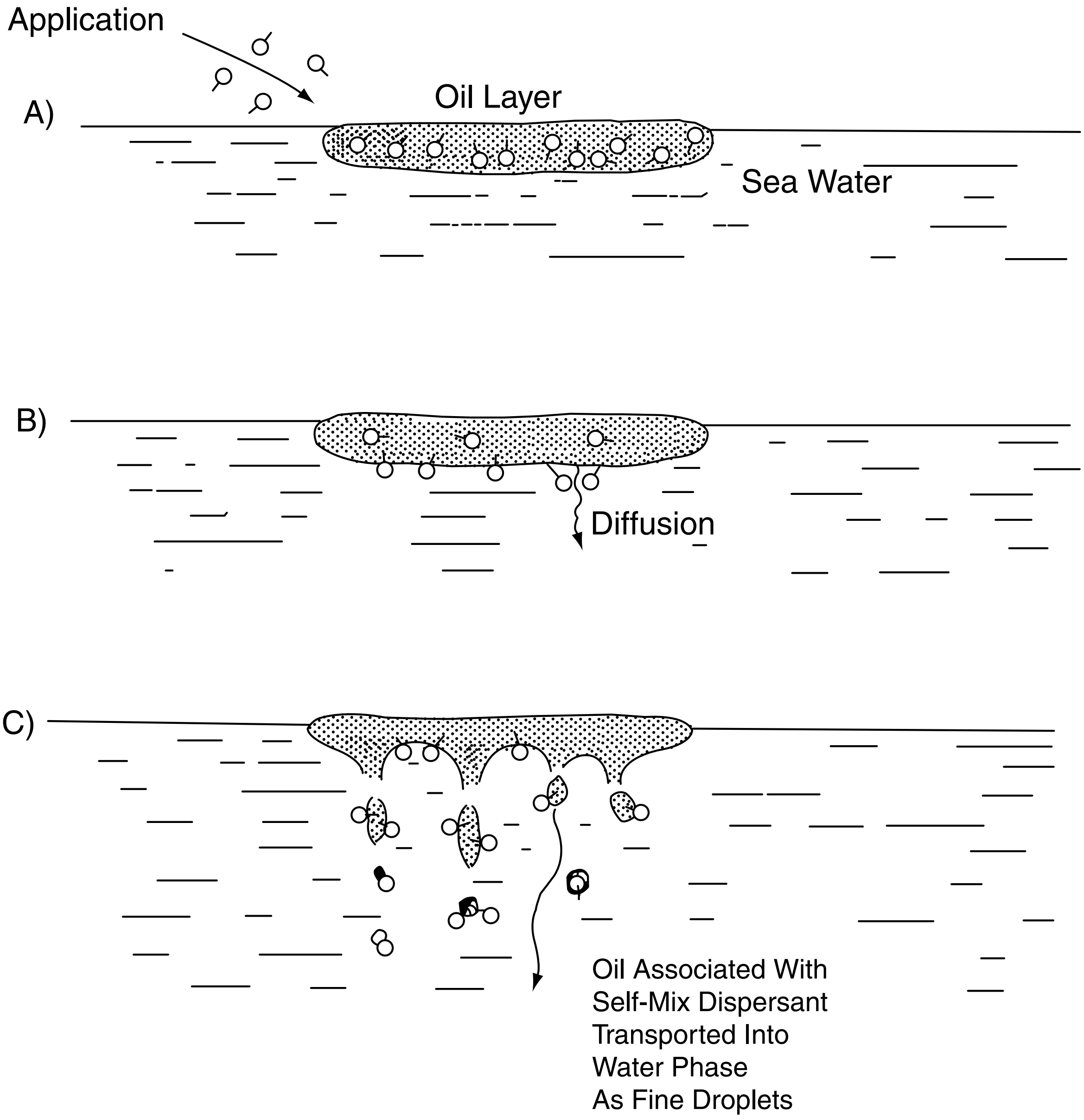


FIGURE 7 Mechanism of self-mix dispersion.

has a strong driving force to diffuse into the water phase. In this transport process, a small amount of oil "associated" with the surfactant is carried into the water phase. A continuation of this process produces a series of fine oil droplets migrating from the oil phase into the water phase as schematically shown by Figure 7.

In the graphical presentation of Figure 7, the surfactant formulation can be seen to be compatible with the crude oil phase as shown in (A). However, due to the nature of the specific compounds, there is a driving force for part of the formulation of diffuse into the water phase when it contacts an oil/water interface (B). During this diffusion, some oil associated with the surfactant as fine oil droplets is carried along with the surfactant into the water column as shown in (C). In essence, a three component system—oil + water + surfactant is formed at the interface. As the surfactant diffuses into the water phase, the associated oil is thrown out of solution.

The migration of the surfactant from the oil into the water phase-in essence, the source of energy for spontaneous emulsification comes from the redistribution of materials. It can be seen that for this system to work in the field as an oil slick dispersant, the surfactant must be brought into contact with the oil phase initially.

It is also interesting to observe that as the surfactant diffuses through the interface, a reduction in interfacial tension occurs. Over the entire oil/water interface, there are dissimilar values of interfacial tension due to the somewhat random diffusion of the surfactant at varying sites along the interface. Any difference in interfacial tension produces a spreading pressure, II, which causes rapid movement of the interface. This interfacial turbulence also aids in the dispersion of the oil into the water phase.

Field Tests Support the Role of Chemical Dispersants to Minimize Oil Spill Impact

In summary, there is an increased awareness and recognition that there is a role for chemical dispersants in minimizing damage from oil spills. The improved effectiveness afforded by the self-mix dispersant system has been demonstrated.

Over the past 10 years, there have been a number of major field tests that have demonstrated under real life conditions the effectiveness and biological safety of this approach. These have been reviewed and summarized in a study by the National Research Council.[41]

In order to establish that the transient, rapidly diluting concentrations of dispersed oil are not harmful, actual measurements of the biological effects were made during several controlled oil spills.

For examples, on August 19, 1981 a field experiment was carried out in Long Cove, Searsport, Maine, which simulated the dispersal of oil slicks in the nearshore zone.[42] The object of this experiment was to obtain quantitative information on the fate and effects of dispersed and non-dispersed oil in the nearshore area. An upper and lower intertidal sampling are within a 60 × 100 meter test plot were exposed to dispersed oil in water resulting from the discharge of 250 gallons of oil premixed with 25 gallons of COREXIT 9527 dispersant. Release of treated oil was around high-water slack tide on the surface of the water. The maximum water depth over the test areas was 3.5 meters. Untreated crude oil (250 gallons) was released on an ebbing tide within a separate, boomed-off 60 × 100 meter test plot. A third test plot served as an oil-free reference plot. To evaluate the effects on the intertidal infaunal community structure, chemical and biological analyses were carried out concurrently throughout the pre- and post-spill periods. The conclusions reached by the Bowdoin College scientists are quoted as follows:

- No evidence of any adverse effects was observed on infaunal community structure from the exposure of intertidal sediments to dispersed oil under real spill treatment conditions.
- There is clear evidence that the undispersed oil treatment caused some mortality of a commercially important bivalve and increased densities of opportunistic polychaetes.
- The results seen in the test plot that received untreated oil, are consistent with studies of real-world oil spills.

REFERENCES

1. Corino, E.R., E.F. Broderick and G.P. Canevari, Method of gelling tanker cargoes, US Patent 3,634,050. Issued January 11, 1992.
2. Corino, E.R., Chemical gelling agents and dispersants. Paper presented to the Third Joint Meeting of the American Institute of Chemical Engineers and Puerto Rican Institute of Chemical Engineers, May 20, 1970.
3. Department of US Navy, The recovery of bunker C fuel oil from the sunken tanker, SS ARROW, Navships 0994–008–1010, March 1970.
4. Lehr, W.E. and J.O. Scheren, Jr., Design requirements for booms, Proc. of API and FWPCA Joint Conference on Control of Oil Spills, NYC, New York, December 1969.
5. Hoult, David P., Containment and collection devices for Oil slicks, Oil on the Sea, Plenum Press, 1969.
6. Hoult, David P., Containment of Oil Spills by Physical and Air barriers, paper presented on the Third Joint Meeting of the American Institute of Chemical Engineers and the Puerto Rican Institute of Chemical Engineers, May 20, 1970.
7. Struzeski, E.J., Jr. and R.T. Dewling, Chemical treatment of oil spills, Proc. of API and FWPCA Joint Conference on Control of Oil Spills NYC, December 15–17, 1969.
8. Schatzbertg, Paul and K.V. Nagy, Sorbents for oil spill removal. Proc. of API and EPA Joint Conference on Prevention and Control of Oil Spills, Washington, DC, June.
9. Subcommittee on Marine Pollution IMCO, National arrangements for dealing with oil pollution preparation of a manual for the guidance of governments, March 2, 1992.
10. Chemical Oil Spill Treating Agents MSRC. Technical Report Series 93–015, 1993.
11. Oil Spill Response Manual, Exxon Production Research Co. page 77, 1992.
12. Allen, Alan, Comparison of Response Options for Offshore Oil Spills, 11th Annual AMOP Seminar Vancouver British Columbia, June 7–9, 1988.
13. Nelson Smith, A., Effects of oil on plants and animals, Proc., Seminar on Water Pollution by Oil, Aviemore, Scotland, May 4–8, 1970.
14. McCaull, Julian, The black tide, Environment, November 1969.
15. Blumer, M., G. Souza, and J. Sass, Hydrocarbon pollution edible shellfish by an oil spill, Marine Biology, 1970.

16. Murphy, Thomas A., Environmental aspects of oil pollution, paper presented to the Session on Oil Pollution Control, ASCE, Boston, Massachusetts, July 13, 1970.
17. Baker, Norman, The life and death of the good ship RA, Sports Illustrated, April 20, 1970.
18. Sea Secrets. International Oceanographic Foundation, 14, No. 4, p. 2, July–August 1970.
19. Horn, Michael H., John M. Teal, and Richard H. Backus, Petroleum lumps on the surface of the sea, Science, 168, pp. 245–246, April 10, 1970.
20. Canevari, Gerard P., The role of chemical dispersants in oil cleanup, Oil on the Sea, Plenum Press, 1969.
21. Garrett, William D., Confinement and control of oil pollution on water with monomolecular surface films, Proc. of API and FWPCA Joint Conference on Control of Oil Spills, NYC, NY, December 15–17, 1969.
22. Cochran, Robert A. and Paul R. Scott, The growth of oil slicks and their control by surface chemical agents, *J. Petroleum Technology,* July 1971.
23. Bancroft, W.D., *J. Phys. Chem.,* **17**, p. 501, 1913; **19**, p. 275, 1915.
24. Becker, P., Emulsions: Theory and Practice, Reinhold Publishing Corp., NY, 1957.
25. Canevari, G.P., Some basic concepts regarding the separation of oily water mixtures, ASLE Transactions, pp., 190–198, July 1968.
26. Poliakoff, M.Z., Oil dispersing chemicals, Water Pollution Control Research Series ORD-3, Washington, DC, May 1969.
27. Canevari, Gerard P., The role of chemical dispersants in oil cleanup, Oil on the Sea, Plenum Press, pp. 29–51, 1969.
28. Canevari, Gerard P., General dispersant theory, Proceedings of Joint Conference on Prevention and Control of Oil Spills, API/FWQA, New York City, New York, Dec. 1969.
29. Canevari, Gerard P., Oil spill dispersants-Current status and future outlook, Proc. of API and EPA Joint Conference on Prevention and Control of Oil Spills, Washington, DC, June 15–17, 1971.
30. First Report of the President's Panel on Oily Spills, Executive Office of the President, Office of Science and Technology, Washington, DC, 1970.
31. ZoBell, Claude E., The occurrence, effects and fate of oil polluting the sea, Int. Journal Air Water Pollution, pp. 173–198, Pergamon Press, 1963.
32. Robichaux, T.J. and H.N. Myrick, Chemical enhancement of the biodegradation of oil pollution, paper presented at the Offshore Technology Conference, Dallas, Texas, April 19–21, 1971.
33. Smith, J.E., Torrey Canyon Pollution and Marine Life, Cambridge University Press, 1968.
34. Blumer, Max, The extent of marine oil pollution, Oil on the sea, Plenum Press, pp. 29–51, 1969.
35. Portmann, J.E., The toxicity of 120 substances to marine organisms, Fisheries Laboratory, Burnham-on-Crouch, Essex, England, September 1970.
36. Sprague, John B. and W.G. Carson, Toxicity tests with oil dispersants in connection with oil spill at Chedabucto Bay, NS Fisheries Research Board of Canada, St. Andrews, NB, 1970.
37. Spooner, M.F. and G. Malcolm Spooner, The problems of oil spills at sea, Marine Biological Association of the UK. Plymouth, England 1968.
38. Spooner, M.F., Preliminary work on the comparative toxicities of some oil spill dispersants and a few tests with oils and COREXIT, Marine Biological Association of the UK, Plymouth, England 1968.
39. A biological evaluation of six chemicals used to disperse oil spills, Department of Natural Resources, State of Michigan (1969).
40. Mitchell, Charles T., Einar K. Anderson, Lawrence G. Jones, and Wheller J. North, What oil does to ecology, journal WPCE, 42, No. 5, Part 1, May 1970, pp. 812–818.
41. Marine Board Commission on Engineering and Technical Systems National Research Council. "Using on Spill Dispersant on the Sea" National academy Press, 1989.
42. Gilfillan, E.S., D. Page, S.A. Hanson, J.C. Foster, J.R.P. Gelber, and S.D. Pratt. 1983. Effect of spills of dispersed and non-dispersed oil on intertidal infaunal community structure. Proc. 1983 Oil Spill Conference Washington, D.C: API. pp. 457–463.

GERARD P. CANEVARI
G.P. Canevari Associates

P

PARTICULATE EMISSIONS

EMISSION STANDARDS

Allowable levels of particulate emissions are specified in several different ways, having somewhat different methodologies of measurement and different philosophies of important criteria for control. Permissible emission rates are in a state of great legislative flux both as to the definition of the suitable measurement and to the actual amount to be allowed. This section summarizes the various types of quantitative standards that are used in regulating particulate emissions. For a detailed survey of standards, the reader should consult works by Stern,[1] Greenwood *et al.*,[2] and the Public Health Service.[3]

A recent National Research Council report proposes future studies on the nature of particulate emissions, their effect on exposed populations and their control[4]. Friedrich and Reis[5] have reported the results of a 10-year multinational European study on characteristics, ambient concentrations and sources of air pollutants.

The following paragraphs give an overview of standards for ambient particulate pollution and source emission. The precise and practical methodology of making accurate and/or legally satisfactory measurements is beyond the scope of this article. Books such as those by Katz,[6] Powals *et al.*,[7] Brenchly *et al.*,[8] and Hawksley *et al.*[9] should be consulted for detailed sampling procedures. In the Federal Register USEPA announced the implementation of the PM-10 regulations (i.e., portion of total suspended particulate matter of 10 μm or less particle diameter).[40,41]

Ringlemann Number

Perhaps the first attempt at quantifying particulate emissions was developed late in the 19th century by Maximilian Ringlemann. He developed the concept of characterizing a visible smoke plume according to its opacity or optical density and originated the chart shown in Figure 1 as a convenient scale for estimation of opacity. The chart consists of four grids of black lines on a white background, having fractional black areas of 20, 40, 60 and 80% which are assigned Ringlemann Numbers of 1–4. (Ringlemann 0 would be all white and Ringlemann 5 all black.) For rating a smoke plume, the chart is held at eye level at a distance such that chart lines merge into shades of grey. The shade of the smoke plume is compared to the chart and rated accordingly. The history and use of the Ringlemann chart is covered by Kudlich[8] and by Weisburd.[9]

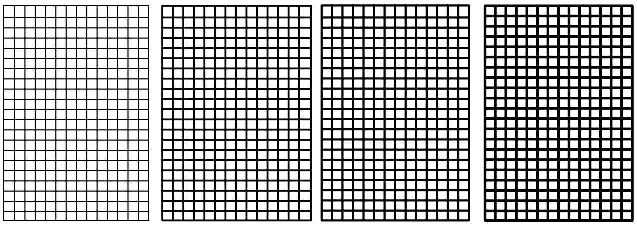


FIGURE 1 Ringlemann's scale for grading the density of smoke.

In actual practice, opacity is seldom determined by use of the chart, although the term Ringlemann Number persists. Instead, observers are trained at a "smoke school."[10] Test plumes are generated and the actual percentage of light attenuation is measured spec-trophotometrically within the stack. Observers calibrate their perception of the emerging plume against the measured opacity. Trained observers can usually make readings correct to ± 1/2 Ringlemann number.[11,13] Thus, with proper procedures, determination of a Ringlemann Number is fairly objective and reproducible.

The Ringlemann concept was developed specifically for black plumes, which attenuate skylight reaching the observer's eye and appear darker than the sky. White plumes, on the other hand, reflect sunlight and appear brighter than the background sky so that comparison to a Ringlemann chart is meaningless. The smoke school approach is quite applicable, however. Observations of a white plume are calibrated against the measured light attenuation. Readings of white plumes are somewhat more subject to variation due to relative locations of observer, plume, and sun. It has been found that observations of equivalent opacity taken with the observer facing the sun are about 1 Ringlemann number higher[13] than those taken in the prescribed method with the sun at the observer's back. Nevertheless, when properly made, observations of Ringlemann numbers are reproducible among observers and agree well with actual plume opacity.

Opacity regulations specify a maximum Ringlemann number allowable on a long-term basis but often permit this to be exceeded for short prescribed periods of time. For instance, a typical requirement specifies that emissions shall not exceed Ringlemann 1, except that for up to 3 min/hr emissions up to Ringlemann 3 are permitted. This allowance is of considerable importance to such operation as soot blowing or rapping of electrostatic precipitator plates, which produce puffs to smoke despite on overall very low emission level.

Federation regulations of the Environmental Protection Agency[14] specify that opacity observations be made from a point perpendicular to the plume, at a distance of between two stack heights and one quarter of a mile, and with the sun at the observer's back. For official certification, an observer under test must assign opacity readings in 5% increments (1/4 Ringlemann number) to 25 plumes, with an error not to exceed 15% on any single reading and an average error (excluding algebraic sign of individual errors) not to exceed 7.5%. Annual testing is required for certification. In view of previous studies,[11,13] this is a very high standard of performance and probably represent the limits of visual quantification of opacity.

Perhaps the greatest advantage of the Ringlemann Number approach is that it requires no instrumentation and very little time and manpower. Readings can usually be made by control authorities or other interested parties without entering the premises of the subject source. Monitoring can be done very frequently to insure continual, if not continuous, compliance of the source. Finally, in terms of public awareness of particulate emissions, plume appearance is a logical candidate for regulation. Air pollution is, to a great extent, an aesthetic nuisance affecting the senses, and to the extend that plume appearance can be regulated and improved, the visual impact of pollution is reduced.

The Ringlemann Number concept has drawbacks reflecting its simple, unsophisticated basis. Most serious is that, at present, there is no really quantitative relationship between stack appearance and the concentration of emissions. Additional factors; such as particle size distribution, refractive index, stack diameter, color of plume and sky, and the time of day, all have a marked effect on appearance. On a constant weight concentration basis, small particles and large smoke stacks will produce a poor Ringlemann Number. Plumes that have a high color contrast against the sky have a very strong visual impact that does not correspond closely to the nature of the emissions. For example, a white plume may be highly visible against a deep blue sky, but the same emission can be practically invisible against a cloudy background. As a result, it is often difficult to predict whether or not proposed control devices for a yet unbuilt plant will produce satisfactory appearance. Certain experience factors are presented in Table 1 for emissions, measured on a weight concentration basis, which the Industrial Gas Cleaning Institute has estimated will give a Ringlemann 1 or a clear stack.

A second objection is that Ringlemann number is a purely aesthetic measurement which has no direct bearing on physiological effects, ambient dirt, atmospheric corrosion, or any of the other very real and costly effects of particulate air pollution. There is some concern that regulations of very low Ringlemann numbers will impose very costly control measures upon sources without producing a commensurate improvement in the quality of the environment. Thus a high concentration of steam will produce a visually prominent plume, but produce virtually no other undesirable effects. Opacity restrictions are usually waived if opacity is due entirely to steam but not if any other particles are present, even if steam may be the major offender.

Instrumental Opacity

Many factors affecting the visual appearance of a smoke plume are external variables, independent of the nature of the emissions. In addition, visual reading cannot be taken at all at night; and manpower costs for continuous daytime monitoring would be prohibitive. For these reasons, instrumental measurements of plume opacity are sometimes desirable.

A typical stack mounted opacity meter is shown in Figure 2. It consists, basically, of a light source, an optical path traversing the smoke stack, and a phototube receiver which responds to the incident light intensity and, hence, to the light attenuation caused by the presence of smoke. Various techniques including beam splitting, chopper stabilization, and filter comparison are used to maintain stable baselines and calibrations. At present, however, there is no way to distinguish between dust particles within the gas stream and those which have been deposited on surfaces in the optical path. Optical surfaces must be clean for meaningful measurements, and cleanliness is difficult to insure for long periods of time in dusty atmosphere. The tendency, therefore, is for such meters to read high, indicating more smoke than is actually present. For this reason, and because of reluctance to have a continuous record of emissions, there has not been a very strong push by industries to supplant Ringlemann observations with opacity meters.

Stack mounted opacity meters, of course, will not detect detached plumes, which may contribute to a visual Ringlemann observation. Detached plumes are due to particles formed by condensation or chemical reaction after gas leaves the stack and are thus beyond detection of such a meter.

At present, Texas is the only state with emissions control regulations based on use of opacity meters,[15] as described by McKee.[11] The Texas regulations is written so that smoke of greater optical density (light attenuation per unit length of light path) is permitted from low velocity stacks or small diameter ones. Basically, a minimum transmittance of 70% is allowed across the entire (circular) stack diameter if the stack has an exit velocity of 40 ft/sec, and adjustment equations are provided for transmittance and/or optical path length if non-standard velocity or path length is used.

Perhaps the greatest dissatisfaction with emission regulations based either on visual observation number or on instrumental opacity is due to the fact that there is presently no

TABLE 1
Industrial process emissions expected to produce visually clear (or near clear) stack

Industrial classification	Process	Grains/ACF @Stack exit temp. (°F)
Utilities and industrial power plant fuel fired boilers	Coal—pulverized	0.02 @ 260–320
	Coal—cyclone	0.01 @ 260–320
	Coal—stoker	0.05 @ 350–450
	Oil	0.003 @ 300–400
	Wood and bark	0.05 @ 400
	Bagasse Fluid	0.04 @ 400
	Fluid code	0.015 @ 300–350
Pulp and paper	Kraft recovery boiler	0.02 @ 275–350
	Soda recovery boiler	0.02 @ 275–350
	Lime kiln	0.02 @ 400
Rock products—kiln	Cement—dry	0.015 @ 450–600
	Cement—wet	0.015 @ 450–600
	Gypsum	0.02 @ 500
	Alumina	0.02 @ 400
	Lime	0.02 @ 500–600
	Bauxite	0.02 @ 400–450
	Magnesium oxide	0.01 @ 550
Steel	Basic oxygen furnace	0.01 @ 450
	Open hearth	0.01–0.015 ≈450–600
	Electric furnace	0.015 @ 400–600
	Sintering	0.025 @ 300
	Ore roasters	0.02 @ 400–500
	Cupola	0.015 @ 0.02 ≈250–400
	Pyrites roaster	0.02 @ 400–500
	Taconite roaster	0.02 @ 300
	Hot scarfing	0.025 @ 250
Mining and metallurgical	Zinc roaster	0.01 @ 450
	Zinc smelter	0.01 @ 400
	Copper roaster	0.01 @ 500
	Copper reverberatory furnace	0.015 @ 550
	Copper converter	0.01 @ 500
	Aluminum—Hall process	0.075 @ 300
	Soderberg process	0.003 @ 200
	Ilmenite dryer	0.02 @ 300
	Titanium dioxide process	0.01 @ 300
	Molybdenum roaster	0.01 @ 300
	Ore beneficiation	0.02 @ 400
Miscellaneous	Refinery cataly stregenerator	0.015 @ 475
	Incinerators—Municipal	0.015 @ 500
	Apartment	0.02 @ 350
	Spray drying	0.01 @ 400
	Precious meal—refining	0.01 @ 400

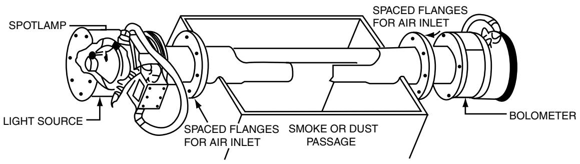


FIGURE 2 Stack mounted opacity meter (Bailey Meter Co.).

quantitative procedure for design of equipment to produce complying plumes. Equipment vendors will usually guarantee collection efficiency and emission concentrations by weight, but they will not give a guarantee to meet a specified opacity. This is indeed a serious problem at a time when a large precipitator installation can cost several million dollars and take twenty months to fabricate and install. Overdesign by a very conceivable factor of two can be very expensive in unneeded equipment. Underdesign can mean years of delay or operation under variance or with penalty payments.

Some progress has been made in applying classical theories of light scattering and transmission to the problem of predicting opacity. This effort has been greatly hampered by paucity of data giving simultaneous values of light attenuation, particle size distribution, and particle concentration in a stack. Perhaps the most comprehensive work to date has been that of Ensor and Pilat.[16]

Weight Limits on Particulates

Perhaps the least equivocal method of characterizing and specifying limits on particulate emissions is according to weight, either in terms of a rate (weight of emissions per unit time) or in terms of concentration (weight per unit volume). Measurement of emission weights must be done by isokinetic sampling of the gas stream, as outlined in the following section on measurement. Although the principles of such measurement are simple, they are difficult and time consuming when applied with accurate methodology to commercial installations. For this reason, such measurements have not previously been required in many jurisdictions and are almost never used as a continual monitoring technique.

Limits on weight rate of emissions are usually dependent on process size. Los Angeles, for instance, permits emissions to be proportional to process weight, up to 40 lbs/hr particulates for a plant processing 60,000 lbs/hr of material. Larger plants are limited to 40 lbs/hr. For furnaces, the determining factor is often heat input in BTU/hr rather than process weight. In cases where a particular plant location may have several independent units carrying out the same or similar processes, regulations often require that the capacities be combined for the purposes of calculating combined emissions.

Concentration limits are usually independent of process size. For instance, the EPA specifies incinerator emission of 0.08 grains particulates per standard cubic foot of flue gas (0.18 gm/NM^3) Dilution of the flue gas with excess air is usually prohibited, or else correction must be made to standard excess air or CO$_2$.

Ground Level Concentrations of Suspended Particulates

A limit on ground level concentration of particulates is an attempt to regulate emissions in accordance with their impact on population. A smoke stack acts as a dispersing device, and such regulations give incentive to build taller stacks in optimum locations.

In theory, ground level concentrations can be measured directly. Usually, however, emissions are measured in the stack, and plume dispersion equations are then used to calculate concentration profiles. Plume dispersion depends on stack height, plume buoyancy (i.e. density relative to ambient air), and wind velocity, and wind patterns. In addition, plumes are never stationary but tend to meander; and correction factors are usually applied to adjust for the sampling time at a fixed location. Dispersion calculations are usually easier than direct ground level measurements; and in cases where many different sources are present, calculation offers the only practical way to assess the contributions of a specific source. A recent evaluation of plume dispersion models is given by Carpenter *et al.*[15]

In some states, a plume dispersion model is incorporated into a chart which gives an allowable weight rate of emissions as a function of effective stack height and distance from property lines. An example of this approach is shown in Figure 3.

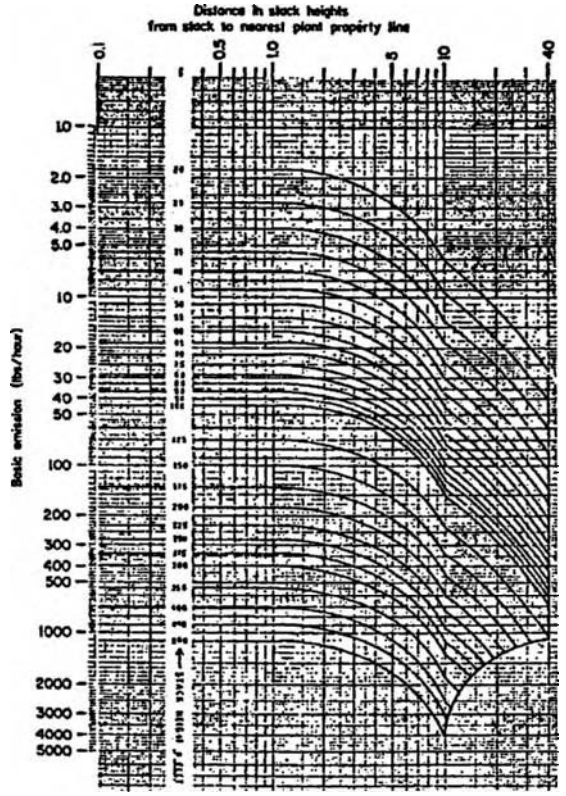


FIGURE 3 Emission requirements for fine particles based on plume dispersion model (New Jersey Air Pollution Code).

The particular regulation shown also accounts for differing toxicity of certain particulates and allocates the emission factors of Table 2 accordingly.

Very often permissible ground level concentrations are set according to other sources in the area. Thus a plant would be allowed greater emissions in a rural area than in a heavily industrialized neighbourhood.

Dust fall

A variant on the ground level concentration limit is a dustfall limit. This basically superimposes a particle settling velocity on ground level concentration to obtain dustfall rates in weight per unit area per unit time. This is a meaningful regulation only for large particles and is not widely legislated at present.

Federal Clean Air Statutes and Regulations

The major federal statutes covering air pollution are PL 88–206 (The Clean Air Act of 1963), PL 90–148 (The Air Quality Act of 1967) PL 92–157, PL 93–115, PL 95–95 (The Clean Air amendments of 1977), and PL 95–190, Administrative standards formulated by the Environmental Protection Agency (EPA) are given in the Code of Federal Regulations Title 40, Chapter 1, Sub-chapter C, with regulations on particulates in parts 50, 51, 52, 53, 58, 60, 61, and 81.

TABLE 2
Emission effect factors (for use with Fig. 3) (New Jersey Air Pollution Control Code)

Material	Effect factor
Fine Solid Particles	
All materials not specifically listed hereunder	1.0
Antimony	0.9
A-naphthylthiourea	0.5
Arsenic	0.9
Barium	0.9
Beryllium	0.003
Cadmium	0.2
Chromium	0.2
Cobalt	0.9
Copper	0.2
Hafnium	0.9
Lead	0.3
Lead arsenate	0.3
Lithium hydride	0.04
Phosphorus	0.2
Selenium	0.2
Silver	0.1
Tellurium	0.2
Thallium	0.2
Uranium (soluble)	0.1
Uranium (insoluble)	0.4
Vanadium	0.2

The EPA has established National Ambient Air Quality Standards (NAAQS). For suspended particulate matter the primary standard (necessary to protect the public health with an adequate margin of safety) is 75 μg/M^3 annual geometric mean with a level of 260 μg/M^3 not to be exceeded more than once per year. All states have been required to file state implementation plants (SIP) for achieving NAAWS. It is only through the SIP's that existing pollution sources are regulated.

The EPA requires no specific state regulations for limits on existing sources, but suggestions are made for "emission limitations obtainable with reasonable available technology." Some of the reasonable limits proposed for particulates are:

1) Ringlemann 1 or less, except for brief periods such as shoot blowing or start-up.
2) Reasonable precautions to control fugitive dust, including use of water during grading or demolition, sprinkling of dusty surfaces, use of hoods and vents, covering of piles of dust, etc.
3) Incinerator emission less than 0.2 lbs/100 lbs refuse charged.
4) Fuel burner emissions less than 0.3 lbs/million BTU heat input.
5) For process industries, emission rates E in lbs/hr and Process weight P in tons/hr according to the relationships:

$$E = 3.59\,P^{0.62} \text{ for } P \leqslant 30 \text{ tons/hr.}$$
$$E = 17.31\,P^{0.16} \text{ for } P \geqslant 30 \text{ tons/hr.}$$

"Process weight" includes all materials introduced to the process except liquid and gaseous fuels and combustion air. Limits should be set on the basis of combined process weights of all similar units at a plant.

In considering what emission limits should be established, the states are encouraged to take into account local condition, social and economic impact, and alternate control strategies and adoption of the above measures is not mandatory. It is expected, however, that such measures will become the norm in many areas.

For new or substantially modified pollution sources, the EPA has established new source performance standards. The standards for particulate emissions and opacity are given in Table 3. Owners may submit plants of new sources to the EPA for technical advice. They must provide ports, platforms, access, and necessary utilities for performing required tests, and the EPA must be allowed to conduct tests at reasonable times. Required records and reports are available to the public except where trade secrets would be divulged. The states are in no way precluded from establishing more stringent standards or additional procedures. The EPA test method specified for particulates measures only materials collectable on a dry filter at 250°F an does not include so called condensables.

In addition to new source performance standards, major new stationary sources and major modifications are usually subject to a "Prevention of Significant Deterioration" review. If a particulate source of more than 25 tons/year is located in an area which attains NAAQS or is unclassifiable with respects to particulates, the owner must demonstrate that the source will not violate NAAQS or PSD concentration increments. This requires modelling and preconstruction monitoring of ambient air quality. If the new or expanded source is to be located in an area which does not meet NAAQS, then emission from other sources must be reduced to offset the new source. The regulation regarding emission offsets and prevention of significant deterioration are relatively recent. A summary of federal regulations as of 1981 has recently been published as a quick guide to this rapidly changing field.[18]

In recent years, regulation of particulate emissions from mobile sources has been initiated. The burden is essentially on manufacturers of diesel engines. Because the emission requirements and test procedures are quite complex and because the target is highly specific, a comprehensive discussion is beyond the scope of this article. Some representative standards are: Diesel engines for urban buses, 0.019 grams/megajoule, and other diesel engines for road use, 0.037 grams/megajoule:[19] Non-road diesel engines, 1 gram/kilowatt-hour for sizes less than 8 kilowatts in tier 1 down to 0.2 grams/kilowatt-hour for units larger than 560 kilowatts in tier 2.[20] Locomotives, 0.36 grams/bhp-hr for switching service in tier 1 down to 0.1 grams/bhp-hr for line service in tier 3.[21] Marine diesel engines, 0.2 grams/KwH to 0.5 grams/KwH, depending on displacement and tier.[22] Note that the emission units above are as specified in the printed regulation.

Particulate emission standards are also being promulgated by agencies other than the Environmental Protection Agency. In general, these are workplace standards. An example would be the standard for mobile diesel-powered transportation equipment promulgated by the Mine Safety and Health Administration. This specifies that the exhaust "shall not contain black smoke."[23]

MEASUREMENT OF PARTICULATE EMISSIONS

As a first step in any program for control of particulate emissions, a determination must be made of the quantity and nature of particles being emitted by the subject source. The quantity of emissions determines the collection efficiency and size of required cleanup equipment. The particle size and chemical properties of the emitted dust strongly influence the type of equipment to be used. Sampling for this purpose has been mainly a matter of industrial concern. A last step in most control programs consists of measuring pollutants in the cleaned gas stream to ensure that cleanup equipment being used actually permits the pertinent emission targets to be met. With increasing public concern and legislation on air pollution, sampling for this purpose is increasingly required by statute to determine compliance with the pertinent emission regulations. To this end the local pollution control authority may issue a comprehensive sampling manual which sets forth in considerable detail the procedures to be used in obtaining raw data and the computations involved in calculating the pertinent emission levels.

Complete and comprehensive source testing procedures are beyond the scope of this paper. References 24–28 give detailed instruction for performance of such tests.

Sampling of gas streams, especially for particulates, is simple only in concept. Actual measurement require specialized equipment, trained personnel, careful experimental and computational techniques, and a considerable expenditure of time and manpower. Matters of technique and equipment are covered in source testing manuals as mentioned above and are briefly summarized later in this paper. Two additional complicating factors are usually present. First is the frequent inaccessibility of sampling points. These points are often located in duct work 50–100 ft above ground level. Scaffolding must often be installed around the points, and several hundred pounds of equipment must be lifted to that level. Probe clearances are often critical, for in order to make a sample traverse on 12 ft dia. stack, a 14 ft probe is needed, and clearance must be available for insertion into the sampling port as well as a means for suspending the probe from above. At least one professional stack sampler is an amateur mountain climber and puts his hobby to good use on the job. A second complicating factor is the adverse physical conditions frequently encountered. A somewhat extreme but illustrative example is a refinery stream recently sampled. Gas temperature was 1200°F requiring special probes and gaskets and protective clothing for the workers. The gas stream contained 10% carbon monoxide creating potential hazards of poisoning and explosion especially since duct pressure was slightly above that of the atmosphere. Temperature in the work area was in excess of 120°F contributing further to the difficulty of the job.

In preparation for a sampling program, work platforms or scaffolding and valved sample ports must be installed. All special fittings for adapting the sampling probes to the ports should be anticipated and fabricated. Arrangements must be made with plant operating personnel to maintain steady operating conditions during the test. The test must be carefully planned as to number and exact location of traverse sample points, and probes should be premarked for these locations. Flow nomographs for sampling nozzles should be made; and all filters, impingers, and other element of sampling trains should be tared. With that advance preparation a 3 man sampling team would require 1–2 days to position their equipment and make gas flow measurements and 2 sample transverses at right angles in a large duct or stack.

Measurement of Gas Flow Rates

A preliminary step in determination of emission rates from a stack is measurement of the gas flow rate. Detailed procedures in wide use including the necessary attention to technique have been published by the ASME,[20] ASTM,[19] the Environmental Protection Agency, referred to as EPA,[21]

TABLE 3
Federal Limits of Particulate Emissions from New Stationary Sources

(Through 2004 Codified in CFR, Title 40. Chapter 1/Part 60)

Subpart	Source	Particulate Emissions	Opacity (%)
D	Fossil fired steam generators	13 ng/j	20*
			(27% for 6 min/hr)
Da	Electric utility steam generators	43 ng/j	20*
			(27% for 6 min/hr)
Db	Industrial/commercial/institutional	22 to 86 ng/j	20*
	steam generators	depending on fuel, size, construction date	(27% for 6 min/hr)
Dc	Small industrial/commercial steam generators	22 to 43 ng/j	20*
		depending on fuel, size	(27% for 6 min/hr)
E	Incinerators	0.18 g/dscm	—
F	Portland cement		
	kiln	0.15 kg/ton	20*
	clinker cooler	0.05 kg/ton	10*
	other facilities	—	10
G	Nitric acid	—	10
H	Sulfuric Acid	0.075 kg/ton	10
I	Hot mix asphalt	90 mg/dscm	20
J	Refinery—fluid catalytic cracker regenerator	1 kg/1000 kg coke burned	30*
			(6 min/hr exception)
L	Secondary lead smelters cupola or		
	reverberatory furnace	50 mg/dscm	20
	pot furnace	—	10
M	Secondary brass and bronze production	50 mg/dscm	20
N	Basic oxygen steel, primary emission	50 mg/dscm	10
	with closed hooding	68 mg/dscm	(20% once per production cycle)
Na	Basic oxygen steel, secondary emissions		
	from shop roof	—	10
			(20% once per production cycle)
	from control device	23 mg/dscm	5
O	Sewage plant sludge incinerator	0.65 g/kg dry sludge	20
P	Primary copper smelters, dryer	50 mg/dscm	20*
	sulfuric acid plant	—	20
Q	Primary zinc smelters, sintering	50 mg/dscm	20*
	sulfuric acid plant	—	20
R	Primary lead smelters, sintering or furnaces	50 mg/dscm	20*
	sulfuric acid plant	—	20
S	Primary aluminum reduction		
	pot room	—	10
	Anode bake plant	—	20
Y	Coal preparation		
	thermal dryer	0.07 g/dscm	20
	pneumatic coal cleaning	0.04 g/dscm	10
	conveying, storage, loading	—	20

(*continued*)

TABLE 3 (*continued*)

Subpart	Source	Particulate Emissions		Opacity (%)
Z	Ferroalloy production			
	control device; silicon, ferrosilicon, calcium silicon or silicomanganese zirconium alloys	0.45 kg/MW-hr		15*
	control device; production of other alloys	0.23 kg/MW-hr		15*
	uncontrolled emissions from arc furnace	—		Not visible
	uncontrolled emissions from tapping station	—		Not visible for more than 40% of tap period
	dust handling equipment	—		10
AA	Electric arc steel plants			
	control device	12 mg/dscm		3*
	shop exit due to arc furnace operation	—		6
	except during charging	—		20
	except during tapping	—		40
	dust handling equipment	—		10
BB	Kraft pulp mills			
	recovery furnace	10 g/dscm		35
	smelt dissolving tank	0.1 g/kg black liquor solids		—
	lime kiln, gas fired	0.15 g/dscm		—
	oil fired	0.30 g/dscm		—
CC	Glass manufacture, standard process	*Gas fuel*	*Oil fuel*	
	container glass	l0.1 g/kg glass	l0.13 g/kg glass	—
	pressed & blown glass, borosilisate	0.5 g/kg	0.65 g/kg	—
	pressed & blown glass, soda lime & lead	0.1 g/kg	0.13 g/kg	—
	pressed & blown glass, other compositions	0.25 g/kg	0.325 g/kg	—
	wool fiberglass	0.25 g/kg	0.325 g/kg	—
	flat glass	0.225 g/kg	0.225 g/kg	—
	Glass manufacture, modified process			
	container, flat, pressed, blown glass, soda lime	0.5 g/kg		*
	container, flat, pressed, blown glass, borosilicate	1.0 g/kg		*
	textile and wood fiberglass	0.5 g/kg		*
DD	Grain elevators			
	column dryer, plate perforation >2.4 mm	—		0
	rack dryer, exhaust screen filter cans thru 50 mesh	—		0
	other facilities	0.023 g/dscm		0
	fugitive, truck unloading, railcar loading/unloading	—		5
	fugitive, grain handling	—		0
	fugitive, truck loading	—		10
	fugitive, barge or ship loading	—		20
GG	Lime rotary kiln	0.30 g/kg stone feed		15*
LL	Metallic mineral processing			
	stack emissions	0.05 g/dscm		7
	fugitive emissions	—		10
NN	Phosphate rock			
	dyer	0.03 g/kg rock		10*
	calciner, unbeneficiated rock	0.12 g/kg rock		10*
	calciner, beneficiated rock	0.055 g/kg rock		10*
	rock grinder	0.0006 g/kg rock		0*
PP	Ammonium sulfate manufacture, dryer	0.15 g/kg product		15

(*continued*)

TABLE 3 (*continued*)

Subpart	Source	Particulate Emissions	Opacity (%)
UU	Asphalt roofing		
	shingle of mineral-surfaced roll	0.04 g/kg	20
	saturated felt or smooth surfaced roll	0.4 g/kg	20
	Asphalt blowing still		
	with catalyst addition	0.67 g/kg	—
	with catalyst addition, #6 oil afterburner	0.71 g/kg	—
	no catalyst	0.60 g/kg	—
	no catalyst, #6 oil afterburner	0.64 g/kg	—
	Asphalt storage tank		0
	Asphalt roofing mineral handling and storage		1
AAA	Residential wood heaters		
	with catalytic combustor	4.1 g/hr	—
	no catalytic combustor	7.5 g/hr	—
OOO	Nonmetallic mineral processing		
	stack or transfer point on belt conveyors	0.05 g/dscm	7
	fugitive emissions	—	10
	crusher fugitive emissions	—	15
PPP	Wool fiberglass insulation	5.5 g/kg	
UUU	Calciners & dryers in mineral industries	0.092 g/dscm	10*

*Continous monitoring by capacity meters required

The above standards apply to current construction. Existing unmodified units may have lower standards.

Many applications require continuous monitoring of operating variables for process and control equipment.

the Lost Angeles Air Pollution Control district, referred to as APCD,[21] and the Western Precipitation Division, referred to as WP.[21] This article will only treat the general procedures and not significant differences between popular techniques.

Velocity Traverse Points Because of flow non-uniformity, which almost invariably occurs in large stacks, the stack cross section in the sampling plane must be divided into a number of smaller areas and gas velocity determined separately in each area. Circular ducts are divided by concentric circles, and 2 velocity traverses are made at right angles. Figure 4 shows a typical example. Location of the sample points can be determined from the formula

$$R_n = D\sqrt{\frac{2n-1}{2N}}$$

where

R_n = distance from center of duct to the "nth" point from the center
D = duct diameter
n = sample point number, counting from center
N = total number of measurement points in the duct. The number of sample points along one diameter is $N/2$.

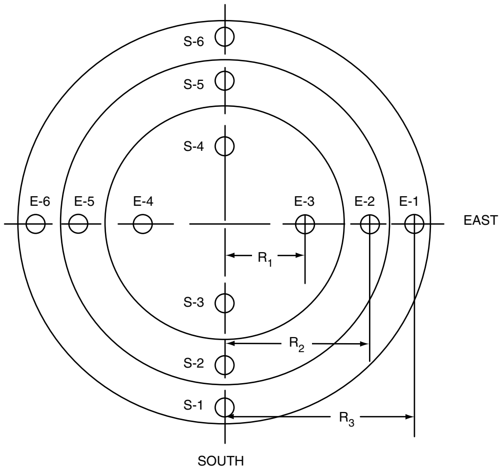


FIGURE 4 Velocity and sampling traverse positions in circular ducts.

For rectangular ducts the cross section is divided into N equal rectangular areas such that the ratio of length to width of the areas is between one and two. Sample points are at the center of each area.

The number of traverse points required is usually specified in the applicable test code as a function of duct area or diameter. Representative requirements are shown in Table 4.

TABLE 4
Required traverse points

Code	Duct sizes	Number of points
EPA[8]	2 ft dia.	12 minimum
More according to Figure 2 if near flow disturbance		
WP[17]	<2 ft^2	4
	>2–25 ft^2	12
	25 ft^2	20 or more
APCD[14]	1–2 ft^2 (rectangular)	4
and	2–12 ft^2	6–24
ASTM[15]	>12 ft^2	24
	1–2 ft dia.	12
	2–4 ft	16
	4–6 ft	20
	>6 ft	24 or more
These numbers should be doubled where only 4–6 duct diameters of straight duct are upstream.		
ASME[16]	<25 ft^2	8–12
	>25 ft^2	12–20
Double or triple these numbers for high nonuniform flow.		

Very often more points are required if the flow is highly nonuniform or if the sampling point is near an elbow or other flow disturbance. Figure 5 shows the EPA adjustment for flow nonuniformity.

Velocity Measurement Velocity measurements in dusty gases are made with a type S (special or staubscheibe) pitot tube, shown in Figure 6, and a draft gage manometer. Gas velocity is given by

$$V = C\sqrt{2gh_L\rho_L/\rho_g}$$

where

V = gas velocity
C = pitot tube calibration coefficient. This would be 1.0 for an ideal pitot tube, but type S tubes deviate considerably.
g = acceleration of gravity
h_L = liquid height differential in manometer
ρ_L = density of manometer liquid
ρ_g = gas density.

It is necessary to measure the temperature and the pressure of the gas stream and estimate or measure its molecular weight in order to calculate density.

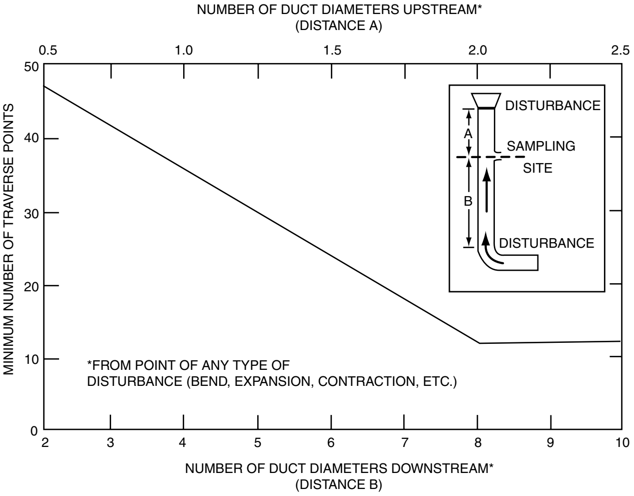


FIGURE 5 Sampling points required in vicinity of flow disturbance (EPA).

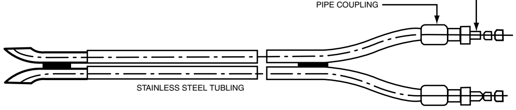


FIGURE 6 Type S Pitot tube for use in dusty gas stream.

Gas Analysis For precise work gas composition is needed for three reasons (1) so that molecular weight and gas density may be known for duct velocity calculations, (2) so that duct flow rates at duct condition can be converted to standardized conditions used for emission specifications. Standard conditions are usually 70°F, 29.91 in. mercury barometric pressure, moisture free basis with gas volume adjusted to 12% CO_2. Some codes differ from this, however. (3) For isokinetic sampling moisture content at stack conditions must be known in order to adjust for the fact that probe gas flow is measured in a dry gas meter at ambient conditions.

Gas analysis for CO_2, CO, and O_2 is almost always done by Orsat analysis. Moisture may be determined gravimetrically by condensation from a measured volume of gas as required by EPA.

Overall Flow Rate Total flow rate is calculated from duct area and average gas velocity as determined by the pitot tube measurements. Pitot tube traverse points are at the center of equal areas so no weighting is necessary to determine average velocity. This gives flow at duct conditions which is usually converted to standard conditions.

Measurement of Particulate Concentrations in Stacks

Standard methods for measuring particulation concentrations in stacks depend on the principle of isokinetic sampling. Since particles do not follow gas streamlines exactly but tend to travel in straight lines, precautions must be taken that the gas being sampled experiences no change in velocity or direction in the vicinity of the sampling point. This is done by using a thin walled tubular probe carefully aligned with the gas flow and by withdrawing gas so that velocity just within the tip of the probe equals that in the main gas stream. Several recent studies[29–31] have measured effects of probe size, alignment, and velocity on accuracy of sampling. The sampled gas is drawn through a train of filters, impingers, and a gas meter by means of a pump or ejector. Typical probes are shown in Figure 7, and several types are commercially available. With these probes the necessary gas sampling velocity must be previously determined by pitot tube measurement, and the gas flow rate at the flow meter is adjusted (taking into account gas volume changes due to cooling and condensation between stack and meter) to equal that velocity. An alternate method is to use a null nozzle, which contains static pressure taps to the outside and inside surfaces of the sample probe as shown in Figure 8. Flow through the probe is adjusted so that the static pressures are equal at which point the velocities inside and outside the probe should be the same. The null nozzle greatly simplifies sampling, but null nozzles require careful periodic calibration and are not generally used for high precision work.

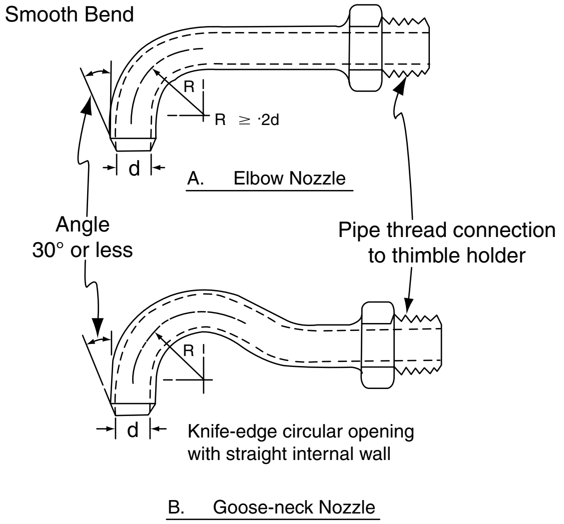


FIGURE 7 Nozzles for particulate sampling.

The sampling train of filters and impingers, which collects the particles, is usually carefully specified in the test method or governmental regulation in force. Differences between sampling trains to some extent reflect different technical solutions to the sampling problem but they also reflect differences in the philosophy of what exactly should be measured.

Perhaps the most widespread train will be that specified by the EPA[14] for testing new emission sources, shown in Figure 9. The original intent was to collect and measure not only particles which actually exist in the stack at stack conditions, but also solids or droplets that can be condensed out of the stack gas as it is cooled to ambient conditions. The filter is heated to avoid condensation and plugging. The first two impingers contain water to collect most of the condensables. The third impinger is empty and serves as an additional droplet tray while the fourth impinger is filled with silica gel to collect residual water vapor. Although the impingers in the train collect condensibles, present regulations are written only in terms of the solid particulates which are collected in the filter.

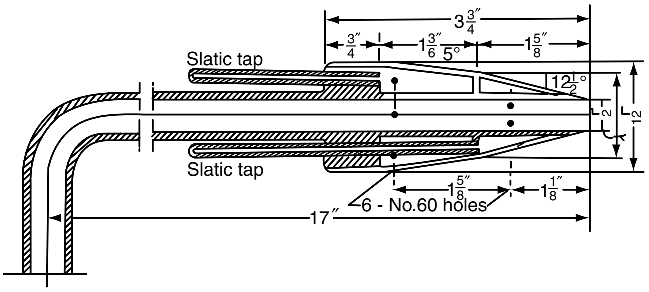


FIGURE 8 Null type nozzle for isokinetic sampling.

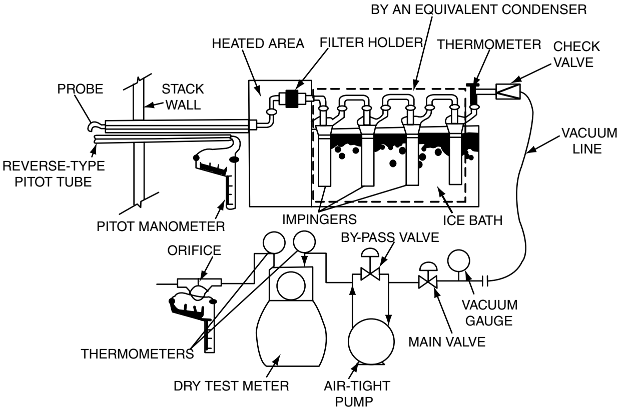


FIGURE 9 Environmental Protection Agency particulate sampling train.

The ASME power test code,[27] in contrast, is designed to measure performance of devices such as precipitators and cyclones, and thus is concerned only with substances which are particulate at conditions prevailing in the equipment. This test usually used a filter assembly with the filter very close to the sampling probe so that the filter may be inserted into the stack avoiding condensation. No impingers are used.

To some extent filter characteristics are determined by process conditions. Alundum thimbles and glass wool packed tubes are used for high temperatures. If liquid droplets are present at the filter inlet, glass-wool tubes are the only useful collection devices, because conventional filters will readily become plugged by droplets. Glass-wool collection greatly complicates quantitative recovery of particles for chemical or size analysis.

In sampling a large duct having several traverse points for flow and particle measurement, particles for all points on a traverse are usually collected in a single filter impinger train, thus giving an average dust concentration. Each sample point is sampled for an equal time but at its own isokinetic velocity. The probe is then immediately moved to the next point and the flow rate adjusted accordingly. Sample flow rate is adjusted by rotameter or orifice readings, but total gas flow during the entire test is taken from a dry meter.

Minimum sampling time or volume is often set by regulation. Examples are:

Bay area[24]—Sample gas volume = 20 $L^{0.8}$, where volume is in standard cubic ft, and L is duct equivalent dia. in ft.

A maximum sampling rate of 3 SCFM is specified and a minimum time of 30 min.

ASME[27]—Minimum of 2hr with at least 10 min at each traverse point through two complete circuits.

APCD[25]—5–10 min/point for a total run of at least 1 hr.

Industrial gas cleaning institute—At least 2 hr or 150 ft^3 sample gas or until sample weight is greater than 30% of filter weight.

Emissions are calculated from test volumes of weight of particulates collected and volume of sample gas through the gas meter. Care must be taken to include particles deposited on tubing walls as well as those trapped by the filter. If condensibles are to be included, the liquid from the impinger train is evaporated to dryness, and the residue is weighed and included with the particulates. Corrections to the gas volume depend on sample train operation and on standard conditions for reporting emissions, and these are spelled out in detail in the specific test codes to be used. Results are usually expressed both as grains per cubic foot (using standard conditions defined in the code) and as lbs/hr from the whole stack.

Measurement and Representation of Particle Size

A determination of the emitted particle size and size distribution is a desirable element in most control programs. Collection efficiency of any given piece of equipment is a function of particle size, being low for small particles and high for large ones, and capital and operating costs of equipment required increased steadily as the dust particle size decreases.

Perhaps the simplest method of particle size measurement, conceptually at least, is by microscope count. The minimum size that can be counted optically is about 0.5μ which is near the wavelength of visible light. Electron microscopes may be used for sizing of smaller particles. Counting is a laborious procedure, and sample counts are often small enough to cause statistical errors at the very small and very large ends of the distribution. This method requires the smallest sample size and is capable of giving satisfactory results. Care must be taken in converting from the number distribution obtained by this method to mass distribution.

A second simple method is sieve analysis. This is commonly used for dry freely flowing materials in the size range above 44μ, a screen size designated at 325 mesh. Using special shaking equipment and very delicate micromesh sieves particles down to 10μ can be measured. Error can be caused by "blinding" of the sieve mesh and sticky or fine particles, incomplete sieving, and particle fragmentation during sieving. A sample size of at least 5–10 g is usually required.

Another class of measurement techniques is based on the terminal falling velocity of particles in a gas (air). The quantity measured is proportional to ρd^2, where ρ is particle density and d is diameter. Hence a separate determination of density is needed. One such device is the Sharples Micromerograph (Sharples-Stokes Division, Penwalt Corporation, Warminster, Pennsylvania). The device records the time for particles to fall through a 2 m high column of air onto the pan of a continuously recording balance. Templates are available to convert fall time to ρd^2. The Micromerograph is mechanically and electrically complex but easy to use. An objection is that a significant fraction of the injected particles stick to the column walls and do not reach the balance pan. This effect can sometimes be selective, and it thus gives a biased size distribution.

A second sedimentation device is the Roller elutriator tube, Figure 10. A powder sample is placed in the tube and air is passed upwards through it for a specified time. A separation is effected with small particles being carried overhead and large ones remaining in the tube. Often a series of tubes of decreasing diameter are connected in a cascade with each successive tube having a lower air velocity and retaining finer particles. The Roller method was used quite widely in the petroleum industry for many years. However, it is slow, requires a large sample, does not give clean particle size cuts, and is sensitive to tube orientation. It is therefore being supplanted by newer methods.

A third sedimentation is centrifugal sedimentation. This is the standard test method of the Industrial Gas Cleaning Institute, and use of such devices of the Bahco type has been standardized by the ASME.[32] The Bahco analyzer consists of a rapidly spinning rotor and a superimposed radial gas flow from circumference to center. Larger particles are centrifuged to the outside diameter of the rotor, while small ones are carried to the center with a cut point determined by gas velocity and rotor speed.

Still another method is the Coulter Counter. In this technique the test powder is dispersed in an electrolyte, which is then pumped through a small orifice. Current flow between electrodes on each side of the orifice is continuously

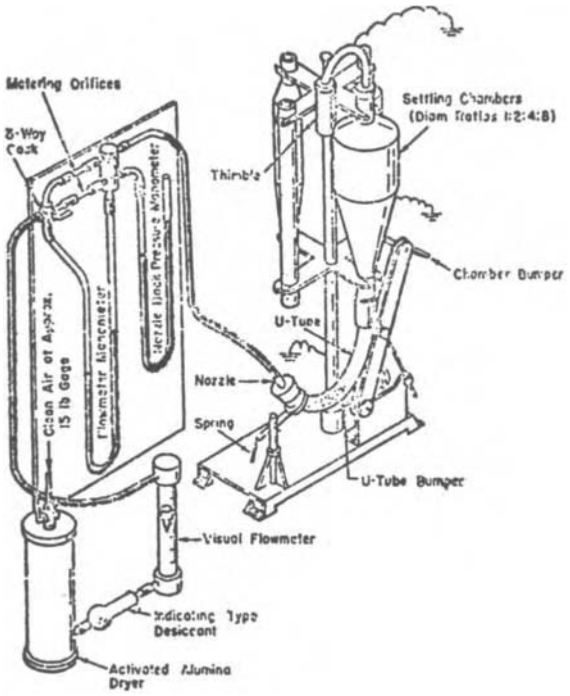


FIGURE 10 Air classifier for subsieve particle size analysis.

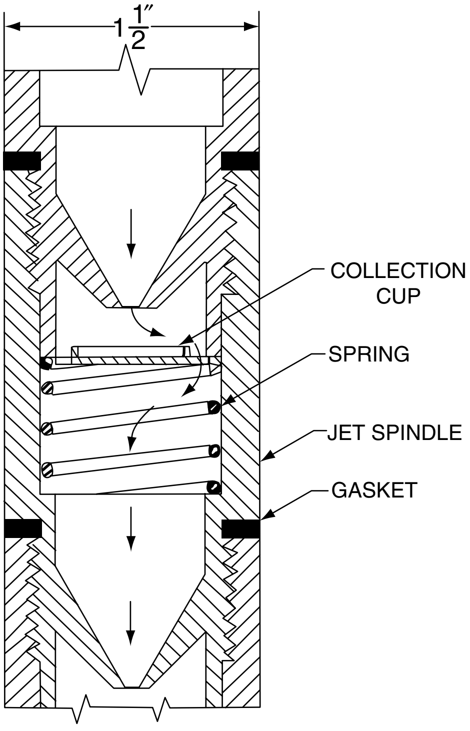


FIGURE 11 Stage of typical cascade impactor (Monsanto).

monitored. Passage of a particle through the orifice momentarily reduces current to an extent determined by particle size. The device electronically counts the number of particles in each of several size ranges, and a size distribution can then be calculated. The method is capable of giving very good results, and newer model counters are very fast.

A novel liquid phase sedimentation analyzer is the Sedigraph (Micrometrics Instrument Corporation, Norcross, Georgia). The particle sample dispersed in liquid is put into a sample cell and allowed to settle. Mass concentration is continuously monitored be attenuation of an X-ray beam, and this is mathematically related to particle size, X-ray location and time. The instrument automatically plots particle diameter as cumulative weight percent. The device can cover the size range from 0.1–10μ in a single operation, a much wider range than can be conveniently analyzed by most analyzers.

Laser optics techniques relying on light scattering, Fraunhofer diffraction, or light extinction are becoming the method of choice in many applications. The Leeds and Northrop "Microtrac" and Malvern Instruments Co. laser particle and droplet sizer are representative of such techniques. These Instruments can measure particles in a flowing gas stream, and thus can theoretically be used on line. More often a collected particulate sample is dispersed in liquid for analysis.

Impingement devices such as the Anderson Impactor, or in the impactors developed by May or Batelle, may be used to measure particle sizes *in situ* in a combined sampling and sizing operation. As is shown in Figure 11 such a device consists of a series of orifices arranged to give gas jets of increasing velocity and decreasing diameter, which jets impinge on collection plates, Successive stages collect smaller and smaller particles, and the size distribution of aspirated particles can be obtained from the weight collected on each stage and the size "cut point" calibration of the stage. Several studies of calibrations have been published,[33-37] and discrepancies have been pointed out.[38] Impactors must be operated at constant known gas flow rate and for this reason are not capable of giving true isokinetic sampling under conditions of fluctuating duct velocity. This is one of the few types of devices which may be applied to liquid droplets, which coalesce once collected. It is capable of size determination well below 1μ (finer than most devices). Because it eliminates recovery of particles from a filter and subsequent handling, it can be useful in measuring distributions at low concentrations.

Most particles emitted to the atmosphere are approximately spherical so that the exact meaning of "diameter" is not usually important in the context. For highly irregular particles a great many different diameters many be defined, each with particular applications. For purposes of particulate control equipment, persistence of airborne dusts, and physiological retention in the respiratory tract, the most meaningful diameter is usually the "aerodynamic" diameter, that of the sphere having the same free fall velocity as the particle of interest. This is the diameter measured by sedimentation, eleutriation, and inertial impaction techniques.

A large number of methods are available for expressing particle size distributions, each having properties of fitting certain characteristic distribution shapes or of simplifying

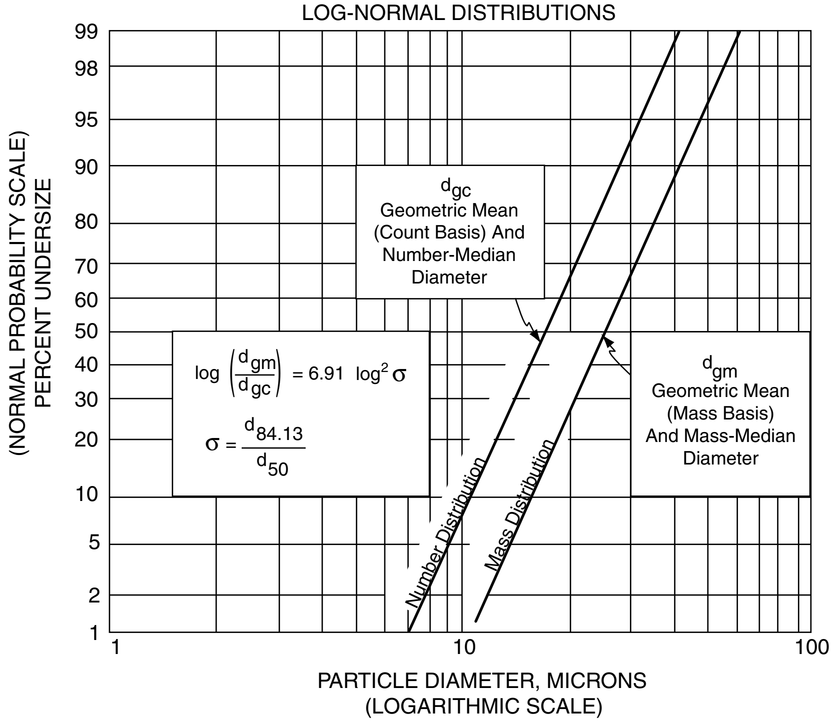


FIGURE 12

certain mathematical manipulations. A comprehensive summary of various distribution functions is given by Orr.[39] The most useful function in emission applications seems to be the long-normal distribution. Commercial graph paper is available having one logarithmic scale and one cumulative normal probability scale. If particle size is plotted vs. cumulative percentage of sample at or below that size, the log-normal distribution gives a straight line. A large percentage of emissions and ambient particulate distributions have log-normal distributions, and plotting on log-probability paper usually facilitates interpolation and extrapolation even when the line is not quite straight. For a true log-normal distribution very simple relationships permits easy conversion between distributions based on number, weight, surface area, and so on, which are covered in Orr.[39] Relationships between weight and number distribution are shown in Figure 12.

REFERENCES

1. Stern, A.C. 1977, Air Pollution Standards, **5**, Chapter 13 in *Air Pollution,* 3rd Edition, Ed. by A.C. Stern, Academic Press, New York.
2. Greenwood, D.R., G.L. Kingsbury, and J.G. Cleland, “A Handbook of Key Federal Regulations and Criteria for Multimedia Environmental Control” prepared for U.S. Environmental Protection Agency. Research Triangle Institute, Research Triangle N.C. 1979.
3. National Center for Air Pollution Control (1968), *A Compilation of Selected Air Pollution Emission Control Regulations and Ordinance,* Public Health Service Publication No. 999-AP-43. Washington.
4. National Research Council ad hoc Committee (vol. 1, 1998, vol. 2, 1999, vol. 3, 2001) “Research Priorities for Airborne Particulate Matter”, National Academy Press, Washington, D.C.
5. Friedrich, R. and Reis, S. (2004) “Emissions of Air Pollutants” Springer, Berlin.
6. Katz, M. ed. “Methods of Air Sampling and Analysis” American Public Health Association, Washington, 1977.
7. Powals, R.J., L.V. Zaner, and K.F. Sporck, “Handbook of Stack Sampling and Analysis” Technomic Pub. Co. Westport Ct., 1978.
8. Brenchley, D.L., C.D. Turley, and R.F. Yarmak “Industrial Source Sampling Ann Arbor Science, Ann Arbor MI 1973.
9. Hawksley, P.G.W., S. Badzioch, and J.H. Blackett, “Measurement of Solids in Flue Gases, 2nd Ed.” Inst. of Fuel, London, 1977.
10. Kudlich, R., Ringlemann Smoke Chart, US Bureau of Mines Information Circular 7718, revised by L.R. Burdick, August, 1955.
11. Weisburd, M.I. (1962), *Air Pollution Control Field Operations,* Chapter 10, US Public Health Service, Publication 397, Washington.
12. Griswold, S.S., W.H. Parmelee, and L.H. McEwen, Training of Air pollution Inspectors, 51st annual meeting APCA, Philadelphia, May 28, 1958.
13. Conner, W.D. and J.R. Hodkinson (1967), *Optical Properties and Visual Effects of Smoke-Stack Plumes,* PHS Publication No. 999-AP-30.
14. Environmental Protection Agency, Standards of performance for new stationary sources, Code of Federal Regulations 40 CFR, Part 60.
15. McKee, Herbert C. (1971), Instrumental method substitutes for visual estimation for equivalent opacity, *Jr. APCA* **21**, 489.
16. Ensor, D.S. and M.J. Pilat (1971), Calculation of smoke plume opacity from particulate air pollutant properties, *Jr. APCA* **21**, 496.
17. Carpenter, S.B., T.L. Montgomery, J.M. Leavitt, W.C. Colbaugh and F.W. Thomas (1931), Principal plume dispersion models, *Jr. APCDA* 21, 491.
18. Air Pollution Control Association Directory and Resource Book pp 143–158. Pittsburgh, 1981.
19. Code of Federal Regulations 40:CFR 86.004–11. US Government Printing Office, Washington 7/1/2004.
20. Code of Federal Regulations 40:CFR 89.112. US Government Printing Office, Washington 7/1/2004.
21. Code of Federal Regulations 40:CFR 92.8, US Government Printing Office, Washington 7/1/2004.

22. Code of Federal Regulations 40:CFR 94.8, US Government Printing Office, Washington 7/1/2004.
23. Code of Federal Regulations 30:CFR 36.2a, US Government Printing Office, Washington 7/1/2004.
24. Wolfe, E.A. (1966), Source testing methods used by bay area air pollution control district, BAAPCD, San Francisco.
25. Devorkin, H. (1963), Air pollution source testing manual, Air Pollution Control District, Los Angeles.
26. ASTM Standards (1971), Standard method for sampling stacks for particulate matter, Part 23, designation D-2928–71. ASTM, Philadelphia.
27. ASME Power Test Codes (1957), Determining dust concentration in a gas stream, Test code No. 17, ASME, New York.
28. Haaland, H.H. (1968), Methods for determination of velocity, volume, dust and mist content of gases, bulletin WP-50, Western Precipitation Division, Joy Manufacturing Company, Los Angeles.
29. Benarie, M. and S. Panof (1970), *Aerosol Sei.* **1**, 21.
30. Davies, D.N., The entry of aerosols into sampling tubes and heads, *Staub-Reinhalt Luft,* **28** (June, 1968), p. 1–9.
31. Raynor, C.S. (1970), Variation in entrance efficiency of a filter sampler with air speed, flow rate, angel, and particle size, *Am. Ind. Hyg. Ass. J.*, **31**, 294.
32. ASME Power Test Codes, Bacho, No. 28.
33. May, K.R. (1945), The cascade impactor, *J. Sci. Inst.*, **22**, 187.
34. Ranz, W.E. and J.E. Wong (1952), Impaction of dust and smoke particles, *Ind. Eng. Chem.*, **3**, 1371.
35. Mitchell, R.J. and J.M. Pilcher (1959), Improved cascade impactor, *Ind. Eng. Chem.*, **51**, 1039.
36. Brink, J.A. (1958), Cascade impactor for adiabatic measurements, *Ind. Eng. Chem.*, **50**, 645.
37. Andersen, A.A. (1958), New sampler for the collection, sizing, and enumerations or viable airborne particles, *J. of Bacteriology,* **76**, 471.
38. Lippmann, S.M. (1959), Review of cascade impactors for particle size analysis, *Am. Ind. Hygiene Assoc. J.* **20**, 406.
39. Orr, C. (1966), *Particulate Technology,* Macmillan, New York.
40. Federal Register Vol. 52, 126, July 1, 1987 pages 24634–24750. Engineering Nov. 1987, p. 60.
41. Rich, G.A., PM-10 Regulations Pollution Engineering Nov. 1987, p. 60.

JOHN M. MATSEN
Lehigh University

OIL SPILLAGE: see MARINE SPILLAGE—SOURCES AND HAZARDS

PARTICULATE REMOVAL

TYPES AND CHARACTERISTICS OF CLEANING EQUIPMENT

Three principal considerations enter into the selection of particulate removal equipment. Cleaning efficiency must be sufficient to meet the desired particulate emission levels, but efficiency much beyond this lower limit is usually unwarranted. Equipment operating characteristics, such as temperature and pressure drop limitations, must be compatible with the specific process application. Economic factors such as purchase and operating cost, delivery time, and reliability must also be considered.

Cleaning performance of specific particulate removal equipment is perhaps most usefully expressed in the form of a grade efficiency curve. This is the relationship of dust particle size (or a related property such as free-fall velocity) versus corresponding collection efficiency for that size particle. Overall collection efficiency is obtained by integrating the product of weight fraction increment times collection efficiency at that particle size over the cumulative weight fraction from 0 to 1. Since collection efficiency is highly dependent on particle size, this approach permits calculation of overall efficiency for any particle size distribution of dust entering the device. This is a more useful way of expressing performance than is the use of adjectives such as high efficiency or low efficiency.

Representative grade efficiency curves are shown in Figure 1. The figure should not be used for design purposes, because for a given type of equipment the exact placement of a curve depends on design characteristics and operating conditions. The figure is useful for preliminary screening and evaluation of relative merit of different types of collectors.

Extensive coverage of operating principles and design methodology is given by Stern,[1] Brauer and Varma,[2] Theodore and Buonicore,[3] and Licht.[4] More recent comprehensive treatments have included those by Cooper and Alley[5], Heumann[6] and Schneller[7]. Manufacturers literature and specifications have been conveniently collected[8] for many types of particulate collectors. The recent NAPCA study[9] is perhaps the most comprehensive attempt at a direct comparison of different types of equipment from the standpoint not only of operation and efficiency, but also of economics. This chapter will attempt to survey the features, performance, and selection of some of the major types of equipment in current pollution control applications.

Cyclones

Cyclones are one of the cheaper and simpler dust collectors available, but they have a relatively low efficiency unless used with coarse dust. A basic cyclone is shown in Figure 2. Dusty gas enters the tangential inlet, typically at velocities of 50–100 ft/sec. Gas flows in a helical path, first downward in an annulus and then upward in the center, passing out the top of the cyclone through the outlet pipe. Particles are driven to the cyclone walls by centrifugal forces on the order of 100 g and exit from the dust outlet at the bottom of the cyclone cone.

In many applications cyclone alone can give satisfactory particulate removal. For more stringent requirements cyclones are often used in series with other equipment with complementary characteristics. For example, cyclones can operate at almost any dust loading that a gas stream can carry and become increasingly efficient at higher loadings so that they are useful as pre-cleaners for equipment such as filters that would be choked by a high dust loading. Cyclone efficiency increases with gas throughput in direct contrast to precipitators, and thus they may serve to damp out effects of changing operating conditions.

Cyclone Performance The conceptual simplicity of cyclone operation has encouraged considerable theoretical analysis of particle collection efficiency. The usual approach is to calculate the acceleration the particle experiences based on inlet velocity and cyclone barrel diameter, and compute a terminal velocity of particles in a radial direction based on Stokes Law. A maximum distance a particle must travel before being collected at the cyclone wall is selected (usually the width of the cyclone inlet), and a gas residence time is calculated. The result is that one can calculate a critical particle diameter, the minimum size that can, according to such oversimplified theory, be collected at 100% efficiency as well as efficiencies for smaller parties. Stern[10] has compared several such calculations.[11–17] A more recent review by Strauss[18] has included references.[19–23]

All of the above mentioned equations employ extremely simplified models of gas flow. They assume instantaneous particle acceleration and no interaction of particles as well as absence of non-idealities such as particle re-entrainment. For such reasons this purely theoretical approach has not been very successful in predicting quantitative cyclone efficiencies. It gives, however, an acceptable basis for scaling of an experimental efficiency to other operating conditions and somewhat different sizes and geometries of cyclone. A semi-empirical

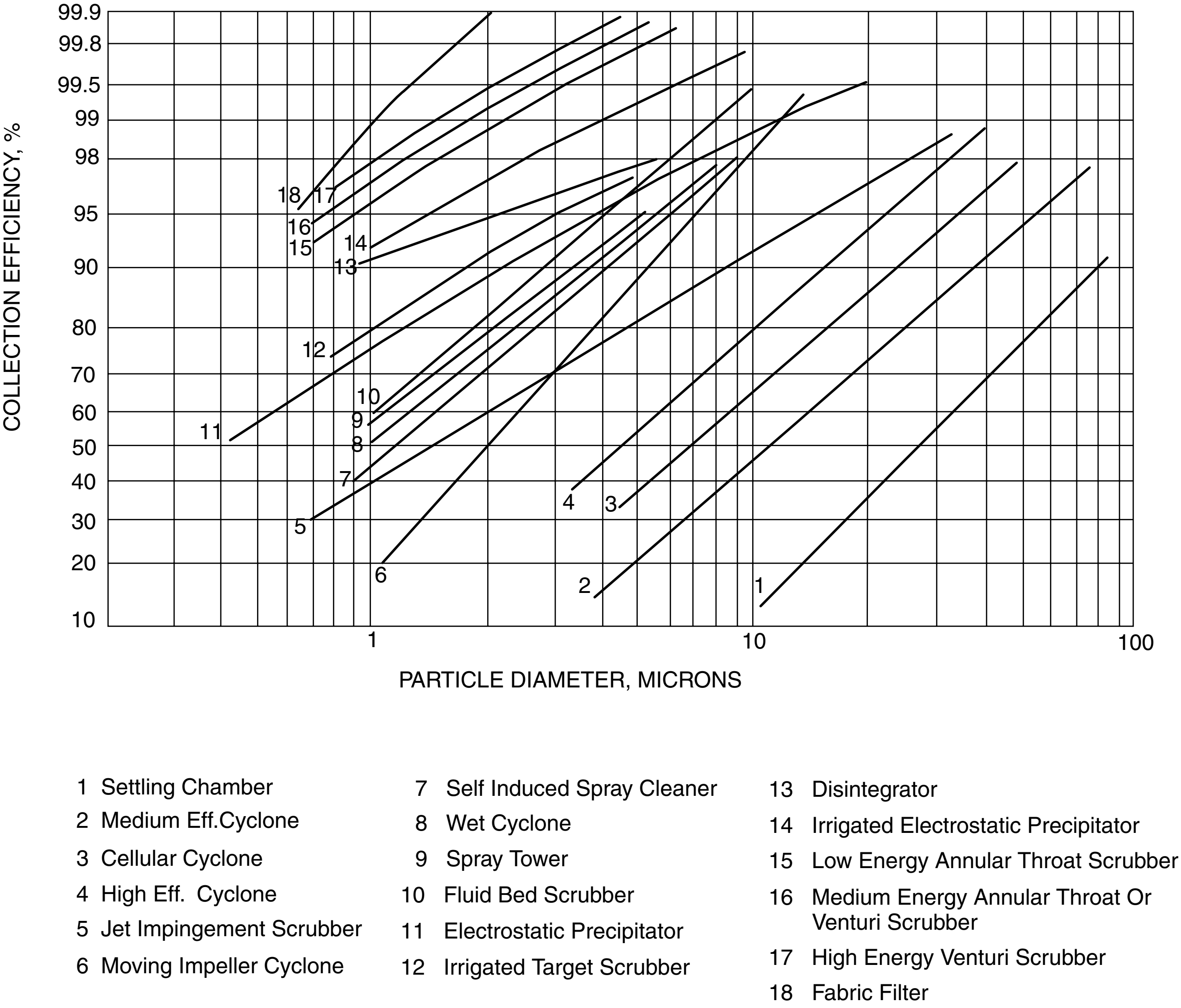


FIGURE 1 Typical grade efficiency curves.

approach can be based on the normalized grade efficiency curve shown in Figure 3. A survey of many experiments on efficiency[21–23] indicates that most cyclone grade efficiency curves have approximately the same shape. Using the theoretical basis for calculating critical particle diameter but introducing an empirical constant, one can calculate , the particle diameter at which collection efficiency is 50% at standard dust loading:

$$d^{\circ}_{50} = 1.16\sqrt{\frac{\mu WD}{\rho U^{2} t}}, \qquad (1)$$

where

μ = gas viscosity
W = cyclone inlet width
D = cyclone barrel diameter
ρ = particle density
U = gas velocity at inlet
t = average gas residence time.

Gas residence time is given by

$$t = \frac{V}{hWU},$$

where

V = cyclone volume
h = cyclone inlet height.

The above equation is applicable at dust concentrations of about 5 grains/ft^3 which is a standard concentration for

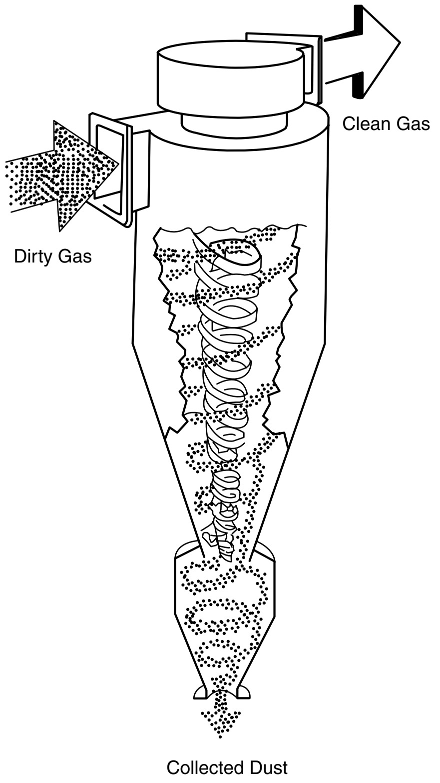


FIGURE 2 Conventional cyclone operation (The Ducon Co.).

cyclone testing. Studies of the effect of loading on grade efficiency have shown[24,25] that cyclone efficiency is very sensitive to dust loadings, and for this reason a correction to equation (1) is necessary for other concentrations:

$$\frac{d_{50}^{\circ}}{D_{50}} = \left(\frac{L}{L^{\circ}}\right)^{0.2}, \tag{2}$$

where the reference loading L° is 5 grains/ft^3. Using Eq. (2) for d_{50}, efficiency for other particle diameters d can be found from Figure 3.

Equation (1) can give directional effects for many geometric changes in design; but if data are available on the actual geometry of interest, it is much more satisfactory to use such data to modify the numerical constant in the equation. It is especially important to note that modifications to outlet pipe diameter to not enter into Eq. (1), and the numerical constant is for cyclones with normal ratios of outlet to inlet area of 1.3 to 1.5. For certain applications area ratios well below 1.0 are now being offered. Such cyclones have higher pressure drop. Unfortunately, suitable published data are not available on this effect to permit quantitative prediction of effect on efficiency.

Although some reviews have compiled "consensus" dimensions for cyclone design, quite a range of relative dimensions exists within the bounds of good design practice. Design is a compromise among efficiency, cyclone throughput, and pressure drop, and improvement in one factor will

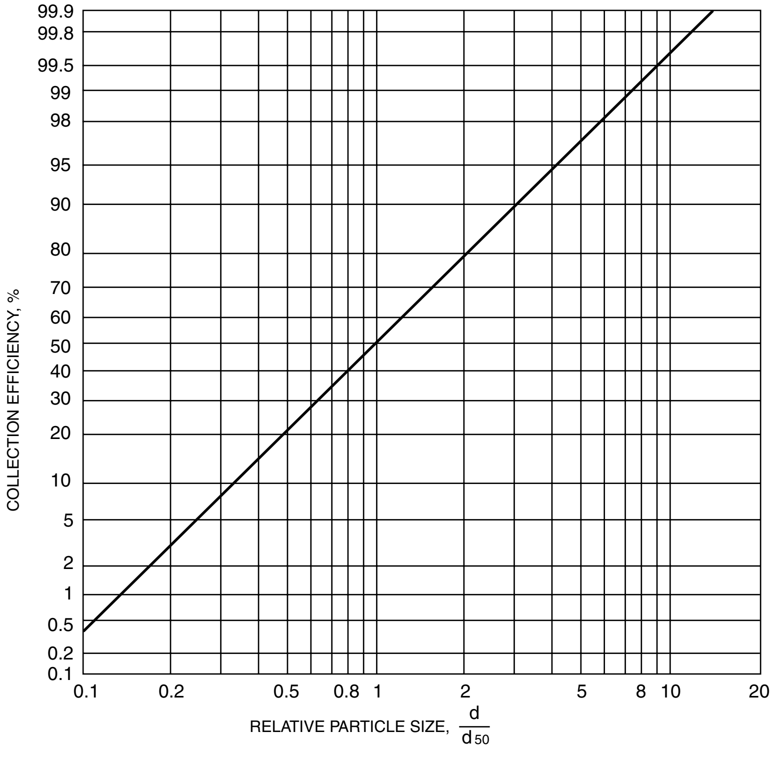


FIGURE 3 Generalized cyclone grade efficiency.

almost always result in debits to another. An example is shown in Figure 4 of three cyclones studied by Van Ebbenhorst Tengbergen.[24] They are drawn to the same scale and sized to have the same gas throughput and the same pressure drop. The relative costs of A : B : C are estimated as 4 : 3 : 9, but the difference in cost is reflected in the grade efficiency curves.

Pressure Drop Although pressure drop is much easier to measure than overall or grade efficiency, prediction of pressure drop from theoretical principles is not in good agreement with experiment. Stern[10] has analyzed the pressure drop predictions of several authors[26–28] and has concluded that a constant value of $X = 16$ in the equation

$$h_n = X\,\frac{H_i W_i}{d_0^2}, \tag{3}$$

where

h_n = pressure drop is expressed as inlet velocity heads
H_i = inlet height
W_i = inlet width
d_0 = outlet pipe dia.

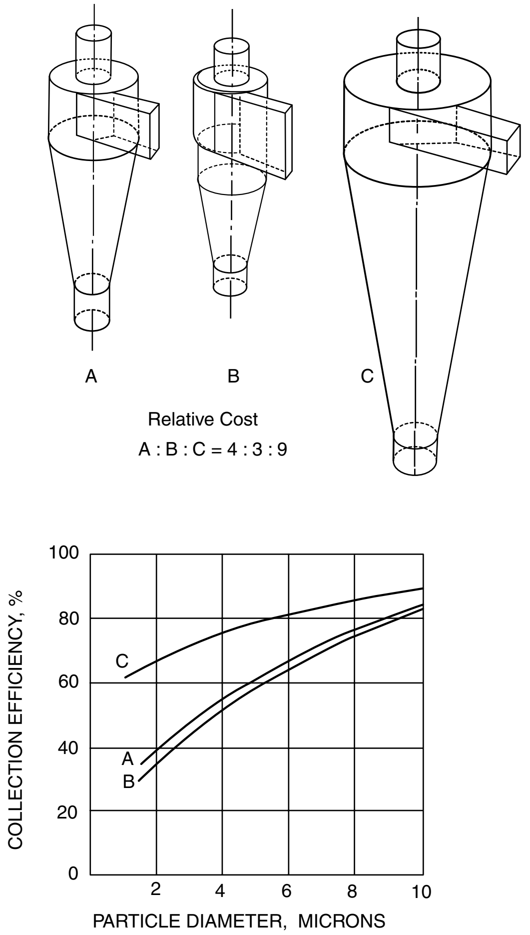


FIGURE 4 Relative sizes, costs, and grade efficiencies of different cyclone designs having same throughput and pressure drop.[21]

gives as good agreement with experimental data as does any equation for X which includes geometric factors. Measures values of X do indeed cover a wide range of values, and use of a constant value of X is merely a best guess in absence of data on the specific cyclone. Fortunately, because measurement is straightforward and simple, most vendors supply reliable information on pressure drop for their own cyclones.

For most cyclone applications the only pressure drop of interest is that from gas inlet to gas outlet. In some applications the pressure drop from gas inlet to the dust outlet at the bottom of the cone is also important. Such is usually the case for cyclones above fluid beds where collected dust is returned directly to the bed through a fluidized standpipe or dipleg. Too great a pressure drop to the dust outlet will preclude a satisfactory pressure balance in the dipleg and prevent proper discharge of particles. As a first approximation pressure drop from inlet to dust outlet can be taken as 13.5 inlet velocity heads.

Cyclones in Series Cyclones are often installed in series either with other cyclones or as pre-cleaners ahead of more efficient dust collecting equipment. As pre-cleaners they are especially attractive because efficiency increases with dust loading and because they seem able to handle almost any dust loading that a gas stream can carry to the inlet. Some cyclones in fluid catalytic crackers operate at inlet loadings of over 1lb dust per actual cubic foot of gas. Two stages of internal cyclones in a cat cracker often operate at an overall collection efficiency of 99.997% on catalyst with a particle density of 75–80 lbs/ft^3 and a mass median dia. of 60 μ. A third stage of cyclone in this application may be expected to operate at about 85% efficiency, however, because that stage sees much finer particles at much lower loading. Cat cracking cyclones typically operate at pressure drops of about 1 psi/stage, higher than is tolerable in many applications.

Optimum Cyclone Size and Parallel Operation The choice of one or a few large cyclones as opposed to many smaller cyclones operated in parallel represents a balance between the lower cost of a few large cyclones combined with easy manifolding against the inherently higher efficiency of the smaller cyclones. Each application is different, but the considerations that apply in determining size of cat cracker cyclones may prove instructive:

1) Cracking catalyst is highly erosive necessitating periodic repair of internal metal surfaces or refractory linings. The cyclones must permit access for such repairs, which sets a minimum barrel dia. of about 3½ ft.
2) Cyclone length is proportional to diameter, and headroom plus dipleg pressure balance often limits its diameter (a 5 ft dia. cyclone is typically 17 ft long). Reactor openings also limit diameter as does reduced efficiency.

The result is that in this application cyclone diameters usually range from about 3½ to 5 ft. A large cat cracker regenerator may have 20 pairs (primary and secondary) of cyclones in parallel. The secondary cyclones discharge into a common plenum head.

Operating Conditions Upper temperature limits for cyclones are limited by materials consideration to about 1700°F in conventional designs. They can withstand pressures of several psi with normal design, but for high pressure operation enclosure or special fabrication is required. Only fairly freely flowing material can normally be discharged, and special design or operating techniques are required for sticky or coking material. Low efficiency on fine particles usually precludes use where a significant fraction of material to be collected is below 10 or 15 μ in dia.

Multicellular Cyclones

Examination of Eq. (1) in the discussion of cyclones shows that d_{50}° becomes smaller and collection efficiency therefore increases as cyclone diameter decreases. Advantage is taken of this effect in multicellular cyclones, which have diameters in the range of 6–10 in. as compared to several feet for conventional cyclones. Because they are small, many cellular cyclones must be operated in parallel, and they are designed to facilitate manifolding into arrays which may number several hundred. Figure 5 shows a typical cellular cyclone. For ease of manifolding it is constructed with an axial inlet with swirl vanes to impart spin to the gas instead of having the tangential inlet of large cyclones. They are usually cast or molded in standardized designs instead of being custom fabricated as large cyclones are. An array of cellular cyclones is shown in Figure 6. This type of collector is designed to operate at low pressure drop, typically less than 4in. of water. This makes it attractive for use for fly ash recovery in power plants, where only slight pressure drops can normally be accommodated.

Arrays of cellular cyclones in parallel suffer an efficiency debit compared to single cyclones of the same design. This is usually attributed to unequal distribution of gas among cyclones abetted by the low pressure drop, and to recirculation patterns of gas down through one cyclone dust outlet and up through the next. The gas upflow causes a loss of efficiency. Many vendors will cite grade efficiency curves for a single cyclone which are, however, not representative of manifolded arrays. The magnitude of the performance debit can be seen in Figure 7.

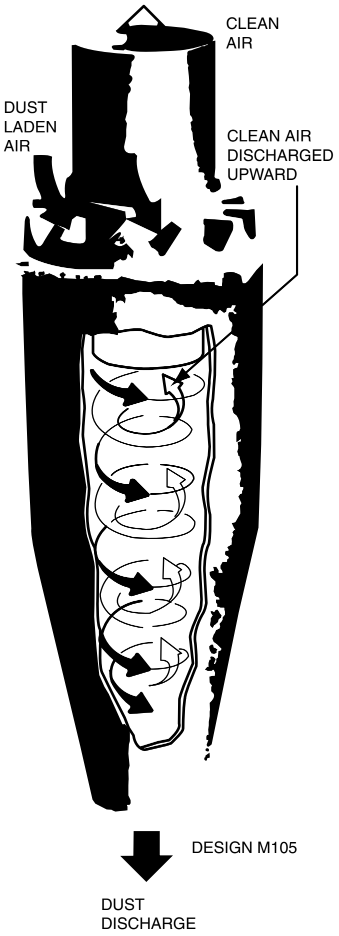


FIGURE 5 Multicellular cyclone with axial inlet vanes. (Courtesy: Wheelabrator Air Pollution Control.)

Electrostatic Precipitators

Electrostatic precipitators enjoy an image of being the ultimate in high efficiency air pollution control equipment. There is an element of reality in such a judgment, but precipitators are extremely expensive in terms of investment, operation, and space requirements, and they are often subject to obscure maladies which may cause collection efficiency to be far less than expected. Thus installation of a precipitator will be costly and may not solve a particulate emission problem.

The strong points of precipitators are (1) very low pressure drop (less than 1 in. of water, compared to about 3in. for cloth filters, 5–30 in. for centrifugal collectors and up to 60 in. for high energy scrubber); (2) sizes to handle very high gas flow rates; (3) potential of high collection efficiency on fine dust; (4) ability to operate at high temperatures (800°F is reached in commercial practice and pilot units have operated at 1700°F) and (5) dry collection of particles avoiding water pollution problem. These factors have been particularly suited to the requirement of coal-fired power plants, which constitute probably the most widespread users of precipitators or any other high efficiency particle collectors.

The standard text on electrostatic precipitator principles and practice is that by White.[29] A more recent work by Oglesby and Nichols[30] also includes an extensive survey of precipitator installations and performances. Other general references include the monographs by Rose[31] and Gottsclich.[32]

FIGURE 6 Typical manifolding arrangement of multicellular cyclones. (Courtesy: Wheelabrator Air Pollution Control.)

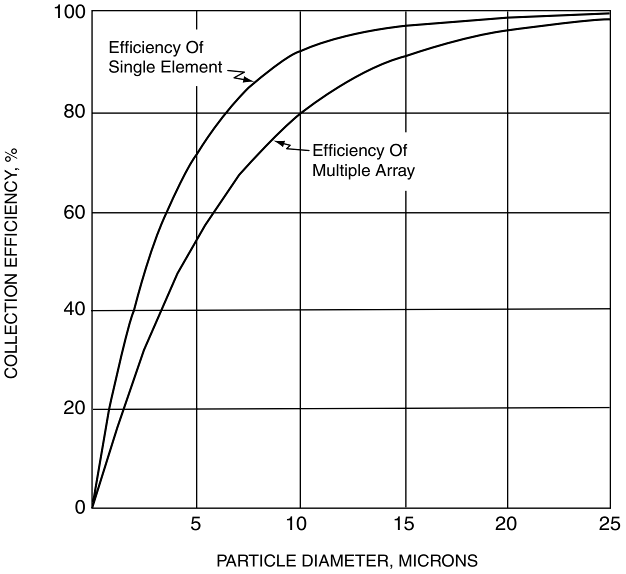


FIGURE 7 Performance debit for manifolding of multicones.

An excellent monograph on practical aspects of precipitator specification, operation, testing, maintenance, and troubleshooting has been recently published by Katz.[33]

Precipitator Operation Electrostatic precipitators have been constructed in a great many configurations. Contemporary commercial practice is fairly standard, however, and only this will be considered here. A typical precipitator, shown in Figure 8, consists of an array of parallel vertical plates, spaced about 9 or 10 in. spacing. Rows of wires hang vertically between the plates. These wires, called discharge electrodes, are maintained at a voltage of 30–60 kV with respect to the grounded plates. The very high electric field gradient near the wire electrode produces a corona discharge generating a high concentration of ions, which in turn charge the dust particles in the flowing gas stream. Under influence of the voltage gradient particles migrate to the plates, also called collector electrodes, where the charge is largely dissipated. A residual charge retains particles on the plates, and they are periodically dislodged by rapping. When dislodged, the particles fall in clumps into collection hoppers beneath the electrodes.

Gas flow across a flat plate surface tends to re-entrain collected dust particles causing an appreciable debit to collection efficiency. Various types of baffles are therefore used to shield the plate surface and provide a stagnant collection zone. The usual baffles are vertical and are perpendicular to the plate surface. They are typically 20 in. apart and extend 1.5 in. from the plate surface. Experimental and theoretical studies have shown that with such baffles re-entrainment of typical dusts by gas flow at the plate surface is not significant at gas velocities below 6 to 8 ft/sec.[34] This is well above recommended precipitator gas velocity.

The power supply for a precipitator consists of a high voltage step-up transformer and a rectifier. Rectifiers of various types have been used over the years, but solid state silicon rectifiers are almost universal in new installations. Unfiltered double half wave rectification is usually used. Full wave rectification is preferable for certain high current applications, and a simple change in connections usually suffices to change between the two wave forms.

Particle collection is enhanced by high voltage and current, but excessive voltage causes undesirable arcing or sparking between the electrodes. An arc is an ionized short circuit which diverts most current away from charging and collecting particles and also causes an undesirable voltage drop. Impingement of the arc on dust collected on the plate electrode can cause reentrainment. Optimum overall collection efficiency usually occurs when power input is adjusted to a spark rate of about 50–100 discharges/min. Modern precipitator control systems will automatically adjust voltage to maintain a preset sparking rate. Saturable core reactor control came into use in the late 1950's and a decade later this was being supplanted by entirely solid state (thyristor, i.e. silicon controlled rectifier or "S.C.R.") controls. SCR control has appreciably lower power losses and much faster response characteristics than the saturable core reactor and is able to operate at appreciably higher voltage and current without excessive sparking.

Large precipitators are usually sectionalized into several chambers in parallel and (within chambers) several electrical fields in series with respect to gas flow. Chambers can usually be independently isolated from gas flow so that one chamber may be shut down for maintenance while the others remain on stream. Use of several electrical fields in series permits optimum voltage control in each field. Outlet

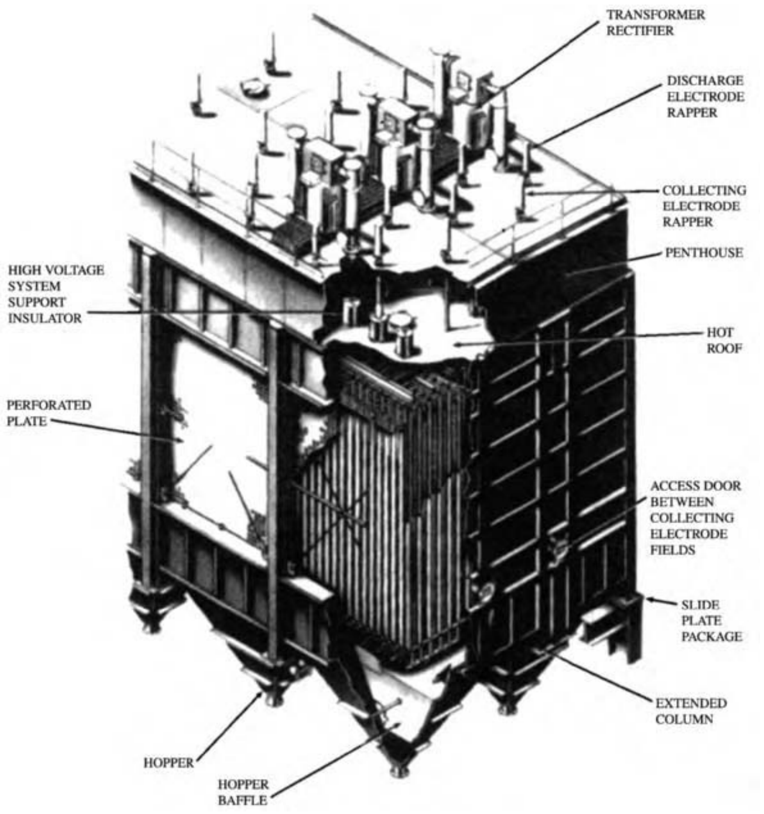


FIGURE 8 Perspective view of electrostatic precipitator showing main features of construction. (Courtesy Research Cottrell.)

fields with low dust concentrations can tolerate higher voltage without excessive sparking than can inlet fields. Further electrical sectionalization within fields is often desirable. A spark discharge momentarily hurts the performance of an entire electrical section, and small sections obviously limit the extent of the disturbance.

Electrodes are rapped with either pneumatic or solenoid rappers. Variable frequency and intensity controls are usually provided so that rapping may be optimized. Ideally that dust layer should be dislodged before becoming so thick as to interfere with electrical performance but after becoming thick enough to drop off in a coherent clump of particles. Rapping to an appreciable extent redisperses collected particles back into the gas stream, and a puff of dust is often seen in effluent gas when electrodes are rapped.

Precipitator Performance Parameters Precipitator collection efficiency is commonly expressed by the Deutsch[35] equation:

$$\eta = 1 - \exp\left(-\frac{AW}{Q}\right), \tag{4}$$

where

η = collection efficiency
A = plate electrode area
Q = volumetric gas flow rate
W = effective migration velocity rate of the dust.

In practice the migration velocity is an empirical constant which depends on the application and dust and gas properties. Typical migration velocities are given in Table 1. Although a considerable range of values is shown for some applications, a large part of such variation can be attributed to known causes and effects and does not represent uncertainty.

White[26] has derived a theoretical equation for migration velocity

$$W = \frac{E_o E_p d_p}{4\pi\mu},$$

where

E_o = charging field gradient, stat V/cm
E_p = collecting field gradient, stat V/cm
d_p = particle diameter, cm
μ = gas viscosity, poise.

Values of W measured on commercial installations are lower by a factor of 2–5 than the theoretical value above. Although the theoretical W is of little use for qualitative

TABLE 1
Representative precipitation rates for various applications[27]

Application	Precipitation rate w ft/sec	
	Average	Range
Utility fly ash	0.43	0.13–0.67
Pulp and paper	0.25	0.21–0.31
Sulfuric acid	0.24	0.20–0.28
Cement (wet)	0.35	0.30–0.40
Smelter	0.06	—
Open hearth	0.16	—
Cupola	0.10	—
Blast furnace	0.36	0.20–0.46

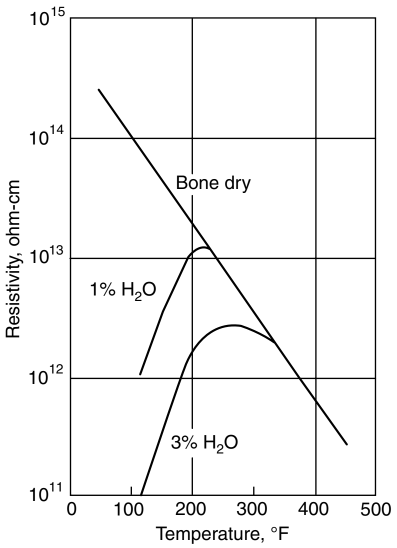


Figure 9 Effect of temperature and humidity on particle resistivity.[29]

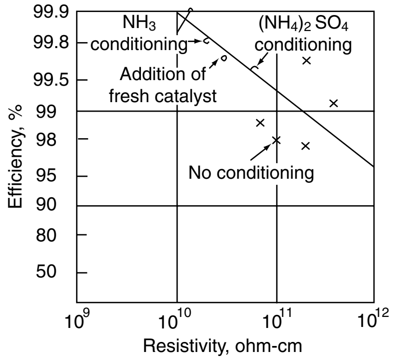


Figure 10 Effect of gas conditioning on efficiency of electrostatic precipitators.[26]

design, it does give a clue as to the influence of operating conditions of W. The term E_oE_p is roughly proportional to the power input to the precipitator, and it is clear that this should be as high as possible. From the dependence of W on d_p it is clear that precipitators have a higher collection efficiency for large particles than small ones. This is seldom taken into account in current design procedures in which an average particle size is included in the empirical W for an application.

Dust resistivity can have a very pronounced, although not quantitatively predictable, effect on precipitator performance. If the *in situ* resistivity of a dust layer collected on the plate electrode is less than about 10^7 Ω-cm, the electrical force holding the dust particles to the electrode will be low, and significant amounts of dust may be reentrained into the gas stream. If the resistivity is above 2×10^{10} Ω-cm, excessive sparking and back corona will occur hurting electrical performance and efficiency. Resistivity is a function not only of the dust but also of temperature and gas composition. At low temperatures surface conduction is predominant due to a thin layer of adsorbed moisture, and this decreases with increasing temperature. Volume conduction is important at high temperatures and decreases with temperature. A maximum resistivity usually occurs at about 400°F due to these opposing tendencies as is shown in Figure 9.

Gas conditioning by addition of moisture or other substances is sometimes practiced in order to improve the dust resistivity and enhance electrical performance of the electrodes. In certain instances addition of 10–20 ppm of ammonia has dramatically improved precipitator performance. Sulfur trioxide has also been used successfully, but it is expensive and creates its own pollution and safety problems. The effect of resistivity change, caused by conditioning, on precipitator efficiency is shown in Figure 10. In some gas streams conditioning agents may be naturally present which enhance precipitator operation. For instance, as indicated in Figure 11, power plants burning high sulfur fuel have improved performance due to the conditioning properties of sulfur combustion products. The present trend towards mandating use of low sulfur fuels in many urban areas will have an adverse effect on particulate removal.

Equation (4) shows the dependence of efficiency on gas flow rate provided flow is uniform. If gas velocities are different throughout the precipitator due to maldistribution of flow, the equation will be applicable only to local efficiency and the overall precipitator efficiency will be considerably lower. For instance a precipitator which could operate at 98% efficiency with even gas distribution would operate at only 96.5% efficiency if the same total gas flow were maldistributed so that half of the precipitator received 30% more than average and the other half received 30% less than average. Sharp turns and rapid expansion and contraction in dust work often make it difficult to achieve good gas distribution. Very often preliminary flow distribution tests are made on $\frac{1}{6}-\frac{1}{16}$ scale clear plastic models using smoke as a tracer. Such model tests are helpful in calling attention to duct work designs which will

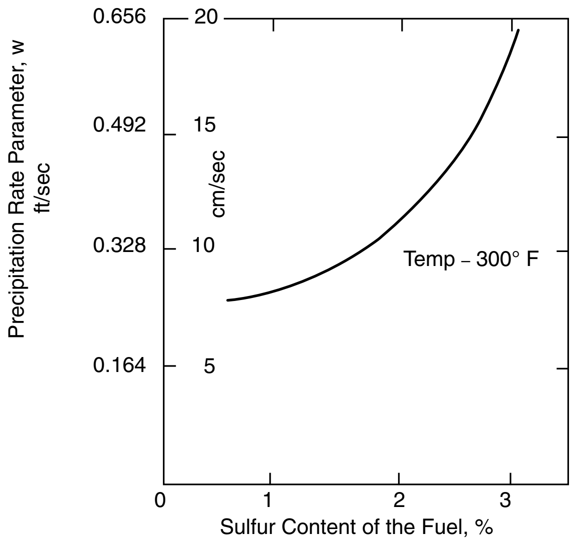


Figure 11 Effect of sulfur in fuel on precipitation velocity.[27]

cause poor gas flow patterns and allowing trials of baffle, diffusers, and turning vanes to reduce the problems.

Fabric Filtration

The process of particulate removal by filtration is one of the oldest methodologies in the gas cleaning field. Fabric filters possess several advantages, the most important one being that dust collection efficiencies of better than 99% are obtainable with sub-micron particles. Fabric filters are reliable, and relatively simple to operate, require moderate capital and operating costs and can be designed to handle a wide range of particulate loadings and gas flows. On the other hand, the choice of a fabric material limits the temperature, and corrosivity characteristics of the gas and high humidities must be avoided in order to prevent cake formation.

Removal of particulates from a gas stream by fabric filtration involves passing the gas through a filter medium made of a porous fabric. The solid is deposited on the fabric with a resultant increase in the flow resistance of the gas. A means of removing the solid matter which is built up on the filter is thus necessary to avoid excessive pressure drop of the gas through the filter cake.

Industrially, the filtration medium is contained in equipment known as a baghouse. These baghouses function to hold a large area of the filter exposed to the gas flow. The filter medium itself is generally in a tubular shape with gas flowing from the inside, through the medium and exiting outside the tubes.

This section will deal with the mechanisms of particulate collection on fabric filters, some of the design parameters of baghouses, performance characteristics of baghouses, and some data on the cost of baghouse equipment and the cost of running a fabric filtration.

Mechanisms of Particle Collection

Particulate collection by fabric filters essentially involves the capture of a particle by a collecting body. The gas stream carries the particles close to the collecting body, where a variety of mechanisms are involved in the actual capture. The mechanism or mechanisms which predominate depend on the size and velocity of the particles, and on the characteristics of the filter element itself.

The three major collection mechanisms involved in fabric filtration are (1) Interception, (2) Inertial Impaction, and (3) Diffusion. In addition to these three mechanisms, gravity, electrostatic, and thermal forces may act to modify filtration performance in special cases, but in general, these effects are negligible.

The mechanism of interception can be visualized as follows. The particle is assumed to follow the streamlines of the gas flow. As the gas flows by the collector, the streamlines are bent. Interception occurs when the streamlines on which the particle travels come close enough to the collecting body such that the particles comes into contact with the body. The efficiency of interception is a function of the ratio of the diameter of the particle to the diameter of the target body and to some extent to the flow characteristics of the stream. Ranz[36] has developed an expression for the efficiency in the case of a cylindrical target

$$\eta = \frac{1}{2.002 - \ln \mathrm{Re}_c}\left[(1+R)\ln(1+R) - \frac{R(2+R)}{2(1+R)}\right] \quad (1)$$

The acrylics (for example Orlon®, Acrilan®, Creslan®, and Crylon®) can be used at temperature up to about 275°F, but have relatively poor resistance to alkalis. Their resistance to acids and moist gases leads to application in the metallurgical industries as well as others such as carbon black, cement, fertilizers, solvents, and galvanizing.

The polyesters (for example Dacron®, Fortrel®, Vycron®, and Kodel®) can be used at temperatures of 250°F and have good resistance to alkalis as low temperatures.

Glass, Nomex®, and Teflon® are the fibers of choice of high temperature application (400–500°F). Glass is superior to the others in terms of high chemical and thermal resistance but suffers from very poor resistance to abrasion. Abrasion resistance can be improved by coating the glass fibers with a lubricant such as silicon oil or graphite. Applications include cement, gray iron cupolas, and electric arc steelmaking. Nomex®, an aromatic polyamide from duPont, can be used at temperatures up to about 400°F and excellent abrasion and flexural resistance.

Cloth Cleaning A large amount of effort in the development of fabric filters has gone into methods of removing the deposited cake from the fabric. Several considerations are important with regard to cloth cleaning.

1) The cleaning time should be much smaller than the filtering time, otherwise effective use is not being made of the filter.

TABLE 2
Relative properties of some common fabrics

Fiber	Max. temp. (dry) °F	Resistance to moist heat	Resistance to abrasion	Resistance to mineral acids	Resistance to alkalis
Cotton	180	G	G	P	G
Wool	200	F	F	F	P
Rayon	200	G	G	P	F
Nylon	200	F	E	P	G
Polypropylene	200	F	E	G	E
Acrylic	260	G	G	G	F
Polyester	275	F	E	G	F
Nomex	400	E	E	F	G
Teflon	400	E	F	E	E
Glass	550	E	P	E	P

E = excellent; G = good; F = fair; P = poor.

2) The cake should be removed uniformly.
3) Some residual cake is necessary after cleaning for efficient filtering.
4) The cleaning operation should not result in a large dispersion of dust because either a refiltration operation would be necessary or the efficiency of the filter would suffer.
5) The cleaning method should not result in excessive bag damage.

In general, the filter cloth is cleaned by shaking, gas flow reversal, or some combination of the two methods.

Mechanical shaking from the upper fastening is the most common method of cleaning the filter bags. Amplitudes vary from less than 1 in. to a few inches, and the frequency is usually several cycles per second. During the cleaning, the filtration is stopped, otherwise the dust will have a tendency to work through the bags and result in internal abrasion. For this purpose, the baghouse is built in several compartments to allow the isolation of one compartment at a time for cleaning. One disadvantage of this method of cleaning is the relatively large amount of fabric flexure involved especially at the top of the bags.

The second method of cleaning the bags is by flow reversal. If the dust cake is loosely adhering to the bag, then only a low-pressure reversal of fast flow will be necessary for cleaning; however, partial collapse of the bag can be done for the more adherent cakes. Although this method avoids some of the problems associated with bag flexure during shaking, it does have several disadvantages. Reverse flow without bag collapse requires that some support structures be placed within the bags (rings, screens etc.) which leads to difficulty in changing bags and chafing effects. Also, the reverse air used for cleaning has to be refiltered which results in a larger baghouse for a given filtration capacity.

A combination of the above two methods has been devised which is known as pulse cleaning. In this method, a pulse of compressed air is released from nozzles which results in a combination of flow reversal and fabric deformation. A unique feature of pulse cleaning is the short time required for cleaning. Therefore pulse cleaning is useful for equipment which must handle very high dust leadings. One of the major disadvantages of the method is its tendency to over-clean, especially for baghouses equipped with woven bags.

Other variations are available such as reverse jet cleaning which employs a jet of compressed air supplied from a ring which travels up and down on the outside of the bag. Occasionally, sonic vibration is used to aid in the cleaning.

Baghouse Construction The type of baghouse construction depends on the moisture condensation and temperature considerations. Below a temperature of about 135°F, an open-sided design can be used (inside filtration) which is the least expensive type. A closed baghouse is used at higher temperatures, and above 160°F, the structure is usually insulated.

Materials of construction are selected with regard to corrosiveness of the gas handled. Steel, aluminium, and asbestos cement are some typical materials of construction.

Performance of Fabric Filters The two parameters which are important in determining the performance of the fabric filters are its efficiency and pressure drop. The efficiency is a measure of how effectively the device is performing its job. The pressure drop is one of the major variables which determines operating cost because it affects the fan power necessary to move the gas through the filter, and also determines the frequency and duration of the cleaning cycle.

There are several factors affecting filter efficiency. The major one is particle size. In general, particles having diameters $>1\ \mu$ will be filtered with a $+99\%$ removal efficiency. Efficiencies are about 90% for particles $<1\ \mu$. In a typical baghouse application, most of the particles passing through the filter are of sub-micron size.

The amount of dust buildup on the filter medium also affects the efficiency of collection, because a particle has a greater probability of being trapped by one of the filtration mechanisms discussed previously if there is a greater amount of cake. Thus, cleaning should not be too thorough.

TABLE 3
Breakdown of capital and operating costs for fabric filtration

A. Capital investment

Item	Percentage of total cost
Planning and design	4.2
Baghouse, FOB	33.6
Freight	2.0
Fan and motor	10.5
Ducting	27.5
Dust disposal	4.2
Instrumentation	2.0
Installation labor	11.8
Start-up	4.2

B. Operating cost

Item	Percentage of total cost
Electric power	11.4
Labor	28.6
Plant overhead	23.8
Cloth	9.5
Depreciation	15.2
Interest	11.5

The dust loading *per se* does not affect the efficiency significantly; however in the case of very low dust loadings (<0.5 gr/ft^3), a low efficiency will be observed for a time because of the low dust buildup on the fabric. To cirumvent this problem, a precoat is often used.

The relative humidity of the gas affects the efficiency somewhat, depending on the material being filtered. For some particles, an increase in relative humidity results in increased particle adhesion force, which tends to increase the rate at which the openings in the filter medium are bridged, resulting in an increase in the collection efficiency. The filtering velocity does not appear to affect efficiency significantly, except that higher velocities would tend to force more dust through defects in the cloth.

The pressure drop of the gas through the filter is a function mainly of the gas velocity, the thickness of the dust cake, and of the particle characteristics which determine the porosity of the cake which has built up. There have been a number of studies on pressure drop through fabric filters and the reader is referred to Billings and Wilder[37] for a more complete treatment of this subject.

Economics of Fabric Filtration

A detailed survey of fabric filtration costs was done by Billings and Wilder.[37] There was a relatively large range of reported costs. Capital investment was found to vary from less than $l per cfm to greater than $10 per cfm. operating costs varied from about $17 per cfm per year Values of about $2.50/cfm capital and $1.10/cfm annual operating can be considered typical. These costs are broken down into component parts into Table 3. It is apparent from an examination of the items contribution to the total economic picture that the costs for an installation will depend a great deal on many factors specific to the installation and gas being handled.

GUARANTEE SPECIFICATIONS

Equipment vendors usually make a guarantee of overall particle collection efficiency as a part of bid proposals, but historically such guarantees have been more an indication of anticipated performance than an enforced or enforceable contract for several reasons. The isokinetic sampling always necessary for a proper efficiency test is difficult and time consuming and in some cases, e.g. for internal cyclones above fluid beds, it is a practical impossibility. Secondly, in the days of less stringent emission statutes the mere presence of control equipment often served to satisfy the spirit if not the letter of the law. Most guarantees in the past were written for a fixed set of operating conditions and were not legally enforceable at the deviant conditions that usually obtained during a test. Finally a legal recourse to poor performance is seldom spelled out in a contract. In the present climate of stringent emission limits and enforcement thereof, of substantial penalties for non-compliance, and of long equipment delivery times, purchasers are obliged to become more insistent that equipment perform according to a strict guarantee. Accordingly, the following suggestions are made for the writing of a performance specification.

A guarantee can only deal with the cleaning efficiency of a piece of equipment on the dusty gas stream actually entering it. Vendors quite understandably will refuse to guarantee an outlet loading, a Ringlemann number, or a dustfall, because these depend directly on the amount and particle size of dust entering the equipment as well as on the inherent capabilities of that equipment. In addition, optical properties of plumes are too little understood for generally acceptable prediction of Ringlemann numbers, even when the concentration and particle size of the emissions are known.

Performance guarantees are often written for a single set of design conditions without any provision for adjustment if, as usually happens, the design and actual test conditions are different. In such a situation the likelihood of proving and receiving compensation for substandard performance is small. It is therefore desirable that a range of operating conditions be specified and a procedure be defined for adjusting the guarantee performance at design conditions to a consistent performance at other conditions. The adjustment should be based on generalized efficiency relationships, e.g. Eq. (1) for cyclones and (4) for precipitators, or else on vendors design procedures. For instance Eq. (1) predicts that for a given cyclone, the response to changes in operating conditions is given by

$$d_{50} \propto \sqrt{\frac{\mu}{\rho U}}.$$

This used with the generalized grade efficiency curve, Figure 1, and the dust loading correction will give a suitable adjustment for changes in operating conditions so that starting from a guarantee performance at design conditions corresponding performance for other conditions may be calculated. In essence this establishes a guarantee envelope or curve to replace guarantee of a single point.

"Outlet stoppers," that is, low outlet dust loadings to be accepted as *prima facie* evidence of satisfactory performance regardless of actual efficiency, are sometimes included in guarantees. The reasons are that isokinetic sampling is laborious and increasingly imprecise at low dust loadings, and that certain types of equipment are thought to have inherent limits on outlet loadings, for example precipitators have rapping losses which may control overall efficiency in the low loading range. It is felt, however, that outlet stoppers should be used only to exclude certain tests of the guarantee and not as evidence that the guarantee has been met. Cases are known in which the purchaser desired an outlet loading of approximately .003 grains/ft^3 in order to meet anticipated regulations, but precipitator vendors specified outlet stoppers of .01 grains/ft^3.

Generally speaking, it is desirable to the purchaser to operate equipment for a number of weeks before making a guarantee acceptance test. Deterioration from factory performance will often occur in such a time span. This time span should be used only as a shakedown period and not a life test.

Consideration should be given to the action to be taken if equipment does not satisfy the guarantee performance, and a guarantee will usually be the better for spelling out such remedies. One frequent clause specifies that the vendor remove below par equipment and refund the purchase price. This gives considerable incentive to the vendor to meet the guarantee, but removal of pollution control equipment may mean complete shutdown or production at a loss of tens of thousands of dollars per day to the user. Another arrangement is the bonus-penalty clause, under which the vendor on a sliding scale receives a bonus or pays a penalty if equipment performs above or below guarantee. At present the bonus payment aspect is unattractive to the purchaser, because he receives no benefit from producing less pollution than the law allows although this would change if a system of effluent taxation were implemented. Most often the most satisfactory approach is to require the vendor to upgrade installed equipment and make additions as needed so that the guaranteed performance level is reached.

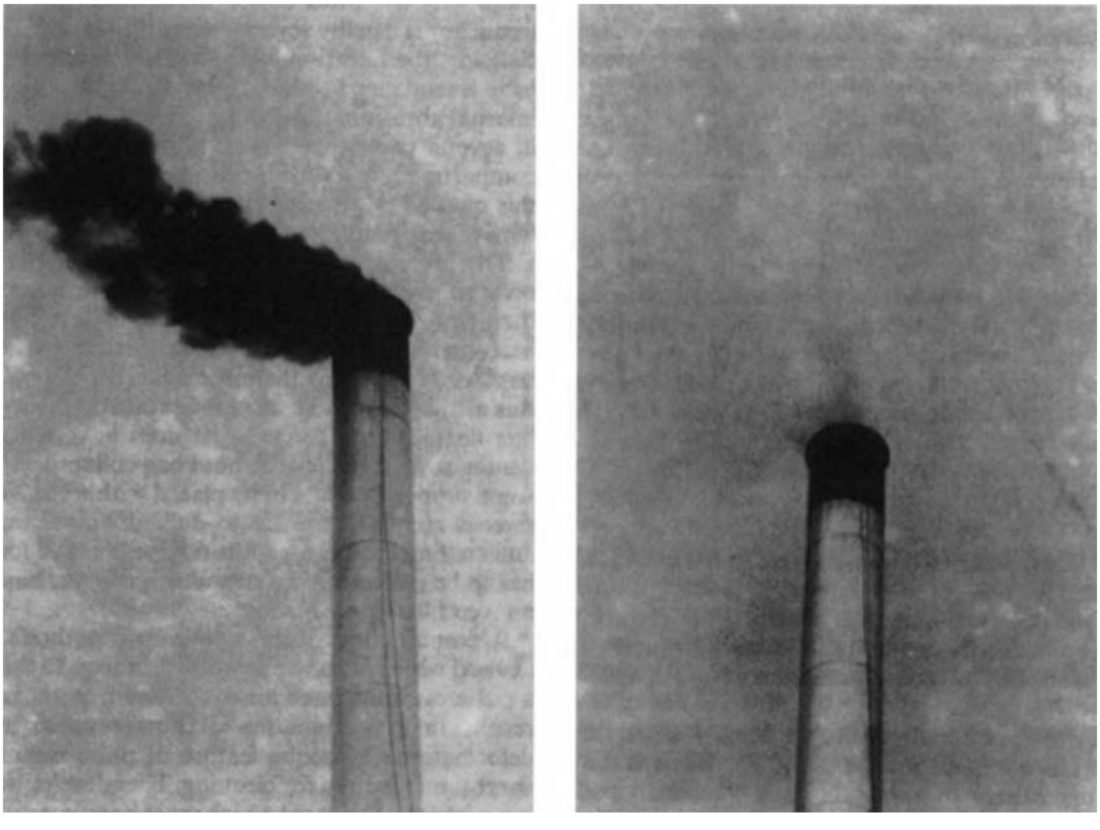


FIGURE 12 Before and after. The photograph on the left shows the typical dense smoke and soot being emitted from the stack of an industrial oil-fired boiler during the daily tube blowing or cleaning process. The photograph on the right shows the same stack during tube blowing with an emission control unit in operation.

REFERENCES

1. Stern. A.C., ed., (1976), *Air Pollution,* **4**, 3rd Edition, Academic Press, New York.
2. Brauer, H. and Varma, Y.B.G. “Air Pollution Control Equipment” Springer Verlag, Berlin, 1981.
3. Theodore, L. and Buonicore, A.J. “Industrial Air Pollution Control Equipment for Particulates” C.R.C. Press, Cleveland, 1976.
4. Licht, W. “Air Pollution Control Engineering” Marcel Dekker Inc, New York, 1980.
5. Cooper, C. D. and Alley, F.C. (2002) *Air Pollution Control: A Design Approach.* Waveland Press, Prospect Heights, IL.
6. Heumann, W. A. (1997) *Industrial Air Pollution Control Systems* McGraw-Hill, New York.
7. Schneller, K. B. and Alley, F. C. (2002) *Air Pollution Control Technology Handbook* CRC Press, Boca Raton, FL.
8. Pazar. C., *Air and Gas Cleaning Equipment,* Noyes Data Corp., Park Ridge, N.J., November 1970.
9. National Air Pollution Control Administration, *Control Techniques for Particulate Air Pollutants,* NAPCA Publication AP-51, Washington, January, 1969.
10. Stern, A.C., K.J. Caplan, and P.D. Bush (1956) *Cyclone Dust Collectors,* American Petroleum Institute, New York.
11. Davies, C.N., (1952), Separation of airborne dusts and particles, *Inst. Mech. Eng. Proc.* (B) **1B** 5, 185.
12. Gardiner, J.E., Tech. Report No. 1-CT/19, Shell Petroleum Co., Ltd., London.
13. Lapple, C.E., (1950), Gravity and centrifugal separation, *Indust. Hygiene Quart.,* **11**, 40.
14. Lapple, C.E., and C.B. Shepherd, (1940), Calculation of particle trajectories, *Ind. Eng. Chem.* **32**, 605.
15. Muhlrad, S.W., (1954), *Chaleur et Ind.,* **35**, 41.
16. Rosin, P., E. Rammler, and W. Intlemann (1932), Theory and limits of cyclone dust removal, (in German), *Zeit. Ver. Deut. Ing.,* **76**, 433.
17. Tarjan, G., (1950), Contributions to the kinematics of the cyclone and the cyclone scrubber, (in Hungarian), *Acta. Tech. Acad. Sci. Hung.*, **1**, 22.
18. Strauss, W. “Industrial Gas Cleaning, 2nd Edition” Pergamon Press, Oxford 1975.
19. Rietema, K., (1961), *Cyclones in Industry,* Elsevier, Amsterdam.
20. Feifel, E., Cyclone dust collection; The ideal eddy sink and its approximation, (in German), *Ver. Deut. Ing. Forschungshefte,* **9**, 68 (1938); **10**, 212, 1939.
21. Stairmand, C.J., (1951), The design and performance of cyclone separators, *Trans. Inst. Chem. Engng.*, London, **29**, 356.
22. Barth, W., (1956), Calculation and interpretation of cyclones on the basis of new studies, (in German), *Brennstoff-Warme-Kraft,* **8**, 1.
23. Ter Linden. A.J., (1953), Studies of cyclone separators, (in German), *Tonindustrie Zeitung,* **22** (iii), 49.
24. Van Ebbenhorst Tengbergen, H.J., Dust cyclones-Law of similarity-Influence of the dust concentration, *De Ingenieur,* Irg 77 No. 2, W1, January 8, 1965.
25. Kriegal, E. (1967), Model studies of cyclone separators, *Techn. Mitt. Krupp Forsch.-Ber.*, **25**, 21.
26. Alexander. R. McK. (1949), Fundamentals of cyclone design and operation, *Aust. Inst. Min. Met. Proc.,* New Series, **152**–3, 202.
27. First, M.W., Cyclone dust collector design, Am. Soc. Mech. Eng. Paper 49-A-127, 1949.
28. Shepherd, C.B., and C.E. Lapple, Flow pattern and pressure drop in cyclone dust collectors, *Ind. Eng. Chem.,* **31**, 972 (1939); **32**, 1246, 1940.
29. White, H.J. (1963), *Industrial Electrostatic Precipitation,* Addison Wesley, Reading, Mass.
30. Oglesby, S., and G.B. Nichols, *A Manual of Electrostatic Precipitator Technology,* Southern Research Institute, Birmingham, August, 1970. Distributed by National Technical Information Service No. PB-196–380.
31. Rose, H.E., and A.J. Wood (1966), *An Introduction to Electrostatic Precipitation in Theory and Practice,* Constable, London.
32. Gottschlich, C.F. (1961), *Electrostatic Precipitators,* American Petroleum Institute, New York.
33. Katz, J. “The Art of Electrostatic Precipitation,” 2nd Edition Precipitator Technology Inc., Munhall PA., 1989.
34. Halow, J.S., Personal Communication.
35. Deutsch, W. (1922), *Ann. der Physik,* **68**, 335.
36. Ranz, W.E., Technical Report No. 8, University of Illinois Engineering Experimental Station, (January 1953).
37. Billings, C.E., and J. Wilder, *Handbook of Fabric Filter Technology,* **I**, Report from GCA Corporation, Bedford, Mass. to National Air Pollution Control Association (EPA), December, 1970.
38. Culhane, F.R. (1968), Production baghouses, *Chem. Eng. Prog.,* **46**(1), 65.

JOHN M. MATSEN
Lehigh University

PCBs AND ASSOCIATED AROMATICS

INTRODUCTION

Polychlorinated biphenyls (PCBs) are a family of chemical compounds which do not exist in nature but which are manmade. Commercial mixtures are clear, pale yellow liquids manufactured by the replacement of hydrogen atoms on the biphenyl molecule by chlorine. The reaction is controlled according to the percentage replacement of hydrogen and results in a product which contain chlorobiphenyls with one or, more usually, several chlorine atoms attached to each molecule.

The family of polychlorinated biphenyls can contain from 1 to 10 chlorine atoms arranged within the molecule in 209 different ways. The different molecular combinations are called congeners. The congeners which comprise the series of compounds with the same number of chlorine atoms are known as an homologous series.

The congeners are named according to the homologous series to which they belong. For example, a PCB containing 3 chlorine atoms would be called a trichlorobiphenyl, or one with 5 chlorine atoms a pentachlorobiphenyl. The homologous series of PCBs and the number of isomers contained in each are listed in Table 1.

The synthesis of biphenyl and its subsequent reaction with chlorine to yield polychlorinated biphenyls is shown in Figure 1.

PCBs have been manufactured by various companies throughout the world and are often referred to by the manufacturer's trade name. Table 2 lists the main producers.

The perceptions of risk associated with ownership and operation of PCB containing equipment is related to the risk of fires, spills, cleanup or noncompliance with regulations. The following time line focuses on the development of PCB risk perceptions as new knowledge grew from the time that PCBs were first used in industrial applications.

1881

PCBs were first described in the chemical literature in 1881 by Schmidt and Schultz.

1929–1931

PCBs were developed by the Federal Phosphorus Company and Swann Research Co. The latter business was bought by Monsanto in the Early 1930's.

The available data on the toxicity of PCBs were first summarised in an article published in 1931 (Smyth, 1931).

1933

The first reported episode of PCB-related toxicity occurred in 1933, at the Swann Chemical Co. Twenty three of 24 men working on the manufacture of PCB almost simultaneously developed skin eruptions that were diagnosed as 'chloracne'. Some of the men also complained of lassitude. The investigators report on the episode concluded that the toxic agent was an unidentified impurity, which had probably appeared

TABLE 1

Chlorobiphenyl	Empirical Formula	# of Isomers	Weight% Cl
mono	$C_{12}H_9Cl$	3	18.8
di	$C_{12}H_8Cl_2$	12	31.8
tri	$C_{12}H_7Cl_3$	24	41.3
tetra	$C_{12}H_6Cl_4$	42	48.6
penta	$C_{12}H_5Cl_5$	46	54.3
hexa	$C_{12}H_4Cl_6$	42	58.9
hepta	$C_{12}H_3Cl_7$	24	62.8
octa	$C_{12}H_2Cl_8$	12	66.0
nona	$C_{12}HCl_9$	3	68.7
deca	$C_{12}Cl_{10}$	1	71.2

Adapted from Reference 1: Hutzinger *et al.*[1]

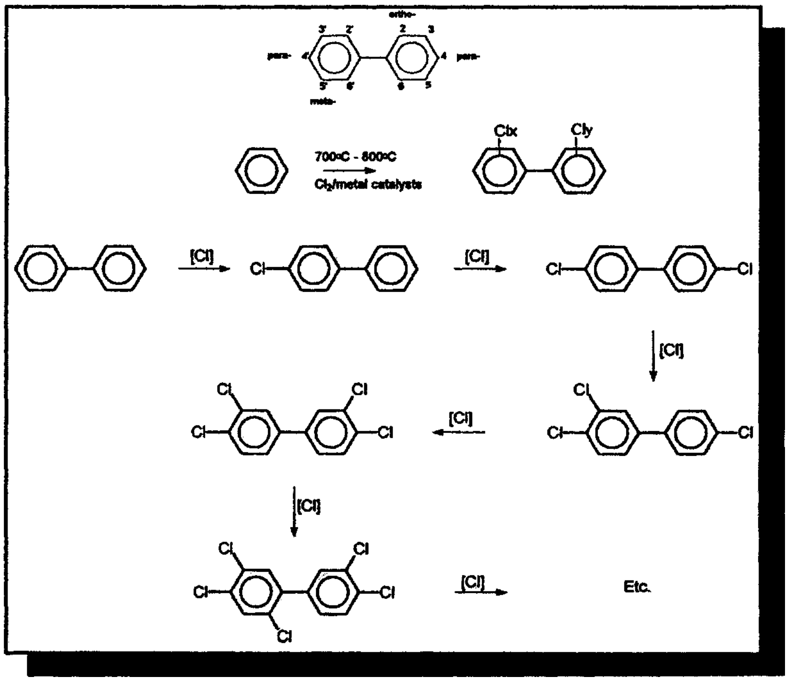


FIGURE 1 The reaction of anhydrous chlorine with biphenyl results in isomers of homologous series of PCBs.

TABLE 2
Commercial PCB mixtures have been produced worldwide

Trade name	Manufacturer	Country
Aceclor	ACEC	Belgium
Apirolio	Caffaro	Italy
Aroclor	Monsanto	United States
Clophen	Bayer	West Germany
Kaneclor	Mitsubishi	Japan
Pyralene	Prodelec	France
Pyranol	G.E.	United States
Pyroclor	Monsanto (U.K.)	United Kingdom

in the PCB when a new source of crude benzene was used for making the biphenyl. When purchase from this benzene source was discontinued, and the process better enclosed, the chloracne problem disappeared.

1936

Halowax acne was a term first coined by Jones and Alden in 1936 to describe chloracne induced by PCBs. PCBs were being manufactured in a distillation process in which biphenyl was extracted from coal tar using benzene, styrene, toluene and xylene. Once the biphenyl was extracted, it was chlorinated with chlorine gas to produce PCB.

Chloracne was identified in workers manufacturing Aroclors [Schwartz, 1936].

1937

An early case report of death by cirrhosis of the liver with acute yellow atrophy may be attributable to PCNs rather than PCBs [Drinker *et al.,* 1937].

1938

Morphological changes were observed in rats dosed orally with PCBs (Aroclor 1265) in concentrations as low as 0.001 mg/Kg/d [Bennett *et al.,* 1938].

1943

Good and Pensky described the formation of acneiform eruptions in workers manufacturing Aroclors.

An investigation of dermatitis among cable strippers in the shipbuilding industry revealed lesions, cysts and keratinization of skin in not only the workers, but also in some spouses and children [Schwartz, 1943].

1944

Skin lesions similar to chloracne in humans were observed in three species of laboratory animals subcutaneously exposed to Aroclor 1242 at levels ranging from 7 mg/Kg/d to 138 mg/Kg/d [Miller, 1944].

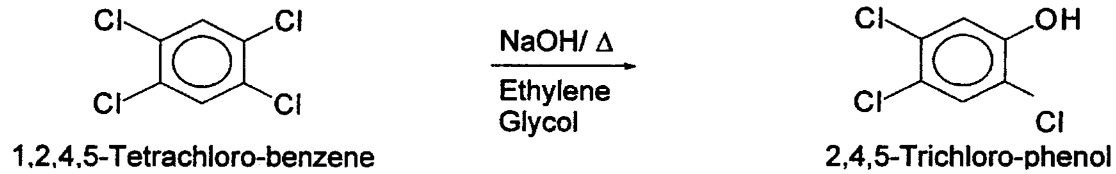


FIGURE 2 The intended reaction produces TCP.

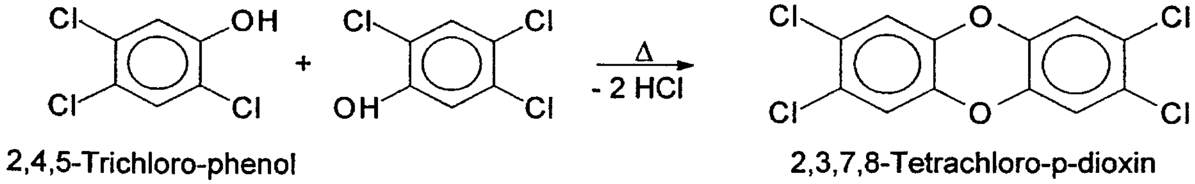


FIGURE 3 The dimerization of chlorophenates yields PCDDs.

1949

PCBs extracted from air were analyzed by Jacobs (1949) before the advent of gas chromatography.

1950

In the 1950–1952 time frame, an episode of PCB poisoning occurred when 7 of 14 people exposed to vapors from a leaking PCB-filled heat exchanger developed chloracne. [Meigs, 1954].

1952

Askarel used in capacitors up to 1952 was Aroclor 1254. After 1952 the type of askarel was changed to Aroclor 1242.

1953

Badische Anilin had an industrial accident on 17th November, 1953 and is one of seven similar accidents which have been recorded. The process involved the manufacture of trichlorophenol from tetrachlorobenzene (Figure 2).

The Seveso incident, which occurred in 1976, is perhaps the best documented industrial accident of this type. Thermographic records at the Seveso plant showed an uncontrolled rise in the temperature of the reaction vessel, following failure of a safety shutdown device. The mixture was heated at 230° to 240°C for 4 to 5 hours and 3 to 4 atmospheres before a safety disc ruptured and released nearly the entire mixture (Figure 3).

1954

Meigs *et al.* estimated that about 1 year was required to develop chloracne after the beginning of exposure to air with a concentration of 0.1 mg PCB/m^3. PCB was used as a heat exchange fluid for a reaction chamber. There was leakage of vapour which exposed a group of 7 to 14 workers.

1956

A Monsanto accident in the manufacture of chlorophenols occurred in the Monsanto Nitro Plant during 1949. A similar accident also occurred at Badische Anilin in Germany on 17th November, 1953.

Monsanto described their experience in a 1956 memorandum with personnel from Badische Anilin: "A thorough systematic investigation has isolated impurities in the trichlorophenol process (or residues) which will cause the same effects in rabbits." The materials mentioned include "pentachlor naphthalene" and "chlorinated diphenyl oxide".

PCDFs and PCDDs can be formed from chlorophenols in three different ways as follows:

1) the dimerization of chlorophenates;
2) the cyclization of polychlorinated diphenyl ethers (PCDPEs);
3) the cyclization of polychlorinated phenoxy phenols, termed pre-dioxins.

1 The dimerization of chlorophenates has been shown in the reaction shown above.

2 Polychlorinated diphenyl ethers are formed in the polychloro-phenol synthesis according to the following reaction (Figure 4).
The formation of PCDFs from the cyclization of polychlorinated diphenyl ethers is as follows (Figure 5).

3 Polychlorinated phenoxy phenols are termed 'pre-dioxins' for the reason that they are intermediates in the cyclization sequence (Figures 6 and 7).

One or more of the steps in the formation of PCDDs has a high activation energy. Activation energy in chemical

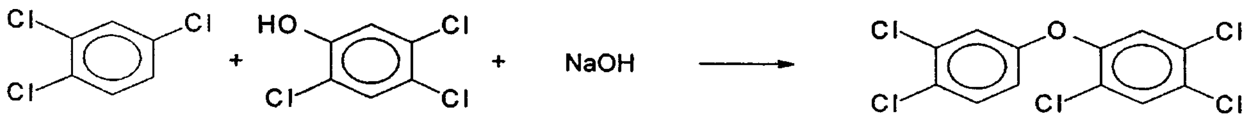


FIGURE 4 Formation of polychlorinated diphenyl ethers (PCDPEs) can lead to the formation of polychlorinated dibenzofurans (PCDFs).

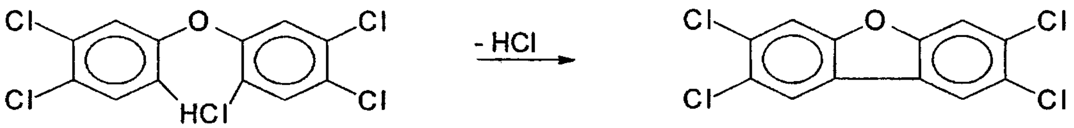


FIGURE 5 Loss of HCl produces cyclization of the diphenyl ether to a dibenzofuran.

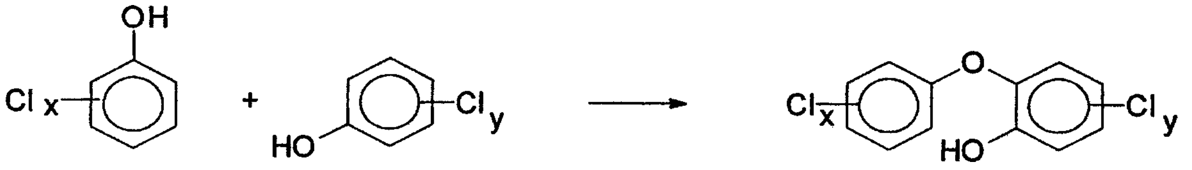


FIGURE 6

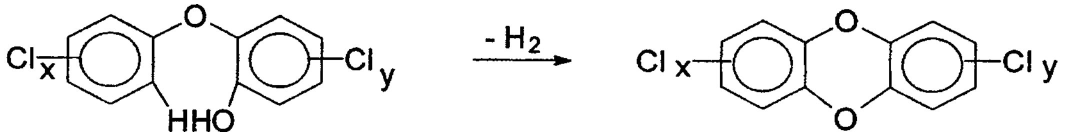


FIGURE 7 Loss of H_2 from the pre-dioxin produces PCDD.

thermodynamics is a concept which describes the energy flow in a reaction from reactants to products. When there is a high activation energy barrier, the reactants require more energy to attain an activated state than when there is a low activation barrier. Once the reactants have gained enough energy to form an activated intermediate they can either return to the reactant form or proceed to a stable product. The balance between the forward and backward reaction directions is an equilibrium condition which may be heavily weighted in favor of either the reactants or products. The high activation energy in the formation of PCDDs contributes directly to their low yields. The residue from the trichloro-phenol distillation in the Monsanto synthesis was reported to be 20 times less effective in producing liver necrosis in rabbits than the residue of the fraction remaining after heating to above 230°C. The boiling points of many dioxin precursors are lower than the reaction temperature range in which PCDDs are formed and therefore there is a loss of precursor molecules before reaction can occur.

The possibility of pressure differences in the formation of the residues are only one of several key parameters which make the comparison difficult. It is, nevertheless, clear that the residue derived from the high temperature distillation was substantially more toxic than the pentachloro naphthalene studied. The health effects of PCNs had been described in the literature some 17 years before but the effect observed on rabbits left in the operating area after the BASF accident was surprising—they died from liver necrosis after a *one week* exposure!

1957

A review of characteristic chloracne lesions appeared in 1957 by Schwartz *et al.*

1962

Threshold limit values for PCBs in workroom air were set in 1962 by the American Conference of Governmental Industrial Hygienists. Under “Chlorodiphenyl—42% Chlorine” it was stated that “Acne, systemic poisoning and even death may result from exposure to chlorinated diphenyls.” The references indicated were [Schwartz, L.: Am. J. Pub. Health; **26**, 586 (1936) and Drinker, C.K., Warren, M.F., Bennett, G.A.: J. Ind. Hyg. & Tox.; **19**, 283 (1937)]. A threshold limit value of 1 mg/m^3 was recommended to offer good protection against severe systemic toxicity but may not guarantee complete freedom from chloracne. In other words, it was recognized that there was a risk of severe systemic toxicity and that chloracne was a separate issue.

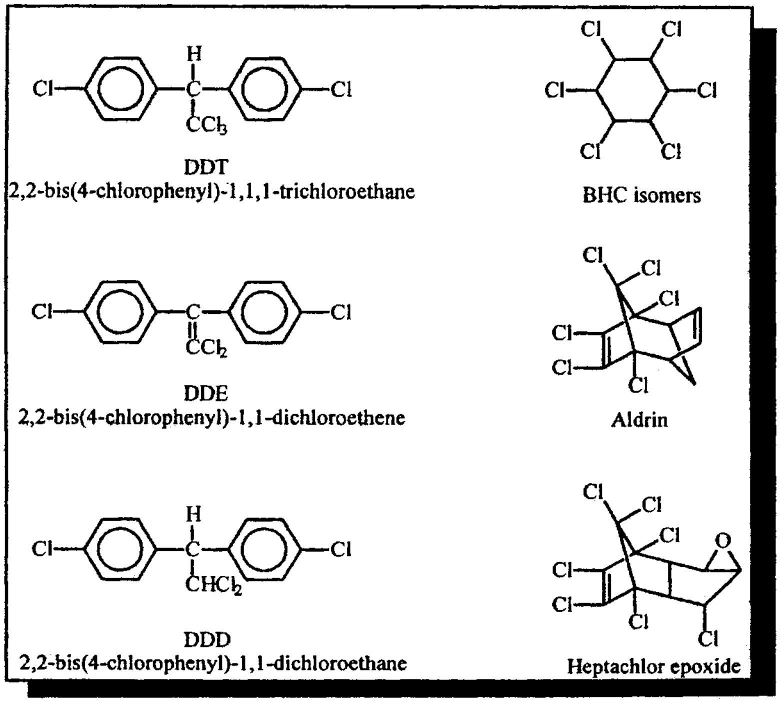


FIGURE 8 Widespread contamination by these compounds complicated the analytical quantitation of PCBs.

Chick edema disease was linked to PCB exposure when chicks were placed in galvanized cages that were coated with an epoxy resin paint. The paint had not hardened prior to placing the chicks in the cages, and within three weeks the chicks showed symptoms resembling those that had been reported as chick edema in incidents resulting from toxic components in chicken feed [McCune *et al.,* 1962].

Chick edema disease was linked to PCB exposure when chicks were placed in galvanized cages that were coated with an epoxy resin paint. The paint had not hardened prior to placing the chicks inside the cages, and within three weeks the chicks showed symptoms resembling those that had been reported as chick edema in incidents resulting from toxic components in chicken feed [McCune *et al.,* 1962]. Aroclor 1242 was determined to be the causative agent. Prompted by this initial finding, further studies on the effects of PCBs in chicks were conducted. Studies on chicks fed Aroclor 1242 showed high mortality, hydropericardium (accumulation of water in the sac that surrounds the heart), hemorrhagic edematous lungs (abnormal swelling from blood and fluid accumulation in the lungs), kidney tubular dilation and lesions, swollen liver, and a dose-dependent pathology [McCune, 1962; Flick *et al.,* 1965]. Growth rates of chicks were slowed in a dose-dependent manner and defeathering and dermatitis were reported [Flick *et al.,* 1965].

1963

A study with chick embryos showed 0% hatchability in high Aroclor 1242-dosed embryos. Embryo's showed beak deformity, edema and growth retardation [McLaughlin *et al.,* 1963].

1966

In 1966 the American Conference of Governmental Industrial Hygienists amended their description of the suggested limit of "Chlorodiphenyl—42% Chlorine—Skin" and stated that "Acne is not an invariable warning sign of impending, more severe, systemic toxicity" (Figure 8).

Jensen identified Aroclors 1254 and 1260 while analyzing for DDT [Jensen 1966]. He detected the presence of certain unidentified peaks in specimens of white-tailed eagle feathers dating as far back as 1942. Since DDT was used widely as a pesticide only after 1945 it was apparent that the pesticide-like chemical found in the samples was not a DDT break-down product, DDE or DDD. In 1967 G. Widmark unambiguously identified the contaminants as PCBs using MS.

The initial work by Jensen (1966, 1969) and of Risebrough (1968) sparked an immediate concern that PCBs had become ubiquitous in the environment. The information was reported in the San Francisco Chronicle of February 24, 1969. In an open letter to Monsanto's customers, the Environmental Health department, gave a cautious overview of the findings on March 3, 1969.

1967

G. Widmark published "IUPAC Commission of Methods of Pesticide Analysis: Possible Interference by PCBs."

Also: Jensen, S. and G. Widmark: OECD pesticide conference report on unintended residues in the environment.

1968

Risebrough *et al.* (1968) published in Nature: "Polychlorinated Biphenyls in the Global Ecosystem". Found that the eggshell thickness of fragments of a peregrine falcon egg was 34% less than the thickness of eggshells present in the same area 21 years previously.

September, 1968: Information sent by Westinghouse on "Inerteen and Environmental Contamination" described that "It has been indicated by several laboratories that measurable amounts of the PCBs are present in our general environment and may have some effect on certain species of wildlife. While askarels are generally regarded as being non-toxic to humans, it is reasonable to assume that very high standards of control will be issued by the Government in the overall program against pollution". The bulletin indicated that Monsanto had arranged for return of scrap Inerteen fluid in sealed drums and indicated that "Inerteen soaked coils, insulation scrap, filter cartridges, and other materials must be stored until appropriate methods can be provided to dispose of them properly."

In October 1968, Japanese health officials became aware of an epidemic disease which came to be known as "Yusho".

The cause of the disease was the ingestion of PCB contaminated rice oil over a period of about eight months [Kuratsune *et al.,* 1972]. The epidemic became the subject of a special issue of the journal Fukuoka Acta Medica in June 1969 [Kuratsune *et al.,* 1969]. The incident is important because it clearly documents the potential for reproductive and transplacental effects in humans.

About 325 patients with acneiform eruptions had been treated in northern Japan between the summer of 1968 and January 1969. The apparent cause of the poisoning was traced to the consumption of a particular brand of rice bran oil.

By May 1970 the total number of cases had risen to 1,291. The common symptoms included acneiform eruptions, hyperpigmentation of the skin, nails and mucous membranes, swelling of the upper eyelids, and hyperemia of the conjunctivae.

The 'Yusho' incident and another very similar accident in Taiwan, known as the 'Yu-Cheng' incident, are fully discussed in a later section.

The Yusho incident prompted the US Food and Drug Administration to initiate a national survey to determine the extent and levels to which PCBs might have made their way into the food chain.

1969

Analytical methods for the detection of PCBs in food became recognised as routine by the Food and Drug Administration (FDA).

Kuratsune *et al.* (1969) published an article entitled "Study on 'Yusho' or Chlorobiphenyls Poisoning".

Among the women with Yusho who were pregnant at the time of the outbreak, some were examined for observations that might indicate reproductive effects upon human exposure to PCBs. Since these case reports are based upon a small sample of Yusho cases, there observations cannot be verified statistically and are therefore only indicators that PCBs in rice oil might result in some reproductive effects that are not inconsistent with those reported in laboratory studies. Out of nine pregnant women, all of whom had characteristic Yusho poisoning, two incidents of stillbirth were reported in which the fetuses manifested Yusho symptoms such as brown staining of skin, nails and gingive [Taki *et al.,* 1969].

February 24, the *San Francisco Chronicle* carried a major feature about "a menacing new pollutant" in marine life research carried out by Dr. Robert Risebrough of the University of California.

March 3, 1969: Information on the San Francisco Chronicle's article is circulated to Monsanto customers in a letter dated March 3, 1969.

On March 7, 1969 the Associate Commissioner for Science and Deputy Associate commissioner for Compliance asked the Bureau of Science to evaluate foods sampled for pesticides for evidence of PCBs. None was indicated using the analytical methods of that time except for traces of PCB in fish from the Great Lakes.

March 10–14, 1969: District pesticide experts were given analytical information concerning PCB and told to be aware of the possibility of its presence in foods. (Pesticide Analytical Workshop 3/10–14/69).

PCB was identified as a contaminant of many of the fish analyzed as part of proceedings against the City Smoke Fish Co. of Detroit, MI (June and July 1969). The fish also contained DDT above 5 ppm.

July 1969: Cow's milk was found to be contaminated with PCBs (July 1969) from cows given feed treated with herbicides that had been diluted with PCBs.

On September 13, 1969 the Baltimore District Office of the FDA reported PCB in West Virginia milk. The source was traced to the use of an Aroclor as a solvent for pesticide spray.

Regulatory action at 0.2 ppm PCB on a whole milk basis was established by the FDA on December 16, 1969.

1970

Aroclor 1260 was discontinued in 1970. 1254 was an available alternative.

Monsanto, the only producer of PCBs in the US alerted its customers to the potential problem of environmental contamination. Major electrical equipment manufacturers were requested to relay the information to end-use customers with advice as to proper destruction of PCBs. Monsanto advised clients from 1970 onwards to be careful not to let water that may contain PCBs get into sumps or drains that would lead into the sewer system.

In 1970, the FDA prepared a summary of the available information on the chemistry and toxicity of PCBs. [US Dept. of Health, Education, and Welfare (1970): "Supplement 1—Status Report on the Chemistry and Toxicology of PCBs of

Aroclors"; Washington, D.C.: Bureau of Foods, Food and Drug Administration, Public Health Service.]

During 1970, the Council on Environmental Quality (CEQ) studied regulatory approaches to the problem of toxic chemicals in the environment. In its report published in 1971, CEQ identified PCBs as a major problem. [Council on Environmental Quality (1971): "Toxic Substances"; Washington, DC.]

Vos *et al.* (1970) published "Identification and Toxicological Evaluation of Chlorinated Dibenzofuran and Chlorinated Naphthalene in Two Commercial PCBs".

Increased levels of porphyrins are consistent with chronic hepatic porphyria that has been associated with exposures to other polyhalogenated aromatics. This condition can range from a subclinical state, where increased urinary porphyrin excretion results in no observed adverse effects, to porphyria cutanea tarda (PCT) which is a skin disease resulting in excessive light sensitivity, scaling, and itching. Increased levels of coproporphyrins and protoporphyrins in faeces and porphyrins in liver, bile and the small intestines of chicks fed Aroclor 1260, Clophen A60 and Phenoclor DP6 were reported by Vos and Koeman (1970).

Morphological changes in monkeys were observed with doses ranging from 1.4 mg/Kg/d to 16.4 mg/K/d and also mice dosed with 12.8 mg/Kg/d [Nishizumi, 1970].

Aroclors 1221, 1232, 1242 and 1248 in high doses, bind to estrogen receptors in the rabbit uterus [Bitman and Cecil, 1970].

Close parallels between the toxic effects of PCBs and PCDFs/PCDDs were established in a series of studies and reported in major scientific reviews such as [Crow, 1970].

Cleanup methods for analysis were devised to separate DDT and its analogues from some of the PCBs [Amour and Burke, 1970].

The Pesticide Regulation Division of the FDA issued PR Notice 70–25 instructing manufacturers and formulators of economic poisons containing PCBs and PCTs to change their formulations to exclude these classes of compounds.

Open letter from GE Electric Utility Sales Division to electric utility customers expresses environmental concerns and starts "As you are well aware the public concern with pollution is rapidly escalating."

February 3, 1970: The FDA initiated a regulatory threshold value of 5 ppm PCB in the edible flesh of fish, February 3, 1970.

August 1970: Fish near the effluent of Monsanto's Annison plant were found to contain PCB in August 1970.

August 1970: Cow's milk was found to be contaminated with PCBs (August 1970) from cows given feed that had been stocked in silos coated with PCB-based products.

September 1970: The Office and Science and Technology panel on Hazardous Trace Substances reviewed cadmium, arsenic and PCBs. Also, the National Swedish Environment Protection Board held a conference on PCBs. One of the sessions of the conference highlighted "environmental problems". This conference, perhaps for the first time, brought together the extent of understanding of the manufacture, use, extent of environmental contamination, analytical methods, and biological effects of PCBs.

September 14, 1970: An open letter from Monsanto, dated September 14, 1970 states: "As a supplier of askarel fluids namely Inerteen, Pyranol, and Aroclors which do contain PCBs, we wish to inform you that the FDA announced a guideline for PCB of 5 ppm in fish and 0.2 ppm in milk. Therefore, control of waste and spills is imperative."

A letter was sent to Monsanto customers requesting information on the quantities of PCB fluids that might be shipped to Monsanto for disposal so that the company could design an appropriate sized incinerator.

October, 1970: At the October Double Client Committee meetings, a subcommittee was formed to investigate PCB disposal problems.

October 12–14, 1970: The International Electrotechnical Commission Technical Committee No. 10; Liquid and Gaseous Dielectrics Subcommittee 10B; Insulating Liquids Other than Hydrocarbon Oils indicated in its minutes of meeting dated October 12–14, 1970: "Agenda–Item 5—To review current concern with the finding of PCBs in some environments and what action should be taken in recommending precautions in the application, processing and effluent disposal of askarels to prevent them becoming environmental contaminations … Copies of IEEE and ASTM documents were circulated by the US Delegation. A French Delegate circulated copies of a publication from Prodelac (a French producer of askarels)."

October 28, 1970: Minutes of the IEEE Insulating Fluids Subcommittee dated October 28, 1970 indicated "Monsanto will, however, discontinue by the end of 1970, the sale of Aroclors for application as plasticizers, heat transfer agents, hydraulic fluids and other non-electrical applications."

November 19, 1970: A NEMA letter to (ANSI) dated November 19, 1970, regarding the proposed committee on the disposal of askarel used in electrical equipment indicates " … The Monsanto Company has already taken steps to limit the availability of askarel to those users where the fluid is essential and is contained in a closed system, such as is the case for transformers and capacitors. With the cooperation of transformer and capacitor manufacturers, steps have been taken to alert all known users on the problem and on corrective measures. Monsanto has also instituted a means for disposing of askarel-soaked waste materials, coils, insulation, scrap, discarded capacitors, etc."

November 1970: Occurrence of PCBs in chickens following absorption from PCB-containing plastic wrappings.

In December 1970, Monsanto sent a letter to end users of PCB products requesting information on the amount and types of solid PCB contaminated materials that might be shipped to Monsanto for disposal. Monsanto was in the process of sizing the design for a kiln to incinerate PCB wastes and intended to make the service generally available.

1971

In 1971 Monsanto began to phase out production of the more highly chlorinated Aroclors: 1232, 1248, 1260, 1262 and 1268.

In 1971 Monsanto introduced Aroclor 1016, which had an average chlorination similar to Aroclor 1242 but with fewer of the more highly chlorinated compounds. At the end of 1975, only Aroclors 1221, 1016, 1242 and 1254 were still being produced.

Up to this point in time (1971) the principal government "control" actions for PCBs were the result of 'persuasion' by the Federal Government rather than by direct regulatory exercise. The Government possessed no real regulatory authority to control the manufacture, distribution or use of PCBs. However, strong influence was brought to bear by the several Government inquiries of the time—especially the PCB Interdepartmental Task Force. The principal recommendation of the PCB Task Force was the discontinuance of all uses of PCBs except in electrical capacitors and transformers. The latter uses were judged to be both essential to industry and represented 'closed' systems. Thus, the major action of PCB manufacturing phaseout and use restrictions were taken 'voluntarily' by Monsanto.

An OECD study on the sampling and analysis of residues in wildlife involved 23 laboratories in 13 OECD member countries between 1966 and 1971 and confirmed that residues of organochlorine pesticides, PCBs and mercury were to be found in a wide variety of species, in different habitats, and in areas where PCBs have not been deliberately released.

The OECD published a case study report of PCB and mercury pollution in Japan in 1976. The general situation regarding PCB pollution was described as follows: "Environmental pollution of PCBs was first recognized in Europe followed by the US then Japan. Results of field investigations demonstrated that PCB pollution in organisms inhabiting the natural environment was at a severe level. In Japan, analytical results indicating concentrations of residues in wild animals were first reported in 1971. In January of 1972, a standard method for the analysis of PCBs was established by the research group of the Ministry of Health and Welfare sponsored by Science and Technology Agency, whereby the uniform analysis of PCB residues in wild animals was realized.... A national survey concerning environmental pollution of PCBs was conducted from May to December, 1972. During this period, the water quality of 1,084 locations and bottom sediments sampled from 1,445 places was inspected."

During 1970, the Council on Environmental Quality (CEQ) studied regulatory approaches to the problem of toxic chemicals in the environment. In its report "Toxic Chemicals" published in 1971, CEQ identified PCBs as a major problem. [Council on Environmental Quality, 1971].

The instructions for handling Westinghouse's Inerteen insulating fluid referenced the ANSI C107.1–1974 for complete information on handling and disposal of Askarels. The document also included a section on "Environmental Considerations" which stated "It has been shown by several laboratories that measurable amounts of the PCBs, particularly those with more than 50% chlorination, are present in the general environment and are a threat to certain species of wild-life.... To be sure that the Inerteen and Inerteen contaminated materials do not contaminate the environment they must be incinerated in approved equipment." A disposal section in the handling instructions describes the disposal of Inerteen liquid, solvent rinses contaminated with Inerteen, solids contaminated with inerteen and the cleaning of contaminated drums.

The American Conference of Governmental Industrial Hygienists retained a threshold limit value of 1 mg/m^3 in 1971 and stated "Exposure to the vapor and fume of chlorodiphenyl may result in acne, irritation of respiratory passages, and injury to the liver.... Several deaths due to atrophy of the liver occurred among workers exposed to the fumes of chlorodiphenyls and chloronaphthalenes."

Transplacental passage of Aroclors 1221 and 1254 in rabbits was observed [Grant *et al.,* 1971] when PCB residues were found in higher concentrations in fetal liver than in tissues of dosed does.

Placental and adipose tissues of stillbirths from Yusho mothers confirmed the presence of Kanechlor components [Kuratsune *et al.,* 1971].

The total quantity of contaminated rice oil ingested by patients in the Yusho incident of 1968, as well as the dose rate of ingestion of PCBs was estimated[38] by Yoshimura, 1971.

Studies of Yusho patients have revealed symptoms of numbness, pain, hypothesia or areflexia of the limbs in about 50% of cases [Murai and Kuroiwa, 1971].

From 1971 to December 1975 a frequent method of disposal of askarel capacitors was to throw them into a dumpster for disposal in various landfills.

The State of Massachusetts enacted a bill in 1971 to prohibit the discharge of any chemical solution containing PCBs into the environment.

A paper was presented at the 1971 Doble Client Conference to report activities of the special Doble subcommittee on askarel disposal problems.

February 23, 1971: ANSI meeting of the committee on Use and Disposal of Askarel of the Power Equipment Division, February 23, 1971. "The committee noted that the purpose of this meeting is to review developments since the August 20, 1970 meeting and develop recommendations to ANSI on a suggested program of action, including priorities for various activities, list of organizations to be invited to participate, etc.

The formation of the ANSI committee had as its objectives to:

"1) Serve as a source of technical information and advice for Federal, State and local authorities and for the information of all others concerned.
2) Encourage the development of suitable disposal facilities and maintain a list of their capabilities and locations, for the information of all concerned.
3) Serve as the advisory group for US participation in CEE, IEC, COPANT, SEGRA and other international organizations."

Between February and October 1971 TCP/PCDD/PCB was sprayed on a horse arena in Missouri. Still bottom residues from TCP manufacture were contained in the waste oil used to control dust in Missouri horse arenas. When the waste oil storage tank was examined, it was found to contain 300 ppm TCDD.

July 1971: A Monsanto PCB bulletin [Monsanto, 1971] was issued in July 1971 entitled 'Handling, Waste Control and Disposal of PCBs'. The document was "*Produced and distributed as a service to industry by Monsanto*". It was pointed out that Pydraul®, a fire resistant hydraulic fluid, was no longer manufactured with PCBs but other products had a place in continued applications provided that common sense and good maintenance procedure were used towards eliminating any threat to the environment from PCBs.

Monsanto offered nine guidelines to prevent the escape of PCBs into the environment. In conclusion, the disposal of solids was suggested in a properly operated landfill not located near any water system.

The Monsanto position was explained in a series of points which included Monsanto's recommendation that PCB contaminated solids could be buried as a satisfactory disposal option provided that the site was dry. Tests by Monsanto and others of PCBs in the environment had shown that the higher chlorinated PCB congeners were being retained in organisms and that there was an absence of PCBs with three chlorines or less. Some initial successes had been obtained using biodegradation of PCBs under lab conditions. The apparent lack of lower chlorinated species in the environment led Monsanto to develop Aroclor 1016 as a more environmentally acceptable product. By this time, Monsanto had made it known that it would supply Aroclors only to the electrical industry since the electrical applications were considered to be enclosed. Information was presented which suggested that the health effects of PCBs were minimal. Interestingly, the health effects observed in the Yusho incident were thought to have been caused by PCDFs as impurities which were typically found 'in some European PCBs'. The correlation between PCBs and PCDFs had been shown by Vos *et al.* (1970) who found that the relative concentrations of the PCDF isomers present in Yusho oil and in two samples of used heat exchanger PCBs (Kanechlor 400 and Mitsubishi-Monsanto T1248) were strikingly similar. The mechanism for the interconversion of PCBs into PCDFs was not published until 1979 [Buser and Rappe, 1979].

By 1971 Monsanto had determined that about 75% of the PCBs in the environment could be attributed to plasticizers, lubricants, hydraulic fluids, incineration of PCB materials and open dumps with the remainder coming from the electrical industry. It was concluded that there was sufficient biological evidence to indicate that PCBs can be deleterious to the health of animal and human life and that, even though it could not be expressed quantitatively, it was inappropriate for electrical equipment manufacturers to ignore the weight of the evidence against PCBs.

Electrical equipment manufacturers were to inform customers about the nature of PCBs, and the hazards associated with them. Instructions were to be provided on the proper disposal of units containing PCBs. PCB management controls were to be initiated in manufacturing facilities. Customers were to be informed that the disposal of PCB liquid waste should be done by sending the material to Monsanto for incineration while failed or non-serviceable units should be sent back to the manufacturer so that the PCB in the unit could be disposed of adequately. All of the maintenance and repair shops were to be alerted concerning the nature of PCBs and, again, a procedure set up for the proper use and disposal of the liquid and solid waste. Equipment manufacturers intended to encourage NEMA and other industrial organizations to establish uniform use, handling and disposal standards. Field records of PCB equipment performance were to be obtained.

June 1971: The National Industrial Pollution Control Council published "The Use and Disposal of electrical Insulating Liquids" in June 1971 and included in their report recommendations to minimize the possibility of entry into the environment of PCBs used in capacitors and transformers. These regulations were later developed by a committee of ANSI.

In July 1971, leakage of a heat exchanger fluid caused contamination of pasteurized fish meal which in turn was fed as a ration to chickens and catfish.

In July 1971: discovery of PCBs in food wrappings due to recycling of copy paper in the production of paper pulp.

Occurrence of PCBs in turkey and chicken fed PCB-containing products (reason unknown).

FDA began recall of 50 tons of meat, 300 tons of fish feed, and 3,300 tons of fish meal contaminated with PCBs caused by a leak in a heat exchanger system.

In August 1971, Westinghouse's Power Transformer Division issued an updated set of 'Instructions for Handling Inerteen® Insulating Fluid P.D.S. 54201 CM and Installation and Maintenance of Inerteen Transformers.' [Westinghouse, 1971], which incorporated the earlier supplemental notice of environmental concerns and waste disposal. In this version of the guide disposal suggestions are much more explicit and contain recommendations for Inerteen liquid, solvent-rinses contaminated with Inerteen and solids, including drums, contaminated with Inerteen. Environmental concerns were also expressed to the equipment purchaser much more explicitly in this document than in the earlier 1968 version: "*… a carefully planned program of waste disposal must be followed at every step of the equipment life. This includes manufacture, repair and final disposition of the fluid and the Inerteen contaminated parts.*"

On August 3, 1971, FDA called an Interagency meeting of spokesmen from several Federal agencies and research laboratories to review PCBs. OST released its report on PCB environmental impact in March 1972.

September 1971, Monsanto introduced Aroclor 1016 as a replacement for Aroclor 1242 in capacitors. The fluid was made by vacuum distillation of Aroclor 1242 to remove the more highly chlorinated PCB homologs since these were recognized as being less biodegradable than the lower chlorinated congeners.

September 1, 1971 representatives of several Agencies of the Federal Government established an Interdepartmental Task Force to coordinate the scientific efforts relative to understanding PCBs. The Agencies involved were USDA, EPA, the Council on Environmental Quality and the Office of Science and Technology. The OST took the lead role and the work was conducted under what became known as the Interdepartmental Task Force on PCBs. The Task Force was announced on September 5, 1971 in a joint FDA–USDA press conference.

September 15, 1971: The OST Panel on Hazardous Trace Substances and the Governmental Task Force met jointly with representatives of the Monsanto Co. The principal agenda item of this meeting was a request for information concerning the amounts and usage, and estimates of losses into the environment.

October 1, 1971: Norway enacted regulations which came into effect on October 1, 1971 which required prior authorization for the use of PCBs.

December 1971: Monsanto adopted a policy not to sell any more dielectric fluids to its customers before they entered into an indemnity agreement.

December 20–21, 1971: National Institute of Environmental Health Sciences symposium on PCBs held at Durham, NC on December 20–21, 1971. Members of the press, especially science writers, were invited to attend. Thus, the aims of the conference were to educate both scientists and the public about the science of PCBs and its interpretation.

1972

PCBs were the subject of an increasing number of papers reported in the scientific literature dealing with the environment between 1970 and 1971. A conference which dealt with the environmental problem of PCBs was held in September 1970 in Sweden and in August 1971, an environmental quality workshop was convened in Durham, New Hampshire by the National Academy of Sciences.

The lay press, prompted by the newsworthiness of several accidental PCB spills and the widespread finding of PCBs in the environment, published many articles in 1970–1971 which dealt with PCB contamination. Many PCB reports involved the contamination of foodstuffs.

The Federal Department of Agriculture made an extensive surveillance of food in cooperation with the Department of Agriculture and established 'Action Levels' for PCBs in milk, poultry and fish. Action levels are temporary thresholds for regulatory decision pending the establishment of regulatory policy and procedure. During 1970 and 1971, the FDA used these guideline values in several seizures of foods found to be contaminated with PCBs.

Following announcements in the public press in 1971 there began to be political enquiries made by Senator McGovern and Governor Miliken of Michigan (Figure 9).

By August 1971, the FDA was being pressured into taking a position in which public pressure and advocacy demanded a stronger position against the use of PCBs. FDA initiated a meeting to discuss the scientific issues concerning PCBs which involved several Government agencies and Federal research laboratories. Following this meeting the FDA asked the President's Science Advisor to provide a focus for the Government's actions and decisions concerning PCBs. The request was made because:

- the PCB issue cut across several Federal agencies;
- the issue was becoming politically stressful for the FDA;
- the Office of Science and Technology (OST) had already started a scientific review of PCBs (October 1970).

The OST review committee became known as the Panel on Hazardous Trade Substances and had as its goal the identification of the needs for information by the Government in making judgments concerning trace hazardous substances.

On September 1, 1971 the FDA met with the USDA, EPA, the Council on Environmental Quality (CEQ) and the OST to explore options concerning PCBs. The OST adopted the lead role and this group became known as the Interdepartmental Task Force on PCBs. The Task Force met with Monsanto, the only US producer of PCBs to request information concerning the amounts of PCBs produced, patterns of distribution and usage, and estimates of losses into the environment.

A model was developed for the rates and routes of environmental transport of PCBs based on Monsanto's information. The model coefficients were tested wherever possible by physical measurement and analytical monitoring. The exercise pointed up gaps in the knowledge base. In particular, analytical methods for PCBs were not well developed and biological effects were not well defined.

The final Interdepartmental Task Force report, in addition to being a review of the scientific information of the time also dealt with the benefits of PCBs derived from a National Bureau of Standards (NBS) report. This review became the basis for the rule-making decision to maintain electrical uses of PCBs and to restrict other applications.

The Task Force report reviewed the Federal regulations for their applicability to PCBs in part to determine the power of the Government to control and limit the manufacture, distribution, use and disposal of PCBs. Regulatory authorities were capable of responding to specific incidents of contamination but the Government had no power to prevent more PCBs from entering the environment.

Despite mounting public pressure calling for an outright ban on PCBs the Commissioner of the FDA held a press conference in which he deferred the decision making process until further review and deliberation of perspectives had been accomplished.

During this time period, the National Institute of Environmental Health Sciences had initiated its own review of scientific knowledge concerning PCBs and it brought together the scientific community and the public in a conference to discuss the findings in December 1971. The result of the information exchange was to reduce the sensationalism surrounding PCBs and to reduce the newsworthiness of the now more scientific issues. The result was that a dramatic decline occurred in the number of sensational newspaper reports concerning PCBs.

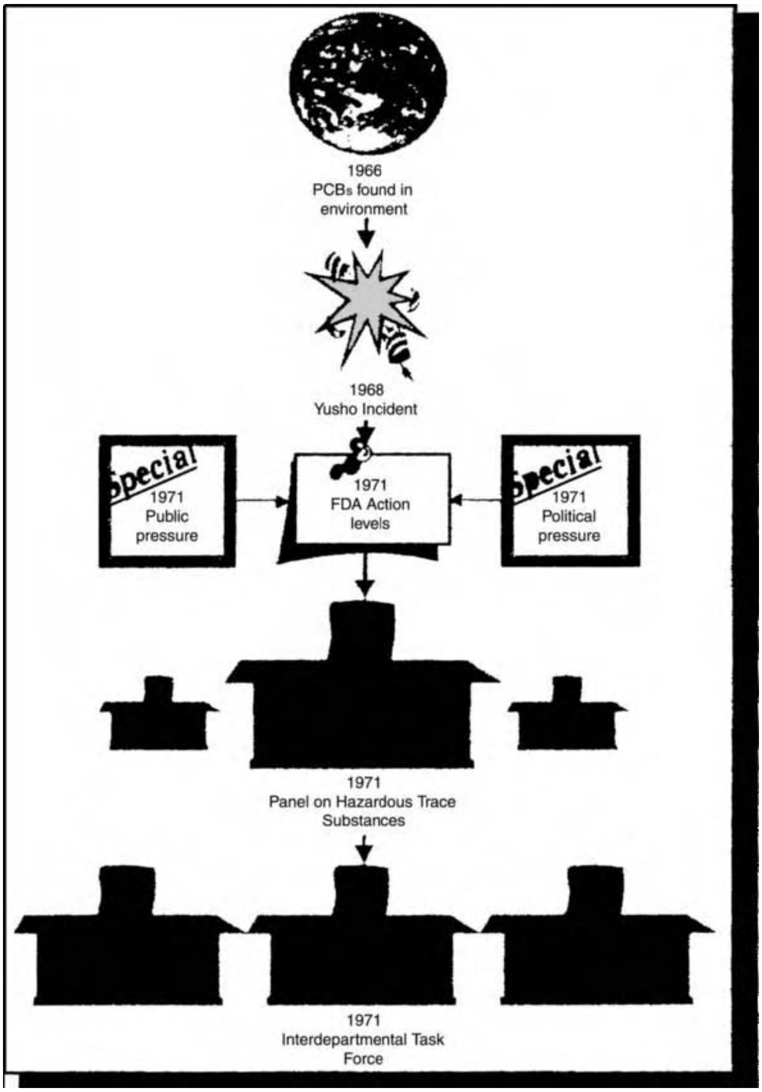


FIGURE 9 Time-line for the development of PCB regulations.

At this point in time (1971), the Government had no direct regulatory authority to control the manufacture, distribution or use of PCBs. However, the Government did possess 'persuasive' influence and, beginning in 1970, Monsanto had begun to reduce the sales of PCBs and 'voluntarily' discontinued sales for use in paints, plasticizers, sealants, adhesives, hydraulic systems and other 'open systems'.

The principal recommendation of the Interdepartmental Task Force was the discontinuance of all uses of PCBs except in 'closed systems' such as transformers and capacitors. This decision derived from an analysis of the benefits of PCBs, the economics of fluid replacement and the availability of substitute fluids. Non-electrical uses in open systems were viewed as either not essential, potentially or actually contributory to the environmental load, or had suitable replacement fluid alternatives.

True regulatory control by the Government in 1971 was limited to persuasive influence. The gap between persuasion and regulation was closed by the enactment of the Toxic Substances Control Act.

In summary, the factors which led to the successful regulation of PCBs were as follows.

- Only one manufacturer was involved and end-users were easily identified.
- A model for the rates and routes of dissipation of PCBs into the environment was developed and the benefits of PCBs were weighed against the risks.
- Scientific information underwent several peer reviews and was widely disseminated.
- The rule-making process was an open one which education both the regulators and the regulated communities.

Kuratsune *et al.* (1972) “Epidemiologic Study on Yusho, a Poisoning Caused by Ingestion of Rice Oil Contaminated with a Brand of PCB.”

Kuratsune (1972) “Results of Lab Examinations of Patients with Yusho Symptoms”.

Enzyme induction in rats had been reported at levels as low as oral doses of 0.03 mg/Kg/d [Litterest *et al.,* 1972] and porphyria had been observed in rats fed PCBs at levels ranging from 6.8 mg/Kg/d to 37.6 mg/Kg/d [Kimbrough *et al.,* 1972].

The first report that adenofibrosis occurred in the livers of rats fed Aroclor 1254 and 1260 was made in 1972 [Kimbrough *et al.,* 1972].

Bioaccumulation in humans was shown in analyses of 4,000 adipose tissue samples from the US Dept. of Health Education and Welfare Human Monitoring Survey which estimated that 41–45% of the US population has PCB levels greater than 1.0 ppm PCB, identifying food chain exposure and occupational exposure as possible sources [Price and Welch 1972].

February 4, 1972: Open letter from Monsanto, dated February 4, 1972 states that customers must sign a ‘Special Undertaking by Purchasers of PCBs’. The document contained wording which put intending buyers on notice of the environmental issues. This was a ‘hold harmless’ agreement.

February 8–9, 1972: Both the capacitor and transformer sub-committees of ANSI committee C-107 on the use and disposal of askarel and askarel-soaked materials met on February 8–9, 1972. The committees discussed analytical methods for PCBs and a series of questions and answers on PCB concerns.

Monsanto discontinued direct sales of PCB products in 1972.

March 1972: OST released its report on PCB environmental impact in March 1972. As well as serving as a review of the scientific aspects of PCBs, the Task Force explicitly reviewed a number of broad aspects of the PCB question. Most important, perhaps, was the exploration of the benefits or utility of PCBs and of the industrial and commercial dependencies built up over the years. This explicit review of benefits of PCBs was of vital important for decision-making on PCBs. The National Bureau of Standards engaged in a review and analysis of the benefits and even the “essentiality” of each of the several uses of PCBs. In this, the NBS received advice from the National Industrial Pollution Control Council especially for electrical uses of PCBs. This review became the basis for the ultimate decision to preserve electrical uses of PCBs (for which there were true dependencies and no satisfactory substitutes) and to restrict other uses. The Task Force Report reviewed all of the pertinent Federal regulatory laws for their applicability to PCBs. The review pointed out that existing regulatory authorities were capable of responding to specific incident of contamination of foodstuffs once they were recognised. However, it acknowledged that the Government’s legal strength was generally “ . . . inadequate to prevent more PCBs from entering the environment.”

Despite the recommendation of the 1972 Interdepartmental Task Force on PCBs [Interdepartmental Task Force on PCBs. “Polychlorinated biphenyls and the Environment”; NTIS (COM-72-10410) May 1972] that discharges of PCBs be strictly controlled, EPA did not effect a policy of stringently curtailing discharges of PCBs from point sources until December 1975.

Comments by GE [GE 1972] on the notice of proposed rule making in the Federal Register of March 8, 1972 included a report entitled ‘The Tole of PCBs in Electrical Equipment’ dated February 4, 1972. The report addressed the risks and alternatives to the use of askarels in transformers. It was pointed out that the US experience over 40 years had resulted in very few cases of nonchronic chloracne or other temporary skin lesions or irritations. By the time that this report was published, however, numerous studies had been done on the Yusho incident and many researchers were in the process of debating the causative agents in the poisoning episode.

An assessment of the risk of ownership of PCB equipment included a consideration of the probability of leaks from originally sealed equipment. Reference was made to the fact that specific control measures had already been instituted by manufacturers and that these would be supplemented by a committee of ANSI.

The distribution of membership in the ANSI committee C-107 on “Use and Disposal of Askarel Used in Electrical Equipment” included members from equipment manufacturers, US government bodies, manufacturing associations, utility representatives, waste disposal companies and consulting engineers. The membership of the ANSI committee is shown in Table 3.

Monsanto advised that the company had adopted certain policies with respect to its further sale and delivery of all PCB products used as dielectric fluids. In particular, Monsanto had decided that it would deliver Aroclors only to manufacturers of electrical equipment who had entered into an agreement with Monsanto to indemnify and hold the company harmless in the use of their products. The letter had copies attached of the proposed agreement and a copy of an article by Gustafson (1970) on PCBs found in the food chain.

The ‘Special Undertaking by Purchasers of PCBs’ contained wording which put intending buyers on notice of the environmental issues viz. “*Buyer acknowledges that it is aware and has been advised by Monsanto that PCB’s tend to persist in the environment; that care is required in their handling, possession, use and disposition; that tolerance limits have been or are being established for PCB’s in various food products.*”

May 1972: The Interdepartmental Task Force (ITF) report on PCBs was issued publicly in May 1972 and was accompanied by a statement of Governmental ‘thinking’ and Governmental ‘action’. Perhaps the major conclusion reached in the report, which became an objective in Government decisions, was that of limited restriction of PCBs. This was a direct outcome of the analysis performed by the National Bureau of Standards of the utility and essentiality of PCBs

TABLE 3
ANSI committee C0-107 membership

# of Representatives	Organization Represented
2	Department of the Army
2	Environmental Protection Agency
1	US Department of Agriculture
1	Tennessee Valley Authority
1	General Services Administration
1	National Bureau of Standards
2	Certified Ballast Manufacturers Association
1	Edison Electric Institute
1	Institute of Electrical and Electronics Engineers
5	National Electrical Manufacturers Association
2	Monsanto Company
2	Commercial Waste Disposal Companies
1	Engineering Consulting Firm
1	Capacitor Manufacturer Serving as an Independent Member

Monsanto released information concerning its PCB production figures from 1930–1971 to the Dept. of Commerce in November 1971 as 4.3×10^5 tons PCB.

which pointed to the possibility of an increase in fires and explosions from encases or enclosed transformers if PCBs could no longer be used—representing the possibility of trading one hazard for a different hazard. In addition, the ITF had reviewed the patterns of environmental dispersal of PCBs which had been lost from their point of application. Electrical applications were seen as 'closed' and were not thought to be contributory to environmental distribution. Other uses of PCBs were reviewed as either not essential, potentially or actually contributory to the environmental 'load', or were found to have suitable available substitutes. These conclusions pointed to an elimination of essentially all other uses of PCBs in heat exchange fluids, hydraulic fluids, and the miscellaneous category of 'plasticizer' uses. This objective of limited restriction became the basis for Governmental persuasion of the Monsanto Company to restrict its distribution and sales of PCBs.

June 1, 1972: The Swedish government enacted regulations effective June 1, 1972 which required prior authorization for the use of PCBs, or compounds containing PCBs, and the PCB content to be indicated on wrappings.

June 1972: Import and production of PCBs were banned in Japan. Continued use of PCBs was to be allowed in dielectric (September 1, 1972) and heat transfer applications (July 1, 1972) when complete recovery could be assured.

September 1972: Instructions for shipment, installation and storage of oil-filled power transformers were made effective in September 1972 at the Westinghouse Sharon transformer division [Westinghouse 1972]. The manual does not mention PCBs specifically but refers to an instruction manual, I.B. 45-063-100, for the proper sampling, testing and handling of the oil.

October 1, 1972: Switzerland enacted regulations which came into effect on October 1, 1972 in which PCBs or PCB-containing products may not be sold to the public or to light industry while heavy industrial use became subject to prior authorization.

Late 1972: A letter from Monsanto was attached to all order acknowledgements by the Monsanto Customer Service Department which read: "As we have previously notified you by way of correspondence since early 1970, the material described on the acknowledgement order contains polychlorinated biphenyls. PCBs are highly stable chemical compounds that are not readily biodegradable. Therefore, when placed in the environment they must be considered contaminants and may adversely affect some species of animals and marine life … You must take every precaution to prevent any entry of polychlorinated biphenyls into the environment through spills, usage, leakage, disposal, vaporization or otherwise."

1973

A 'gentleman's agreement' had been reached in the Netherlands that, as of January 1, 1973, PCBs would no longer be used by manufacturer's of:

—paints, inks and lacquers,
—adhesives, resins, etc.,
—wire and cable coatings,
—lubricating oils,
—hydraulic fluids,
—copy paper.

During 1973 and 1974, the EPA proposed the establishment of water quality criteria for PCBs in industrial discharges as part of a program for establishing such criteria for a larger group of pesticides. However, PCBs were not covered in the effluent standards that were eventually promulgated.

Louw *et al.* "Vapor Phase Chemistry of Arenes. Part II. Thermolysis of Chlorobenzene and Reactions with Aryl Radicals and Chlorine and Hydrogen Atoms at 500°C". The pyrolysis of chlorobenzene was interpreted as a free radical reaction sequence involving $\bullet C_6H_4Cl$, •Cl and •H as carriers.

Kanechlor 500 and Aroclor 1254 were found to be carcinogenic in male mice [Ito 1973].

Dioctyl phthalate was introduced as a non-PCB dielectric fluid for capacitors.

The OECD announced a decision in February 1973 to control the manufacture and trade of PCBs but allow for the production of PCBs to be used in closed systems. It recommended that Member countries give priority attention to the elimination of PCBs in hydraulic liquids (other than in mining), lubricating oils, vacuum pump fluids and cutting oils and request that firms use, as PCB replacements, materials which are less hazardous to human health and the environment than the range of PCBs in use (in 1973).

March 5, 1973: A spill of about 630 gallons of transformer oil occurred on State Highway 58, South of Kingston, TN on March 5, 1973. The clean-up cost $1.7 million.

July 6, 1973: Federal Register lists PCBs in a proposed list of toxics and pollutants.

July 6, 1973: Federal Register **38**, 18096 (1973): FDA issued rules limiting sources by which PCBs may contaminate feed, food and packaging materials. Transformers were specifically exempted from the restrictions.

September 7, 1973: In September 1973 EPA published a list of "toxic pollutants" which included PCBs. The list was unchanged from that published on July 6, 1973.

December 27, 1973: EPA proposed toxic pollutant effluent standards for the nine toxic substances, including PCBs (38 FR 35388).

1974

The ANSI C-107 standard was issued in 1974 and at the time it was approved the committee contained representatives from electrical utilities, as well as IEEE.

The ANSI C-107 guide, in a section on risks, mentions that "*… limited amounts of PCBs can get into the environment during the manufacture, delivery, improper use, maintenance, repair and disposal of transformers and capacitors.*" Under §3.3 'Plant Housekeeping and Employee Safety' it is indicated that procedures and limits are given to be met by manufacturers as well as users.

Large capacitors were to carry a caution label containing the wording "*To avoid possible environmental contamination, it (PCB) should be disposed of only in supervised dry landfill areas meeting state requirements or in incineration facilities designed for disposal of PCBs.*"

It is mentioned in the guide that an askarel-filled transformer scrapped for any reason should be considered as a source of materials requiring special handling and disposal procedures. It was suggested that cellulosic materials derived from the core/coil assembly should be incinerated in a high temperature furnace while coil structures could be solvent washed and subsequently treated as normal scrap. The disposal of askarel waste, including liquid and solid wastes could be done in controlled landfills or deep-wells where permitted by federal, state and local regulations. New transformers were to have a warning label which indicated that care should be taken to prevent entry of PCBs into the environment and that, in case of malfunctions or leaks, that owners should consult the instruction manual or the manufacturer. Westinghouse's instruction manuals already contained [Westinghouse I.B. 45-063-99C effective August 1971] specific instructions on the material disposal requirements of both liquids and solids.

§4.2.3.6 describes two methods for the ultimate disposal of askarel-filled transformers. One method involved complete draining, dismantling and disposal of askarel-soaked components using methods referred to earlier in the standard. The second alternative states that disposition of askarel transformers by means of junk or scrap dealers should be avoided unless the units are first drained and flushed.

The instructions for handling Westinghouse's Inerteen insulating fluid referenced the ANSI C107.1-1974 for complete information on handling and disposal of Askarels. The document also included a section on "Environmental Considerations" which stated "It has been shown by several laboratories that measurable amounts of the PCBs, particularly those with more than 50% chlorination, are present in the general environment and are a threat to certain species of wild-life.… To be sure that the Inerteen and Inerteen contaminated materials do not contaminate the environment they must be incinerated in approved equipment." A disposal section in the handling instructions describes the disposal of Inerteen liquid, solvent rinses contaminated with Inerteen, solids contaminated with inerteen and the cleaning of contaminated drums.

ANSI C-107.1-1974 standard on "Guidelines for Handling and Disposal of Capacitor- and Transformer-grade Askarels Containing Plychlorinated Biphenyls" issued in 1974.

January 28, 1974: EPA held a thirty day evidentiary hearing on the proposed effluent standards for the nine toxic substances, including PCBs.

March 5, 1974: Federal Register **39**, 8325 (1974): Public hearings on effluent standards.

March 15, 1974: A leak of 265 gallons of PCBs into the Duwamish River in Seattle cost $148,000 for partial recovery. Dredging of 50,000 cubic yards of river bottom in 1976 cost about $300,000.

March 21, 1974: Federal Register **39**, 10603 (1974): Correction—Effluent Standards.

August 1974: Two local EPA personnel took samples of fish, sediment and water below the GE outfall at Ft. Edward and found 355 ppm PCB in the fish. The report was not acted upon in EPA for a year until it finally reached a chemist at EPA's National Water Quality Laboratory in Duluth, Minn. A review of the PCB report was written and sent to the EPA regional office in New York.

December 31, 1974: GE is granted a federal permit, with the full blessings of NYDEC, to discharge up to 30 pounds per day into the Hudson River.

1975

1975: Federal Register **40**, 11563 (1975): FDA regulations concerning PCBs in paper/food packaging material.

Despite the recommendation of the 1972 Interdepartmental Task force on PCBs [Interdepartmental Task Force on PCBs. "Polychlorinated biphenyls and the Environment"; NTIS (COM-72–10410) May 1972] that discharges of PCBs be strictly controlled, EPA did not effect a policy of stringently curtailing discharges of PCBs from point sources until December 1975. Thus, GE was allowed to discharge under its NPDES permit, issued under FWPCA, 30 lb. of PCBs per day into the Hudson river. When the extent of pollution was brought to light in 1975, the allowable amount of PCBs discharged was reduced to approximately 1.5 pounds per day and a limit set of 3.5 ounces per day effective June 1, 1977. EPA did not set a toxic pollutant effluent standard for PCBs under authority of section 302 of FWPCA until February 2, 1977.

A report on the 'Handling and Disposal of PCBs from Repairs of Electrical Transformers' was published in December 1975 [Versar, 1975]. The purpose of the study was to review the responses to a questionnaire sent out to 13 companies involved with transformer repair at a total of 131 locations. One of the more pertinent findings was that transformers scrapped by the owner/user may be a serious potential source of pollution. While those transformers which are junked by repair shops are governed by a detailed NEMA specification (ANSI C-107/1974) the disposal of transformers through local junk yards may result in significant pollution because of lack of information on proper disposal procedures concerning PCB contaminated transformers. It was noted that uncontrolled handling of PCB filled transformers by junk yards was thought to have occurred in the past.

September 1975: A five-region council, the Lake Michigan Toxic Substance Committee of Midwest States, passed a resolution urging State agencies to support a ban on all PCB uses except those determined necessary by public hearing.

September 1975: CRC Critical Reviews on Environmental Control: description of the health effects of PCBs with an extensive bibliography of the early literature.

September 8, 1975: Ogden Reid (NYDEC Commissioner) brought action against GE to reduce discharge to 2 lb. per day (from 30 lb. per day) by December 31, 1975 and to zero by September 30, 1976.

October 1975: Wisconsin State officials held public hearings on the need for a statewide PCB ban.

October 6, 1975: EPA issued a draft set of water quality criteria under section 304(a) of the Federal Water Pollution Control Act.

November, 1975: An administrative hearing was begun on the discharge of PCBs into the Hudson river, chaired by a Columbia University Law Professor, Abraham Sofaer, to determine if GE had violated state water quality standards and, if so, what sort of restitution the company should reasonably make. GE presented an 'affirmative defense' which held that it had been granted both state and Federal permits to discharge PCBs. Sofaer agreed that this argument had "more than superficial appeal" but that it was not enough to excuse GE from the general requirements of its permit, which forbade the company to violate state water quality standards even if the precise prohibitions were not spelled out. Accordingly, Sofaer ruled that "GE has discharged PCBs in quantities that have breached applicable standards of water quality."

November 3, 1975: Background information on PCBs was published by the Office of Occupational Health Surveillance and Biometrics National Institute for Occupational Safety and Health, November 3, 1975 to inform the occupational health community of the then current knowledge concerning industrial uses and the toxicity of PCBs. The document contains an extensive bibliography of early literature. The report states that "Prior to the environmental concern surrounding the persistence and ubiquitousness of PCBs, [Interagency Task force on PCBs: PCBs and the Environment COM-72-10419, Washington, DC, March 20, 1972.] they were more widely used in industry as fluids for heat transfer systems, hydraulic systems, gas turbines, and vacuum pumps.... "

November 19–21, 1975: EPA and other Federal agencies held a PCB conference in Chicago, II. Several articles on the environmental effects of PCBs that appeared in the popular press shortly before the conference [Boyle, R.H.: "Of PCB ppms from GE and a SNAFU from EPA and DEC"; Audubon **77**, 127 (1975): and Boyle, R.H.: "Poisoned Fish, Troubled Waters"; Sports Illustrated, September 1, 1975, pp. 14–17 (1975)] caused considerable public interest in the conference and a number of demands that EPA regulate PCBs.

December 12, 1975: EPA announced the inclusion of PCBs in a national monitoring program designed to provide information for future possible regulation under the Safe Drinking Water Act.

December 30, 1975: EPA published a proposed list of hazardous substances for discharge into navigable waters which included PCBs.

1976

In early 1976, EPA requested Versar to expand its scope of work to include a study of wastewater treatment technology that could be used to reduce the concentration of PCBs in industrial effluents. [Versar report, May 16, 1979.]

Nagayama *et al.*[39] (1976) "Determination of Chlorinated Dibenzofurans in Kanechlors and Yusho Oil." The presence of key compounds of concern were identified in Yusho oil (Table 4).

The instructions for handling Inerteen insulating fluid referenced the ANSI C107.1-1974 for complete information on handling and disposal of Askarels. The document also included a section on "Environmental Considerations" which stated "It has been shown by several laboratories that measurable amounts of the PCBs, particularly those

TABLE 4

In Service Age (yr.)	Temperature(°C)	PCBs(ppm)	PCQs*(ppm)	PCDFs(ppm)
14	180–270	968,400	31,000	510
3	200–220	999,000	690	277
0‡	—	999,800	209	33

* PCB = polychlorinated quaterphenyl.

‡ = not used.

with more than 50% chlorination, are present in the general environment and are a threat to certain species of wildlife.... To be sure that the Inerteen and Inerteen contaminated materials do not contaminate the environment they must be incinerated in approved equipment."

A disposal section in the handling instructions describes the disposal of Inerteen liquid, solvent rinses contaminated with Inerteen, solids contaminated with Inerteen and the cleaning of contaminated drums.

February 9, 1976: The NYDEC found GE to have violated two State laws by its discharge of PCBs.

February 17, 1976: Ogden Reid (Comm. NYDEC) announced that he would issue orders closing the Hudson River as a commercial fishery for most species of fish because of PCB contamination.

February 13, 1976: The Hudson River was ordered closed to commercial fishing with the exception of chad.

February 24–25, 1976: The transformer subcommittee of ANSI met on February 24 and 25, 1976 and issued minutes which contained the following: "Transformers with Replacement Fluids: Apparatus that has contained PCBs which have been replaced with another fluid should be labeled as indicated below, and the disposal ultimately should follow the same disposal procedures as those recommended for PCB liquids and PCB-filled apparatus."

February 25, 1976: Versar Task I final report: "It is estimated that approximately 1.5 billion pounds of PCBs have been sold for industrial use in the US since initiation of production around 1930. Of this amount, at least 95% is still in existence; most is in service in capacitors and transformers, but about 290 million pounds are believed to reside in landfills and dumps and about 150 millions pounds are believed to be 'free' in the environment. The magnitude of these values indicates that there is a strong future threat from PCBs in land disposal sites,".... "Although PCB content in industrial wastes can be reduced through various approaches (treatment, substitution, etc.), the large amounts of PCBs already contained in land disposal sites present a severe hazard for the future." ... "Monsanto and portions of the electrical equipment industry which use PCBs have greatly reduced PCB releases to water and land over the past few years, primarily through improvement of plant housekeeping, improved waste collection and handling, and disposal of liquid wastes through incineration."

March 26, 1976: Senator Nelson introduced an amendment to the Toxic Substances Act which would eliminate the remaining uses of PCBs unless EPA finds that there is not a serious health hazard. The amendment was incorporated into TSCA as Section 6(e) and became a legislated requirement when TSCA was signed into law on October 11, 1976. The effective data of TSCA was January 1, 1977.

April 1, 1976: EPA issued recommended procedures for disposal of PCBs by industrial users under the Solid Waste Disposal Act. Recommended options for the disposal of PCB-containing wastes were given in priority order as (1) incineration and (2) controlled land disposal. In the case of land disposal it was mentioned that wastes containing PCBs should not be disposed of with other mixed wastes in a sanitary landfill. At this point in time the characteristics of transport of PCBs through soil were not definitively established. The regulations pointed out that "The ubiquity and persistence of PCBs indicate that their disposal should be carefully controlled until additional data are developed." Non-chemical waste landfills were then strongly contrasted with the following description of a chemical waste landfill. "In general terms, a chemical waste landfill provides complete long term protection for the quality of surface and subsurface waters from hazardous waste deposited therein, and against hazards to public health and the environment. Such sites should be located or engineered to avoid direct hydraulic continuity with surface and subsurface waters. Generated leachates should be contained and subsurface flow into the disposal area eliminated. Monitoring wells should be established and a sampling and analysis program conducted."

June 1976: Instructions for handling Inerteen® insulating fluid and installation and maintenance of inerteen transformers were issued by Westinghouse's Small Power Transformer Division in June 1976 [Westinghouse, 1976]. A section was included on environmental considerations in which PCBs were described as extremely persistent in the environment. In closing the section, Westinghouse goes beyond the recommendation of the ANSI C-107 guideline by indicating that the Inerteen and Inerteen contaminated materials should be incinerated in approved equipment. No mention is made of the second ANSI guideline alternative of draining and flushing the transformer carcass before disposal in a scrap yard. The suggested method of disposal is quite the opposite and leads to the ultimate destruction of PCB contaminated, porous materials such as absorbing clay, filter paper, cartridges, sawdust, rags and insulation while the non-porous metallic transformer components are solvent cleaned and the solvent incinerated.

June 28, 1976: Westinghouse indicated to its managers on June 28, 1976 that it would cease manufacture of Inerteen Transformers after December 31, 1976.

July, 1976: Publication of the "Final Report of the Subcommittee on the Health Effects of polychlorinated and polybrominated biphenyls; Dept. of Health, Education and Welfare, Washington.

July 23, 1976: Federal Register **41**, 30468 (1976): EPA published proposed toxic pollutant effluent standards for PCBs.

August, 1976: New York Times Magazine article on PCBs in the Hudson: "Last August, after months of negotiations, GE agreed to pay the State Department of Environmental Conservation, $3 million to help restore the river and another $1 million for research. The state will chip in $3 million of its own for restoration.... In fact, until very recently, it had a permit issued by the EPA and endorsed by the state–that specifically authorized it to discharge up to 30 pounds of PCBs into the river daily. By contributing $3 million of its own toward restoration as part of the agreement the DEC has tacitly acknowledged its own failure as a watchdog.".... "The August agreement, by contrast, was accomplished by negotiation after a rigorous but comparatively swift ex-judicial hearing sponsored by the DEC and chaired by a law professor from Columbia University. The whole process took less than a year. And while the parties never saw the inside of a courtroom, the agreement nonetheless puts American companies on notice that they may be held responsible for what they do to the environment even when they have not been specifically prohibited from doing it."

Seven California class 1 landfills were listed as approved disposition service companies together with nine other, principally incineration, options.

August 24, 1976: Rep. Gilbert Gude (R-MD) proposed a last-minute amendment to the House version of the Toxic Substances Control Act singling out PCBs for special restrictions. The August 24, 1976 edition of the New York Times reported: "The House of Representatives voted 319–45 to ban within three years the manufacture of PCBs, a class of chemicals that has been linked to cancer and birth defects, water pollution and wildlife contamination ... some House members said during today's debate that they felt that banning PCBs was a moral responsibility to the people ... ".

In October 1976, Congress passed the Toxic Substances Control Act. The continued use of PCBs in totally enclosed systems was allowed.

December 8, 1976: Federal Register **41**, 53692 (1976): Panel discussion/formation of PCB work group.

As of December 31, 1976 the Westinghouse South Boston, Virginia Small Power Transformer Division ceased manufacture of Inerteen transformers.

Instructions for handling Inerteen® insulating fluid P.D.S. 54201 CM and Installation and maintenance of Inerteen transformers were reissued from Westinghouse's Small Power Transformer Division in February 1976 [Westinghouse, 1976a]. Also included was information on actions to be taken when an Inerteen transformer was found to be leaking in transit.

As of December 31, 1976 the South Boston, Virginia Small Power Transformer Division ceased manufacture of Inerteen transformers.

A National conference on PCBs sponsored by EPA was held in November 1975 in Chicago.

1977

January 19, 1977: Federal Register **42**, 3701 (1977): Notice of Jan. 24, 1977 public meeting.

February 2, 1977: Federal Register **42**, 6531 (1977): Effluent Standard Regulations.

Section 307 (a) of the Federal Water Pollution Control Act required the EPA to publish a list of toxic pollutants, and submit proposed and final effluent standards for toxic pollutants. PCBs were one of the nine toxic pollutants contained in EPA's final listing of proposed pollutants on September 7, 1973 but did not publish the final toxic pollutant effluent standards for PCBs until February 3, 1977.

April 1, 1977: The FDA proposed the reduction of temporary tolerances for PCBs in milk, dairy products, poultry, eggs, fish and shell-fish.

April 21, 1977: Federal Register **42**, 20640 (1977): Proposed procedures for rule-making under Section 6 of TSCA.

May 24, 1977: Federal Register **42**, 26564 (1977): Proposed marking and disposal regulations.

July 15, 1977: Federal Register **42**, 36484 (1977): Deadline for reply comment period to proposed marking and disposal regulations of May 24, 1977.

Monsanto terminated the manufacture of PCBs in July 1977.

Effective August 31, 1977 Monsanto stopped accepting PCB waste returns.

1978

February 17, 1978: Federal Register **43**, 7150 (1978): EPA promulgated marking and disposal regulations.

April 21, 1978: Federal Register **43**, 17060 (1978): National Cancer Institute report on the carcinogenicity of Aroclor 1254.

June 7, 1978: Federal Register **43**, 24802 (1978): Proposed PCB ban regulations.

June 7, 1978: Federal Register **43**, 24818 (1978): Requires notification of intent to export PCBs.

July 18, 1978: Federal Register **43**, 30882 (1978): List of approved PCB disposal facilities.

August 2, 1978: Federal Register **43**, 33918 (1978): Corrections to marking and disposal regulations.

August 25, 1978: Federal Register **43**, 38057 (1978): Incorporates hearing record of effluent standard regulations into hearing record for ban regulations.

August 25, 1978: Federal Register **43**, 38087 (1978): List of Approved PCB disposal facilities.

October 26, 1978: Federal Register **43**, 50041 (1978): List of approved PCB disposal facilities.

November 1, 1978: Federal Register **43**, 50905 (1978): Interim rules-applications for exemption from PCB manufacturing ban.

December 20, 1978: Federal Register **43**, 59432 (1978): List of approved PCB disposal facilities.

1979

February 16, 1979: Federal Register **44**, 10266 (1979): Definition of "Discharge" under Clean Water Act.

February 16, 1979: Federal Register **44**, 10271 (1979): Defines reportable quantities of PCBs spilled into waterways, reporting requirements and fines.

March 12, 1979: Federal Register **44**, 13575 (1979): Request for comments on citizens' petition to give Regional Administrators authority to approve alternate disposal methods.

1981

August 15, 1983: Versar report "Exposure Assessment for Incidentally Produced PCBs: Appendix D: Prediction of Human Exposure for Hydraulic and Heat Transfer Authorized Uses:

- The USEPA's May 1979 regulation on PCBs, entitled "Polychlorinated Biphenyls Manufacturing, Processing, Distribution in Commerce, and Use Prohibitions" (40CFR 761.30 Subpart B, Authorizations), authorizes the use of PCBs until July 1, 1984, in hydraulic systems, in heat transfer systems, for research and development (small quantities), and as a microscopy mounting medium.
- After discontinuing production of their heat transfer oil in 1972 (Therminol), Monsanto notified all known Therminol users that hazards were associated with PCB use, offered to accept all fluid drained from PCB-containing heat transfer systems, and suggested that systems be refilled with non-PCB oil. Monsanto did not extend this service to users of PCB hydraulic fluid (Pydraul).
- Die casting hydraulic systems operate at high temperatures and require a fire-resistant hydraulic oil such as the type containing PCBs. PCBs have also been used in other general hydraulic systems (e.g., metal stamping equipment, forges, furnace lid lifts, presses), although not as commonly as in die casting hydraulic systems. It is likely that the PCB contamination of many hydraulic systems results from the reuse and recycling of PCB-containing hydraulic oils.
- Releases to water from hydraulic systems: PCB releases to water from hydraulic systems result from leakage and draining of fluid. Individual hydraulic systems are known to leak amounts of fluid equivalent to 0.25 to 10 volumes annually. This amount varies with machine age, maintenance, operating pressure, and configuration (e.g., fluid lines, seals, fittings). It is estimated that the average system leaks at a rate of 2 volumes/year.
- The leaked fluid recovery rate is the fraction of oil that has leaked from hydraulic systems and has been collected (typically with troughs or collection pans) and recycled into the system as topping fluid. Estimates of fluid recovery range from 30% to 80% for die casting systems and from 10% to 50% for general hydraulic systems. Leaked hydraulic fluid that is not recovered either volatilizes or seeps into the wastewater system.
- The average fluid capacity of a die casting system is 350 gallons; it is assumed that this is also the average volume of a general hydraulic system.
- The PCB regulations of May 1979 require that hydraulic and heat transfer systems that have PCB concentrations greater than 50 mg/Kg be drained and resampled annually until the concentration is reduced below 50 mg/Kg.
- Approximately 20.3 million pounds of PCB-containing heat transfer fluid was produced between 1962 and 1972 in the US.
- An average heat transfer system uses 400 gallons of fluid and has an average leakage of 0.2 volumes of fluid annually.
- PCB releases of heat transfer systems to water: PCB releases to water from heat transfer systems result from leakage and draining of fluid. The leakage rates in typical heat transfer systems range from 0.05 to 0.7 equivalent volumes of fluid annually. These rates are less than those for typical hydraulic systems, since heat transfer systems operate under less pressure, have fewer external fluid lines, and do not perform work on external components (pistons). A typical leak rate is 0.2 equivalent volumes of fluid annually. Recovery of leaked heat transfer fluid ranges from 20% to 40%: the average is estimated to be 25%. Leaked heat transfer fluid that is not recovered either volatilizes or seeps into the wastewater system.
- Occupational PCB air concentrations derived from hydraulic systems:
    - Die casting systems have peak operating temperatures of 100°C and average operating temperatures of 60°C. General hydraulic systems have peak operating temperatures of 82°C and average operating temperatures of 52°C. Portions of some die casting systems can have surface temperatures of 200°C.
    - Hydraulic systems leak continuously, covering the reservoir, fittings, hoses, and portions of the machine and floor below with fluid. These surfaces are therefore sources of volatilized PCBs. The average hydraulic machine has a 350-gallon reservoir with an appropriate outer surface area of 6.2 m^2.

- Occupational PCB air concentrations derived from heat transfer systems:
    - Heat transfer systems have much higher operating temperatures (average 180°C) and much lower leakage rates (0.2 volumes/yr) than do hydraulic systems. Therefore, less fluid will be lost through leakage and a higher proportion of leaked fluid is expected to volatilize from heat transfer systems.

September 1, 1983: Report on "Carcinogenic Risk Assessments of PCBs"; Health and Environmental Review Division of Office of Toxic Substances; September 1, 1983: document summarizes the results of four previous risk assessments for cancer conducted by FDA, OTA, CAG/EPA, and OTS.

October 1983: The Office of Pesticides and Toxic Substances of EPA published "Regulatory Impact Analysis of the Proposed Rule Regulating Inadvertent PCB Generation from Uncontrolled Sources."

1984

March 16, 1984: Chemical Manufactures Association wrote to EPA to request further discussions regarding PCB spill cleanup—"For several years, we have been concerned that various EPA regions have applied different approaches to determine the adequacy of PCB clean-up." In particular, they referred to a decision in which GE was required to clean up "to the lowest level of concentration below 50 ppm which is practically attainable through the use of normal cleanup methods."

August 22, 1984: Tests of new dioxin destruction processes involving chemical, biological and vaporization methods were being developed at Times Beach, Missouri, starting August 22, 1984. Dioxin had been found at more than 30 sites in the state, causing the EPA to relocate residents, in contaminated areas. [Environment Reporter, September 14, 1984].

1985

May 23, 1984: Plant Engineering reported that GE claimed that both aerobic and anaerobic naturally occurring bacteria were degrading the PCBs in the Hudson river to less chlorinated, less toxic forms and that, in deeper sediment samples, estimated to be about 25 years old, the process is about 75% complete.

1987

September 1987: A PCB trial burn report for the USEPA Combustion Research Facility liquid injection incinerator system was published in September 1987. It describes a trial burn of PCB laden oil and No. 2 fuel oil (about 40% A1260) injected into the liquid incinerator. Destruction efficiencies in excess of 99.99999% were achieved.

1988

July 22, 1988: EPA list of approved PCB disposal companies: 7 incinerators, 2 alternate thermal, 11 chemical, 4 physical separation, 1 biological, 8 chemical waste landfills, 9 companies with PCB disposal method applications.

Ballard, J.G. and S.H. Hawthorne; 'Solvent Decontamination of PCB Electrical Equipment'; Canadian Electrical Association Engineering and Operating Division; Part 1, 81-A-66 (1981).

Bryant, J.G.; open Monsanto letter; September 15, 1970.

Buser, H.R. and C. Rappe; 'Formation of PCDFs from the Pyrolysis of Individual PCB Isomers.'; *Chemospherei,* **8**, 157 (1979).

Dakin, T.W. and L. Mandelcorn; 'A Status Report on Chlorinated Diphenyls, Regarding Environmental Pollution'; Proprietary report with strictly limited distribution; dated December 27, 1971.

GE; open letter to electric utility customers from the Electric Utility Sales Division; 1970.

GE; 'The Use and Disposal of Electrical Insulating Liquids. Report to the Electrical and Nuclear Sub-Council of the National Industrial Pollution Control Council; revised June 1971.

GE; Comments on notice of proposed rulemaking, *Fed. Reg.,* **37** (54), 5705, March 18, 1972; 'The Role of PCBs in Electrical Equipment'; Feb **4**, 1972.

Gossage, T.L.; Marketing Director, Specialty Products Group, Monsanto; letter to Purchasing Agents indicating discontinuance of direct sales of PCBs; February 28, 1972.

Gustafson, C.G.; *Environ. Sci. Technol.,* **4**, 814 (1970).

Jensen, Soren: "Report of a New Chemical Hazard"; *New Scientist,* **32**, 612 (1966).

Keiser D.P.; Westinghouse open letter to industry Purchasing Executives re. PCB-in-oil contamination; November 22, 1976.

Council on Environmental Quality; "Toxic Substances"; Washington, D.C. (1971).

Meigs, J.W., J.J. Albom and B.L. Kartin; "Chloracne from an unusual exposure to Aroclor."; *JAMA,* **154**, 1417 (1954).

Monsanto PCB Bulletin; 'Handling, Waste Control & Disposal of PCBs'; July 1971.

Munson, T.O., H.D. Palmer and J.M. Forns; 'Transport of Chlorinated Hydrocarbons in the Upper Chesapeake Bay'; EPA-560/6-75-004 March 1976.

Olson, D.A; Monsanto letter; February 18, 1970.

Papageorge, W.B.; Monsanto letter; September 1, 1970.

Papageorge, W.B.; letter to A.M. Salazar of NEMA re. PCB symposium information; January 4, 1972a.

Papageorge, W.B.; Letter to Monsanto PCB customers re. confidentiality of sales lists; April 1972b.

Smyth, H.F.; "Toxicity of certain benzene derivatives and related compounds"; *J. Ind. Hyg. Toxicol.,* **13**, 87 (1931).

Vos, J.G., J.H. Koeman, H.L. van der Mass, M.C. Ten Noever de Braun and R.H. de Vos; 'Identification and Toxicological Evaluation of PCDDs and PCNs in Two Commercial PCBs'; *Fd. Cosmet. Toxicol.,* **8**, 625 (1970).

TABLE 5
State properties of PCB homologous series

PCB Homologous Series	Physical State	Melting Point (°C)	Boiling Point (°C)	Vapor Pressure (Pa at 25°C)
Biphenyl	S	71	256	4.9
Monochlorobiphenyl	S/L	25–77.9	285	1.1
Dichlorobiphenyl	S/L	24.4–149	312	0.24
Trichlorobiphenyl	S/L	28–87	337	0.054
Tetrachlorobiphenyl	S/L	47–180	360	0.012
Pentachlorobiphenyl	S/L	76.5–124	381	2.6×10^{-3}
Hexachlorobiphenyl	S/L	77–150	400	5.8×10^{-4}
Heptachlorobiphenyl	S/L	122.4–149	417	1.3×10^{-4}
Octachlorobiphenyl	S/L	159–162	432	2.8×10^{-5}
Nonachlorobiphenyl	S/L	182–206	445	6.3×10^{-6}
Decachlorobiphenyl	S	305.9	456	1.4×10^{-6}

S = solid.
L = liquid.

Versar Inc.; 'The Handling and Disposal of PCBs from Repairs of Electrical Transformers'; EPA contract No. 68–01–3259, December 5, 1975.

Westinghouse; 'Instructions for Inerteen® Insulating Fluid P.D.S. 54201 CM and Installation and Maintenance of Inerteen Transformers'; Power Transformer Division, Sharon, PA; IB 45-063-99B, September 1968.

Westinghouse; Supplement to 'Instructions for Inerteen® Insulating Fluid P.D.S. 54201 CM and Installation and Maintenance of Inerteen Transformers' entitled 'Inerteen and Environmental Contamination'; Power Transformer Division, Sharon, PA; Supplement to IB 45-063-99B; September 1968.

Westinghouse; 'Instructions for Handling Inerteen® Insulating Fluid P.D.S. 54201 CM and Installation and Maintenance of Inerteen Transformers'; Power Transformer Division; I.B. 45-063-99C, August 1971.

Westinghouse; Sharon Division; 'Instructions for Shipment Installation & Storage of Oil-Filled Power Transformers'; Sharon units publication number I.L. 48-069-43., September 1972.

Westinghouse; 'Instructions for Handling Inerteen Insulating Fluid P.D.S. 54201 CM and Installation and Maintenance of Inerteen Transformers'; Small Power Transformer Division; I.B. 45-063-99D; February 1976a.

Westinghouse; 'Instructions for Handling Inerteen Insulating Fluid P.D.S. 54201 CM and Installation and Maintenance of Inerteen Transformers'; Small Power Transformer Division; I.B. 45-063-99E; June 1976b.

Wheeler, E.P.; Monsanto letter; March 3, 1969.

Wiener, G.W.; Westinghouse confidential internal memo Re. 'Minutes of PCB Status'; December 28, 1971.

TABLE 6
Fluid properties of PCB homologous series

PCB Homologous Series	Specific Gravity	Viscosity mPa.s at 25°C	Viscosity Universal Saybolts at 25°C
Biphenyl	1.0	17	78
Monochlorobiphenyl	1.1	20	80
Dichlorobiphenyl	1.3	28	100
Trichlorobiphenyl	1.4	56	190
Tetrachlorobiphenyl	1.5	200	610
Pentachlorobiphenyl	1.5	1.5×10^{-3}	4.4×10^{3}
Hexachlorobiphenyl	1.6	2.9×10^{4}	8.2×10^{4}
Heptachlorobiphenyl	1.7	$>10^{6}$	$>10^{6}$
Octachlorobiphenyl	1.7	$>10^{6}$	$>10^{6}$
Nonachlorobiphenyl	1.8	$>10^{6}$	$>10^{6}$
Decachlorobiphenyl	1.8	$>10^{6}$	$>10^{6}$

PROPERTIES OF PCBs

The more chlorine used in the manufacture of PCBs the greater the viscosity of the product and the greater its resistance to heat, moisture and electrical degradation. The properties of PCBs which are the most important from the point of view of usage and environmental exposure risk are as follows:

- PCBs are heat stable and have no flash or fire point;
- very chemically stable and unreactive;
- excellent dielectric fluids;
- only slightly soluble in water, very soluble in oils and organic solvents;
- Very low volatility in air;
- highly lipophilic. (Table 5, 6, 7, 8)

TABLE 7
Some of the environmentally important properties of PCB homologous series

PCB Homologous Series	Molecular Weight (g/mol)	Water Solubilityat 25°C (g/m^3)	Log K_{ow}	Bioconcentration Factor in Fish	Evaporation Rateat 25°C (g/(m^{2}h))
Biphenyl	154.2	9.3	4.3	1000	0.92
Monochlorobiphenyl	188.7	4.0	4.7	2500	0.25
Dichlorobiphenyl	223.1	1.6	5.1	6300	0.065
Trichlorobiphenyl	257.5	0.65	5.5	1.6×10^4	0.017
Tetrachlorobiphenyl	292.0	0.26	5.9	4.0×10^4	4.2×10^{-3}
Pentachlorobiphenyl	326.4	0.099	6.3	1.0×10^5	1.0×10^{-3}
Hexachlorobiphenyl	360.9	0.038	6.7	2.5×10^5	2.5×10^{-4}
Heptachlorobiphenyl	395.3	0.014	7.1	6.3×10^5	6.2×10^{-5}
Octachlorobiphenyl	429.8	5.5×10^{-3}	7.5	1.6×10^6	1.5×10^{-5}
Nonachlorobiphenyl	464.2	2.0×10^{-3}	7.9	4.0×10^6	3.5×10^{-6}
Decachlorobiphenyl	498.7	7.6×10^{-4}	8.3	1.0×10^7	8.5×10^{-7}

TABLE 8
Some Aroclor properties of environmental importance

	Aroclor				
Property	1016	1242	1248	1254	1260
Vapor Pressure at 25°C (Pa)	0.10	0.091	0.023	6.7×10^{-3}	6.4×10^{-4}
Viscosity at 25°C (mPa.s)	45	69	280	2.0×10^3	1.9×10^5
Specific gravity at 25°C	1.4	1.4	1.4	1.5	1.6
Fluid density at 25°C	1,370	1,381	1,445	1,539	1,621
Solubility in water at 25°C (g/m3)	0.84	0.75	0.32	0.14	0.035
$\mathrm{Log}_{10}\ K_{ow}$	4.4–5.8	4.5–5.8	5.8–6.3	6.1–6.8	6.3–7.5
Bioconcentration	$1.3 \times 10^3 - 3.2 \times 10^4$	$1.6 \times 10^3 - 3.2 \times 10^4$	$3.2 \times 10^4 - 1.0 \times 10^5$	$6.3 \times 10^4 - 3.2 \times 10^5$	$1.0 \times 10^5 - 1.6 \times 10^6$
Factor in Fish	—	—	—	—	—
Evaporation rate at 25°C (g/(m2h))	0.031	0.029	8.3 × 10-3	2.7 × 10-3	2.9 × 10-4

COMMERCIAL APPLICATIONS OF PCBs

Because of their physical properties, PCBs found extensive applications in electrical equipment such as power transformers and capacitors, as well as in hydraulic machinery, vacuum pumps, compressors and heat-exchanger fluids. Other uses include:

- lubricants
- fluorescent light ballasts
- paints
- glues
- waxes
- carbonless copy paper
- inks, including newspapers
- dust-control agents for dirt road
- solvents for spreading insecticides
- cutting oils.

Table 9 illustrates the diversity of applications in the United States of different Aroclors.

A numbering system used in Aroclor products describes, firstly the number of carbon atoms in the molecule by the first two digits (12) and, secondly, the last two digits indicate the percentage of chlorine in the product. For example, Aroclor 1254 is biphenyl with 6 carbon atoms in each of two phenyl rings and 54% chlorine by weight.

Aroclor 1016 is an exception to the rule. It was not manufactured with 16% chlorine but rather is the name given to a product obtained by distilling Aroclor 1242. The distillate product contained 41% chlorine and the percentage of the relatively low boiling, low chlorine isomers, with 3 or 4 chlorine

TABLE 9
A variety of Aroclors were used throughout industry

Application	Types of Aroclor Used
Transformers	1242, 1254, 1260
Capacitors	1016, 1221, 1242, 1254
Vacuum pumps	1248, 1254
Hydraulic fluids	1232, 1242, 1248, 1254, 1260
Heat exchangers	1242
Waxes	1242, 1254, 1268
Dust-control agents	1254, 1260
Inks, lubricants, cutting oils	1254
Carbonless copy paper	1242

TABLE 10
The number of chlorine atoms in the series tends to increase as the amount of chlorine in the PCB increases

	% of Homologous Series in Aroclors			
Chlorobiphenyl	1242	1248	1254	1260
$C_{12}H_9Cl$	3	—	—	—
$C_{12}H_8Cl_2$	13	2	—	—
$C_{12}H_7Cl_3$	28	18	—	—
$C_{12}H_6Cl_4$	30	40	11	
$C_{12}H_5Cl_5$	22	36	49	12
$C_{12}H_4Cl_6$	4	4	34	38
$C_{12}H_3Cl_7$	—	—	6	41
$C_{12}H_2Cl_8$	—	—	—	8
$C_{12}HCl_9$	—	—	—	1

atoms per molecule, also increased, relative to Aroclor 1242. The product is more biodegradable and less toxic than Aroclor 1242. Aroclor 1242 was the principal capacitor impregnant in the U.S. prior to about 1971 when it was replaced by Aroclor 1016.

Table 10 shows the percentage distribution of isomers amongst different homologous series.

Table 11 shows the IUPAC numbering and substitution pattern for PCB congeners.

The numbering system for PCB congeners is important because it allows isomers to be referred to easily and without ambiguity. In particular, instances in which particular congeners are recognized, such as in chromatographic analyses of commercial PCB products, and in considerations of toxic isomers, the IUPAC numbering system is extremely valuable.

Receptor-mediated responses are characterized by the stereoselective interaction of ligands with receptors. Structure-activity relationships for PCBs have been extensively investigated and are thought of as the basis for the role of the Ah receptor in initiating the toxic effects observed

TABLE 11
The International Union of Pure and Applied Chemistry numbering system for PCB congeners

No.	Structure
	Monochlorobiphenyls
1	2
2	3
3	4
	Dichlorobiphenyls
4	2,2′
5	2,3
6	2,3′
7	2,4
8	2,4′
9	2,5
10	2,6
11	3,3′
12	3,4
13	3,4′
14	3,5
15	4,4′
	Trichlorobiphenyls
16	2,2′,3
17	2,2′,4
18	2,2′,5
19	2,2′,6
20	2,3,3′
21	2,3,4
22	2,3,4′
23	2,3,5
24	2,3,6
25	2,3′,4
26	2,3′,5
27	2,3′,6
28	2,4,4′
29	2,4,5
30	2,4,6
31	2,4′,5
32	2,4′,6
33	2′,3,4
34	2′,3,5
35	3,3′,4
36	3,3′,5
37	3,4,4′
38	3,4,5′
39	3,4′,5
	Tetrachlorobiphenyls
40	2,2′,3,3′
41	2,2′,3,4′
42	2,2′,3,4′

(*continued*)

TABLE 11 (*continued*)
The International Union of Pure and Applied Chemistry numbering system for PCB congeners

No.	Structure	No.	Structure
43	2,2′,3,5	88	2,2′,3,4,6
44	2,2′,3,5′	89	2,2′,3,4,6′
45	2,2′,3,6	90	2,2′,3,4′,5
46	2,2′,3,6′	91	2,2′,3,4′,6
47	2,2′,4,4′	92	2,2′,3,5,5′
48	2,2′,4,5	93	2,2′,3,5,6
49	2,2′,4,5′	94	2,2′,3,5,6′
50	2,2′,4,6′	95	2,2′,3,5′,6
51	2,2′,4,6′	96	2,2′,3,6,6′
52	2,2′,5,5′	97	2,2′,3′,4,5
53	2,2′,5,6′	98	2,2′,3′,4,6
54	2,2′,6,6′	99	2,2′,4,4′,5
55	2,3,3′,4	100	2,2′,4,4′,6
56	2,3,3′,4′	101	2,2′,4,5,5′
57	2,3,3′,5	102	2,2′,4,5,6′
58	2,3,3′,5′	103	2,2′,4,5′,6
59	2,3,3′,6	104	2,2′,4,6,6′
	Tetrachlorobiphenyls	105	2,3,3′,4,4′
60	2,3,4,4′	106	2,3,3′,4,5
61	2,3,4,5	107	2,3,3′,4′,5
62	2,3,4,6	108	2,3,3′,4,5′
63	2,3,4′,5	109	2,3,3′,4,6
64	2,3,4′,6	110	2,3,3′,4′,6
65	2,3,5,6	111	2,3,3′,5,5′
66	2,3′,4,4′	112	2,3,3′,5,6
67	2,3′,4,5	113	2,3,3′,5′,6
68	2,3′,4,5′	114	2,3,4,4′,5
69	2,3′,4,6	115	2,3,4,4′,6
70	2,3′,4′,5	116	2,3,4,5,6
71	2,3′,4′,6	117	2,3,4′,5,6
72	2,3′,5,5′	118	2,3′,4,4′,5
73	2,3′,5′,6	119	2,3′,4,4′,6
74	2,4,4′,5	120	2,3′,4,5,5′
75	2,4,4′,6	121	2,3′,4,5′,6
76	2′,3,4,5	122	2′,3,3′,4,5
77	3,3′,4,4′	123	2′,3,4,4′,5
78	3,3′,4,5	124	2′,3,4,5,5′
79	3,3′,4,5′	125	2′,3,4,5,6′
80	3,3′,5,5′	126	3,3′,4,4′,5
81	3,4,4′,5	127	3,3′,4,5,5′
	Pentachlorobiphenyls		Hexachlorobiphenyls
82	2,2′,3,3′,4	128	2,2′,3,3′,4,4′
83	2,2′,3,3′,5	129	2,2′,3,3′,4,5
84	2,2′,3,3′,6	130	2,2′,3,3′,4,5′
85	2,2′,3,4,4′	131	2,2′,3,3′,4,6
86	2,2′,3,4,5	132	2,2′,3,3′,4,6′
87	2,2′,3,4,5′	133	2,2′,3,3′,5,5′

(*continued*)

TABLE 11 (*continued*)
The International Union of Pure and Applied Chemistry numbering system for PCB congeners

No.	Structure	No.	Structure
134	2,2′,3,3′,5,6	173	2,2′,3,3′,4,5,6
135	2,2′,3,3′,5,6′	174	2,2′,3,3′,4,5,6′
136	2,2′,3,3′,6,6′	175	2,2′,3,3′,4,5′,6
137	2,2′,3,4,4′,5	176	2,2′,3,3′,4,6,6′
138	2,2′,3, 4,4′,5′	177	2,2′,3,3′,4′,5,6
139	2,2′,3, 4,4′,6	178	2,2′,3,3′,5,5′,6
140	2,2′,3, 4,4′,6′	179	2,2′,3,3′,5,6,6′
141	2,2′,3,4,5,5′	180	2,2′,3,4,4′,5,5′
142	2,2′,3,4,5,6	181	2,2′,3,4,4′,5,6
143	2,2′,3,4,5,6′	182	2,2′,3,4,4′,5,6′
144	2,2′,3,4,5′,6	183	2,2′,3,4,4′,5′,6
145	2,2′,3,4,6,6′	184	2,2′,3,4,4′,6,6′
146	2,2′,3,4′,5,5′	185	2,2′,3,4,5,5′,6
147	2,2′,3,4′,5,6	186	2,2′,3,4,5,6′,6′
148	2,2′,3,4′,5,6′	187	2,2′,3,4′,5,5′,6
149	2,2′,3,4′,5′,6′	188	2,2′,3,4′,5,6,6′
150	2,2′,3,4′,6′,6′	189	2,3,3′,4,4′,5,5′
151	2,2′,3,5,5′,6	190	2,3,3′,4,4′,5,6
152	2,2′,3,5,6,6′	191	2,3,3′,4,4′,5′,6
153	2,2′,4,4′,5,5′	192	2,3,3′,4,5,5′,6
154	2,2′,4,4′,5,6′	193	2,3,3′,4′,5,5′,6
155	2,2′,4,4′,6,6′		Octachlorobiphenyls
156	2,3,3′,4,4′,5	194	2,2′,3,3′,4,4′,5,5′
157	2,3,3′,4,4′,5′	195	2,2′,3,3′,4,4′,5,6
158	2,3,3′,4,4′,6	196	2,2′,3,3′,4,4′,5,6′
159	2,3,3′,4,5,5′	197	2,2′,3,3′,4,4′,6,6′
160	2,3,3′,4,5,6	198	2,2′,3,3′,4,5,5′,6
161	2,3,3′,4,5′,6	199	2,2′,3,3′,4,5,6,6′
162	2,3,3′,4′,5,5′	200	2,2′,3,3′,4,5′,6,6′
163	2,3,3′,4′,5,6	201	2,2′,3,3′,4,5,5′,6′
164	2,3,3′,4′,5′,6	202	2,2′,3,3′,5,5′,6,6′
165	2,3,3′,5,5′,6	203	2,2′,3,4,4′,5,5′,6
166	2,3,4,4′,5,6	204	2,2′,3,4,4′,5,6,6′
167	2,3′,4,4′,5,5′	205	2,3,3′,4,4′,5,5′,6
168	2,3′,4,4′,5′,6		Nonachlorobiphenyls
	Heptachlorobiphenyls	206	2,2′,3,3′,4,4′,5,5′,6
169	3,3′,4,4′,5,5′	207	2,2′,3,3′,4,4′,5,6,6′
170	2,2′,3,3′,4,4′,5	208	2,2′,3,3′,4,5,5′,6,6′
171	2,2′,3,3′,4,4′,6		Decachlorobiphenyl
172	2,2′,3,3′,4,5,5′	209	2,2′,3,3′,4,4′,5,5′,6,6′

with PCBs and related compounds. The PCB binds to a receptor protein and is transported into the nucleus, occupies a nuclear binding site and turns on gene expression similar to the way steroid hormone-receptor complexes turn on genes. The most toxic PCB congeners, namely isomers #77, 81, 126 and 169 are dioxin-like in their coplanar conformation. The four coplanar PCBs are approximate isostereoisomers of 2,3,7,8-tetrachloro-p-dioxin, as shown in the following diagram. The most biologically active compounds are substituted at the para- and at least one meta- position of both phenyl rings and do not contain any orthochloro substituents. These compounds, fortunately, do not occur as major constituents in commercial PCB mixtures. Substitution patterns such as the (2,6), (3,5), (2,4,6) and (3,4,5) are not commonly found in commercial mixtures. The number of chlorines on the two rings seldom differs by more than one for an odd number of chlorines in the molecule, and is usually equal for an even number of chlorines. Because of this, the number of congeners found in technical PCBs is far less than the theoretical number.

The introduction of a single ortho-chloro substituent into the biphenyl ring results in decreased coplanarity between the two phenyl rings due to steric interactions. Nevertheless, the isomers bind to the Ah receptor and cause dioxin-like responses although their potency is less than the coplanar PCBs. The mono-ortho- PCBs are shown in Figure 12.

Di-ortho- substitution of the four coplanar PCBs results in 13 possible di-ortho-coplanar PCBs. This substitution pattern would result in more steric hindrance to coplanarity than the mono-ortho- congeners and it would therefore be expected that they would exhibit weaker dioxin-like effects. The di-ortho- compounds are shown in Figure 14.

In more general terms, PCBs can be separated into six different structural classes. Class I and Class II compounds are the coplanar and mono-ortho-coplanar PCBs, respectively. Class III are mono-ortho-coplanars, which lack a chlorine in a paraposition. Class IV PCBs are the di-ortho-coplanar compounds. Classes I to IV are dioxin-like in decreasing order of potency. Class V and Class VI are the tri- and tetra-ortho-PCBs which are not coplanar and do not elicit significant dioxin-like activity. The six classes are shown with representative examples in Figure 15.

In summary, the coplanar PCB congeners are toxic and bind with high affinity to the Ah receptor and are highly dioxin-like. The mono-ortho-coplanar PCBs exhibit less toxicity than the coplanar isomers and interact with the

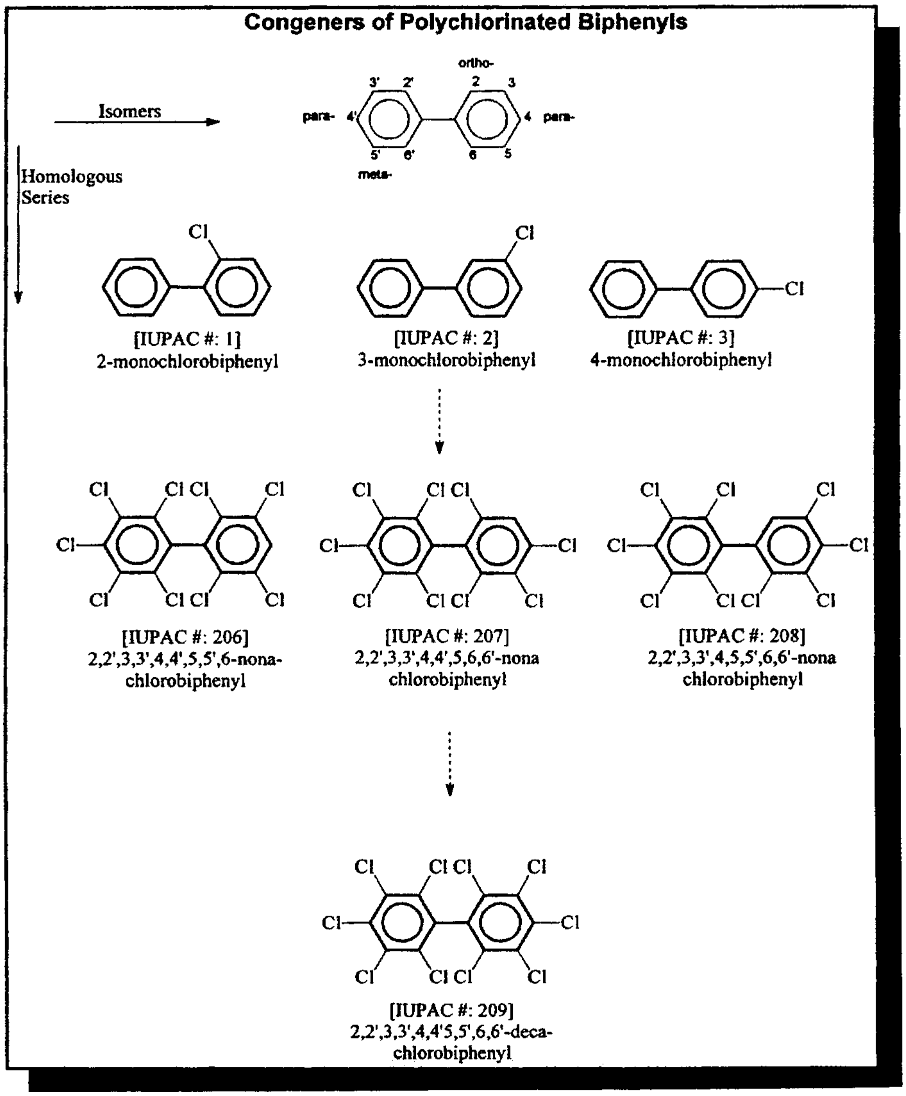


FIGURE 10 PCB congeners are isomers contained in homologous series.

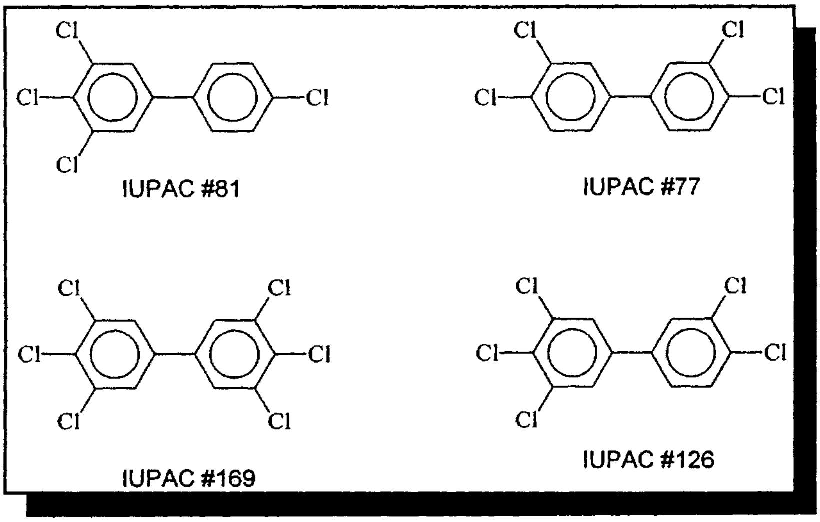


FIGURE 11

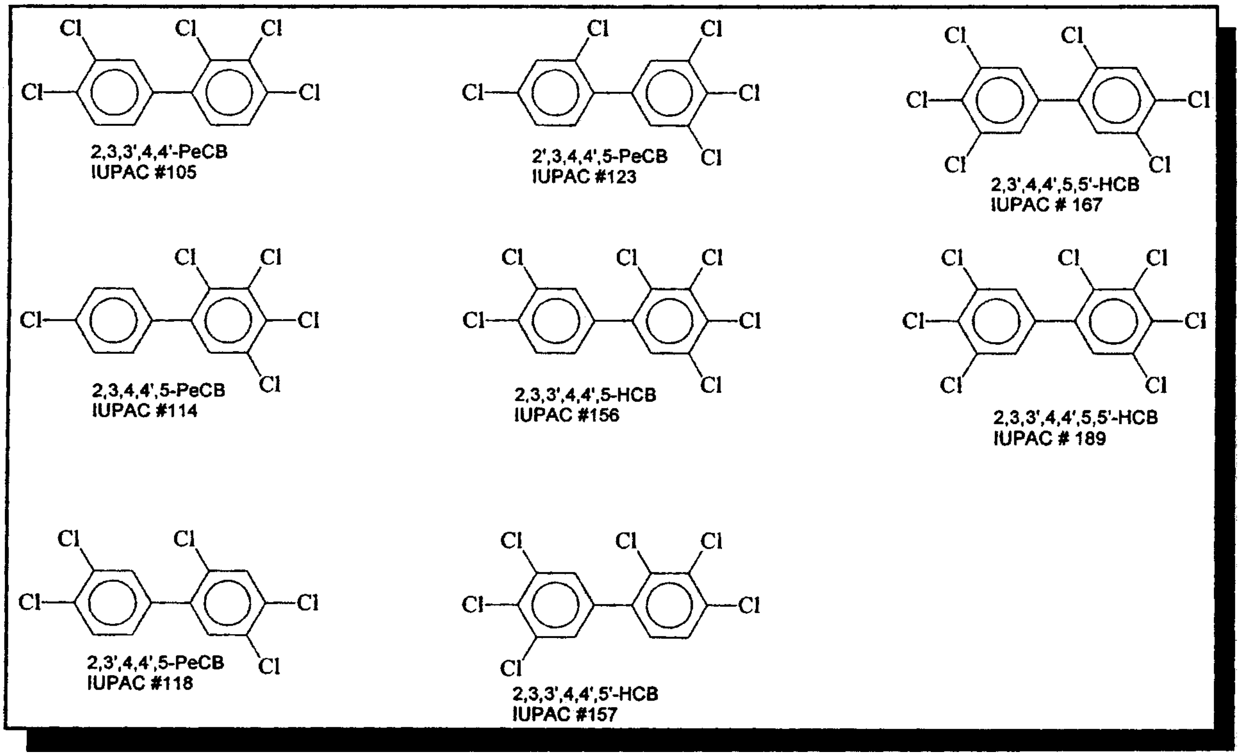


FIGURE 12 Mono-ortho-chlorobiphenyls.

Ah receptor with much lower affinity, but elicit dioxin-like responses. The di-ortho-coplanar PCBs are very insoluble so that receptor-binding activity cannot be readily measured, but at sufficiently high doses, dioxin-like responses have been observed. PCB congeners which lack a para-chloro substituent also produce weak dioxin-like effects.

The main mechanism of PCB metabolism in birds and mammals involves the insertion of oxygen into two adjacent carbon atoms on the ring by hepatic microsomal oxidases (HMOs) to form a transient arene oxide intermediate which then becomes degraded to hydroxylated PCBs. For example, the metabolism of 4,4'dichlorobiphenyl

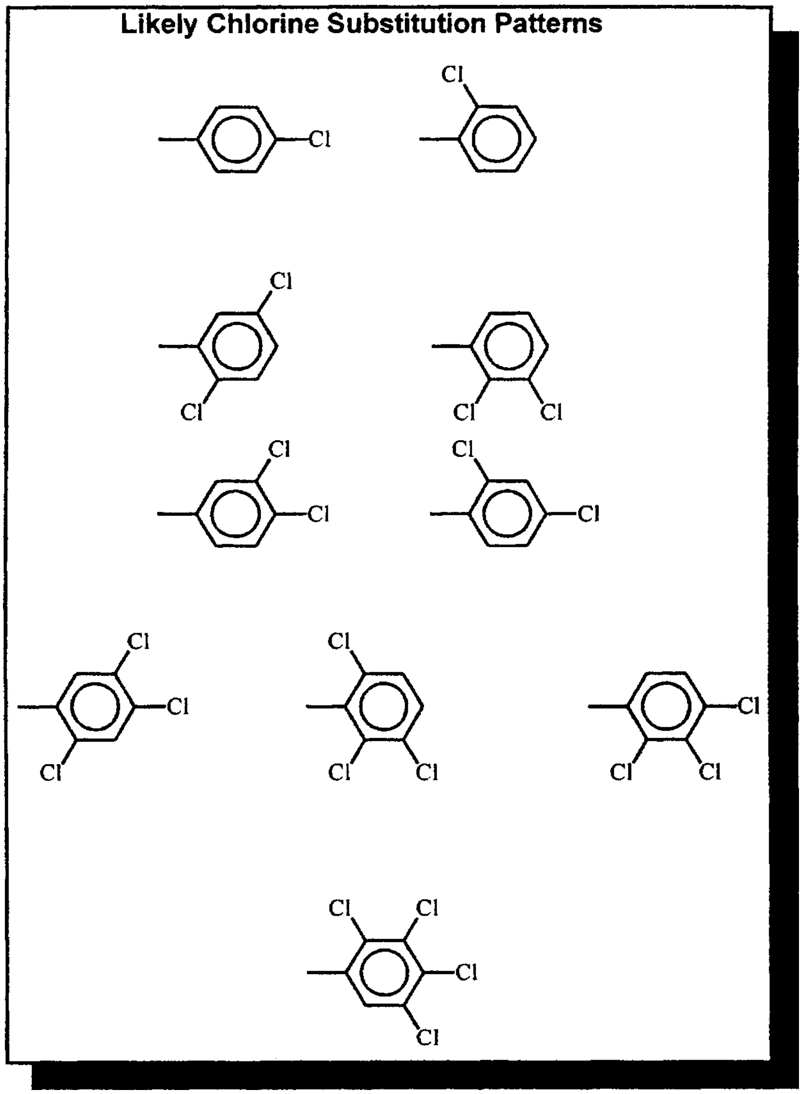


FIGURE 13 Likely substitution patterns.

in the rat yields four monohydroxy-, four dihydroxy-, and two trihydroxy-metabolites. The mechanism is shown in the figure below as epoxidation of the biphenyl nucleus followed by epoxide ring opening accompanied by a 1, 2-chlorine shift. Subsequent cyclisation of the hydroxylated derivatives produced either by metabolism or as a result of photo degradation in the environment or as a reaction by-product of a PCB disposal method, can lead to the formation of chlorinated dibenzofurans. The transformation of only 0.002% of a major constituent of an Aroclor mixture to the corresponding chlorinated dibenzofurans would produce concentrations in the mixture corresponding to the values reported by Vos *et al.* [Vos, 1970] as toxicologically significant.

Table 12 shows the concentrations of toxic PCB congeners which have been found in commercial PCB mixtures.

The toxic isomers of interest are shown in Figure 17.

The manufacturing process for PCBs caused the chlorination of the hydrocarbons present to yield a number of polysubstituted products. In some cases, the initial reaction mixture was contaminated with trace quantities of dibenzofurans and naphthalenes. The polychlorinated dibenzofuran (PCDFs) and polychlorinated naphthalene (PCNps) products are compounds of concern even in trace quantities. Based on a vast amount of published information in scientific journals the causative agent of health effects attributed to PCBs is most likely due to the presence of PCDFs-see the section on “The Perception of PCB Health Effects; the Yusho and Yu-Cheng Poisonings.”

Table 13 indicates the concentrations and types of polychlorinated dibenzofurans in commercial products.

TABLE 12
Data from Larsen *et al.*, Gordon and Breach 1993. (Ref. 155).

PCB #	A1232	A1242	A1248	A1254	A1260	A1262	Askarel
60	0.46	0.66	1.32	0.56	0.011	0.027	0.039
74	0.86	1.37	2.65	1.77	0.023	0.061	0.070
77	—	0.50	0.30	ND	ND	—	—
81	—	ND	—	ND	ND	—	—
105	0.17	0.43	1.00	4.71	0.045	0.0079	0.32
114	0.0080	0.0098	0.019	0.043	0.0014	0.0003	0.0011
118	0.32	0.74	1.69	9.09	0.57	0.25	1.94
123	0.024	0.038	0.085	0.33	ND	ND	ND
126	0.0013	0.0037	0.011	0.027	0.0004	ND	ND
156	0.059	0.026	0.083	1.07	0.48	0.59	0.56
157	0.0013	0.0026	0.011	0.026	0.024	0.0078	0.085
167	ND	ND	0.0014	0.045	0.030	0.017	0.059
169	ND	ND	ND	ND	ND	ND	ND
189	ND	ND	0.012	0.031	0.13	0.052	0.077

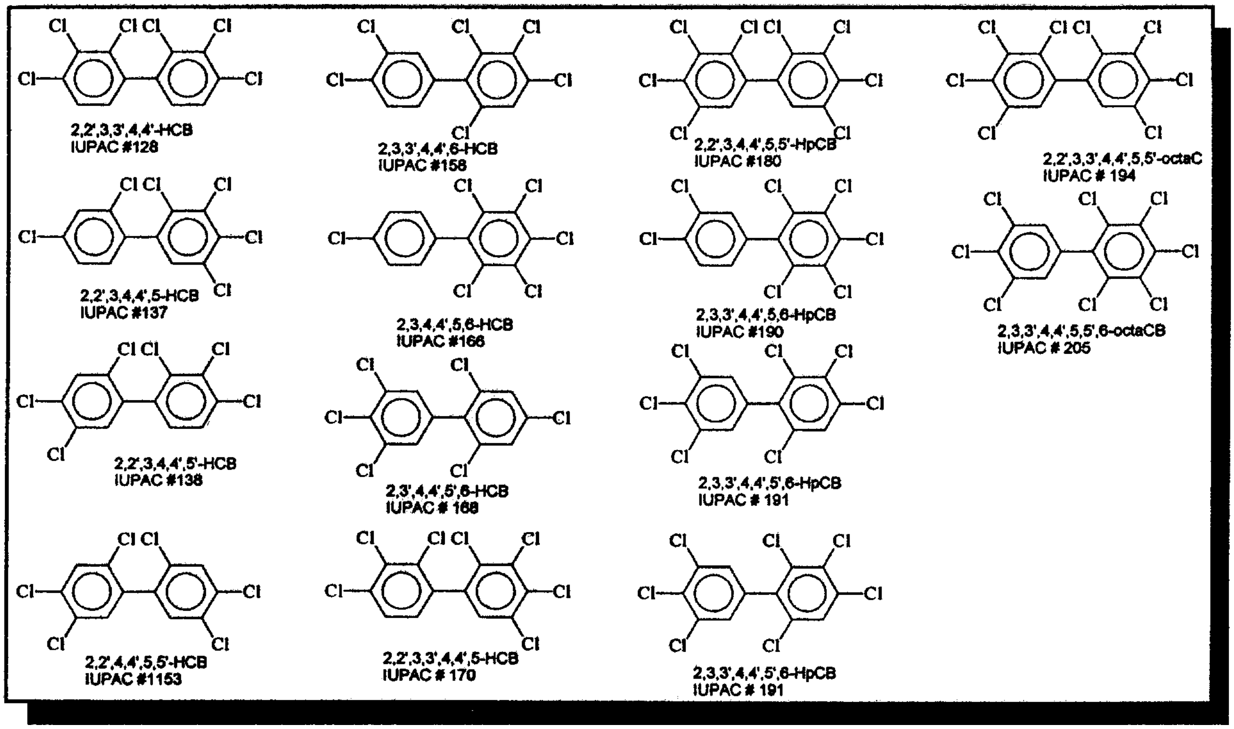


FIGURE 14 Di-ortho-chlorobiphenyls.

COMPOUNDS OF CONCERN

In 1968 a group of workers at Kyushu University in northern Japan determined that a group of patients with acneiform eruptions similar to chloracne had ingested contaminated rice oil used in cooking. The symptoms of the poisoning suggested that it was due to organochlorine compounds. The word "Yusho" or "rice oil" disease is used to describe the incident. The number of recognized victims over a period of about ten years rose to a total of approximately 1800 people.

The rice oil contamination was determined to be Kanechlor 400 which was produced as an approximately

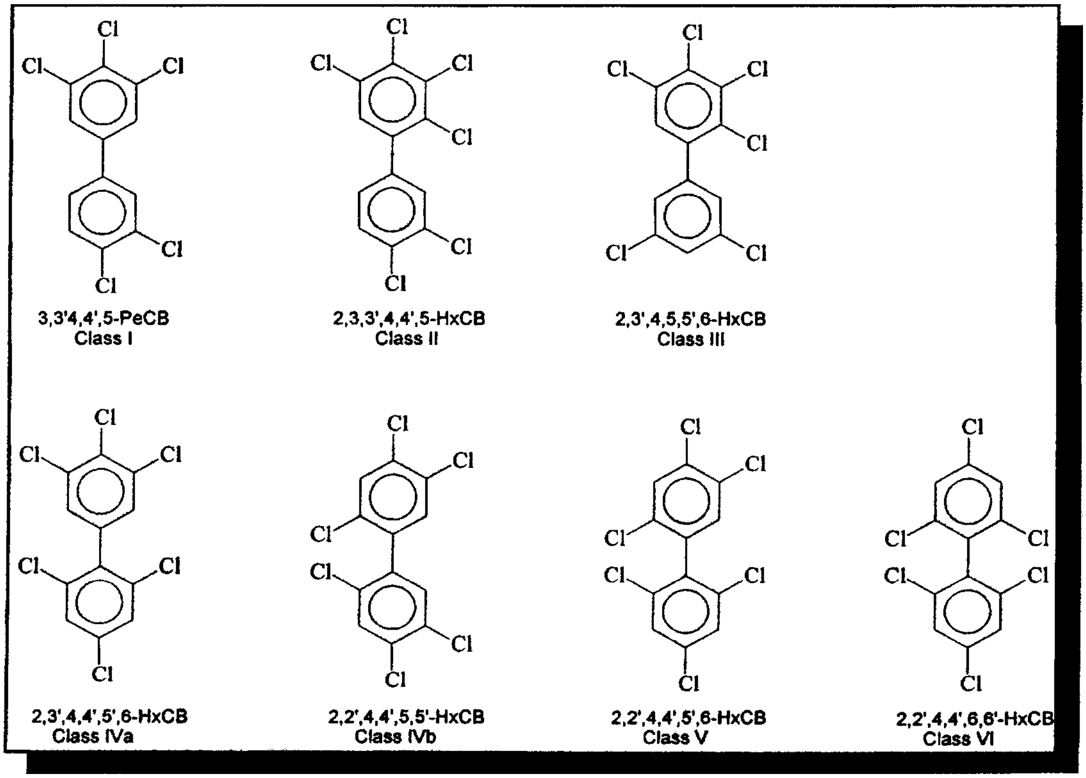


FIGURE 15 Structural categories of PCBs.

TABLE 13

PCB products manufactured in the United States fortunately have low concentrations of PCDF impurities

Commercial Product	Chlorinated Dibenzofurans						
	Di-	Tri-	Tetra-	Penta-	Hexa-	Hepta-	Total
(1) 1016	ND	ND	<0.001	<0.001	—	<0.001	—
(1) 1248	”	”	0.5	1.2	0.3	—	2.0
(1) 1254	”	”	0.1	0.2	1.4	—	1.7
(1) 1254	”	”	0.2	0.4	0.9	—	1.5
(1) 1260	”	”	0.1	0.4	0.5	—	1.00.8
(1) 1260	”	”	0.2	0.3	0.3	—	8.4
(2) A-60	”	”	1.4	5.0	2.2	—	13.6
(3) DP-6	”	”	0.7	10.0	2.9	—	17–18
(4) K-400	”	”	—	—	—	—	—

(1) Aroclor; (2) Clophen; (3) Phenoclor; (4) kanechlor. Data from Bowes *et al.*[2] and Nagayama *et al.*[3]

48% by weight chlorine substituted biphenyl. The fluid was used by the rice oil manufacturer as a fire hazard reduction measure in an indoor heat exchanger.

The PCB concentration in Yusho oil was found to be about 1000 ppm PCB (Kuratsune *et al.*[4]). As analytical capabilities increased there has been a parallel increase in the speculation that the causative agents found in the rice oil such as polychlorinated dibenzofurans (PCDFs) and polychlorinated quaterphenyls (PCQs) (Masuda and Yoshimura[5]) had much greater relative toxicities. Original estimates of the PCB contamination were in the range of2000–3000 ppm PCB but were found to be in error primarily because of the analytical interference presented by PCQs.

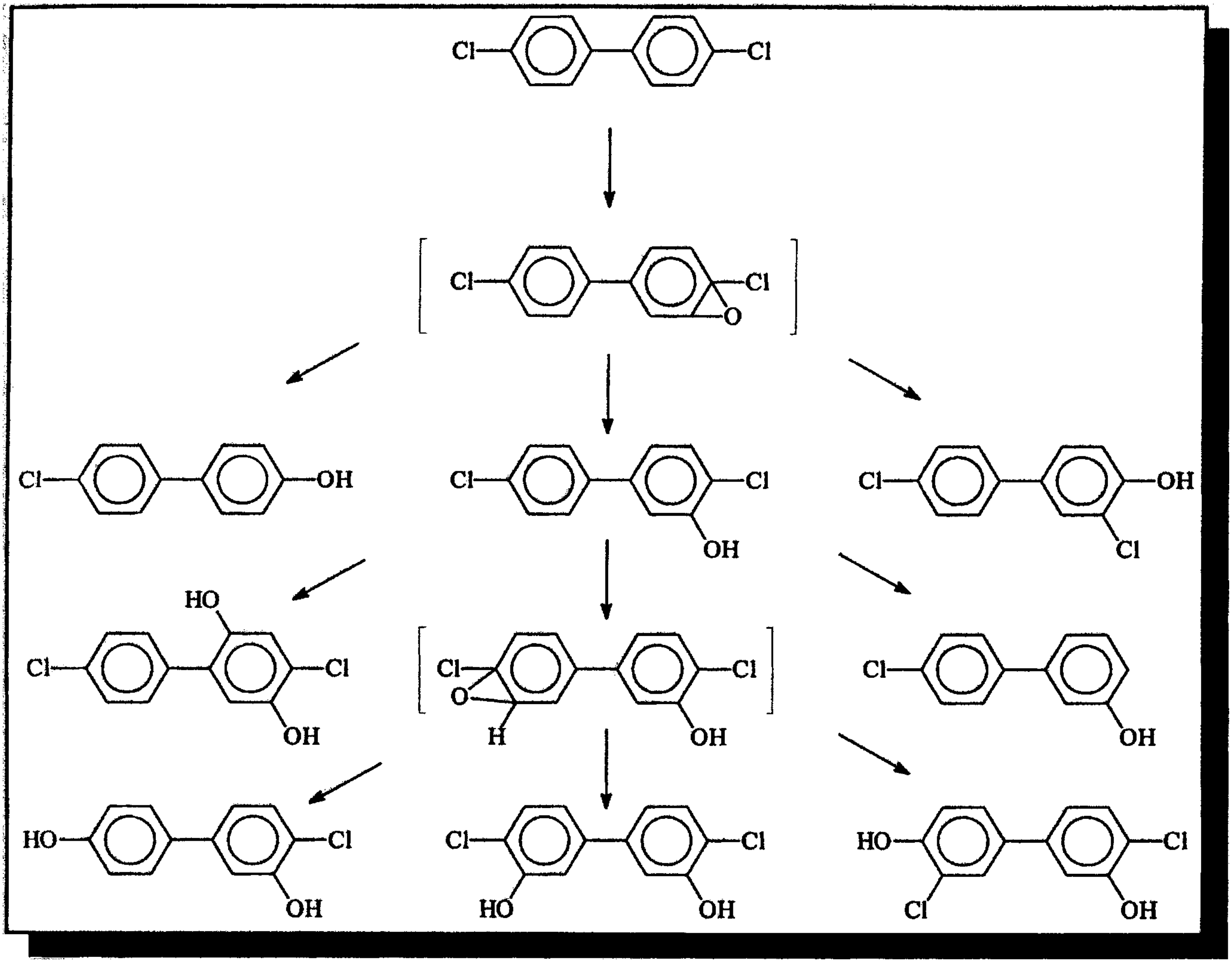


FIGURE 16 PCB metabolism in birds and mammals.

A poisoning very similar to the Japanese Yusho incident occurred in Taiwan in 1978–1979 and is known as the "Yu-Cheng" incident. In this case also, a leaking heat exchanger caused the contamination of rice oil during the manufacturing process with a mixture of Kanechlor 450 (48% Cl) and Kanechlor 500 (54% Cl). 1,843 cases of poisonings were reported during the period 1978–1979 (Masuda and Yoshimura;[5] Kunita *et al.*;[6] Chen *et al.*[7]). The PCB concentration in the oil was less in this incident than in Yusho but the higher dose rates to the victims resulted in approximately the same total dose.

Hsu *et al.*[8] have estimated that the minimum effective dose of PCBs in the Yu-Cheng incident was 0.3 to 0.5 g. The ratio of PCDFs to PCBs in the Yu-Cheng rice oil was found by Chen *et al.*[7] and Kunita *et al.*[6] to be approximately 1 : 300 and therefore the minimum effective total dose for PCDFs is of the order of 1 mg.

The assessment of potential toxicity is exacerbated by the fact that not all PCB, PCDF, PCDD, and so on, isomers are equally toxic. It is therefore not sufficient to analyze for the presence of a class of compounds but also necessary to quantitate the biologically active isomers within a class.

There are 209 PCB isomers contained within the 10 homologous series of PCBs, see Table 10. The series result from the number of chlorine atoms which substitute hydrogen on the biphenyl rings. A commercial arocior mixture may contain 60 or more different isomers.

PCBs are absorbed in the fatty tissues of humans. The ultimate fate of the PCB isomers depends upon their pharmacokinetics so that the analysis of isomers remaining in a tissue may show little resemblance to the isomers originally ingested. Less chlorinated PCBs tend to be metabolized more readily than their higher chlorinated congeners and therefore arociors such as aroclor 1260 or 1254 tend to bioconcentrate more than aroclor 1242.

There have been several studies which have concluded that past exposure to PCBs is associated with excess mortality or morbidity from cancer in human populations (Bertazzi

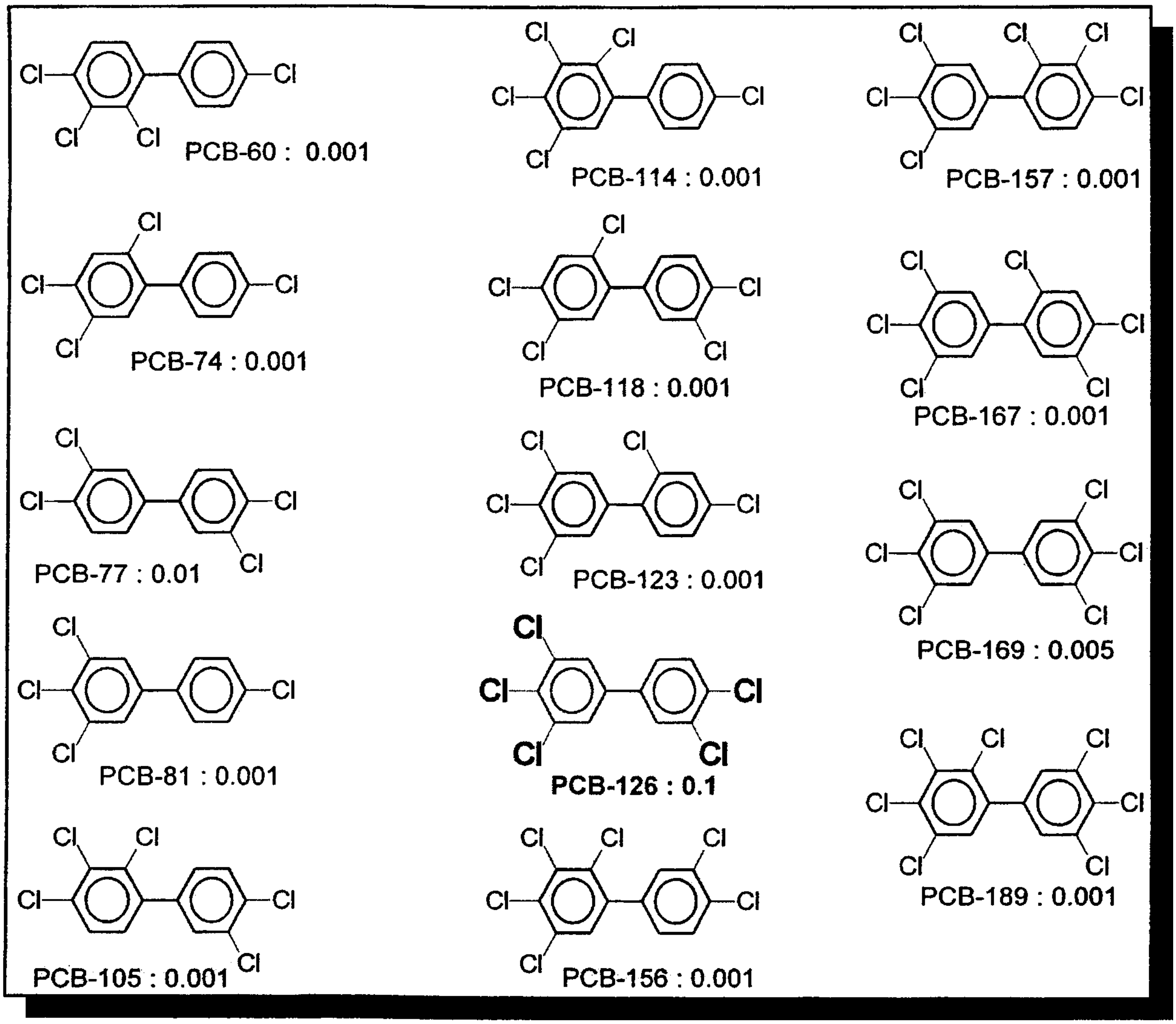


FIGURE 17 The dioxin-like isomers of PCBs.

et al.;[8] Humphrey;[10] Unger *et al.*[11]). However, the health effects in occupationally exposed workers are typically complicated by such factors as incomplete exposure information, job turnover and low chemical specificity.

The most prominent and consistent health effect associated with exposure to PCBs are chloracne and dermatitis. These effects usually clear up quickly and, if they were the only effects, would reduce the need for concern. It is beyond the scope of this paper to investigate all the health effects associated with PCBs but there is little doubt that PCBs are compounds of concern. The concentration at which this occurs is much higher than that of many other classes of chlorinated aromatic compounds. The risk associated with exposure to PCBs can be separated into a consideration of the biological activity of PCB isomers contained in the PCB mixture and the biological activity and probability of formation of degradation products and oxidation products.

Estimates of the dosage of toxic compounds received by victims of the Yusho and Yu-Cheng poisonings were somewhat dependent upon the chemical specificity of the analytical methods used. There is no doubt that the PCB involved had under gone partial oxidation and consequently, the overall toxicity of the fluid was the combined effect of compounds of concern.

In 1970 Vos *et al.*[12] showed a correlation between the toxic effects of European PCBs and the concentration levels of PCDFs. The relative concentrations of the PCDF isomers present in Yusho oil and in two samples of other used heat exchanger PCBs (Kanechlor KC400 and Mitsubishi-Monsanto T1248) were found to be strikingly similar. This fact underscores the work of Kunita *et al.*[6] who have concluded that not only are PCDFs orders of magnitude more toxic than PCBs but claimed also that there is a synergistic action between them. The evidence suggests that PCDFs were a major causative agent in the Yusho and Yu-Cheng incidents.

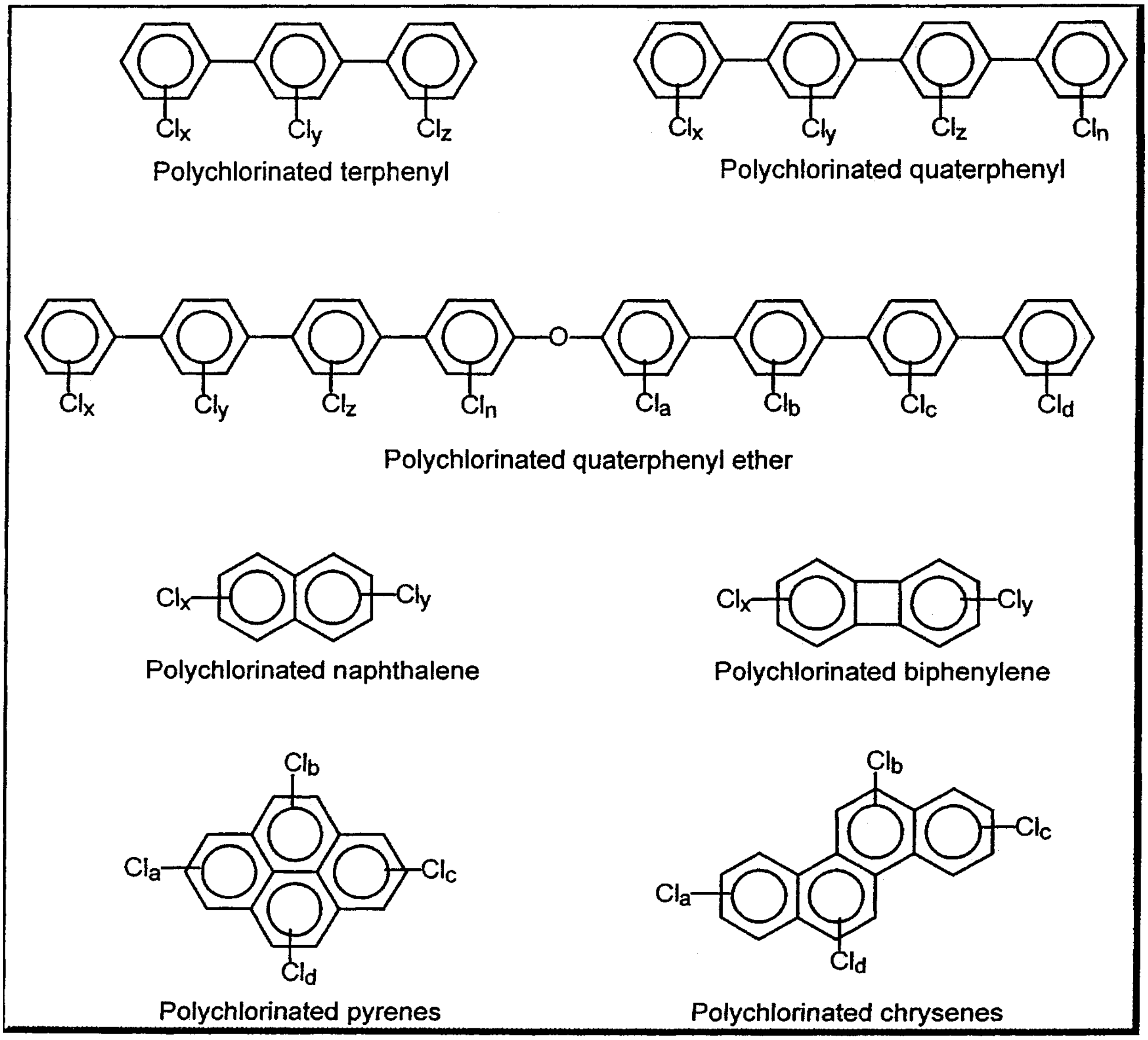


FIGURE 18 Compounds of concern.

It is not yet possible to estimate the sensitivity of humans to the toxic effects of PCDDs. It has been suggested that humans are resistant to the toxic effects of PCDDs. The supposition is not presently verifiable because of the lack of characterization of human exposure. There is no doubt, however, that the 2,3,7,8-tetrachlorodibenzo-dioxin (2,3,7,8-TCDD) isomer is a potent toxin which has caused impaired reproduction and/or birth defects in every species tested and has been described as the most toxic compound known to man.

In addition to PCDFs and PCDDs mentioned above there are several classes of polychlorinated aromatic hydrocarbons which may be contained as impurities or degradation products of PCBs. These classes include polychlorinated terphenyls (PCTs), polychlorinated quaterphenyls (PCQs), polychlorinated quaterphenyl ethers (PCQEs), polychlorinated naphthalenes (PCNs), polychlorinated biphenylenes (PCPs), polychlorinated pyrenes (PCPYs) and polychlorinated chrysenes (PCCYs).

Polychlorinated terphenyls have been found (Allen *et al.*[13]) to have qualitatively similar enzyme inducing effects to those of PCBs but appear to be less potent (Goldstein[14]). Mixtures of PCBs and PCTs have been marketed but were not widely used.

Polychlorinated quaterphenyls were found as contaminants of Yusho and Yu-Cheng oils but for a long time were an unrecognized interference in the quantitation of PCB contamination in these incidents. Studies by Hori *et al.*[15] and Kunita *et al.*[6] showed that the toxic effects of PCQs are similar to PCBs.

Polychlorinated quaterphenyl ethers (PCQEs) have been identified in Yusho and Yu-Cheng oils. This class of compounds does not seem to have been tested for toxicity but structure/activity relationships do not predict that they should be particularly toxic. Nevertheless, cyclization to a phenyl substituted dibenzofuran possibly could yield a toxic product. The likelihood of the cyclization reaction provides a measure for the concern to be accorded the presence of ether precursors.

Polychlorinated naphthalenes (PCNs) have been characterized in commercial PCBs and probably resulted from the presence of naphthalene as an impurity in the biphenyl raw material used for the manufacturing process. Typical concentrations in PCB fluids are less than 600 ppm. Results obtained by Kimbrough[16] and Goldstein,[14] indicated that it is unlikely that PCNs would add significantly to the toxicity of PCBs at the concentration levels typically observed in commercial mixtures, although the stereochemical requirement of planarity to include enzyme activity is fulfilled.

Polychlorinated biphenylenes (PCPs), because of their structure, are expected to be as toxic as the correspondingly substituted PCDDs. According to Poland and Glover,[17] 2,3,6,7-tetrachlorobiphenylene is approximately as potent as 2,3,7,8-TCDD. PCPs have been found to result from the reductive conditions which occur in the early stages of askarel transformer fires and are therefore compounds of concern (Smith *et al.;*[18] Rappe *et al.;*[19] Williams *et al.*[20]) (Figure 18).

Polychlorinated pyrenes (PCPYs) and polychlorinated chrysenes (PCCYs) have been identified as components of the soot from an askarel transformer fire (Rappe *et al.*[21] Williams *et al.*[20]). The toxicities of these classes of compounds have not been determined but their concentrations are so low that they probably do not add significantly to the overall toxicity of the degraded PCB.

THE PERCEPTION OF PCB HEALTH EFFECTS: THE YUSHO AND YU-CHENG POISONINGS

As was mentioned above, a large part of what is currently known about the effects of human exposure to PCBs is derived from two incidents of poisoning which occurred in Japan in 1968 and in Taiwan in 1978. Both episodes involved the contamination of rice oil used for cooking purposes, both involved approximately the same number of people and both involved the same total dose of PCBs (Table 14).

The poisoning in Japan in 1968 became the subject of a special issue of the journal Fukuoka Acta Medica in June 1969.[27] Many of the papers which have been written on various aspects of the poisonings have been published in series and there are now over 200 of them. Most of these papers are in Japanese with English abstracts but there have also been reviews written in English.[27–37]

About 325 patients with acneiform eruptions had been treated in northern Japan between the summer of 1968 and January 1969. The apparent cause of the poisoning was traced to the consumption of a particular brand of rice bran oil.

By May 1970 the total number of cases had risen to 1,291. The common symptoms included acneiform eruptions, hyperpigmentation of the skin, nails and mucous membranes, swelling of the upper eyelids, and hyperemia of the conjunctivae.

PCB concentration in the rice oil varied with the production batch. The highest concentration was found to be 3,000 ppm PCB in oil shipped on February 5, 1968 and only traces were found in oils produced after February 19, 1968. The oil was consumed between March and October 1968 and the first known clinical examination of a Yusho patient occurred on June 7, 1968. The onset of Yusho symptoms occurred in the majority of patients in June, July and August of 1968. The total quantity of contaminated rice oil ingested by patients as well as the dose rate of ingestion of PCBs has been estimated by several workers (T. Yoshimura *et al.*[38]) (Table 15).

A typical quantity of oil consumed was 800–1200 mL containing about 2g of PCB. For victims who had ingested more than 720 mL of oil the attack rate of Yusho symptoms was 100%. The PCB content of Yusho oil was about 1,000 ppm PCB as Kanechlor 400 (J. Nagayama *et al.*[39]). Clinical features of Yusho poisoning are listed below in Table 16.

Undegraded Kanechlor 400 has been found to contain about 18 ppm PCDFs while the Yusho oil contained 5 ppm PCDFs and 1,000 ppm PCBs. In addition, polychlorinated quaterphenyls, PCQs, were found in the oil at concentrations similar to the PCBs. The presence of these compounds initially caused the estimates of PCB contamination to be stated at the 2000–3000 ppm PCB level. Table (14) shows the concentrations of PCDFs and PCBs in samples of Yusho oil.

The presence of PCDFs in the Kanechlor 400 mixture is of great concern because if PCBs are not the major toxic constituents of the mixture then either the ingested rice oil contained contaminated PCBs or degraded PCBs. Several authors have investigated the presence of PCDFs and other toxic contaminants in commercial products prior to use. Table 14 indicates the PCDF concentration in a number of different rice oil products compared with the analysis of a sample of the type used in the heat exchanger. Table 13 indicates the levels of PCDFs in different Aroclors as well as German and Japanese products. The data do not support the hypothesis that the PCDFs were present in the Kanechlor 400 as manufactured because of their high relative concentrations in Yusho oils. Instead, chemical degradation of the PCBs must have occurred, albeit in extremely low yield, to produce degradation products of concern at concentrations of concern. The production of PCDFs is markedly affected by factors such as temperature, time, the presence of oxygen and catalytic metals such as copper or iron. Miyata and Kashimoto[41] have demonstrated that steel acts as an effective catalyst when Kanechlor 400 is heated to 360°C for 20 days. The effect of different catalysts is shown in Table 15 (Table 17).

It would be unusual to find a commercial PCB in an application which would involve such high temperatures unless the system was faulting in some way. The kinetics of chemical reactions, as a rule of thumb, increase by a factor of 2 for every 10°C rise in temperature. In other words, the quantity of PCDFs produced in twenty days at 350°C would be expected to be about half the amount formed at 360°C under

TABLE 14
The most highly PCB contaminated Yusho oils also contained the highest relative concentrations of PCDFs

Sample	PCB Concentration (ppm)	PCDF Concentration (ppm)	PCDF:PCB ratio
Kanechlor 400	1,000,000	33	1:30,000
Kanemi-1	154	2.2	1:70
Kanemi-2	0.9	0.0021	1:492
Kanemi-3	1,029	5.3	1:194
Kanemi-4	1,025	4.9	1:209
Kanemi-5	1,023	5.4	1:189
Kanemi-6	2.7	0.0031	1:871
Kanemi-7	2.7	0.0021	1:1,286

From reference 40: Miyata, H., T. Kashimoto and N. Kunita; “Detection and determination of polychlorodibenzofurans in normal human tissues and Kanemi rice oils caused Kanemi Yusho”; *J. Food Hyg. Soc. Japan;* **18**, 260 (1977).

TABLE 15
The total concentration of PCDFs increase by a factor of up to 6.4 in the presence of catalysts

Temperature 0°	Duration (days)	Catalyst	Total PCDFs (mg/kg PCBs)	Total TCDFs (mg/kg PCBs)
Ambient	0	none	33	26.2
360	20	none	65	52.1
360	20	steel	77	32.7
360	20	water	212.1	152.4
360	20	steel + water	72.1	46.6

Miyata H. and T. Kashimoto.[4]

similar conditions. Only a very small fraction of the PCDFs produced at 360°C would be expected to be formed at 250°C. The presence of PCDFs in heat exchanger fluids operated in the range 200°–270°C was investigated by Miyata and his results are shown in Table (16). The PCDF concentration in the fluids increased by a factor of up to 15.5.

Nagayama *et al.*[42] have confirmed the hypothesis that PCDFs as well as PCQs are produced in the degradation process within heat exchangers. They found that the concentration of PCDFs increased by as much as 40 to 50 times the original level i.e., up to about 740 ppm PCDF by heating the PCB mixture to 300°C for 7 weeks followed by heating to 350°C for 4 weeks in the presence of iron.

The concentration of PCDFs in service aged transformer askarels has been shown to occur by Chittim *et al.*[43] It was found that a correlation existed between the concentration of tetrachlorodibenzofurans (T_4CDFs) and the length of time the fluid had been in service. The small number of samples tested was insufficient to obtain a definite correlation between TCDF concentration and transformer load. While the results indicated that transformer load is probably a major factor in the formation of PCDFs, the effects of discharging or arcing appeared to be negligible.

It has been observed that some PCB products were manufactured with trace PCDF contamination caused by the presence of dibenzofuran in the biphenyl starting material (see Table 12). Fortunately, North American starting materials seem to have been purer than those available in Japan and very small amounts of PCDFs are found in undergraded US PCB products. An alternative route of formation might be derived from the use of sodium hydroxide to remove traces of HCl produced during the PCB synthesis. Miyata and Kashimoto[41] found that in the presence of catalytic metals, such as copper and iron found in heat exchangers and transformers, there is up to a 15.5 fold increase in the PCDF concentration in a temperature range of 180 to 270°C (Tables 18 and 19).

Brown *et al.*[45] have also investigated the levels of bioactive PCDF congeners in PCB dielectric fluids obtained from capacitors and transformers. Both service aged and unused equipment was investigated and no indication was found of a higher level of PCDFs caused by PCB degradation. None of the equipment had malfunctioned and failed due to overheating but had all either never been used, or had been operated for a documented period of up to 30 years. The distinction between operating and failed units was not made, nor was

TABLE 16

Percent distribution of symptoms of Yusho reported by 189 patients examined before October 31, 1968

Symptoms	Males N = 89	Females N = 100
Dark brown pigmentation of nails	83.1	75.0
Distinctive hair follicles	64.0	56.0
Increased sweating at palms	50.6	55.0
Acne-like skin eruptions	87.6	82.0
Red plaques on limbs	20.2	16.0
Itching	42.7	52.0
Pigmentation of skin	75.3	72.0
Swelling of limbs	20.2	41.0
Stiffened soles in feet and palms of hands	24.7	29.0
Pigmented mucous membrane	56.2	47.0
Increased eye discharge	88.8	83.0
Hyperemia of conjunctiva	70.8	71.0
Transient visual disturbance	56.2	55.0
Jaundice	11.2	11.0
Swelling of upper eyelids	71.9	74.0
Feeling of weakness	58.4	52.0
Numbness in limbs	32.6	39.0
Fever	16.9	19.0
Hearing difficulties	18.0	19.0
Spasm of limbs	7.9	8.0
Headache	30.3	39.0
Vomiting	23.6	28.0
Diarrhea	19.1	17.0

Source: Kuratsune *et al.,* 1972.

TABLE 17

The increase in PCDF concentration in service aged Kanechlor 400 heat exchanger fluids is greater even than the increase found in high temperature catalyzed fluids

In Service Age (yr.)	Temperature (°C)	PCBs (ppm)	PCQs* (ppm)	PCDFs (ppm)
14	180–270	968,400	31,000	510
3	200–220	999,000	690	277
0‡		999,800	209	33

* PCB = polychlorinated quaterphenyl.

‡ = not used.

there sufficient data to observe any possible trends which may exist between PCDF formation and temperature/load/dissolved oxygen conditions. It can only be surmised that the onset of a hot-spot which is severe enough to cause degradation of cellulosic or Class A dielectric insulation would result in the failure of the equipment in a time which would be too short to allow the formation of PCDFs.

An EPRI report on the degradation of PCBs under electrical discharge has shown that PCDFs are not formed.[46] The PCB components of Yusho oils were primarily tetra- and pentaCBs with lesser quantities of tri-, hexa- and hepta-CBs and small quantities of octa- and nonCBs. In contrast to the predominance of tetra- and pentaCBs in the rice oil, Yusho patients showed predominantly hexa- and hepta-CBs in their blood which indicates that a selective metabolism of the lower CBs had occurred over a period of about 7 years. This is a very significant finding when considering the relative toxicity of different PCB formulations. In the case of Aroclor 1016, for example, the predominant components are mono- and di-CBs with $<0.1\%$ of hexa-CBs. Aroclor 1016, therefore, would not be expected to bioaccumulate to the same extent as a PCB mixture containing a larger percentage of chlorine such as Aroclor 1254.

PCDFs appear to have a different distribution in the body from that of PCBs. There tends to be accumulation in the liver. About forty PCDF congeners have been identified in Yusho oil but, of these, only 10 were found above the detection limits in the blood of patients. Two tetra-, two penta- and two hexa-CDFs were identified.[47] The distribution of PCQs among blood, liver and adipose tissue of Yusho patients has been found to be similar to that of PCBs although the rate of elimination of PCQs appears to be similar to, or even a little faster than that of PCBs.[48]

TABLE 18
The relative ratios of toxic contaminants in Yusho and Yu-Cheng oils are constant

Type of oil	Concentrations in Oil (ppm)			
	PCBs	PCQs	PCDFs	PCDF : PCB
Yu Cheng:				
Chen *et al.* (1984)	74	40	0.27	1:274
Kunita *et al.* (1984)	44	17	0.14	1:314
Masuda *et al.* (1982)	80	135	0.09	1:888
Yusho:				
Kunjita *et al.* (1984)	430	630	3.8	1:113
Nagayama *et al.* (1976)	1,000	—	5	1:200
Masuda *et al.* (1982)	900	800	2	1:400

P.H.-S Chen *et al.*[49]

TABLE 19
The most likely blood PCB level in Yu-Cheng victims was in the range 51–100 ppb

Blood level (ppb)	Number	Percent
<10	29	5
11–50	143	23
51–100	272	44
101–150	91	15
151–200	28	5
201–300	24	4
301–400	17	3
401–500	4	1
>500	5	1

Cull and Dobbs[44] analysed two service aged PCB transformers but did not find an unusually high concentration of PCDFs.

A similar incident to the Yusho poisoning occurred in Taiwan in 1978–1979 and is now known as the "YuCheng" incident. There are many coincidental similarities between the two poisonings. Approximately 2,000 victims were identified in the Yu-Cheng incident compared with a total of about 1,800 Yusho victims. The sources of poisoning, the toxic constituents and their health effects are very similar.

The source of the toxic constituents in the Yu-Cheng incident was derived from Kanechlor 400 and 500 used in a heat exchanger for the production of rice oil. The concentration of polychlorinated aromatic hydrocarbons in the oil was lower in Yu-Cheng than in Yusho but because the rate of ingestion was higher in Yu-Cheng the total doses were approximately the same. The amount of PCBs ingested by victims in the two incidents was about 0.8 to 1.8 g.

Table 14 shows the distribution of toxic constituents contained in the Yusho and Yu-Cheng oils and from the ratios of PCBs: PCDFs it is possible to calculate the minimum effective dose of PCDFs about 1 mg. The blood PCB levels found in Yu-Cheng victims measured within the first year of exposure range between 3 and 1,156 ppb although 44% of the cases had 51–100 ppb in their blood as shown in Table 19.

The relative rates of elimination of PCBS, PCQs and PCDFs from the blood has been established to be PCQ > PCDF > PCB although the differences between them are small.[6] The prominent clinical features of Yu-Cheng have been correlated with exposure to Yu-Cheng oil as well as to estimates of intake and blood levels of PCBs. Features include pigmentation of the skin, gingiva and nails, chloracne, meibomian gland enlargement, and eye discharge. Subjective complaints against Yu-Cheng patients include increased eye discharge, edema of eyelids, vision disturbances, general malaise, numbness of joints, headache and dizziness, pain in joints and muscles and other less frequently occurring ailments.

Several investigators have concluded that PCDFs were the primary causative agents in both Yusho and Yu-Cheng poisonings. Toxicity studies with PCQ mixtures similar to those in Yusho oil suggest that PCQs are less toxic in rats and cynomolgus monkeys than either PCBs or PCDFs.[6]

A blood concentration of 40 ppb in Yu-Cheng victims was found to be a level above which severe skin and ocular effects became frequent. This is in contrast with the blood levels of PCB exposed populations who have as much as 300 ppb in their blood without there being evidence of severe health effects. This observation points to the relative effect of other constituents in the Yusho and Yu-Cheng oils as well as the possibility of PCB isomer differences. For example, highly toxic PCB isomers such as 3,3',4,4'-tetraCB are 2,3,3',4,4',5-hexa-CB were more prominent in the tissues of Yusho and Yu-Cheng victims than are normally found in samples of Kanechlor 400.[52,53]

Blood levels of PCBs in Yusho and Yu-Cheng victims retuned to background levels over a period of about 7 years but the signs and symptoms of Yusho have persisted for over 25 years. The above data suggest that even though the ratio of PCBs: PCDFs in both Yusho and Yu-Cheng poisoning incidents were about 200:1 the greater toxicity of PCDFs seem to be more than sufficient to offset the difference. No single line of evidence is conclusive but, taken together, they do provide a compelling argument for the conclusion that PCDFs were the primary causative agents.

THE FORMATION OF COMPOUNDS OF CONCERN

Compounds of concern can be contained in PCB fluids for a variety of reasons. For example, they may be formed from impurities in the feedstock used to manufacture PCBs or from the cyclization of PCBs induced by heat, or even from phenolic or ether precursors.

Polychlorinated terphenyls, quaterphenyls and napthalenes could result from feedstock contamination by traces of the aromatic hydrocarbons before the mixture was chlorinated. PCNs have been identified as a pyrolysis product from an askarel transformer fire. Their formation has been explained (Buser[54]) in a laboratory study by invoking the formation of benzyne intermediates or the re-arrangement of intermediates formed between an ortho-chloro phenyl radical with a chlorobenzen (Figure 19).

The overall effect of radical reactions on the product distribution during askarel degradation will be effected by both temperature and the availability of oxygen. Chittim *et al.*[43] have analyzed unused North American askarels and found that they contained less than 0.05 ppm of PCDFs. It is therefore of concern that these workers found a correlation between the concentration of TCDFs in used askarel with the length of time the fluid had been in service. The degradation of PCBs in electrical equipment is discussed above in the section on the perception of PCB health effects.

The Power Transformer Guide ANSI-C57.92-1981 lists a maximum top oil limit of 110°C which may occur when the maximum hot spot conductor temperature is at 180°C due to short term loading. The Distribution Transformer Guide ANSI C57.91-1974 lists a maximum top oil temperature of 120°C which may occur at maximum conductor hot spot temperature of 200°C. Maximum fluid temperatures in a transformer could exceed values in the overload guide if units are subjected to more severe loads. Units with extreme overloads would be expected to overheat and fail. The most probable cause of eventful transformer failure is violent rupture (Doble 1982) and the quantities of any compounds of concern which may be produced is determined by the availability of oxygen in an oxygen depleted, high temperature reaction zone of short duration (Figure 20).

Buser[54] has shown that pyrolysis of chlorobenzenes at 600°C in the presence of excess air yielded about 1% of tetra- to octa-CDFs and tetra- to octa-CDDs. It has also been shown (Buser and Rappe[55]) that the pyrolysis of PCBs occurs intramolecularly by four alternative reaction routes to yield different isomeric PCDF products.

The thermal degradation of PCBs can result in a complex set of reactions which may produce compounds of concern under uncontrolled conditions. The mechanism of the formation of dioxins under pyrolysis conditions is discussed later in this section. Temperature and residence time relationships have been extensively studied to establish the conditions necessary for satisfactory destruction (Mescher *et al.*[56]). It was found, for example, that with a Is residence time, most PCB decomposition occurred in a temperature range between 640°C and 740°C. Commercial incineration equipment for the destruction of PCBs must therefore be designed so that the energy input to disrupt the molecule is made available either by supplying a very high temperature or a satisfactorily long residence time. Several types of incinerator are found useful for this purpose and include rotary kilns, high temperature fluid wall reactors, plasma pyrolysis units, circulating bed combustors and so on.

Uncontrolled reactions which lead to the formation of chlorophenols as a side reaction product have the potential to generate PCDFs and/or PCDDs in the following ways:

1) The dimerization of chlorophenates;
2) The cyclization of polychlorinated diphenyl ethers (PCDPEs);
3) The cyclization of polychlorinated phenoxy phenols.

The Seveso accident occurred in 1976 and was one of six similar occurrences in which uncontrolled reaction conditions caused the dimerization of trichlorphenol produced from tetrachlorobenzene and sodium hydroxide in ethylene glycol to yield PCDDs. The reactions are shown in an early section in explanation of the chemistry observed in 1956. In the case of liquid systems, reactants are retained in the reaction zone for periods of time which are long compared to the time required for the formation of product. In a gaseous reaction, on the other hand, such as in a flame, the reaction zone is relatively short-lived and the yield of product is therefore less. The importance of even statistically unlikely reactions lies in the toxicity of the products. The pyrolysis of PCDPEs follows two competitive reaction pathways viz., reductive dechlorination or ring closure to PCDFs.

The cyclization of polychlorinated phenoxy phenois is also a bimolecular reaction which yields PCDDs under the influence of heat. The above discussion has considered which compounds are of concern and how they may be generated in uncontrolled reactions. There have been many methods developed for the PCB decontamination of materials and these will be discussed more fully later. Incineration methods have already been mentioned. In general, dechlorination reactions of PCBs involve an initial addition of an electron to the aromatic molecule. The end result of the reaction should be that there is no remaining organically bound chlorine. In that case none of the chlorinated toxic compound classes discussed above can be present in the reaction residue. However, depending upon the chemistry involved and the stoichiometric excess of reagent, toxic intermediates can be produced which, while satisfying the requirements for the removal of PCBs, can produce a severe problem where only a relatively slight hazard existed before the process was applied.

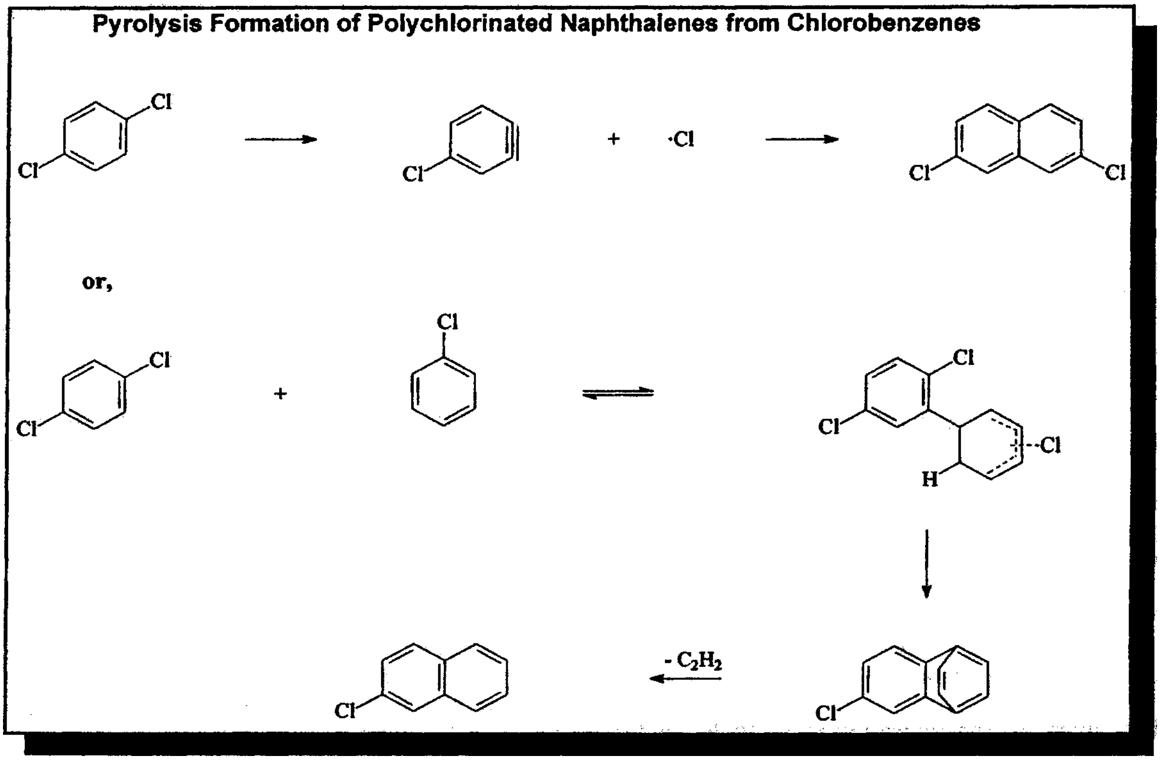


FIGURE 19 Benzyne or other free radical formation of polychlorinated naphthalenes.

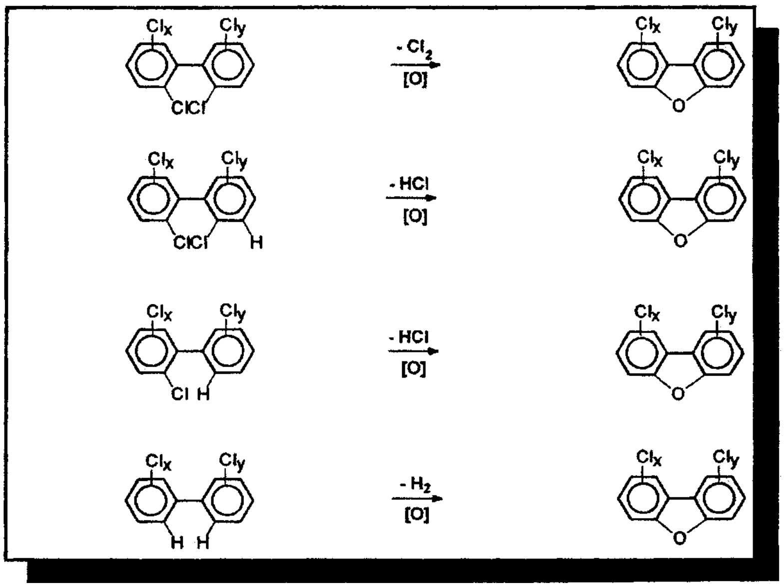


FIGURE 20 Pyrolysis formation of PCDFs from PCBs.

From a chemical point of view there is little difference between whether the electron is derived from an alkali metal or an organometallic reagent or from radiation induced electrons. The gamma-ray radiolysis of halogenated benzenes was investigated by Choi and Willard in 1962[57] using a high specific activity Co^{60} source. Aria *et al.*[58] were the first to apply the method to PCB solutions and in 1974 their work was extended by Sawai *et al.*[59]

Sawai *et al.* used Co^{60} gamma-rays to induce electrons in deoxygenated solutions of PCBs in alkaline isopropanol. They found that the presence of small amounts of electron acceptors resulted in an inhibition of dechlorination. Also, the formation of chloride ion and the consumption of hydroxide ion were approximately equal at each dose. The results paralleled the study by Webber 1983 in which transformer askarels were irradiated in dielectric insulating oil with Sr^{90} β-particles. It was found that the dechlorination reaction did occur but that, in the absence of potassium hydroxide and isopropanol, the degradation required impracticably long irradiation times.

The reaction scheme which explains the experimental data is shown in Figure 21.

- The beta particle radiolysis of water produces hydroxyl radicals and solvated electrons, reaction (1).
- The solvated electrons react with PCBs to give a PCB radical anion, reaction (2).
- The anion stabilizes itself by eliminating chlorine as chloride to leave a PCB radical, reaction (3).
- In the presence of isopropanol the PCB radical can easily abstract hydrogen to become a PCB molecule with one less chlorine than at the beginning of the reaction sequence, reaction (4).
- The isopropanol radical formed in the hydrogen abstraction step reacts with the hydroxyl ions from the potassium hydroxide in the system to produce acetone radical ions, reaction (5).
- The acetone radicals become stabilized as acetone after interaction with PCBs to yield PCB radical anions, reaction (6).
- The reaction sequence continues in this way until total dechlorination has occurred.

Hydroxyl radicals are extremely reactive species and are produced in very small amounts relative to the isopropanol present. It is consequently more likely that reaction will take place with isopropanol than with a small amount of PCB or an even smaller concentration of PCB radicals. Also, the concentration of PCB radicals is so small relative to the macro amount of isopropanol present that coupling to produce a polychlorinated quaterphenyl is unlikely. Certainly mixtures of PCBs are destroyed by β-particle radiolysis most efficiently when the solution is free of oxygen and contains an alkaline hydrogen donor to allow free radical chain propagation. However, in the absence of isopropanol and

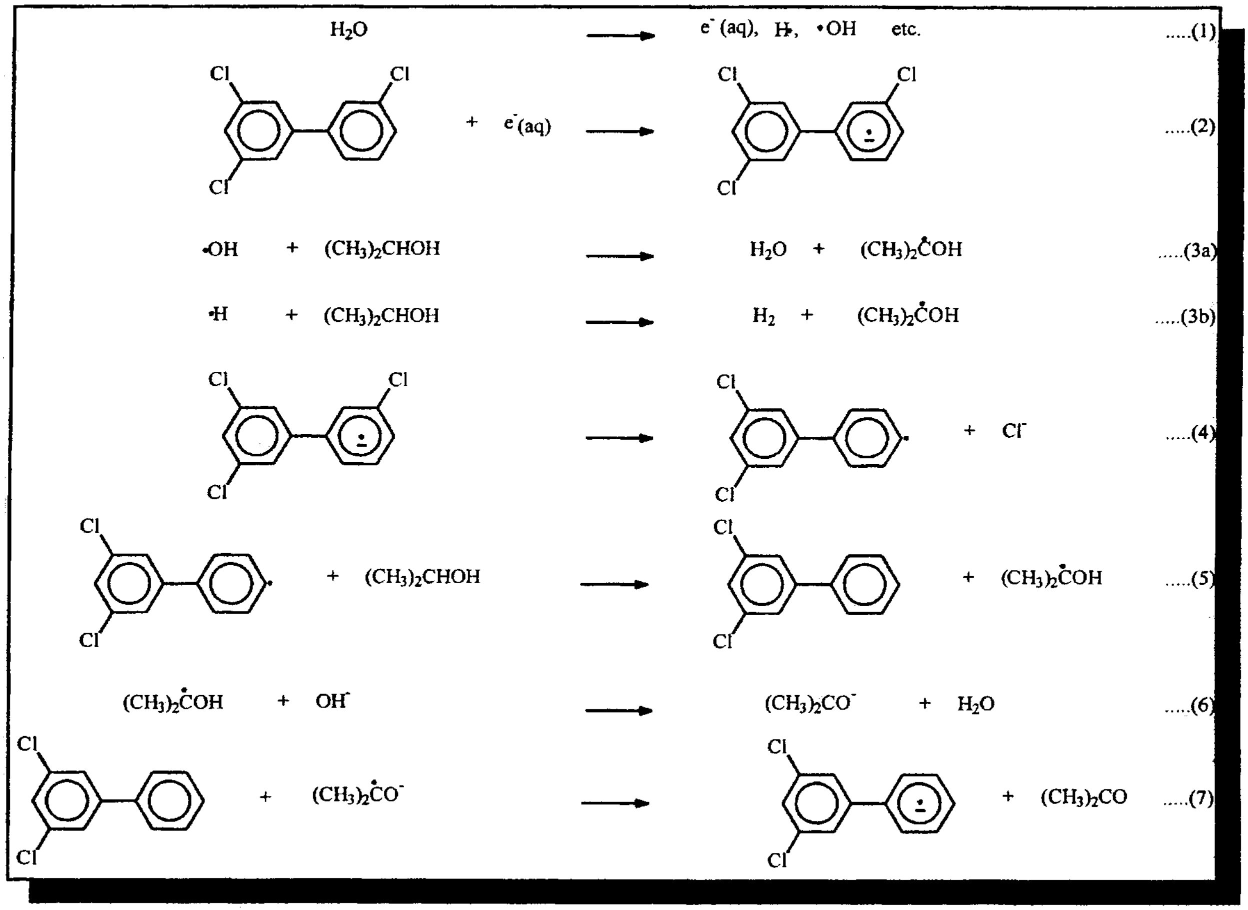


FIGURE 21 The radiolysis of PCBs with 2 MeV Sr^{90} β-particles.

hydroxyl anions the radiolysis reaction requires very much larger irradiation doses and hence very much longer reaction times. Under these circumstances the concentration of hydroxyl radicals which are produced is similar to the concentration of PCB radicals and hydroxylation occurs to give a polychlorinated phenol. Intramolecular cyclization of the hydroxylated intermediates can occur as discussed earlier to yield PCDDs.

As a method for the PCB decontamination of oils the process is impractical but it serves to illustrate how the reaction conditions may *induce* rather than *reduce* a potential hazard.

The analysis of polychlorinated non-oxygenated toxic compounds is made extremely difficult because of the lack of suitable standards. The presence of such classes of compounds as PCQs, PCBPs, PCPYs and PCCYs has nevertheless been established when askarel has been degraded under anaerobic conditions.

Sodium dispersions are effective in the dechlorination of PCBs and were first applied by Japanese workers in 1973 who used it to decontaminate kerosene extracts of PCB contaminated sewage sludge. The method was applied in the decontamination of electrical insulating oils by Webber *et al.,*[60] who found that even high concentrations of askarels would react to give a non-toxic polyphenyl sludge. The mechanism was investigated by Pilgrim and Webber[61] and Webber and Wilson.[62] It was postulated that chlorinated biphenyl radical anions are produced by interaction of the PCB molecule in solution with metallic sodium in suspension. The radical anion eliminates chlorine as chloride to form a chlorobiphenyl radical which, in turn, abstracts available hydrogen from oil components to yield a partially dechlorinated PCB. The reaction continues in the presence of a stoichiometric excess of sodium until total dechlorination is achieved. The reaction sequence is discussed in more detail in the section below on PCB disposal methods.

In the absence of a readily available quantity of abstractable hydrogen the biphenyl radicals tend to couple and form quaterphenyls and higher polyphenyls. This reaction is relatively less likely to occur that the dechlorination reaction when an excess of sodium is present and consequently the quaterphenyls which are formed either have few remaining chlorine atoms on the rings or are completely dechlorinated.

Quaterphenyls and hexaphenyls are slightly soluble in mineral insulating oil and become totally dechlorinated. The sludge is predominantly hexa- and polyphenyl compounds. Work done by Brown and Webber[62] on the NMR of polymerized biphenyls has indicated that electrical conductivity through the polymer chains is dependent upon the extent of electron orbital overlap between the ring constituents of the polymer. Thus, when a small excess of alkali metal reagent is maintained in contact with insulating oil for a long time, in order to minimize the use of reagent for example, polyphenyls are produced which are dissolved in the oil. The power factor of the processed oil then increases dramatically to $>10\%$ (ASTM D924) compared with that of useful oil of $<0.1\%$. An increase in power factor corresponds to an increase in the concentration of polar constituents in the oil.

The concentration of polyphenyls in the oil and the extent of their chlorination was not investigated, but, since PCQs are not expected to be more toxic than PCBs, and because they are only produced in trace quantities, the practical limitation to their presence lies in the power factor of the product oil rather than their toxicity.

In 1970 Vos 79 (Vos *et al.,* 1970) showed a correlation between the toxic effects of European PCBs and the concentration levels of PCDFs. The major PCDF components contained in Yusho oil were the highly toxic 2,3,7,8-tetra-CDF and 2,3,4,7,8-penta-CDF. The relative concentrations of the PCDF isomers present in Yusho oil and in two samples of used heat exchanger PCBs (Kanechlor KC 400 and Mitsubishi-Monsanto T 1248) were found to be strikingly similar[74] (Nagayama *et al.,* 1976). This fact underscores the findings of other workers in this field that there is a connection between the toxicity of degraded PCB fluids, including tri-/tetrachlorinated benzene/PCB blends, and the concentration of PCDFs. The overall toxicity of the fluid may then be attributable, as Vos indicates, to the two particular PCDF isomers mentioned earlier.

In addition to the formation of PCDFs and PCDDs from PCBz/PCB fluids there are also several classes of polychlorinated aromatic hydrocarbons which may be contained as impurities or degradation products of PCBs. These classes are shown in Figure 22 and include:

- polychlorinated terphenyls (PCTs),
- polychlorinated quaterphenyls (PCQs),
- polychlorinated quaterphenyl ethers (PCQEs),
- polychlorinated naphthalenes (PCNs),
- polychlorinated biphenylenes (PCPs),
- polychlorinated pyrenes (PCPYs) and
- polychlorinated chrysenes (PCCYs).

Polychlorinated quaterphenyls were found as contaminants of Yusho and Yu-Cheng oils and for a long time were an unrecognized interference in the quantitation of PCB contamination in these incidents. Studies 88 (Hori *et al.,* 1982) show that the toxic effects of PCQs are similar to PCBs.

Polychlorinated quaterphenyl ethers (PCQEs) have been identified in Yusho and Yu-Cheng oils. This class of compounds does not seem to have been tested for toxicity but structure/activity relationships do not predict that they should be particularly toxic. Nevertheless, cyclization to a phenyl substituted dibenzofuran possibly could yield a toxic product. The likelihood of the cyclization reaction provides a measure for the concern to be accorded the presence of ether precursors.

Polychlorinated naphthalenes (PCNs) have been characterized in commercial PCBs and may have resulted from the presence of naphthalene as an impurity in the biphenyl raw material used in the manufacturing process. Typical concentrations in PCB fluids are less than 600 ppm. Results obtained by 89 (Kimbrough, 1981); 90 (Goldstein, 1980)

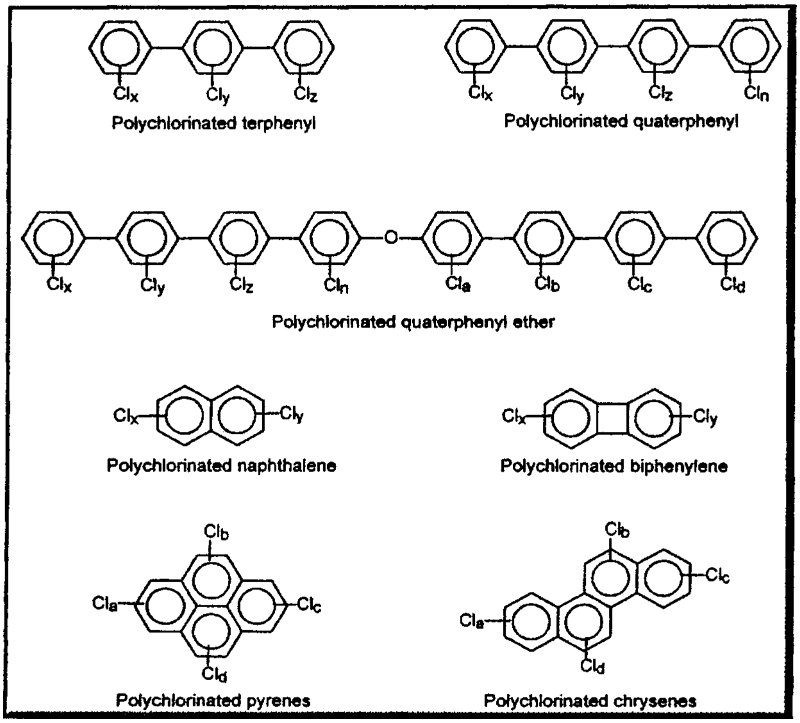


FIGURE 22 Compounds of concern.

indicated that it is unlikely that PCNs would add significantly to the toxicity of PCBs at the concentrations which are typically observed.

Polychlorinated biphenylenes (PCPs), because of their structure, are expected to be as toxic as the correspondingly substituted PCDDs. Accordingly to Poland and Glover (1977) 2,3,6,7-tetrachloro-biphenylene is approximately as potent as 2,3,7,8-TCDD. PCPs have been found to result from the reductive conditions which occur in the early stages of askarel transformer fires and are therefore compounds of concern [see Smith *et al.* (1982); Rappe *et al.* (1982); Rappe *et al.* (1983)].

Polychlorinated pyrenes (PCPYs) and polychlorinated chrysenes (PCCYs) have been identified as components of the soot from an askarel transformer fire (Rappe *et al.,* 1983).

Rubey[64] was studied the thermal decomposition of 2,3′,5,6′-pentaCB in different atmospheres and temperatures with a residence time of 2s. The decomposition temperatures range from about 700°–950°C. Similarly, a plot of PCB degradation in which the formation and decomposition of PCDFs is superimposed on the destruction of 2,3′,4,4′,5-pentaCB shows that the formation of PCDFs occurs at 700°–800°C and their yield increases up to about 900°C. Above 900°C there is rapid decomposition of both the PCB and the PCDFs.

Several physico-chemical factors influence incineration and are listed here to indicate the complexity of the process.

- exposure temperature
- the oxygen composition of flame and non-flame atmospheres
- gas phase residence times in different heated zones
- oxygen concentrations and associated gradients
- pressure
- flame contact time
- spatial and temporal variations in temperature.
- thermodynamic and kinetic properties of the compounds involved.

An extensive treatment of the quantitative prediction of the formation of PCDFs from PCBs under pyrolysis conditions is beyond the scope of this paper but a review of the literature does allow a semi-quantitative assessment of the likelihood of PCDF formation and decomposition in flames under a variety of conditions. In addition, thermodynamic factors will influence the enthalpies of activation and the frequency factors for decomposition of PCDFs depending upon their degree of chlorination (Tables 20, 21).

The chemical reactions and mechanisms which take place in combustion systems can be extremely complex and different systems of reactions, such as free radical decompositions, can occur simultaneously. The intractability of the mathematics used to model such systems depends to some degree upon the number and type of simplifying assumptions which are applied. For example, a steady-state approximation, in which the rate of formation and decomposition of a species is constant, can be applied in certain situations. Then, the equation which relates the rate of change of a species in its steady-state with the concentrations of the other species, can be solved in terms of the concentrations of the other species by setting the differential equal to zero.

TABLE 20
PCDFs in new and used commercial PCB mixtures

PCB	Polychloro dibenzo furans Cl^4	Cl^5	Cl^6	Total
New:				
Aroclors 1254, 1260(51)	—	—	—	<0.05
Clophen A-60(2)	1.4	5.0	2.2	8.4
Phenoclor DP-6(2)	0.7	10.0	2.9	13.6
T 1242(55)	2.3	2.2	—	4.5
T 1248(55)	0.5	2.3	—	2.8
T 1254(55)	0.1	3.6	1.9	5.6
T 1260(55)	0.8	0.9	0.5	2.2
Kanechlor 300(55)	6.7	1.6	—	8.3
Kanechlor 400(55)	12.2	10.4	0.9	23.8
Kanechlor 500(55)	1.7	1.1	3.1	6.1
Kanechlor 600(55)	0.2	0.5	0.4	1.1
Used:				
T 1248(55)	5.8	5.6	0.7	12.4
Aroclor 1254(51)	—	—	—	0.55
Aroclor 1260(51)	—	—	—	1.1
Yusho Oil (KC-400)	520	1330	810	2680

TABLE 21
Chlorobenzene pyrolysis

PCBz studied	Chlorobenzene pyrolysis Process	Product	Yield	Ref.
Mixture of tri-/tetra-CBz; PC_5Bz	Pyrolysis in air	PCDD, PCDF, PCB, higher chlorinated benzenes, Cl-phenols	2,000 ppm	156
Monochlorobenzene	Pyrolysis in N_2	Mono- and di-Cl-biphenyl	50 ppm	96
Monochlorobenzene	Pyrolysis	di-Cl-biphenyl	10%	157

A second simplification can sometimes be made in good approximation when one or more reactions occur much faster than other reactions in the system. Under these conditions, partial-equilibrium is assumed. This allows the expressions which describe forward and backward reactions to be equated. Further simplifications can be made using sensitivity analysis of solution variables in which numerical solutions are derived for small changes in input variables to determine the magnitude of induced change. When reactions or species are found which have little overall effect, the equilibrium equations and terms which involve the species can be dropped and the conservation equations are thereby simplified.

A reaction mechanism for the formation of dioxins has been suggested by Altwicker *et al.* (1990):

$$\text{Precursor}_{(g)} + \text{Precursor}_{(ads)} \rightarrow \text{PCDD}_{(ads)}$$

$$\text{PCDD}_{(ads)} \rightarrow \text{PCDD}_{(g)}$$
$$\text{PCDD}_{(ads)} \rightarrow \text{non-toxic product}$$
$$\text{PCDD}_{(ads)} \rightarrow \text{non-toxic decomposition product.}$$

In generalized terms, formation and destruction can also occur by a mechanism such as the following:

Polychlorinated phenol → polychlorinated phenoxy
radical
↓
polychlorinated
2-phenoxyphenols
(pre-dioxins)
↓

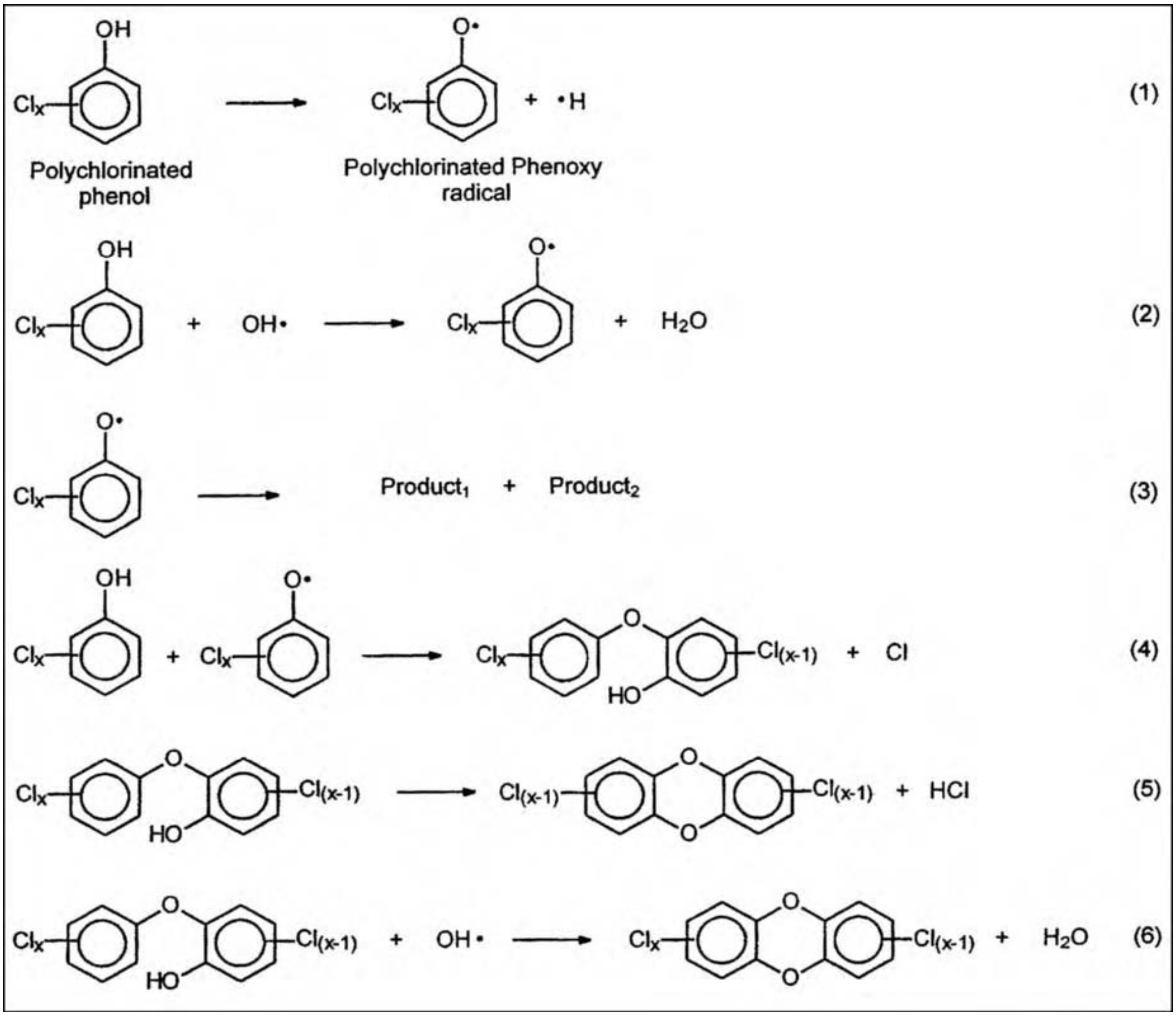


FIGURE 23 Mechanism of dioxin formation in flames [Schaub and Tsang, 1983].

PCDDs
polychlorinated phenoxy radical → non toxic combustion product
PCDDs → non-toxic combustion product.

Increasing substitution by chlorine by chlorophenols would be expected to result in slower rates of attack by hydroxyl radical and therefore the more highly chlorinated dioxins are likely to be less reactive towards bimolecular decomposition with hydroxyl radicals than the lesser chlorinated congeners. One would therefore expect that the dioxin isomers distribution would tend to be skewed toward higher chlorinated species.

The unimolecular rupture of the weakest bond yields a polychlorinated phenoxy radical (Figure 25) .

The phenoxy radical is also produced when hydroxyls radicals cause hydrogen atom abstraction. If, alternatively, •OH addition occurs, then the adduct may undergo further reaction with oxygen, or some other species, to give ring opening and thereby reduce the likelihood of dioxin formation.

When the phenoxy radical decomposes by eliminating CO, it forms non-dioxin products.

Polychlorinated 2-phenoxy-phenols are termed predioxins because, in this configuration, the molecule is stereochemically able to cyclize into the dioxin. The bimolecular reaction can also occur in other positions around the ring which would then not produce pre-dioxins.

The unimolecular elimination of HCl from the predioxin yields dioxin.

The reaction is actually two reactions in one. The hydrogen abstraction reaction is similar to reaction (2) and will ultimately depend upon the equilibrium concentration of •OH radicals. The polychlorinated 2-phenoxy-phenoxy radical produced reacts very fast unimolecularly to yield a dioxin and eliminated either H or Cl.

R represents a gaseous organic fuel molecule. It is assumed to allow hydrogen abstraction to form a fuel free radical species which is more stable than the original phenoxy radical.

Both species are free radicals and therefore are likely to react very fast. The products indicated are not meant to

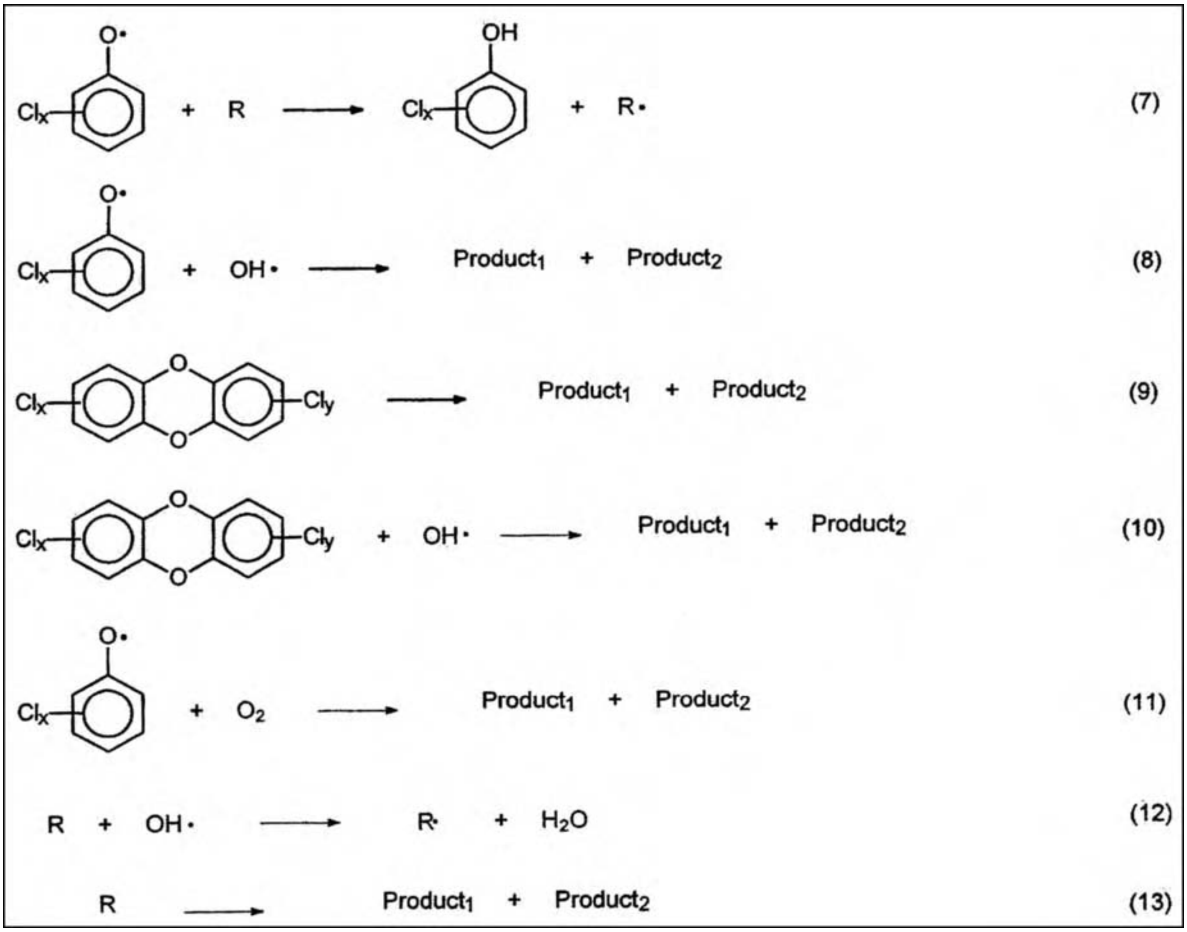


FIGURE 23 (continued) Mechanism of dioxin formation in flames [Schaub and Tsang, 1983].

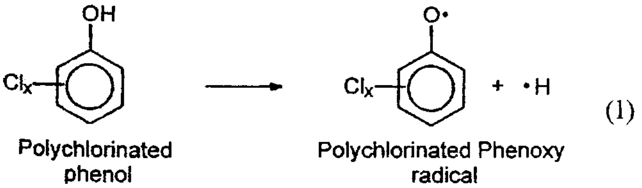


FIGURE 24

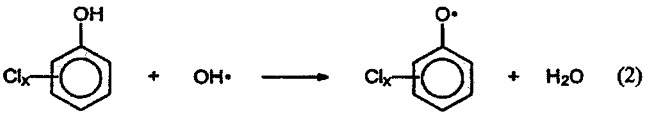


FIGURE 25

FIGURE 26

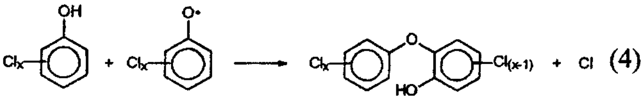


FIGURE 27

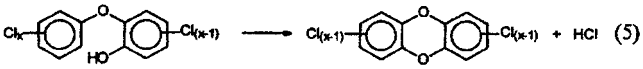


FIGURE 28

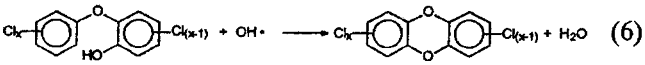


FIGURE 29

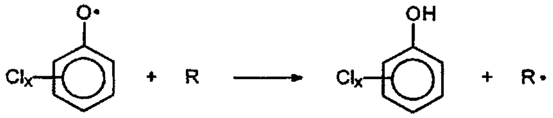


FIGURE 30

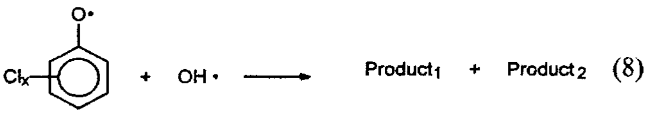


FIGURE 31

FIGURE 32

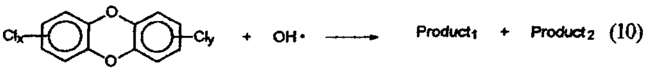


FIGURE 33

represent the same species as in reaction (3) or reactions (9), (10), (11) or (13).

The unimolecular dissociation of dioxin takes place when the molecule is heated sufficiently.

This reaction represents the disruption of the ring system with an hydroxyl radical to yield non-dioxin products.

A major reaction pathway is expected to be the bimolecular decomposition of phenoxy radicals with molecular oxygen when oxygen is present in excess, as is usually the case in incineration, as opposed to pyrolysis in a reductive atmosphere.

The production of fuel radicals by hydrogen abstraction from fuel is likely to be slower than the reaction between hydroxyl radicals and the phenoxy radicals of reaction (8) but this reaction will dominate especially in areas where there are fuel-rich pockets in the incineration process.

The decomposition of fuel into products describes a wide range of reactions.

Schaub and Tsang derived estimates of the rate constants for each of the steps in the reaction sequence by comparing literature values for similar chemistries and calculated four situations governed by different mole fractions. The first case, for example, assumed the condition that no O_2 or fuel is available in the region in which the reaction mechanism is taking place. This was modeled so that the formation of dioxin could be exaggerated.

In the second case, concentrations were calculated on the basis of there being oxygen available for reaction but at an initial mole fraction which corresponded to about 60% of the value needed for stoichiometric combustion. Under normal conditions incinerators are run with excess oxygen or at least a stoichiometric amount. This condition deemphasizes the rate of loss of chlorophenoxy radical in reaction 11 and consequently biases the calculation in favor of dioxin formation. The third scenario models a fuel-rich pocket in a post-combustion zone environment where no oxygen is available for the reaction sequence.

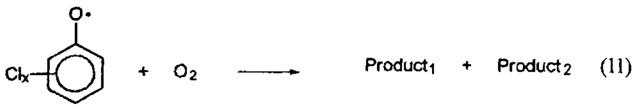


FIGURE 34

$$\mathrm{R} + \mathrm{OH}\cdot \longrightarrow \mathrm{R}\cdot + \mathrm{H_2O} \tag{12}$$

FIGURE 35

$$\mathrm{R} \longrightarrow \mathrm{Product_1} + \mathrm{Product_2} \tag{13}$$

FIGURE 36

Case four is similar to case three but includes the availability of oxygen (Figures 37, 38, 39).

The correlation between combustion temperature and the concentration of dioxin in the example scenarios is illustrated in the above figures. The three curves were not plotted together because the axes extend over significantly different ranges. While the absolute concentrations predicted by the model are relatively unimportant, it is significant that, under post-combustion mixing, in an intermediate temperature zone, unburned chlorophenol is predicted to react with hydroxyl radicals to produce PCDDs. Clearly, high temperatures, and a sufficient quantity of air and fuel are able to provide conditions which will combust chlorophenols into products which are not PCDDs. Conditions can occur during operation of the incinerator in which the amount of PCDD produced is extremely small but, most importantly, some conditions will allow the formation of PCDDs at concentrations which are 5 to 10-orders of magnitude higher than those produced under 'designed' operating conditions.

The rate of destruction of organic material is determined from the kinetic differential equation which relates concentration at a given moment of time to the rate constant for the reaction:

$$\frac{\mathrm{d}C}{\mathrm{d}t} = -k \cdot C \tag{1}$$

where

C = concentration at time, t
k = rate constant.

Upon integration,

$$t = \frac{1}{k} \cdot \ln\left(\frac{C_0}{C}\right) \tag{2}$$

where,

C_0 = initial concentration at time $t = 0$.

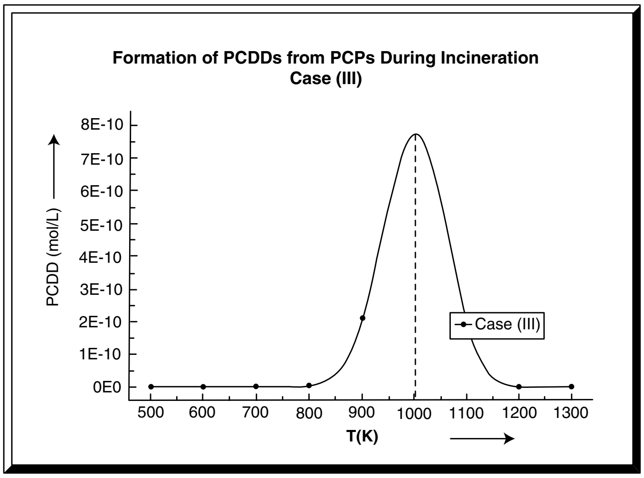


FIGURE 37

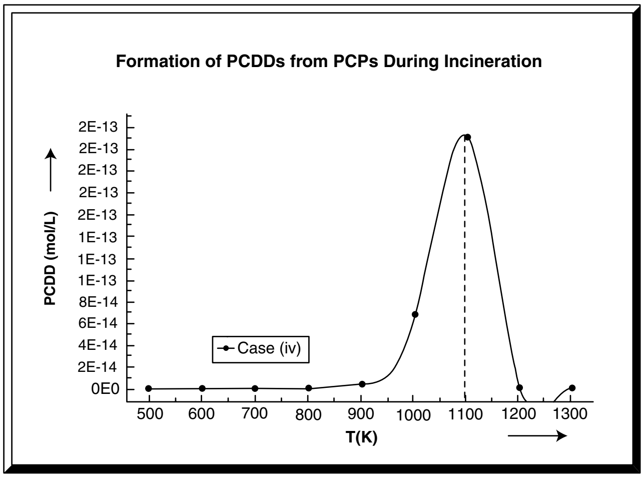


FIGURE 38

Rate constants can be expressed in terms of an Arrhenius expression:

$$k = A\cdot e^{(-(E/R\cdot T)} \tag{3}$$

where

A = frequency factor
E = activation energy
R = Universal gas constant.

From statistical mechanics, the term exp[$-E/RT$] can be taken to mean the fraction of molecular collisions which have an energy E or greater. The pre-exponential term was applied to introduce into the theory a steric collision term intended to describe the need for an orientation of the colliding molecules to produce reaction.

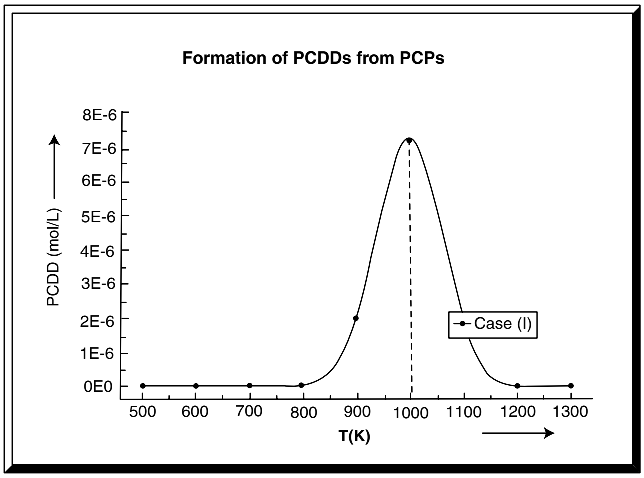


FIGURE 39

Substituting for k in equation (2) yields:

$$t = \frac{1}{A\cdot\exp[-(E/R\cdot T)]}\cdot\ln\left(\frac{C_0}{C}\right)$$

$$\ln\left(\frac{C_0}{C}\right) = A\cdot t\cdot\exp\left(-\frac{E}{R\cdot T}\right)$$

$$\therefore\left(\frac{C_0}{C}\right) = \exp\,[A\cdot t\cdot\exp\left(-\frac{E}{R\cdot T}\right)$$

$$\therefore\frac{C_0-C}{C} = \text{ Destruction Efficiency, } (N)$$

$$= \exp\left[A\cdot t\cdot\exp\left(-\frac{E}{R\cdot T}\right)-1\right]$$

$$\therefore (N+1) = \exp\left[A\cdot t\cdot\exp\left(-\frac{E}{R\cdot T}\right)\right].$$

$$\text{Rearranging, } T = -\frac{E}{R\cdot[\ln\,(\ln(1+N))-\ln A-\ln t]} \tag{4}$$

where

N = destruction efficiency
A = frequency factor (second^{-1})
E = activation energy (cal/g-mole)
R = universal gas constant
= 1.987 (cal/g-mole-°K)
T = combustion temperature (°K).

Table 22 shows the destruction kinetics for several components of transformer askarels and the calculated temperatures necessary to achieve a destruction efficiency of 99.9999% with a range of residence times.

TABLE 22
Destruction Kinetics During the Incineration of Askarel

Compound	Frequency Factor(1/sec)	Activation Energy (cal/g-mole)	Destruction Temperature(°K)/Residence Time (sec)			
			0.5	1	2	7
Monochlorobenzene	$8.0 \cdot 10^4$	23,000	1,056	993	937	851
Dichlorobenzene	$3.00 \cdot 10^8$	39,000	1,023	987	954	899
Trichlorobenzene	$2.20 \cdot 10^8$	38,000	1,013	977	944	889
Tetrachlorobenzene	$1.90 \cdot 10^6$	30,000	1,068	1,019	973	900
Hexachlorobenzene	$1.90 \cdot 10^6$	72,600	983	965	948	918
Pentachlorobenzene	$1.10 \cdot 10^6$	70,000	962	944	927	898

Residence time calculations can be made from consideration of an ideal plug-flow reactor in which the temperature, composition and velocity of the gases are considered to be uniform within any cross-section of the furnace.

The mean residence time,

$$t = \frac{V}{Q} \cdot 60 \tag{5}$$

where
t = mean residence time in seconds
V = furnace volume, cu. Ft
Q = volumetric flow rate, scfm.

The actual flow rate is proportional to temperature,

$$Q' = Q \cdot T \tag{6}$$

where,
Q' = actual flow rate
T = temperature

from equations (5) and (6),

$$t = \frac{V}{Q \cdot T}. \tag{7}$$

The flow through the incinerator is not isothermal and so is not well approximated by the plug flow reactor unless the time-temperature distribution is taken into account as follows.

The differential form of the residence time equation is:

$$\mathrm{d}t = \frac{\mathrm{d}V}{Q'}$$

$$\mathrm{d}V = A_\mathrm{f} \cdot \mathrm{d}x$$

where
A_f = the arc of the furnace
$\mathrm{d}x$ = an element of the axial length of the furnace.

Then, from equation (7),

$$dt = \frac{A_\mathrm{f}}{Q \cdot T} \cdot \mathrm{d}x. \tag{8}$$

If a linear temperature gradient is assumed along the incinerator axis, and the temperature at a point x_m is T_m and the exit temperature T_e is at a distance x_e then,

$$\frac{(T - T_\mathrm{m})}{(x - x_\mathrm{m})} = \frac{(T_\mathrm{e} - T_\mathrm{m})}{(x_\mathrm{e} - x_\mathrm{m})}.$$

Rearranging,

$$T = \frac{1}{(x_\mathrm{m} - x_\mathrm{e})} \cdot [x \cdot (T_\mathrm{m} - T_\mathrm{e}) - T_\mathrm{m} \cdot x_\mathrm{e} + T_\mathrm{e} \cdot x_\mathrm{m}].$$

Substituting into equation (8),

$$\mathrm{d}t = \frac{A_\mathrm{f}}{Q} \cdot \frac{(x_\mathrm{m} - x_\mathrm{e})}{[x \cdot (T_\mathrm{m} - T_\mathrm{e}) - T_\mathrm{m} \cdot x_\mathrm{e} + T_\mathrm{e} \cdot x_\mathrm{m}]} \cdot \mathrm{d}x$$

$$\int \mathrm{d}t = \frac{A_\mathrm{f}}{Q} \cdot \int \frac{(x_\mathrm{m} - x_\mathrm{e})}{[x \cdot (T_\mathrm{m} - T_\mathrm{e}) - T_\mathrm{m} \cdot x_\mathrm{e} + T_\mathrm{e} \cdot x_\mathrm{m}]} \cdot \mathrm{d}x$$

$$\therefore (t - t_\mathrm{m}) = \frac{A_\mathrm{f}}{Q} \cdot \ln\left(\frac{T}{T_\mathrm{m}}\right) \cdot \frac{(x_\mathrm{m} - x_\mathrm{e})}{(T_\mathrm{m} - T_\mathrm{e})} \tag{9}$$

If T_m is close to the furnace entrance, then

$$x_\mathrm{m} \ll x_\mathrm{e}.$$

If $t_\mathrm{m} = 0$, by definition, then

$$t = \frac{A_\mathrm{f}}{Q} \cdot \ln\left(\frac{T}{T_\mathrm{m}}\right) \cdot \frac{(-x_\mathrm{e})}{(T_\mathrm{m} - T_\mathrm{e})}$$

The furnace volume, $V = A \cdot x_e$

$$\therefore t = \frac{1}{Q} \cdot \frac{V}{(T_e - T_m)} \cdot \ln\left(\frac{T}{T_m}\right) \tag{10}$$

that is, the residence time, *t,* is the time spent by the gas above a temperature, *T.*

In partial plug flow kinetics, there is some longitudinal mixing which allows some of the gas to exit the hot zone along a 'fast path'. If the fast path is assumed to be half of the mean residence time, t_f, then,

$$t_f = \frac{V}{2 \cdot Q \cdot (T_e - T_m)} \cdot \ln\left(\frac{T}{T_m}\right). \tag{11}$$

A plot of Eqs. (10) and (11) are shown in Figure 40.

The mean residence time of a material has been calculated on the assumption that it is passing through a 113,000 L incinerator with a gas velocity of 400,000 L/min, a maximum temperature of 1600°C and an exit temperature of 650°C.

A 99.9999% combustion efficiency is a mandated requirement for an incinerator burning PCBs. The PCB contaminated material introduced into the combustion zone must, according to 40 CFR 761.40, be maintained for 2-seconds at 1200°C (±100°C), that is, 2200°F (±200°F), with 3% excess oxygen in the stack gas, or, 1.5 sec residence time at 1600°C (±100°C), that is, 2900°F (±200°F) with 2% excess oxygen in the stack gas.

The figure indicated that a 99.9999% destruction and removal efficiency could be obtained for tetrachlorobenzene in the example incinerator when the mean residence time is 0.7 sec or less. In the case of pentachlorobiphenyl, the residence time would have to be less than about 1.1 sec. If either the tetrachlorobenzene or the PCB passed through the hot zone on a fast-path, the model predicts that the required destruction efficiency would not be achieved since the curves are predominantly above the fast-path temperature-time curve. Note that a residence time of 2.0 sec at 2200°F (1200°C) predicts that a 6–9's DRE might be achieved for both substances, but that these conditions are a lower limit for the incinerator to achieve this level of destruction. A temperature of 2900°F (1600°C) for 1.5 sec would seem to be a much surer set of conditions to achieve 6–9's destruction because then both of the curves for the compounds of interest would lie substantially below the mean residence time and fast-path residence time curves in the model.

Toxic Equivalency Factors and Human Exposure Many of the PCB congeners yield several different PCDF isomers upon pyrolysis. Consequently, in addition to the discussion above on the toxicity of individual PCB isomers present in Aroclor 1016, the following consideration of PCDF isomer toxicity is made of evaluate the risk of generation of PCDFs in a worst case scenario. The analysis of PCDFs in individuals exposed to the Yu-Cheng poisoning showed that the most persistent congeners were the 2,3,7, 8-tetra- 1,2,3,7,8-penta-, 2,3,4,7,8-penta-, and 1,2,3,4,7,8-hexachlorodibenzofuran. Mixtures of these PCDFs with 6 of the most toxic PCBs found in the contaminated oil were prepared as a reconstituted mixture and tested to determine the relative potencies of PCBs and PCDFs (Safe *et al.*[65]). It was shown that PCDFs are from 680 to 2,210 times more toxic than the PCBs. It was therefore suggested that PCDFs were the major etiologic agents in the Yusho and Yu-Cheng poisonings (Table 23).

The relative potencies of PCB and PCDF congeners were measured in terms of their enzyme induction. For example, ED_{50} values are defined as the concentration of chemical which results in 50% of the maximal enzyme induction observed in the dose-response procedure used for each PCDF. The smaller the value for the induction potency or the receptor binding affinity of an isomer, the more toxic the compound. Table 23 lists the relative potencies of some of the most toxic PCDFs. There are considerable differences in the observed activities of PCDFs which differ with respect to Cl substitution at the C-2 and C-3 or C-7 and C-8 positions. In all cases the C-3 or C-7 substituted isomers are the most active. PCDF isomers which differ only in their Cl substitution at C-1 (or C-9) and C-4 (or C-6) positions show that the enzyme induction and receptor binding activities of the C-1 isomers were all lower than the corresponding C-4 substituted isomers.

The toxicity of PCDFs in animals has been reviewed by several authors and while there is very little isomer specific toxicity data for congeners other than 2,3,7,8-TCDF, the generalization has been made that there is a close similarity between the biological activity of PCDFs and that of the structurally similar PCDDS. This is illustrated in Table 18, which shows the relative potency of some PCDF isomers relative to TCDD.

Nagayama *et al.*[67] noted that strains of mice can be separated into two groups after treatment with 3-methylcholanthrene (3-MC) according to whether or not they were AHH responsive. Similarly, the toxicity of PCDFs and PCDDs to mice has been shown to correlate with AHH response (Nagayama *et al.*[68]). The authors therefore attempted to determine the presence of a similar susceptibility in humans by correlating the AHH activity in the lymphocytes of Yusho patients and their skin symptoms (Nagayama *et al.*[69]).

Interpretation of the data given above leads to the conclusion that when PCBs are pyrolyzed, only the formation of PCDF isomers which contain 4 or more Cl substituents need be of concern. In particular, tetra- and higher chlorinated congeners with lateral substitution are the most significant. It should be noted that PCDDs are not generated at all by the pyrolysis of PCBs (Table 24).

The four mechanisms of ring closure illustrated in Figure 20 can be used to predict the isomer distribution of PCDF products although not their relative amounts. Table 25 indicates the PCDF isomers with more than three chlorines which can be obtained from Aroclor 1016. The potentially most toxic mixture of compounds which could be derived from Aroclor 1016 in a worst case scenario would, from the above prediction of the isomers formed, be of very low toxicity. The PCDF isomers which could theoretically be produced would

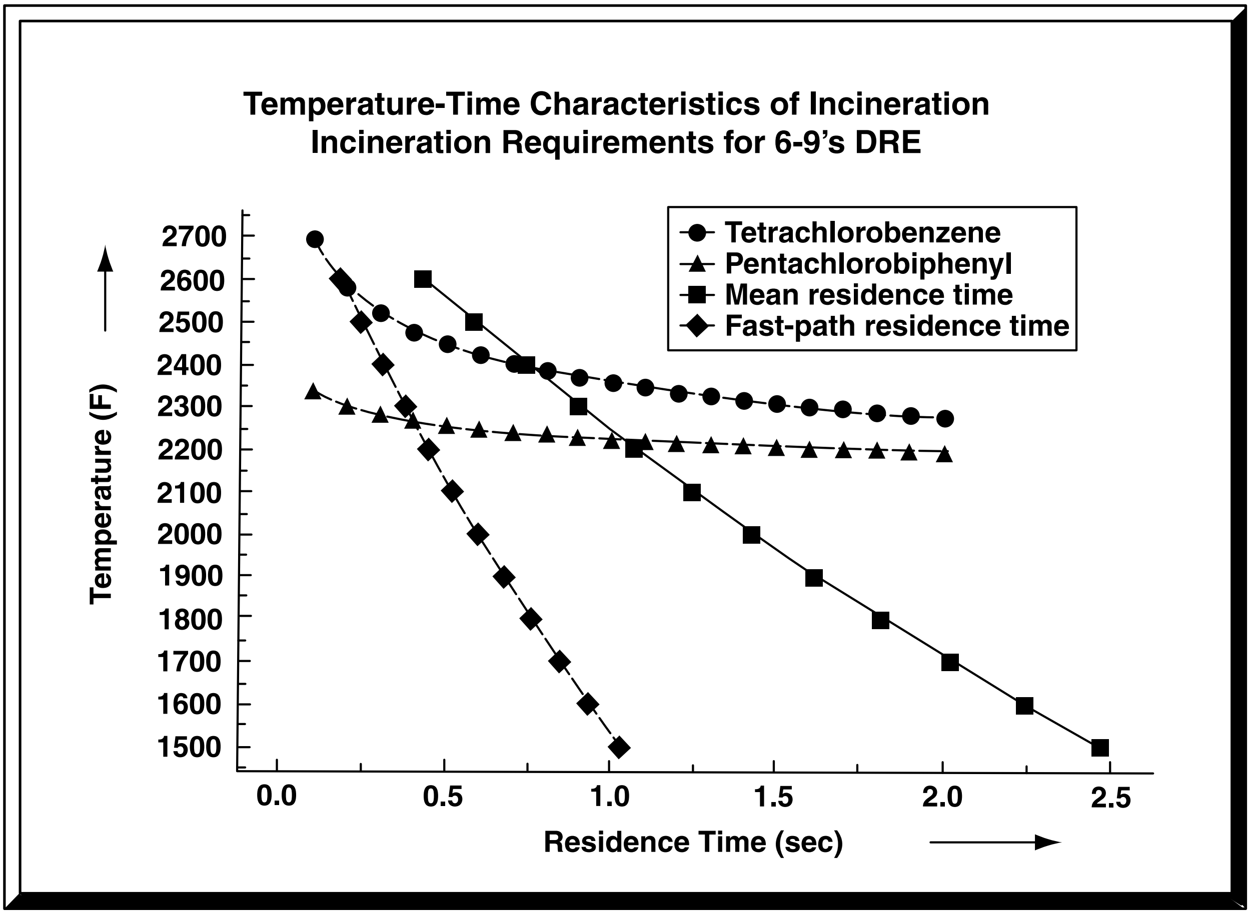


FIGURE 40 The area above the curves shows the temperature-time correlation needed for 99.9999% DRE.

TABLE 23
PCDF toxicity decreases with decreasing lateral Cl substituents

PCDF	Induction potency AHH	Receptor binding affinity
2,3,4,7,8-PeCDF	2.6×10^{-10}	1.5×10^{-8}
1,2,3, 7,8-PeCDF	2.5×10^{-9}	7.5×10^{-7}
1,3,4, 7,8-PeCDF	1.6×10^{-8}	2.0×10^{-7}
1,2,3,7,9-PeCDF	9.0×10^{-8}	4.0×10^{-7}
2,3,4,7,9-PeCDF	7.9×10^{-9}	2.0×10^{-7}
1,2,4,7,8-PeCDF	1.1×10^{-7}	1.3×10^{-6}
2,3,4,7-TCDF	1.8×10^{-8}	2.5×10^{-8}
2,3,4,8-TCDF	4.1 × 10-8	2.0 × 10-7
1,2,3,7-TCDF	2.7 × 10-5	1.1 × 10-7

only amount to the conversion of a total of 5.7% of the PCB present. The quantity of active PCDF isomers which could be generated is extremely low and, typically, conversion from PCB to PCDF takes place under optimum pyrolysis conditions only to the extent of 4% and usually much less. The conversion factor for the yield of slightly biologically active PCDF isomers derived from Aroclor 1016 is therefore a maximum of 0.002, that is, a PCDF : PCB ratio of 1 : 450. Not only is the relative ratio of PCDFs : PCBs one half of the 1 : 200 typical of Yusho and Yu-Cheng oils but, in addition, none of the toxic PCDF isomers found in the Yusho and Yu-Cheng incidents could possibly be present when Aroclor

TABLE 24
The most active compounds are those with Cl substitution at the 2,3,7, and 8 positions

	Relative potency	
PCDF	TCDF	TCDD
Dibenzofuran	$<10^{-4}$	$<10^{-6}$
2,8-diCDF	$<10^{-4}$	$<10^{-6}$
2,4,6-triCDF	3×10^{-5}	3×10^{-7}
2,3,8-triCDF	4.5×10^{-3}	4.8×10^{-5}
1,4,6,8-tetraCDF	$<10^{-4}$	$<10^{-6}$
2,3,6,8-tetraCDF	2.0×10^{-4}	2.1×10^{-6}
2,4,6,8-tetraCDF	$<10^{-4}$	$<10^{-8}$
2,3,7,8-tetraCDF	1	9×10^{-2}
1,3,4,7,8-pentaCDF	4.8×10^{-1}	5.2×10^{3}
1,2,4,7,8-pentaCDF	3.0×10^{-3}	3.2×10^{-5}
1,2,3,4,6,8,9-heptaCDF	3.8×10^{-3}	4.1×10^{-5}

Source: Reference 66: Bradlaw, J. and Casterline, J., Jr.; Induction of enzyme activity in cell culture: PA rapid screen for detection of planar polychlorinated organic compounds; *J. Assoc. Off. Anal. Chem.* **62**, 904 (1979).

1016 is pyrolysed because of the PCB isomer distribution present in the undegraded product.

Erickson *et al.*[23] have determined PCDF formation from PCBs under fire conditions when PCB contaminated oils are burned under a variety of conditions. The results indicated that the optimum conditions for PCDF formation from PCBs occurred near 675°C for 0.8 s or longer, with 8% excess oxygen. The residence time of the PCBs in the hot zone did not significantly affect the PCDF formation in the range 0.3 to 1.5 s. A linear relationship was found for PCDFs formed versus the amount of PCBs present. A 4% conversion efficiency of PCBs was observed.

The New York State Department of Health has developed the concept of "2,3,7,8-TCDD equivalents" as a measure of the effective toxicity of polychlorinated aromatic chemicals present in a mixture. The parameter is used as a determinant for the assessment of exposure and risk. The algorithm used to derive the exposure equivalents takes into consideration the relative toxicity of individual compounds. In short, if the heat exchanger fluids involved in the Yusho and Yu-Cheng incidents had been Aroclor 1016 rather than more highly chlorinated PCBS, it is very unlikely that the events would not have been noticed.

PCB concentrations in the blood of two groups of patients were statistically correlated,[31] in the period 1973–1980 and 27 patients in 1981–1984. The AHH activity in the 1973–1980 group showed a positive correlation with blood PCB concentration but the correlation was less statistically significant for the 1981–1984 group. Both group AHH activities showed a correlation to their former PCB in blood concentrations but much less correlation to the later PCB concentrations. Consequently, AHH activity of Yusho patients at the onset of symptoms was probably higher than in its later stages since PCBs and PCDFs had been eliminated from the body. Yusho patients with higher AHH activity showed more serious skin lesions than patients with lower AHH activity.

Furst *et al.*[70] have analysed the distribution of PCDF and PCDD isomers in human milk samples from West Germany and found that all of the samples contained a "typical" pattern of PCDFs and PCDDs. The first study of this type was reported by Rappe and his co-workers[71] on the identification of specific isomers in human milk samples derived from Sweden and West Germany. In addition, the levels of the congeners identified in all of the samples were of nearly the same proportion in each.

However, there was relatively little correlation between either the PCDFs or PCDDs with the concentration of PCBs. This suggests that the sources of contamination may not be the same. Interestingly, the correlation coefficients for isomers which differed by one chlorine are strikingly higher than those for compounds which differ by more than one chlorine. The PCDFs show a similar consistency. No explanation was offered for the observation but it would be interesting to extend the study to draw possible correlations between the isomer distributions found in foods and those which are excreted in breast milk.

The selective accumulation of tetra- to octa-chlorinated dibenzo-p-dioxins and dibenzofurans in foods and human tissues in northern Japan was reported by Y. Takizawa and H. Muto in 1987.[72] They found that the distribution of isomers was similar in food samples and in human tissue samples. Foods with a high fat content such as fish/shellfish and meats contained high concentrations of PCDFs and PCDDs relative to non-fat foods such as vegetables. The following table illustrates the levels of contamination in daily foods in northern Japan (Table 26).

TABLE 25
Aroclor 1016 contains 0.4% by weight of PCBs with more than 4 Cl substituents

PCB isomer	Weight %	PCDF isomers produced by pyrolysis
2,3,4,3′,5′	0.16	1,2,3,6,8-
		2,4,6,7-
2,3,6,2′,3′,	0.01	1,2,8,9-
		1,2,6,9-
		1,4,6-
2,3,6,3′,5′	0.01	2,4,6,9-
		1,2,6,8-
2,3,6,2′,5′	0.18	1,2,6,9-
		1,4,6,9-
		1,2,8-
		1,4,8-
2,4,6,2′,3′	0.04	1,2,7,9-
		1,3,8,9-
		1,3,6-
2,4,3′,4′	0.14	1,3,7,8-
2,4,5,4′	1.35	1,3,4,7-
2,4,2′,4′	1.81	1,3,7,9-
2,3,2′,3′	0.18	1,2,8,9-
2,3,4,2′	2.00	1,2,3,9-
2,3,2′,5′	1.14	1,2,6,9-
2,4,2′,5′	3.48	1,3,6,9-
2,4,5,2′	1.41	1,3,4,9-
2,3,2′,6′	0.33	triCDFs
2,3,5,2′	0.47	1,2,4,9-
2,3,2′,6′	0.33	triCDFs
2,3,5,2′	0.47	1,2,4,9-
2,4,6,4′	2.40	triCDFs
2,5,2′,5′	4.35	1,4,6,9-
2,5,2′,6′	1.07	triCDFs
2,3,6,2	1.00	triCDFs
2,6,2′,6′	0.19	triCDFs

Tables 27 and 28 show the calculated daily intake of PCDDs and PCDFs from foods. The values used for the daily intake of different types of food were derived from "Nutrition Survey Data, Japanese Agency of Health and Welfare, Daiichi Press, Tokyo, 1982" as reported by Y. Takizawa and H. Muto[72].

Table 29 compares the ratio of PCDD : PCDF isomers found in both the calculated values from foods and the observed ratios in human tissues.

There seems to be a good correlation between the dietary intake of PCDDs and PCDFs and the amounts found in each of the tissue types except for octachlorinated isomers. The reason for this may be that the source of O_8CDDs in tissues may be other than food. The best correlation is found between the penta-hepta isomers. Since there are a large number of factors which have not been considered in these tables, such as obesity and pharmacokinetics, it is somewhat surprising that the correlations are as good as they are.

People in industrialized countries have an adipose tissue level of PCDDs of about 1000–1200 ppt whereas people in non-industrialized countries have much lower levels. PCDF levels in industrialized countries are not usually below 50 ppt.

The estimation of health risk to humans was initially based on the amount of 2,3,7,8-TCDD/F present. In order to better estimate the risk caused by exposure to a mixture of different PCDD and PCDF isomers the concept of "TCDD Equivalents" have been introduced. Several models have

TABLE 26
The distribution of PCDDs and PCDFs in northern Japanese foods and human tissues (ppt)

	T_4CDDs	P_5CDDs	H_6CDDs	H_7CDDs	O_8CDDs
Sample	T_4CDFs	P_5CDFs	H_6CDFs	H_7CDFs	O_8CDFs
Tuna	1674/311	1227/459	1259/300	901/t	t/t
Flatfish	t/N.D.	3120/66	1626/379	1101/t	t/t
Mackerel	t/27	1413/218	926/113	1051/697	N > D./t
Chicken	1171/55	1509/634	539/323	1006/t	t/t
Egg	282/14	164/N.D.	326/103	N.D./124	N.D./1256
Milk	N.D./36	121/43	86/119	N.D./136	N.D./167
Pork	714/159	N.D./361	N.D./250	N.D./416	N.D./68
Rice	t/118	208/80	171/54	N.D./508	N.D./222
Flour	t/69	56/67	59/61	15/112	N.D./113
Potato	N.D./19	17/12	15/122	t/53	N.D./14
Peanuts	N.D./227	204/102	88/29	50/470	N.D./181
Soybeans	11/104	16/281	t/62	26/219	N.D./213
Miso	N.D./N.D.	18/1662	14/326	21/241	13/23
Spinach	36/20	26/36	30/40	18/85	t/131
Radish	N.D./35	13/28	t/36	t/51	N.D./t
Mushroom	t/t	16/t	14/t	t/t	t/48
Carrot	t/32	38/84	40/27	20/123	t/66
Apple	N.D./18	28/234	21/38	11/453	t/t
Lung	1272/143	890/530	635/495	241/487	144/272
Liver	891/150	822/508	342/717	188/298	190/263
Kidney	441/184	1387/519	403/394	394/266	59/t
Adipose tissue	492/129	1658/434	1096/434	301/483	70/356

t = trace, < 10 ppt, N.D. = not detected.

been developed. In addition to Eadon's model,[73] derived from data on soot samples taken from the Binghamton state office building fire, there are also models proposed by Danish and Swiss groups as well as the USEPA[74]. A re-examination of the USEPA model was presented by J.R. Olsen *et al.*[75]

Table 30 shows the "TCDD Equivalent" factors used in different models. The TCDD equivalents of a mixture of isomers can be calculated by multiplying the amount of each isomer with its TCDD equivalent from Table 30 and then summing the individual isomer contributions to the mixture.

Samples of human breast milk taken from areas in Vietnam which were sprayed with Agent Orange have shown markedly increased 2,3,7,8-TCDD levels even 10–15 years after cessation of the contamination. Some of the isomers of PCDDs and PCDFs found in the mother's milk were indicative of other contamination sources such as pentachlorophenol wood and paper preservative as well as, possibly, incineration products. The importance of such studies can be considered in terms of health risks to infants.

Schecter *et al.*[76] have observed that the body burden of 2,3,7,8-TCDD alone in nursing mothers from the south of Vietnam in 1973 was close to the adverse effect level calculated to cause chloracne in occupationally exposed workers based upon the "tolerable daily intake" of 30 pg/Kg of body weight over a 70 year lifetime, nursing infants in both North America and South Vietnam ingest more than a lifetime dose of PCDD/F in their first year.[77]

The levels of 2,3,7,8-TCDD in human adipose tissue in North America have been reported for populations in St. Louis, Atlanta, Utah and in Canada by R.L. Sielken.[78] He pointed out that the random level of 2,3,7,8-TCDD in human tissue is more likely to have a lognormal distribution than a normal

TABLE 27
Flatfish, chicken and rice contribute more than 90% towards the northern Japanese dietary intake of PCDDs

Foods	Food intake (g/d capita)	PCDDs (ng/d/capita)				
		T_4CDDs	P_5CDDs	H_6CDDs	H_7CDDs	O_8CDDs
Flatfish	90.2	—	281	147	99	—
Chicken	70.8	83	107	38	71	—
Egg	40.0	11	7	13	—	—
Milk	124.2	—	15	11	—	—
Rice	218.2	—	45	37	—	—
Flour	95.9	—	5	6	—	—
Potato	61.0	—	1	1	—	—
Peanuts	1.4	—	0.3	0.1	0.1	—
Miso	67.2	—	1	1	1	1
Spinach	58.7	2	2	2	1	—
Radish	192.2	2	3	—	—	—
Mushroom	8.9	—	0.1	0.1	—	—
Apple	159.7	—	5	3	2	—
Total Daily Intake		96	472	259	174	1

TABLE 28
Flatfish, chicken and rice contribute only 40% towards the northern Japanese dietary intake of PCDFs

Foods	Food intake (g/d capita)	PCDFs (ng/d/capita)				
		T_4CDFs	P_5CDFs	H_6CDFs	H_7CDFs	O_8CDFs
Flatfish	90.2	—	6	34	—	—
Chicken	70.8	4	45	23	—	—
Egg	40.0	1	—	4	15	50
Milk	124.2	5	5	15	17	21
Rice	218.2	26	18	12	111	48
Flour	95.9	7	6	6	11	11
Potato	61.0	1	1	7	3	1
Peanuts	1.4	0.3	0.1	—	0.7	0.3
Miso	67.2	—	112	22	16	2
Spinach	58.7	1	2	2	5	8
Radish	192.2	7	5	7	10	—
Mushroom	8.9	—	—	—	1	0.4
Apple	159.7	3	37	6	72	—
Total Daily Intake	—	55	237	138	262	142

TABLE 29
There is a strong correlation between the observed ratios of PCDDs and PCDFs in tissues in northern Japan with the ratios found in foods

Tissue	$(T_4)_t$	$(P_5)_t$	$(H_6)_t$	$(H_7)_t$	$(O_8)_t$
Lung	0.2	1.2	1.5	1.4	0.02
Liver	0.3	1.2	3.9	1.1	0.01
Kidney	0.7	0.8	1.8	0.5	—
Adipose	0.5	0.5	0.8	1.1	0.05

$(T_4)_i$ = ratio of T_4 isomers in the daily food intake
$(T_4)_t$ = $(T_4CDDs/T_4CDFs)_t$ = ratio of T_4 isomers in tissue.

TABLE 30
TCDD equivalents of different PCDD/PCDF congeners

Congeners	USEPA	Swiss	Danish	German	Eadon
2,3,7,8-TCDD	1	1	1	1	1
1,2,3,7,7-P_5CDD	0.5	0.1	0.1	0.1	1
1,2,3,4,7,8-H_6CDD	0.04	0.1	0.1	0.1	0.033
1,2,3,6,7,8-H_6CDD	0.04	0.1	0.1	0.1	0.033
1,2,3,7,8,9-H_6CDD	0.04	0.1	0.1	0.1	—
1,2,3,4,6,7,8-H_7CDD	0.001	0.01	0.01	0.01	—
O_8CDD	—	—	—	0.001	—
2,3,7,8-TCDF	0.1	0.1	0.1	0.1	0.33
2,3,4,7,8-P_5CDF	0.1	0.1	0.2	0.1	0.33
1,2,3,4,7,8-H_6CDF	0.01	0.1	0.1	0.1	0.02
1,2,3,6,7,8-H_6CDF	0.01	0.1	0.1	0.1	0.02
2,3,4,6,7,8-H_6CDF	0.01	0.1	0.1	0.1	0.02
1,2,3,4,6,7,8-H_7CDF	0.001		0.01	0.01	—
O_8CDF	—	—	—	0.001	—

distribution since there was a positive correlation between the dioxin level and age. Sielken's correlation for the St. Louis population was

$$\log_e[\text{TCDD}] = 1.02 + 0.023 + \text{age}; \; r = 0.72.$$

The observation might be partially explained by persons ingesting a contaminated diet but does not support an initially high exposure of the population to contaminated mother's milk. Schecter *et al.*[79] have compared the results from several workers on the concentrations of PCDDs and PCDFs in human milk derived from north and south Vietnam and the United States. They then went on to calculate the 2,3,7,8-TCDD equivalents contained in the milk and to derive the ratio between the first year intake and the recommended intake values. The data showed that in south Vietnam, and a small percentage of the population in the US are exposed to 2,3,7,8-TCDD concentrations in mothers milk which are sometimes very much greater than the limits set by both the USEPA and the United States Public Health Services Centers for Disease Control (USCDC). The calculations are extended in Table 31 to include the data reported by Ogaki *et al.*[80] on Japanese foodstuffs.

Table 32 shows the estimated daily intake of total 2,3,7,8-TCDD equivalents by nursing infants in South and North Vietnam, Japan and the United States. It has been assumed that an average consumption of milk is 800 mL/day and that the body is 10 Kg. The "Acceptable Daily Intake" or ADI value for the ingestion of 2,3,7,8-TCDD set by the EPA for an assumed 70 year lifespan is 1 pg/Kg/d. Adults are assumed to be 55 Kg for Vietnamese residents and 70 Kg for North Americans.

The data in Table 31 were used to calculate 2,3,7,8-TCDD equivalents in Table 32. The daily intake of a nursing baby was derived on the assumption that an average of 800 mL of milk were consumed per day. The first year intake is then 365 times the daily intake corrected for an assumed body weight of 10 Kg. It is further assumed that there is no further ingestion or excretion of PCDDs/PCDFs during the 70 year lifespan of the individual. Under such a condition,

TABLE 31

Concentration of PCDDs and PCDFs in human milk on a wet weight or whole milk basis

Population	Isomer	Concentration	
		Mean	Range
South Vietnam (1970)	2,3,6,8-TCDD	11.4	ND–55
	2,3,7,8-TCDD	0.68	ND–3.1
South Vietnam (1984)	1,2,3,7,8-P_5CDD	0.44	ND–1.3
	H_6CDD	3.0	0.7–10.4
	1,2,3,4,6,7,8-H_7CDD	5.3	0.41–21.3
	O_8CDD	39.8	4.2–102.3
	2,3,4,7,8-P_5CDF	0.63	0.13–1.4
	H_6CDF	1.7	0.41–5.0
	1,2,3,4,6,7,8-H_7CDF	0.9	0.13–2.2
	2,3,7,8-TCDD	0.039	ND
North Vietnam	1,2,3,7,8-P_5CDD	0.24	ND–0.11
	H_6CDD	0.59	ND–0.72
	1,2,3,4,6,7,8-H_7CDD	2.8	ND–1.7
	O_8CDD	0.31	ND–6.2
	2,3,4,7,8-P_5CDF	0.32	ND–0.88
	H_6CDF	0.09	ND–0.6
	1,2,3,4,6,7,8-H_7CDF		ND–0.32
	2,3,7,8-TCDD	0.48	0.33–0.70
Japan[†]	H_6CDD	0.56	0.26–1.11
	H_6CDD	2.81	1.59–5.07
	1,2,3,4,6,7,8-H_7CDD	5.55	3.96–7.36
	O_8CDD	29.2	20.98–37.41
	2,3,7,8-TCDF	0.63	0.19–1.70
	2,3,4,7,8-P_5CDF	1.26	0.81–1.70
	1,2,3,4,7,8-H_6CDF	0.63	0.41–0.85
	1,2,3,4,7,8,9-H_7CDF	0.26	0.07–0.56
	O_8CDF	0.78	3.40–9.40
	2,3,7,8-TCDD	0.19	ND–0.38
United States	1,2,3,7,8-P_5CDD	1.2	ND–6.8
	H_6CDD	2.7	0.3–12.4
	1,2,3,4,6,7,8-H_7CDD	3.3	ND–14.6
	O_8CDD	21	2.4–46.9
	2,3,7,8-TCDF	0.34	ND–1.2
	2,3,4,7,8-P_5CDF	1.0	ND–2.9
	H_6CDF	0.68	0.1–1.3
	1,2,3,4,6,7,8-H_7CDF	0.68	ND–2.1
	O_8CDF	2.3	ND–13.5

[†] The Japanese data have been recalculated from data presented on a liquid basis. The assumption was made that the liquid content of the milk was an average of about 3.7%—see values reported by Schecter *et al.*, 1987, reference (76).

TABLE 32
2,3,7,8-TCDD equivalent for PCDDs and PCDFs derived from the data in Table 31

Population	Isomer	2,3,7,8-TCDD equivalents	
		Mean	Upper value of range
South Vietnam (1970)	2,3,6,8-TCDD	11.4	55
Total Equivalents:		**11.4**	**55**
South Vietnam (1984)	2,3,7,8-TCDD	0.68	3.1
	1,2,3,7,8-P_5CDD	0.22	0.65
	H_6CDD	0.12	0.42
	1,2,3,4,6,7,8-H_7CDD	0.005	0.02
	O_8CDD	0	0
	1,2,3,4,7,8-P_5CDF	0.06	0.14
	H_6CDF	0.02	0.05
	1,2,3,4,6,7,8-H_7CDF	0.0009	0.0022
Total Equivalents:		**1.11**	**4.38**
North Vietnam	2,3,7,8-TCDD	ND	
	1,2,3,7,8-P_5CDD	0.02	0.055
	H_6CDD	0.0096	0.029
	1,2,3,4,6,7,8-H_7CDD	0.0006	0.0017
	O_8CDD	0	0
	1,2,3,4,7,8-P_5CDF	0.031	0.088
	H_6CDF	0.0032	0.006
	1,2,3,4,6,7,8-H_7CDF	0.00009	0.00032
Total Equivalents:		**0.065**	**0.18**
Japan†	2,3,7,8-TCDD	0.48	0.70
	P_5CDD	0.28	0.56
	H_6CDD	0.11	0.20
	1,2,3,4,6,7,8-H_7CDD	0.01	0.01
	O_8CDD	0	0
	2,3,7,8-TCDF	0.06	0.017
	2,3,4,7,8-P_5CDF	0.13	0.17
	1,2,3,4,7,8-H_6CDF	0.01	0.01
	1,2,3,4,7,8,9-H_7CDF	0.0003	0.0006
	O_8CDF	0	0
Total Equivalents:		**1.08**	**1.82**
United States	2,3,7,8-TCDD	0.19	0.38
	1,2,3,7,8-P_5CDD	0.6	3.4
	H_6CDD	0.11	0.5
	1,2,3,4,6,7,8-H_7CDD	0.0033	0.015
	O_8CDD	0	0
	2,3,7,8-TCDF	0.034	0.12
	2,3,4,7,8-P_5CDF	0.1	0.29
	H_6CDF	0.0068	0.013
	1,2,3,4,6,7,8-H_7CDF	0.00068	0.0021
	O_8CDF	0	0
Total Equivalents:		**1.04**	**4.72**

the body mass would serve to dilute the contaminant concentrations ingested during the first year. It is assumed that a North American adult grows to 70 Kg and an individual in Vietnam and Japan grows to about 55 Kg.

The ratio of the first year intake to the recommended intake values of EPA and CDC are shown in the last two columns of Table 33. For a cancer risk of 1-6, the 70 year lifetime recommended intake values for 2,3,7,8-TCDD are 160 and 16000 pg/Kg for the EPA and CDC, respectively. According to these criteria, a child in North America will have consumed between 189 and 858 times the lifetime limit recommended by the EPA. If there is no further ingestion or excretion of contamination then the adult will have received 27 to 124 times the EPA recommended limit. If one considers the CDC values instead than a North American child will have received between 1.8 and 8.2 times the CDC recommended dose or between 0.27 and 1.2 times the CD recommended dose over the person's lifetime.

There is clearly a substantial difference between the 2,3,7,8-TCDD equivalents found in south Vietnam in 1970 and those found in 1984. The groups tested were different and reflect the use of Agent Orange for military purposes. The derived data for the industrial nations of south Vietnam, Japan, and North America are very close and are contrasted with data for north Vietnam where there is infrequent use of synthetic industrial and agricultural chemicals.

The "Acceptable Daily Intake" or ADI method for the determination of health risk from chemicals presupposes that there is a threshold level below which there is no biological response. A safety factor is then applied to account for variations in individual susceptibility. Current ADI values for 2,3,7,8-TCDD are predicated on a threshold value of 1 pg/Kg/d derived by Kociba *et al.* in 1978. Danish authorities chose a safety factor of 200 to define a "Tolerable Daily Intake", or TDI, value of 5 PPq.

A second method of risk assessment is the cancer statistical method which calculates for a given risk level of usually 10^{-6}, the amount of chemical exposure which can occur over a 70 year lifespan. The model assumes a constant dose level in a population and an incremental unit risk.

The EPA has estimated that the incremental unit risk derived from the ingestion of 1 ppq of 2,3,7,8-TCDD in the body over a lifetime may lead to an additional 156 cancer incidents per million people exposed.[81] At this point in time it is not clear what the health risks of the effects of PCDDs and PCDFs are to humans.

TABLE 33
North Americans may consume between 189 and 858 times the lifetime limit of PCDDs/PCDFs recommended by the USEPA during the first year of limit

Population	Daily intake (pg/Kg/d)	First year intake (pg/Kg)		Ratio of first year intake to ADI uncorr./corr.	
South Vietnam	908	332,491	EPA	2077	376
(1970)	(4377)	(1,602,221)		(10,013)	(1,817)
South Vietnam	88.7	32,389	CDC	20.1	3.6
(1983)	(35)	(127,612)	EPA	202	38
				(796)	(149)
			CDC	2.1	0.4
				(8.3)	(1.6)
North Vietnam	5.1	1,853	EPA	11	2.1
	(14.0)	(5,113)		(31)	(5.7)
			CDC	0.08	0.03
				(0.22)	(0.08)
Japan	86	31,536	EPA	197	36
	(146)	(53,144)		(332)	(60)
			CDC	2.0	0.4
				(3.3)	(0.6)
United States	83.1	30,330	EPA	189	27
	(337)	(137,699)		(858)	(124)
			CDC	1.8	0.27
				(8.2)	(1.2)

PCB Degradation in Building Fires The formation of pyrolysis products will ultimately depend upon factors such as the following:

- the time that the askarel mixture is at a temperature which allows a reaction yield of significance,
- the volume and surface area of droplets of askarel emitted in an eventful failure,
- the availability of oxygen and
- the effect of soot particles on the dissipation of heat and the availability of oxygen.

Real-world experience has shown that the eventful failure of electrical equipment can be all too frequently catastrophic. In particular, it is irrefutable that transformer fires do produce askarel oxidation products in significant quantities. As discussed earlier, studies have indicated that the breakdown of askarel transformer dielectric systems due to multi-stress aging may produce toxic products at concentrations of concern. The yield and distribution of products can be expected to depend upon the availability of oxygen. The major quantity of toxic products will undoubtedly be derived after the failure of the transformer has resulted in a fire situation.

When a transformer is retrofilled to <500 ppm PCB, one would expect that the rate of conversion of PCBs would be very much less than in the case of an askarel fire. It becomes increasingly less probable that a large enough quantity of pyrolysis products would be produced as the concentration of PCBs and chlorobenzenes is reduced. These qualitative expectations were realized when a fire occurred in the basement transformer vault of the Electrical Engineering laboratory of the University of Manitoba in 1982. The six transformers involved in the fire contained mineral oil contaminated with 250 ppm PCB. An electrical failure caused one of the transformers to arc and catch fire. All six transformers were damaged. During fire-fighting attempts a transformer exploded and distributed PCB contaminated soot onto the walls. No PCDFs were found. PCDDs, which are typically formed at an order of magnitude less concentration than PCDFs, were not detected either.

The fire which occurred on February 5, 1981 in the Binghamton State Office Building in Binghamton, New York, involved a PCB filled transformer located in the basement. It has been estimated that between 50 and 100 gallons of PCB dielectric fluid were involved in the fire. The building was closed and sealed shortly after the fire and remains closed today in 1996. PCBs and oxidation products were distributed throughout the 18-story structure via two ventilation shafts. Analysis of soot samples revealed high concentrations of PCBs, PCDFs, PCDDs and also other classes of toxic chlorinated aromatic compounds (Table 34).

The concentration of compounds of concern in the soot particles varied between floors in the building. A composite sample was analysed and found to contain three orders of magnitude more of 2,3,7,8-TCDF (270 ppm) and 2,3,7,8-TCDD (2.8 ppm) than fireplace soot. The average PCB concentration in the air was 1.48 $\mu g/m^3$ while dust samples in the spaces above ceilings averaged 1,995 $\mu g/m^2$.

Exposed horizontal surfaces in offices had 162 $\mu g/m^2$ while vertical surfaces and walls had about 7 $\mu g/m^2$. PCB contamination was found inside closed filing cabinets and

TABLE 34
There is a growing list of major PCB-related fire incidents

Location	Date
Toronto, Ontario, Canada	December 1977
Boston, MA (Saniford Street)	October 25, 1981
Boston, MA (Hancock Building)	February 1983
Boston, MA (Quincy Market)	January 14, 1980
Binghamton, NY (BSOB)	February 5, 1981
Pittsburgh, PA (Main Post Office)	September 4, 1981
Tulsa, OK (Beacon Building)	December 29, 1983
Jersey City, NJ (Railway)	June 2, 1982
San Francisco, CA (One Market Plaza)	May 15, 1983
Chicago, IL (First National Bank)	September 28, 1983
Syracuse, NY (Chimes Building)	December 21, 1983
Jacksonville, FL (American Electric)	January 29, 1984
Miami, FL (New World Tower)	May 29, 1984
Duval, FL (Country Courthouse)	June 5, 1984
Florida (J.C. Penney)	September 4, 1984
Montreal (Hydro Quebec Research Lab.)	November 20, 1984
Grand Basile, Quebec, Canada (Warehouse)	August 23, 1988
New York, NY	March 4, 1990

desk drawers at levels of about 75 μg/m^2. The current cost of the cleaning operation is estimated at $46,000,000 and exceeds the cost of construction. The expense to dispose of the building and replace was estimated at about $100,000,000 in 1981. Approximately $1.2 billion in law suits have been entered against the State in this incident (Table 35).

The initial cleanup activity was to discard all the furniture and inexpensive articles covered by soot. Some of the larger items which were destroyed were as follows:

- 930 desks;
- 522 tables;
- 850 file cabinets;
- 110 map files;
- 200 typewriters;
- 15 postage machines;
- 15 computer terminals;
- 1,950 chairs;
- 325 bookcases;
- 310 storage cabinets;
- 292 benches, stools, couches,
- 5 microfiche readers.

A 26,000 gallon water treatment plant was installed in the basement to handle the PCB-contaminated rinse waters. In addition, a pollution control system was constructed to eliminate the flow of potentially contaminated air from the building to the outside environment.

The magnitude and scope of the remedial activities were as follows:

- inventory of all equipment and records;
- removal of furniture and office equipment;
- open all duct shafts for cleaning;
- shred and bale all paper materials;
- removal of all carpet;
- removal of all bathroom and corridor accessories;
- installation of temporary lighting on 16 floors;
- vacuum exhaust dust shafts in bathrooms;
- removal of exhaust dust shafts from bathrooms;
- removal of duct work in basement;
- removal of insulation from terminal boxes in basement;
- removal of records and shelving from storage in basement;
- washing and vacuuming all office areas;
- removal, cleaning and replacement of all lighting fixtures and terminal boxes;
- removal of vinyl and ceramic tile flooring;
- removal of kitchen equipment;
- washing and vacuuming of all floors, stairways and chases;
- washing of 2,160 floor lights, 130 lobby lights, over 9,600 restroom accessories, 3,800 ceiling accessories and 2,500 heating accessories;
- washing and vacuuming of the basement with removal of insulation and plaster ceiling;
- washing and vacuuming elevator shafts and pits;
- removal of fire-proofing from restrooms;
- sealing ventilation chases; and
- disposal of 2,646 cubic yards of carbon (Table 36).

A PCB transformer fire also occurred in the transformer vault at One Market Plaza, San Francisco on May 15, 1983. The vault contained three transformers which contained Aroclor 1242 as the dielectric fluid, undiluted by chloro-

TABLE 35

PCDD and PCDF concentrations in soot from the state office building in Binghamton, NY

Compound	Concentration μg/g		Concentration μg/g
Tetra-CDF (total)	28	Tetra-CDD (total)	1.2
2,3,7,8-CDF	12	2,3,7,8-CDD	0.6
Penta-CDF (total)	670	Penta-CDD (total)	5.0
1,2,3,7,8-CDF	310	1,2,4,7,8-CDD	2.5
2,3,4,7,8-CDF	48		
Hexa-CDF (total)	965	Hexa-CDD (total)	4.7
1,2,3,4,7,8-CDF	310		
1,2,3,6,7,8-CDF	150		
2,3,4,6,7,8-CDF	10		
Hepta-CDF (total)	460	Hepta-CDD (total)	7.0
Octa-CDF	40	Octa-CDD	2.0
Total PCDF	2163	Total PCDD	19.9

benzenes. A faulty electrical cable caught fire and fell onto an energized askarel transformer. The transformer circuit breaker failed and allowed the unit to remain energized for at least three hours. Approximately 30 gallons of Aroclor 1242 leaked from the transformer during the fire.

Soot laden smoke issued from the sidewalk grating adjacent to the building for about three hours and contaminated the switch gear room and some areas of the adjacent parking garage. The fire activated a ventilating air system within the building which was designed to draw air from the ground level intake to the first and second basements and to various retail stores on the plaza level as well as to floors 2 through 6 of the Steuart Street Tower. Dry wipe samples showed contamination of about 720 μg TCDF/100 cm^2 and the building was closed.

The maximum concentrations of PCBs and PCDFs in air were calculated on the basis of measured component ratios. For example, the ratio of 2,3,7,8-TCDF to PCBs was found to be about 1 : 240,000 in air inside the vault but 1:160,000 in air on the 21st floor. Note that PCDDs were not formed because there were no chlorobenzenes used in the dielectric liquid. Analysis of the soot collected in the vault showed the presence of the same classes of compounds as found in the Binghamton incident. The conversion of PCBs to PCDFs seemed to have occurred with approximately the same efficiency as in Binghamton.

The cost of cleanup was close to $20,000,000 with one additional million dollars spent to replace the electrical equipment in the vault. The largest civil suit was that brought by the Fire Department for $100 million.

In September 1983, a fire occurred in a transformer vault under the plaza on the same block as the First National Bank building in Chicago. Smoke issued from a sidewalk grating for about 45 min. but, fortunately, significant PCB contamination was limited to the vault. It was estimated that 1–4 gallons of dielectric fluid were pyrolyzed but, interestingly, no PCDFs or PCDDs were detected in the soot. Cleanup costs were estimated to be $50,000–$100,000 and no civil damage suits were filed.

The known costs of the fire related incidents described in Table 34 has led the EPA to assess an average clean-up cost per incident of $20,000,000. PCDFs and PCDDs are recognized as ubiquitous environmental contaminants. The following Table 37 shows the background contamination levels in two office buildings in Boston, MA. Neither building had been involved in an electrical transformer failure. Two sets of data are shown, one for high skin contact areas and one for elevated horizontal surfaces. A high skin contact area is taken to mean one in which a person might have prolonged direct dermal contact such as desks, tables, counters and file cabinets. Elevated horizontal surfaces are those six feet or more above the floor such as the tops of storage cabinets or bookshelves. Background levels of

TABLE 36

Determination of PCB : PCDF ratios were used to calculate maximum air concentrations

Location	PCB μg/m^3	Total TCDF[1] pg/mg^3	2,3,7,8-TCDF[2] pg/m^3
Fan Room (B2)	2.7	100	11
Switchgear Room (B2)	16	580	67
Transformer Vault	556	20,000	2,300
Garage (B2)	3.2	120	13
Woodshop (B1)	3.0	110	12
Outside Vault			
Perimeter	1.3	48	5.4
Clean Areas	<1.0	<37	<4

[1] Based on a worst-case ratio of PCB : Total TCDF = 27,000.

[2] Based on a worst case ratio of PCB : 2,3,7,8-TCDF = 240,000.

TABLE 37

Background concentrations of PCDFs and PCDDs in Office Buildings

	PCDFs ng/m^2		PCDDs ng/m^2	
	Mean	Range	Mean	Range
High Skin Contact	0.98	0.5–1.9	12	4.1–42
Elevated Horizontal	2.4	0.47–5.4	40	2.8–146
Overall	1.7	0.47–5.4	26	2.8–146

J.R. Kominsky and C.D. Kwoka.[87] Background levels of PCBs have been investigated by NIOSH. Their data are shown in Table 32 for comparison.

PCBs in non-manufacturing buildings have been investigated by NIOSH. Their data are shown in Table 38 for comparison.

Reported eventful failures of PCB contaminated electrical equipment have been well documented only in the United States, Canada, Finland and Sweden. Table 39 lists some of the major PCB fires in Scandinavia. The toxic Substances Control Act was the first regulatory authority which the EPA had concerning PCBs. Section 6(e) of TSCA requires proper disposal of PCBs, and prohibits the manufacture, processing, distribution in commerce, and use of PCBs. Under section 6(e)(2) of TSCA the EPA allows the use of PCBs in "a totally enclosed manner" such as transformers, capacitors and electromagnets based upon considerations of cost and risk reduction. However, when the August 1982 rule was promulgated it was based upon the assumption that the principal route of release of PCB-containing dielectric fluid into the environment was from leaks and spills. Since that time, it has been found that fires involving transformers occur frequently enough to present a significant risk and this has led the EPA to reconsider the August 1982 rule and place the following additional restrictions and conditions on the use of PCB transformers (40 CFR Part 761:50 FR 29170) July, 1985:

1) High secondary voltage PCB transformers (480 V and above, including 480/277 V) configured in a network fashion and used in or near commercial buildings must be removed from use, placed into storage or disposal, disposed, or reclassified to PCB contaminated or non-PCB status by Oct. 1, 1990.
2) PCB transformers can no longer be installed in commercial buildings after Oct. 1, 1985.
3) PCB transformers used in or near commercial buildings (other than high secondary voltage network PCB transformers) must be equipped with enhanced electrical protection, by Oct. 1, 1990, to avoid failures and fires from sustained electrical faults.
4) All transformers must be registered with appropriate emergency response personnel and with building owners by Dec. 1, 1985.
5) All PCB transformer locations must be cleared of stored combustible materials by Dec. 1, 1985.
6) All PCB transformer fire-related incidents must be immediately reported to the National Response Center, and measures must be taken as soon as practically and safely possible to contain potential release of PCBs and incomplete combustion products to waterways.

EPA defines commercial buildings to include all types of buildings other than industrial facilities and would include locations such as office buildings, shopping centers, hospitals, and colleges. A PCB transformer located in or near a commercial building is located on the roof of, attached to the exterior wall of, in the parking area of, or within 30 meters of a commercial building.

The EPA has had to consider both the benefits of PCBs as well as the availability of substitute materials balanced against the costs of regulatory control measures. It was concluded that the removal or retrofill of PCB transformers is both the most effective and the most costly measure to reduce the frequency of serious transformer fires. It was suggested that a less costly, but also less effective, alternative could be represented by providing better electrical protection of the equipment. The effectiveness of increased electrical protection is expected to approach that of phaseout/retrofill but EPA recognizes that electrical protective devices are also subject to malfunction and that PCB transformer fires can result from less common mechanisms of failure. At this point it would be useful to clarify what is meant by the retrofill and reclassification of PCB transformers. The retrofill of a PCB filled transformer involves the replacement of the original PCB dielectric fluid with a substitute oil. The main advantage of retrofilling is that an owner's liability is minimized at minimum cost. In addition, record keeping and reporting requirements are reduced or eliminated and servicing is allowed.

There are three categories of PCB transformers:

- A PCB transformer is defined as one which contains more than 500 ppm PCB.
- A PCB contaminated transformer is one which contains PCBs in the concentration range 50–500 ppm PCB.
- A non-PCB transformer contains less than 50 ppm PCB.

If the flushed carcass is buried in an approved chemical waste landfill there is the possibility of long term liability associated with PCB leaching. Alternatively, carcasses can be cleaned to less than 10 μg PCB/100 cm^2 and the clean metal returned to the secondary metals market.

PCB transformers may not receive servicing which requires the removal of the core/coil assembly. Inspections must be made for leaks and records kept. When the unit is to be replaced, the fluid must be burned in an approved incinerator and the transformer carcass either incinerated or flushed. PCB contaminated transformers (50–500 ppm PCB) do not need to be inspected for leaks and may be repaired or serviced as necessary. EPA reaffirmed its August 1982 determination that the continued use of PCB contaminated transformers does not present unreasonable risks to public health and the environment.

Non-PCB transformers, containing less than 50 ppm PCB, may be used with almost no restrictions since the unit is then not under TSCA purview. The may 1979 PCB ban rule prohibits rebuilding of PCB transformers but does allow for their reclassification. Under the provisions for reclassification the equipment must be put back into service for three months before the bulk oil is tested for its PCB concentration. The final rule clarifies the definition of in-service use for transformers by specifying a minimum dielectric fluid temperature of 50°C to correspond to a condition of light electrical loading. If a unit is retained as a PCB transformer an immediate cash outlay is avoided but this might not be the lowest-cost option in the long term.

TABLE 38
Background level of PCBs found in swipe samples of surfaces in non-manufacturing buildings

Location	Aroclor	n/N1	PCB (ng/m2)	
			Range	Mean2
Cincinnati, OH	1242	31/60	0.5–4.5	1.3
St. Paul, MN	1260	14/18	0.5–5.0	2.0
Boston, MA	1254	30/32	0.1–3.2	0.7
Pine Bluff, AK	1254	4/4	3.5–7.2	5.1
TOTAL:		79/114	0.1/7.2	1.4

[1] n = number of samples above detection limit.
N = total number of samples.

[2] Mean = mean of detected values only.

TABLE 39
Major PCB fires in Scandinavia have involved mostly capacitors

Location	Date
Stockholm, Sweden	August 25, 1981
Stovde, Sweden	March 19, 1982
Arvika, Sweden	May 1982
Imatra, Finland	August 2, 1982
Helsinki, Finland	August 1982
Surahammar, Sweden	September 23, 1982
Hallstahammar, Sweden	November 8, 1982
Locmotive, Sweden	Winter 1982/1983
Kisa, Sweden	April 25, 1983
Halmstad, Sweden	August 15, 1983

For example, if the unit should require servicing at some time in the future it will have, instead, to be scrapped. If tests showed that the transformer was about to fail it would not be possible to retrofill it and service the unit as a PCB contaminated or PCB free unit until it had been set back in operation for 90 days after the retrofill. This may be neither possible nor feasible, depending upon the condition of the transformer. If an owner makes the decision not to continue to operate a unit as a PCB transformer he has then to choose between replacement with a new unit (retrofit) or retrofill of the old one. The age and performance history of the transformer become factors as well as how long the unit is likely to be needed.

TRANSFORMER RETROFILL OPERATIONS

The retrofill operation, as it is conventionally practiced today, consists of multiple stages, each of which involves draining, flushing and filling. The process can apply to two different situations, viz. (1) contaminated mineral oil transformers with >500 ppm PCB and (2) askarel filled transformers in which the coils are impregnated with PCB.

(1) In the case of a PCB contaminated mineral oil transformer it would be possible to reduce the PCB concentration of the working fluid by first draining the transformer and refilling it with non-contaminated oil. Since the PCBs are contained predominantly in the oil, and are not impregnated into the porous insulation, it is possible to retrofill the transformer to a required concentration level by repeated washings, as necessary. The effectiveness of the washings becomes increasingly important as the concentration of the original contamination increases since repeated dilution of residual contamination produces several times the original volume of the transformer tank for disposal. In addition, the presence of trapped, high concentration oil in the core/coil assembly, which is not easily removed by conventional flushing procedures, causes the bulk oil to become re-contaminated.

(2) In the case of an askarel transformer which is impregnated with PCBs, a simple flushing or vapor cleaning process leaves about 2.5% of the original askarel contained in the core/coil assembly. For a typical 250 gal. askarel transformer the impregnated fluid amounts to about 6 gal. A PCB concentration of 500 ppm PCB in a 250 gal. transformer is equivalent to little more than about one coffee mug in volume.

The equilibrium distribution of PCBs between transformer solid materials and the liquid dielectric in a 500 KVA transformer has been found to be about 97.5% in the bulk oil and 2.5% distributed unevenly among the paper, core steel and wire. The vast majority of the PCBs in the solid materials of an impregnated transformer is contained in the paper.

The diffusion of PCBs under the action of a strong concentration gradient from the interior of the paper towards the very low concentration in the bulk retrofill fluid is a slow process governed by Fick's laws. If the retrofilled transformer is not put into operation and remains at ambient temperature, then the viscosity of the askarel may be high enough to prevent diffusion at a significant rate. However, the EPA requires that the reclassification procedure should put the transformer under a normal load or, alternatively, that the bulk oil is heated to more than 50°C for a period of 90 days following the completion of the retrofill operation(s). The power losses of the transformer under load conditions produce heat which, in turn, reduces the viscosity of the impregnated askarel and increases its diffusivity. If the concentration level of PCBs can be reduced to such an extent that subsequent leaching does not yield a concentration which is above the 500 ppm PCB limit imposed by the EPA 90 days after the process has been completed then, according to the letter of the law, the unit can be reclassified as "PCB contaminated". However, experimental data on units which have been retrofilled with a simple vapor cleaning process have shown that the PCB leaching rate under normal load conditions is in the range of 1.5–2.5 ppm PCB per day. If the trapped PCBs are able to leach out into the bulk retrofill fluid at a rate of as little as 1 ppm PCB per day, it would need approximately one year before the fluid concentration had climbed above 500 ppm PCB and the unit would have to be regarded once again as a PCB transformer.

The following cost elements enter the business alternatives of replace vs. retrofill:

Replacement

- Basic cost of transformer and dielectric coolant.
- Miscellaneous fittings: stress cones, etc.
- Possible rental of mobile substation during changeover.
- Crane rental.
- Labor (probably at overtime rates).
- PCB packaging, transport and disposal.
- Purchase cost of flushing solvent to prepare transformer carcass for landfill disposal.
- Packaging, transport and disposal of flushing solvent.
- Transport to and disposal in EPA-approved landfill for transformer carcass.

Retrofill

- An initial, fundamental decision must be made concerning whether to change the classification of the transformer to "PCB contaminated" i.e., <500 ppm PCB, or to <50 ppm PCB to remove the unit from TSCA purview.
- Dielectric fluid costs.
- Solvent flushing fluid, depending upon the retrofill method.
- Transport of oil, solvent and materials to and from the site.
- Gaskets, parts, hosing, fittings etc.
- PCB disposal, solvent disposal.
- Downtime 4 h to 48 h, depending upon the level of decontamination required and the process used.
- Repeated electrical outages, depending upon the retrofill process chosen.
- PCB-in-oil analyses, dielectric fluid properties report.

The following cost comparison compares the options of retrofill or replacement of askarel filled transformers. The comparison is complicated by the many variables which are involved and therefore a range of costs has been considered as they apply to two different sizes of PCB filled transformers.

EPA has completed an analysis of the costs of retrofilling PCB transformers to reduce the PCB concentration to below 500 ppm PCB and has estimated that the cost will vary in the range from $15,500 for a 50 KVA transformer to $32,000 for a 3,000 KVA transformer. The example above is for an "average" or "typical" size transformer or about 225 gal. oil capacity and shows that the cost savings over replacement will be about $20,000 per unit. Clearly, the larger and more inaccessible a transformer the more cost effective retrofilling becomes (Table 40).

Despite the extensive work which has been done over the last few years on the validation of different fire hazard assessment methods for askarel replacement fluids there is still no consensus of opinion. Experience derived from the

TABLE 40
Transformer description

	Transformer 1	Transformer 2
kVA	500	1500
Location	Indoors	Indoors
Primary Voltage (110 kV BIL)	12470Δ	13800Y/7970
Secondary Voltage	480Y/227	480Y/277
°C Rise	65°C	55/65°C
Oil Cooled	OA	OA
Winding Temperature Indicator	Yes	No
HV Connection	Throat	Throat
LV Connection	Throat	Throat
Pressure Relief Device	Cover mounted	Cover mounted
Sudden Pressure Relay	Cover mounted	Cover mounted
Taps	2 ± 21/2%	2 ± 21/2%
Dimensions	4′ × 6′ × 6′	5′ × 7′ × 7′
# of Gallons Oil	190	225

(*continued*)

TABLE 40 *(continued)*
Transformer description

Retrofill Costs ($)	Transformer 1	Transformer 2
○ Askarel removal, Vapor phase degreasing, Webber reclassification process	8,000	13,250
○ Transportation	1,000	1,000
○ Food/Lodging	500	500
○ Oil	1,906	2,257
○ Testing	600	600
○ Supplies	1,500	1,500
○ Equipment lease	133	133
○ Insurance	200	200
○ Contract coordination	1,459	1,459
○ Askarel disposal	2,000	2,400
○ Transportation charges of PCB waste disposal	1,000	1,000
○ Miscellaneous	1,000	1,000
Sub-Total	**$19,300**	**$25,300**
○ Inspection savings	−800	−800
TOTAL	**$18,500**	**$24,500**

Transformer Replacement ($)	Transformer 1	Transformer 2
○ Replacement cost	$14,376	$24,734
○ Change out flexible leads, parts, etc.	1,320	1,650
○ Labour of transformer changeout	2,240	2,240
○ Crane rental	1,200	1,200
○ Disposal of PCB liquid waste	1,600	2,000
○ Tank and winding solvent	855	1,000
○ Flushing solvent disposal	1,600	2,000
○ Labour of flushing PCB transformer	420	420
○ Disposal of PCB transformer carcass	2,880	4,900
○ Transportation	1,500	1,500
○ Fluid transportation for disposal	1,000	1,000
○ Cost of drums and drum disposal	500	640
○ Protective clothing miscellaneous supplies	1,000	1,000
TOTAL	**30,500**	**$44,284**

COST SAVINGS	Transformer 1	Transformer 2
○ Transformer Replacement	$30,500	$44,300
○ Transformer Retrofill	18,500	24,500
COST SAVINGS	**$12,000**	**$19,800**
% COST SAVINGS	**39**	**45**

use of less flammable fluids, however, points to the fact that both hydrocarbon and silicone oils do perform satisfactorily and are suitable as askarel replacement fluids.

The decision to retrofill rather than replace a contaminated transformer depends, as discussed above, upon a number of different factors. For example, the cost of new oil and the cost of labor have to be weighed against the benefits of reduced maintenance, service and disposal costs, the impact of public perception, and reductions in the possible costs of leaks, spills or fires. Once a decision has been made to retrofill, the usefulness of the action is governed by the effectiveness of the process to mitigate the problem of PCB leaching from the impregnated coil.

A cleaning process for the decontamination of PCB contaminated mineral oil in electrical equipment has been described in the form of guidelines by the Ontario Ministry of the Environment (1978). The stages involved are as follows:

- The transformer is emptied of bulk PCB by opening the drain valve. The unit is allowed to drain into an appropriate container for at least 24 h.
- The drain valve is then closed or the unit sealed and filled with solvent and left filled for at least 24 h.
- The solvent should then be drained and the transformer rinsed two more times by filling and draining with clean solvent.
- The second and third rinses may be retained for use as the first rinse in decontaminating other equipment.

The PCB solute in the mineral oil is distributed throughout the liquid and solid phases of the transformer. The PCB concentration remaining in the oil at equilibrium can be calculated from the partition coefficients for PCB between the transformer solid materials and the oil and the residual amounts of flushing liquid remaining in the tank between each stage.

Considerable work has been done on the solvent decontamination of electrical equipment and results have been reported by S.H. Hawthorne[91]. It was found that draining alone removed 91–96% of the PCB while a combination of draining and cleaning with trichloroethylene reduced the PCB concentration by 99.76%. Soaking the drained transformer with hot solvent followed by vapor phase degreasing enabled 99.10% of the PCB to be removed. When an extended soak of 36 days was given to the transformer after the cleaning procedure had been completed the amount of PCB removed increased from the above figures to 99.04%, 99.96%, and 99.72% respectively, that is, the remaining PCB concentrations were 600 ppm, 400 ppm and 2,800 ppm.

Hawthorne's data allow some important conclusions:

- A simple transformer flushing procedure is effective in reducing the initial PCB concentration of an askarel filled transformer to less than 500 ppm.
- The PCB remaining after each cleaning process resided almost exclusively in the interstices between the steel core plates and in the paper and insulation of the copper coils, with little or no PCB remaining on the internal surfaces of the case and radiators.
- After the cleaning procedure had been completed the transformers were soaked for 36 days with the solvent from the third soak to remove the most rapidly leached PCBs. During this time it was found that the PCB concentration levelled out to a pseudo-plateau at 3500 ppm PCB in trichloroethylene after 10 days. In other words, 1000 ppm PCB had been leached out of the core/coil assembly in 10 days (100 ppm PCB/day). Also, the PCB concentration did not increase more than about 140 ppm PCB in the 26 days following the start of the plateau. This means that the leaching rate during this period is about 5 ppm PCB/day.
- The use of hot solvent was found to be more effective than cold solvent. Thus, when the retrofill had been completed and the transformer is put back under load, the heated oil can be expected to leach PCBs at a greater rate than if the transformer were not in operation. According to Hawthorne's results the initial leaching rate should be at least 5 ppm PCB/day.
- The lowest PCB concentration attained before the transformer was put back into service was 400 ppm PCB. If the leaching rate remained steady at 5 ppm PCB/day for the 90 days required by EPA, then the concentration would climb above 500 ppm PCB and the unit would remain a PCB transformer, by definition. This assumes that the leaching rate is 5 ppm PCB/day. It is also assumed that the leaching rate remains constant at its initial, and therefore probably highest, rate. Even so, the results illustrate that a simple solvent flushing technique is able to reduce the PCB concentration into the range where the transformer, in the absence of leaching from the core/coil, could be reclassified as a PCB contaminated unit.

An alternative method of cleaning involves vapor degreasing the unit by spraying the cold internal surfaces with a solvent vapor such as trichloroethylene generated by boiling the solvent in an external vessel. The principle of operation is not unlike the soxhlet extraction of materials.

The process offers numerous advantages:

- the volume of contaminated solvent is kept to a minimum;
- the solubility of PCBs in the hot solvent is greater than in cold liquid and therefore the extraction process is less time consuming;
- the low viscosity solvent is able to penetrate the capillaries of the paper and core laminations and remove the PCBs from all the surfaces within the transformer.

The action of the vapor phase degreasing process is to initially dissolve the PCB on surfaces. Beyond this, capillary penetration takes place and the PCB in the outer layer of paper is replaced with solvent. When the transformer is filled with oil and put back into operation, the non-viscous solvent held in the paper quickly diffuses into the much larger quantity of more viscous dielectric fluid. During this time, the PCB concentration of the absorbed solvent is high because the concentration gradient which it sees is between an initial value of zero and about 700,000 ppm PCB in the interior of the paper. The initial leaching of absorbed solvent into the oil therefore gives a rapid rise in bulk oil PCB concentration. When the outer layer of trapped solvent has been replaced by slightly contaminated bulk oil there is again an almost constant concentration gradient but now the PCBs must diffuse through a medium which is about 450 times more viscous than the originally trapped solvent. Diffusion rate is inversely proportional to viscosity. Overall, the observed leaching rate of PCB will depend upon the relative viscosities of the solvent and oil (and hence temperature) and the depth of penetration of the solvent i.e., the distance the PCB must diffuse.

The depth of penetration attainable by vapor degreasing alone is unknown. The results which have so far been obtained indicate that the leaching rate in the first 90 days is sufficient to increase the PCB concentration to more than the allowed 50 ppm PCB limit and, depending upon the efficiency of the operation, can be greater than 500 ppm PCB after 90 days. Thus the depth of penetration of the degreasing solvent, using existing techniques, can only be very small and the main action of the process is to remove bulk PCB from tank walls and core/coil surfaces.

In the Webber process, which is unique, the leaching effect of residual impregnated PCBs is mitigated by a physicochemical process. The method has been proven in a pilot plant to allow the permanent reclassification of transformers in a single operation. The recontamination of the replacement fluid can be either totally eliminated or reduced to such an extent that the transformer stays reclassified for the remainder of its useful life. Further tests are planned to extend and further validate the data so that warranties can be made concerning the future reclassified usage of the retrofilled equipment.

Figure 41 shows a curve relating the increase in residual PCB concentration with time of transformer operation after application of a vapor phase cleaning apparatus to a 750 kVA network transformer which originally contained about 270 gal. of askarel. The curve consists of two main regions, an initially steep slope leading to a pseudo-plateau region which extends for more than 250 days. The initial slope of the curve is about 6 ppm PCB/day for the first 90 days after the transformer was put back in operation and about 1.3 ppm PCB/day thereafter.

In an isotropic medium, Fick's first law of diffusion states that the amount of a substance diffusing through a medium is proportional to the concentration gradient acting as the driving force and a constant for the medium which is the diffusivity, that is,

$$X = -D\frac{\partial C}{\partial x} \qquad (12)$$

where

X = mass of diffusing substance,
D = diffusivity,
C = concentration of diffusing substance,
x = distance.

Fick's second law of diffusion describes the flux of diffusing substance:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial x}\left[D\cdot\frac{\partial C}{\partial x}\right]$$

where

$\frac{\partial C}{\partial x}$ = concentration change measured in the x-direction

For the sake of approximation and mathematical simplicity, assume that D, the diffusion coefficient, is constant.

Then,

$$\frac{\partial C}{\partial t} = D\cdot\frac{\partial^2 C}{\partial x^2} \qquad (13)$$

The diffusion of PCBs through insulating paper into the bulk oil of the transformer can be modeled as follows.

Assume that the x–y dimensions of an impregnated paper are large compared with its thickness so that edge effects can be ignored. Assume that the paper is homogeneous and of thickness $2d$, and also that the bulk oil is maintained homogeneous.

The space occupied by the paper is $-d \leqslant x \leqslant +d$
The space occupied by the oil is $-d - a \leqslant x \leqslant -d$
and, $d \leqslant x \leqslant +a$

where 'a' is the depth of the bulk oil in the x-direction.

The boundary conditions are:

$t = 0$, $C = 0$ at $-d < x < -d$
and, $C = 0$ otherwise.

At equilibrium, the rate at which PCBs leave the paper is equal to the rate at which PCBs enter the paper, that is,

$$a\cdot\frac{\partial C}{\partial t} = \pm D\cdot\frac{\partial^2 C}{\partial x^2} \qquad x = \pm d,\ \mathrm{t} > 0 \qquad (14)$$

The solution to a problem of this sort is given by J. Crank (1956) in terms of an infinite series as follows:

$$\frac{X_t}{X_\infty} = (1+\alpha)\cdot\left[1 - \frac{\alpha}{(\Pi^{1/2}\cdot\beta^{1/2})} + \frac{\alpha^3}{(2\cdot\Pi^{1/2}\cdot\beta^{1/2})} - \frac{\alpha^5}{(4\cdot\Pi^{1/2}\cdot\beta^{1/2})} + \ldots\right] \qquad (15)$$

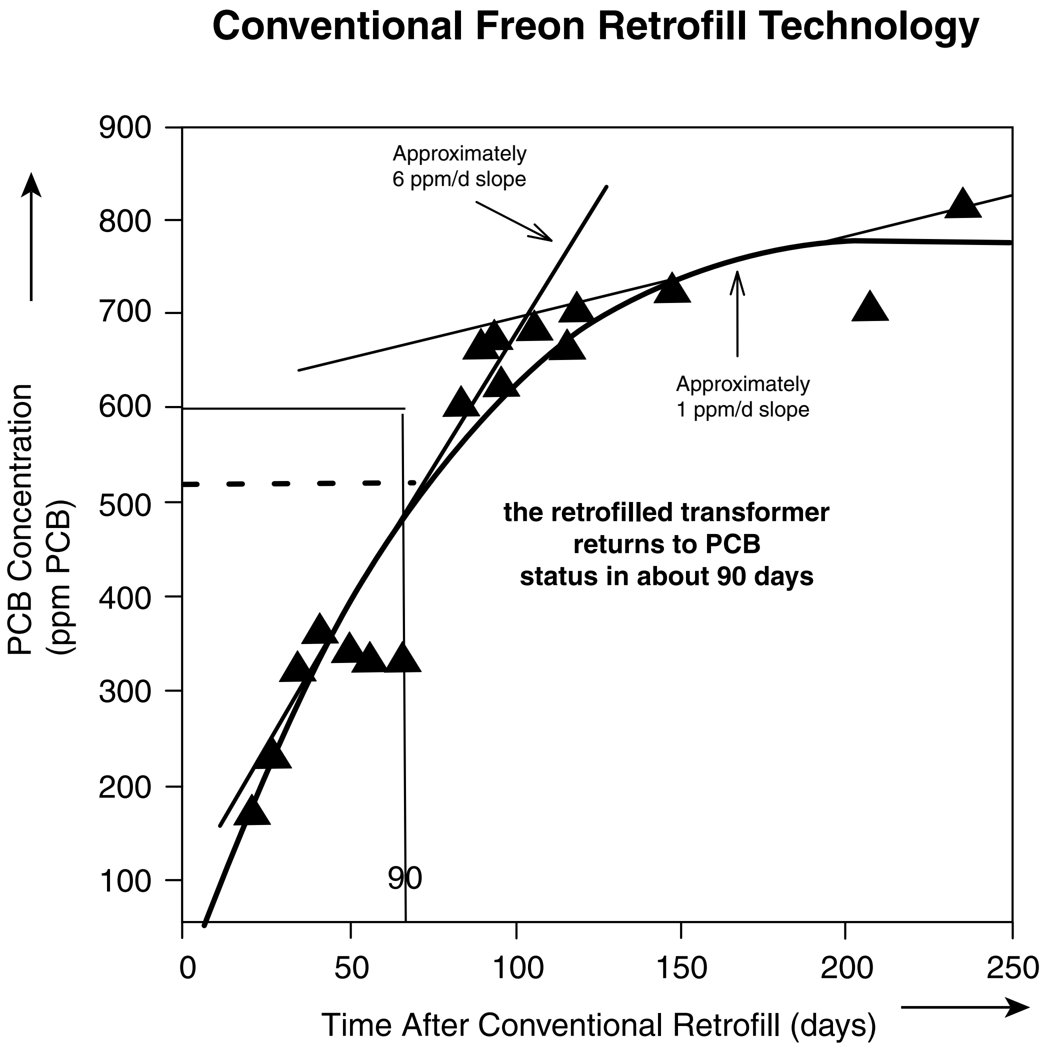


FIGURE 41 Conventional retrofill processes allow continued PCB leaching from the core/coil assembly.

where

X_t = mass desorbed at time, t

X_∞ = mass desorbed at infinite time.

$$\alpha = \frac{1 - \frac{X_\infty}{X_s}}{\frac{X_\infty}{X_s}}$$

where (X_s) is the total solute in the system.

(X_s) and (X_∞) are experimental quantities and therefore (α) can be obtained directly.

A function of time and diffusivity can be defined as:

$$\beta = D \cdot \frac{t}{d^2}. \tag{16}$$

Knowing (α), a series of curves can be drawn relating (X_t/X_∞) with ($\beta^{1/2}$).

The experimental value of (α) is matched to the corresponding theoretical curve of (X_t/X_∞) vs. $\beta^{1/2}$ and since $\beta = D \cdot t/d^2$ a value can be obtained for the diffusion constant, D.

Measurements of the PCB content of impregnated dielectric papers show that a typical paper will contain 58 ± 3% by weight of an askarel such as Inerteen 70/30. Then, to calculate the total quantity of askarel absorbed in the paper one needs to approximate the amount of paper in a typical design.

For a 500 KVA transformer, the approximate quantities of the different types of materials are as follows:

Paper	~	27 lb
Core steel	~	947 in^2
Oil	~	85 gal.

Thus, the amount of askarel impregnating the paper

$= 12.3 \times 10^2 \times 0.58$ g
$= 7{,}100$ g askarel $= 5{,}000$ g PCB.

If it is assumed that the pseudo-plateau shown in the figure above does allow a reasonable approximation of the total PCB leached into the bulk oil within the service lifetime of the transformer then,

$$X_\infty \approx \{X_s - [785 \text{ (ppm PCB)} \times 10^{-6} \times (0.878 \times 8.34) \text{ (lb/gal)} \times 270 \text{ (gal.)} \times 454 \text{ (g/lb)}]\}.$$

If the paper quantity is assumed, for the purpose of calculation, to be proportional to fluid volume from one design to another then,

$$\begin{aligned} X_s &= 12.3 \times 10^3 \times 0.58 \times 270/84 \\ &= 22{,}931 \text{ g} \end{aligned}$$

therefore,

$$\begin{aligned} X_\infty &= 22{,}931 - 705 \\ &= 22{,}226 \text{ g} \end{aligned}$$

that is,

$$X_\infty / X_s = 0.969.$$

The calculation puts an upper bound on the extent of the leaching problem since the model used in the calculation assumes that sufficient time has been allowed to achieve equilibrium. In the experimental arrangement used, the time to reach equilibrium was not a limiting factor since this was long compared with the time required for the diffusion of PCBs into and out of single, as opposed to bundled, sheets of paper. A transformer, however, presents a more complicated situation: the time to achieve true equilibrium may be beyond the service live-expectancy of the transformer since the migration of PCBs through different thicknesses of layered paper is the rate determining factor which governs the total fractional uptake.

The data shown in the above figure for a retrofilled askarel transformer suggest that PCBs will leach at the rate of about 1.3 ppm PCB/day for a period of perhaps years. It is therefore not possible to determine with any accuracy what mass of PCBs migrate into the oil when a final equilibrium is reached. If one assumes that, in the *best* case, a saturation level has been reached after 240 days, then an approximate 'best case' correlation curve can be calculated between $\beta^{1/2}$ and X_t/X_∞ as follows.

From above, $X_\infty/X_s = 0.969$

$$\begin{aligned} \alpha &= \frac{1 - \dfrac{X_\infty}{X_s}}{\dfrac{X_\infty}{X_s}} \\ &= 3.2 \times 10^{-2} \end{aligned}$$

$$\beta = D \cdot \frac{t}{d^2}.$$

Assume that $D = 7.6 \times 10^{-10}$ (cm²/sec)
$d = 1$ (cm)
then, $\beta = 7.6 \times 10^{-10} \cdot t$ (sec).

If the transformer is to be reclassified as 'PCB contaminated' then it must not be greater than 500 ppm PCB after 90 days of operation after the retrofill,

that is, $$\frac{X_t}{X_\infty} = \frac{500}{785} = 0.637.$$

By interpolation,

$\beta^{1/2} = 0.036$ when $X_t/X_\infty = 0.638$
that is, $0.0362 = 7.6 \times 10^{-10} \cdot t$
hence, $t = 19.7$ days.

In other words, an initial clean layer of 1 cm of paper will allow the bulk oil to reach 500 ppm PCB in about 20 days.

The correlation shown in Figure 42 has been calculated between the depth of the cleaned layer and the time required for the PCB concentration of the bulk oil to reach 500 ppm PCB using different values for the diffusion coefficient.

The experimentally determined diffusion coefficient for the migration of Inerteen 70/30 through kraft paper is about 7.6×10^{-10} cm²/sec. Under these conditions, if the cleaned layer is about 4.3 cm deep, the transformer will become a PCB transformer one year after the retrofill has been completed.

If the transformer is to be reclassified as 'non-PCB,' then it must not be greater than 50 ppm PCB 90 days after the retrofill procedure has been completed,

i.e., $$\frac{X_t}{X_\infty} = \frac{50}{785} = 0.064$$

hence, $t \simeq 7$ days.

The correlation between the depth of the cleaned layer and the time of reach 50 ppm PCB is shown in Figure 43 for different values of the diffusion coefficient. Clearly, the time to reach a bulk oil concentration of 50 ppm PCB is much shorter for a given depth of cleaned layer and is critically dependent upon the diffusion coefficient.

Figure 44 shows how the depth of the cleaned layer must vary for different values of the diffusion coefficient in order to obtain a defined reclassification status. If the diffusion coefficient becomes larger than about 10^{-8} or 10^{-9} cm²/sec then the correlation shows that the depth of the cleaned layer must become much greater than it is practicable to achieve.

It is interesting to note that if the cleaned layer is 4.3 cm and the diffusion coefficient is the experimentally measured value of 7.6×10^{-10} cm²/sec, then there is little doubt that the transformer would be reclassified after 90 days. However, after one year of operation it would become a PCB transformer

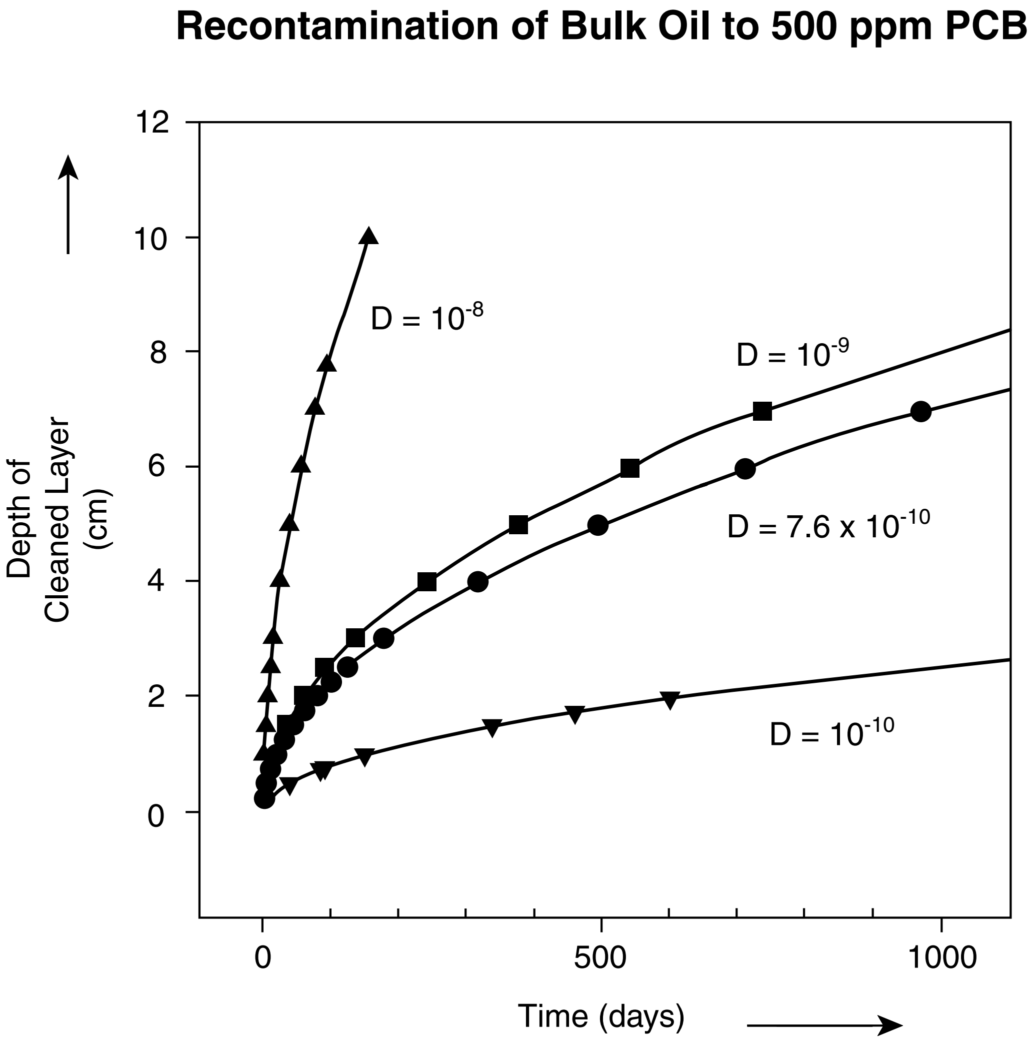


FIGURE 42 The time for the bulk retrofill oil to reach 500 ppm PCB depends on the extent of cleaning and the type of paper used.

again (by EPA definition) *even if it had been initially reclassified as 'non-PCB'*.

From the experimental curve of Figure 41, at $t = 90$ days,

$$\frac{X_{90}}{X_{\infty}} = \frac{560}{785} = 0.713$$

hence, $\beta^{1/2} = 0.0617$

$D = 7.6 \times 10^{-10}\ \text{cm}^2/\text{sec}$(by experiment)

hence, $d = 1.25$ cm.

If the transformer is left in operation long enough to allow the viscosity of the residual impregnating fluid to increase due to the preferential migration of chlorobenzenes then, eventually, the diffusion coefficient will shift from

$$\begin{gathered} D = 7.6 \times 10^{-10}\ \text{cm}^2/\text{sec} \\ \rightarrow D_{\infty} \\ \sim 10^{-10}\ \text{cm}^2/\text{sec.} \end{gathered}$$

The transformer will need to be drained and the oil decontaminated to less than 2 ppm PCB (EPA requirement) when these new conditions are met. An alternative would be to dispose of the contaminated oil as PCB fluid and use new oil to refill the transformer. There is considerable doubt that the diffusion coefficient will become low enough to merit this technique. A further drawback in the method lies in the fact that, during the long waiting period which would

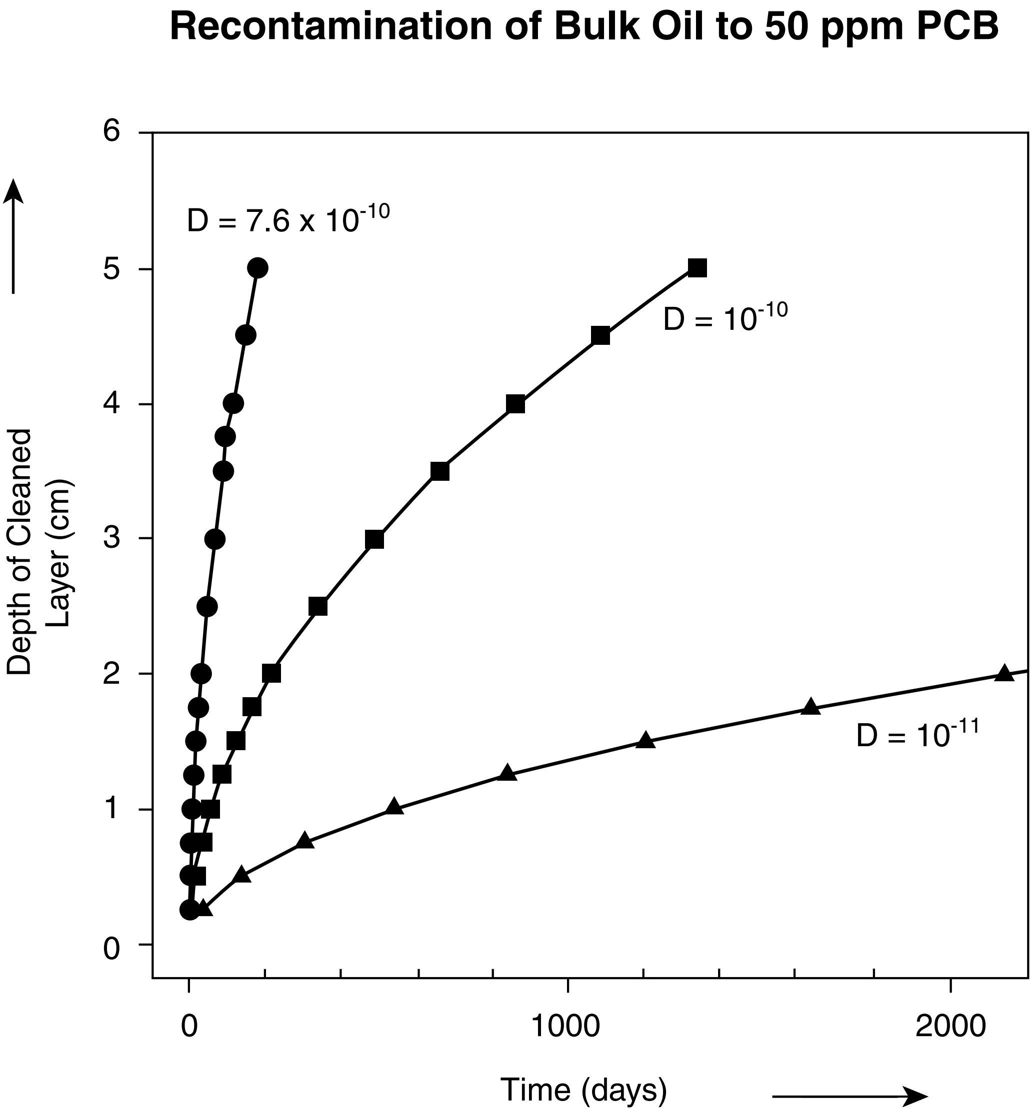


FIGURE 43 Correlations between the depth of the cleaned layer and the time to reach 50 ppm PCB in the bulk retrofill fluid.

be necessary to observe a reduction in the rate of diffusion, the transformer would have to be considered PCB filled as though no retrofill operation had taken place.

The continued leaching of PCBs with a lowered diffusion coefficient of ~10^{-10} cm^2/sec will cause the PCB concentration of the bulk oil to increase above 500 ppm PCB in a period of time which can be calculated as follows:

$$\frac{X_t}{X_\infty} = \frac{500}{785} = 0.637$$

hence, $\beta^{1/2} = 0.036$ as found earlier
hence, $t = 234$ days.

Thus, even if the transformer is reclassified at the end of the EPA mandated 90 day period by a process which involves simple vapor cleaning, leaching and re-retrofill, the technique cannot provide continued, reclassified use of the transformer for its remaining useful life. Indeed, it is not likely to be able to maintain the unit's reclassified status as a PCB contaminated transformer for as long as one year.

If it is hoped that the transformer can be reclassified as a non-PCB unit then,

$$\frac{X_t}{X_\infty} = \frac{50}{785} = 0.0637$$

then, $\beta^{1/2} = 0.0215$
and, $t = 84$ days.

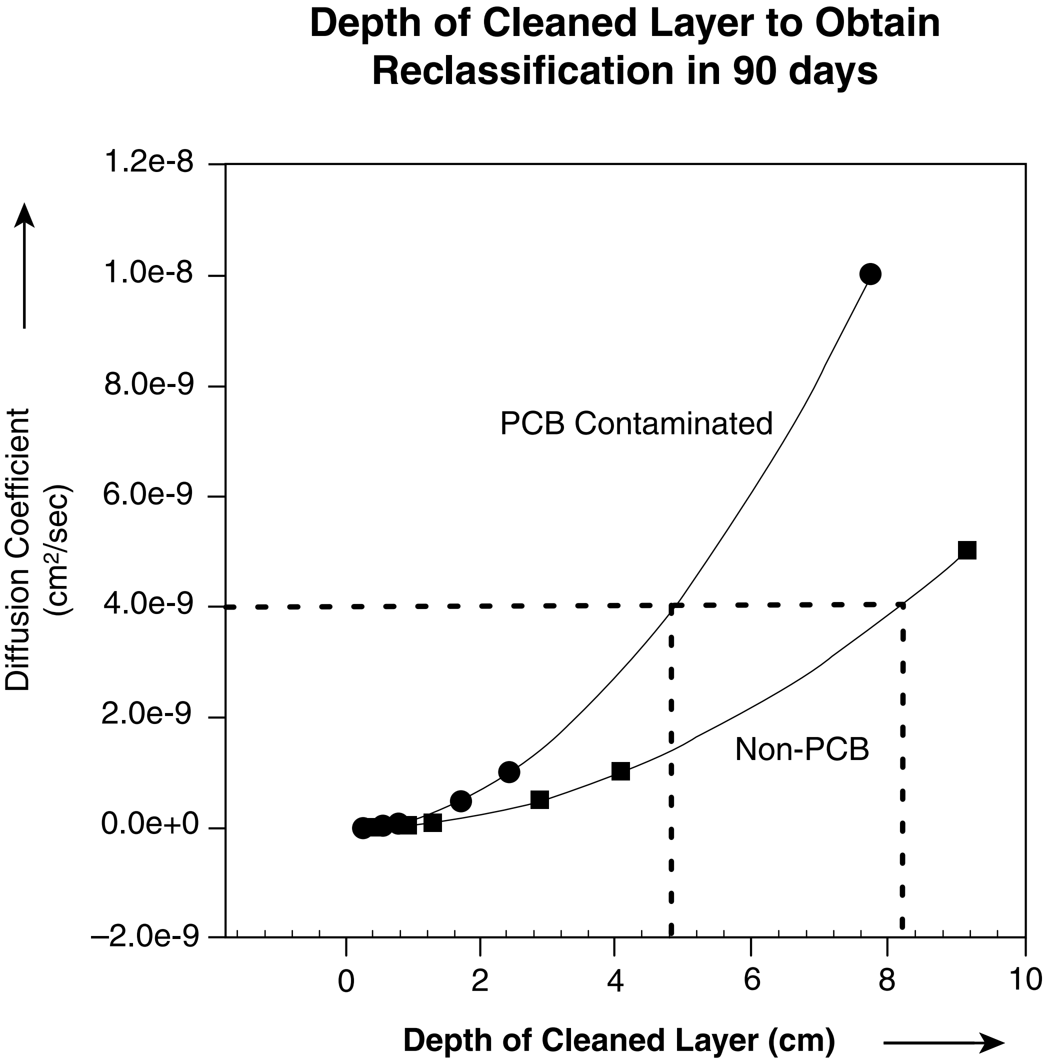


FIGURE 44 For diffusion coefficients $>10^{-8}$ cm²/sec the depth of the cleaned layer to prevent bulk oil contamination becomes impossibly large.

In other words, even if the transformer can be reclassified as non-PCB at the end of 90 days operation, it is likely that it will have to be considered as PCB contaminated very soon afterwards and that it will become a PCB transformer again in less than one year.

The relative importance of the adsorption of PCBs to their absorption can be assessed as follows.

The Freundlich adsorption isotherm is represented by the equation,

$$\frac{X}{M} = k \cdot C^n \tag{17}$$

where X and M are the masses of adsorbate and adsorbent, respectively and 'k' and 'n' are Freundlich coefficients. C is the equilibrium value of the PCB in the bulk oil.

Experiments were conducted at 50°C since this is the temperature required by the EPA when a retrofilled transformer is to be reclassified and also at 110°C representative of the top oil temperature. The following table shows results obtained for two different types of dielectric papers (Table 41).

TABLE 41
Freundlich Adsorption Coefficients

Paper	Temperature (°C)	Freundlich coefficients k	n
Dicy	50	0.49	1.05
	110	2.75	0.77
Diamond	50	1.32	0.78
	110	0.04	1.22

$X/M = 1.32.C^{0.784}$ for diamond paper at 50°C

X = 200 g.

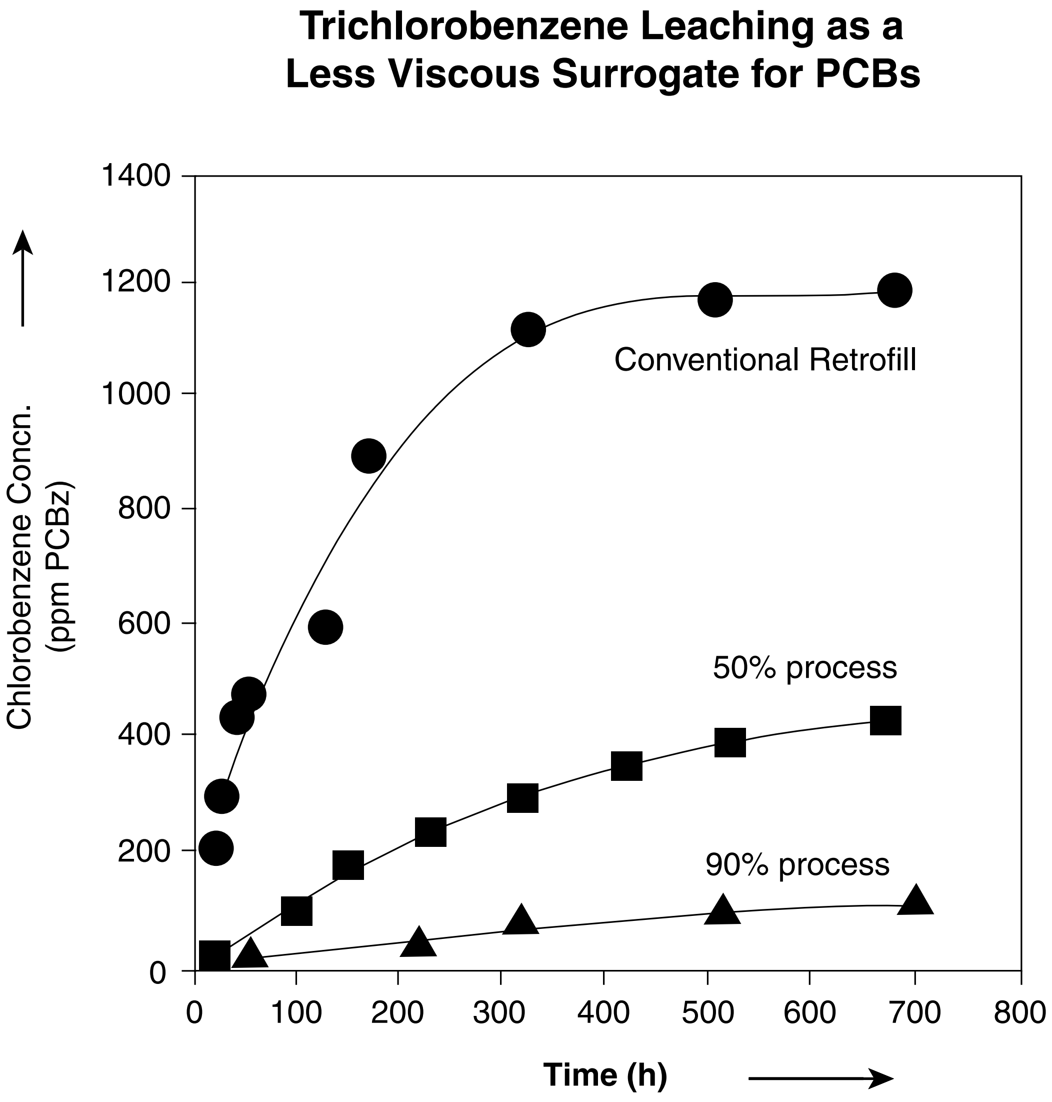


FIGURE 45 The completed process allowed the permanent reclassification of the bulk retrofill fluid.

Thus, the quantity of PCBs adsorbed on the paper is extremely small relative to the total quantity of PCBs adsorbed by the paper.

Figure 45 illustrates the effect which processing can have on mitigating the leaching problem. The data were obtained using chlorobenzenes as a non-viscous surrogate for PCBs in order to more easily quantitate any observed leaching. The completed process allowed permanent reclassification of the bulk retrofill fluid.

The model used virgin transformer materials impregnated with chlorobenzenes. The bulk oil was homogenized before samples were taken and the temperature of the tank adjusted to 50°C. Samples of impregnated kraft paper were wound into a coil and immersed in the oil. Several different kinetic runs were made with different thicknesses of paper as well as different numbers of layers forming the overall wad of paper.

The experimental conditions shown in Figure 45 were found convenient because a series of complete leaching curves could be compared over a period of about 30 days rather than the 300 days found for a 750 KVA transformer. It is clear that sufficient time had elapsed to be able to compare a conventional vapor cleaning retrofill system result with the processing necessary to avoid the leaching problem.

It was found that processing beyond simple vapor cleaning could be continued to the point where leaching was completely stopped. In actual practice, however, such a rigorous level need not be achieved since, if the leaching rate is reduced by, say, an order of magnitude from the measured value of 1.3 ppm PCB/day then, instead of about one year of reclassified use, the unit would retain its classification for ten years. In this example, if the transformer's life-expectancy was only for a further five years then clearly it would remain reclassified for the remainder of its useful life. The extent of processing can be tailored to the level of reclassification required and the life-expectancy of the transformer.

A method which is often used for the PCB decontamination of mineral oil transformers is to drain the fluid into drums for disposal, flush and drain the transformer with a flushing oil, and then refill the unit with uncontaminated mineral oil. The volume of flushing oil needed to reduce the PCB concentration of the retrofilled transformer to less than 50 ppm PCB depends upon the initial concentration of the oil and the amount of residual oil left after each flushing operation.

The PCB concentration after a drain and flush sequence can be predicted according to the following equation:

$$C_1 = C_0 \cdot \left[\frac{R}{100} + \frac{P_s}{P_1} \right] \cdot \left[1 - \frac{P_s}{100} \right] \qquad (18)$$

where

C_1 = PCB concentration in oil after draining and refilling.
C_0 = PCB concentration in oil before draining and refilling.
R = Percent of fluid remaining in the transformer after draining.
P_s = Percent of PCB in solid materials before draining.
P_1 = Percent of PCB in oil before draining.

The mathematics assumes that the oil in the tank is homogeneous and therefore represents a 'best-case'situation. In a transformer, however, the volume of oil trapped in the interstices of the windings are not likely to be removed by a simple drain and flush procedure. Then, when the unit is put back into service, the trapped, highly PCB contaminated, original oil, gradually convects into the retrofilled bulk oil to give an apparent leaching effect. If the trapped, original oil had a PCB concentration of 2000 ppm PCB and becomes diluted by a factor of 40 by convection, then the final concentration of the retrofilled fluid will be 50 ppm PCB.

The higher the PCB concentration in the mineral oil, the more likely it will become necessary to remove trapped oil in order to achieve a retrofilled concentration of less than 50 ppm PCB. The areas which must be cleaned in particular are the core/coil assembly and other regions which contain flow restrictions.

It has been mentioned above that a vapor cleaning system has been proven successful in removing surface PCB contamination from areas previously considered inaccessible. The method has the advantage that a minimum quantity of PCB contaminated fluid is derived for disposal in contrast to the drain and flush method which produces several times the volume of the transformer tank. Also, as the level of PCB contamination in the mineral oil increases, the vapor cleaning method become increasingly necessary and eventually becomes the only viable, cost-effective procedure.

THE VIABILITY OF PCB REPLACEMENT FLUIDS

The need to reduce or eliminate PCBs from the dielectric fluid of transformers is partly because of the generation of compounds of concern in a fire incident. However, it is also important to consider that the original reason for the use of transformer askarels was because they were intended to provide a measure of fire safety.

Fire hazard factors for replacement askarel transformer fluids are neither well defined nor easily estimated, but include:

- the ignition sources available:
- the availability of fire protection systems:
- the ease of ignition of the transformer:
- the close environment of the transformer:
- the effects of burning liquid on the surroundings:
- the depletion of oxygen by burning liquid:
- the smoke evolution from burning liquid:
- the toxicity of combustion products:
- the temperature at which the liquid is used.

The test methodology used to describe a system as complicated as a real transformer fire is complex, and while the factors which affect fire hazards can be listed in broad terms, their interdependent description in an overall assessment is difficult using the rudimentary test methods presently available. More importantly, the interpretation of such results to provide a quantitative measure of the overall fire hazard should be treated with caution.

Ideally, a fire hazard assessment should be predicated on incidence studies in actual usage. Such an assessment is difficult for "less flammable" fluids, i.e., fluids with a fire point above 300°C, because of the lack of statistics on fires in which these fluids are involved.

Section 450-23 of the National Electrical Code specified in the past that the controlling provision for fluids used in indoor transformer applications should be that the insulating liquid have a 300°C fire point. The required temperature has some validity in that it is known that flaming will not persist in a wood slab not subjected to supplemental heating unless the average temperature within the slab is greater than about 320°C.

In the absence of primary current limiting fusing, the intensity and duration of a high energy arc are limited by the recovery capability of the liquid. When there is sufficient energy to produce an explosion, liquid aerosol and gaseous decomposition products are expelled from the transformer as a hot plume. The magnitude and probability of potential loss will be partially dependent upon whether the plume ignites. Subsequently, the effect of the transformer failure will depend upon the ignitability of the fluid and the material surrounding the transformer.

A report was issued by the National Electrical Manufacturers Association (NEMA) in 1980 concerned with "Research on Fire Safety Test Methods and Performance Criteria for Transformers Containing PCB Replacement Fluids". The scenario used by NEMA to describe an eventful failure is as follows:

1) Incipient fault induction period.
2) Growth of incipient fault to large arc.
3) Failure of electrical protection devices to remove the transformer from the line before tank rupture.

4) Tank rupture with scattering fluid, gaseous decomposition products, solid insulation, steel components and molten conductor leading to ignition of nearby combustible building materials and furnishings by the transformer.
5) Scattered materials in static condition within minutes after tank rupture resulting in ignition of combustible building materials and furnishings.

NEMA has recognized that the principal underlying cause of transformer failures involves insulation breakdowns. A typical failure was described as beginning as an insulation failure in the high voltage winding which initiates a low current, high impedance fault. This turn to turn or layer to layer fault progresses rapidly, involving additional turns and layers. The coil impedance decreases and the fault current increases. An alternative failure model begins with a high impedance path to ground along the surface of insulation (tracking).

The following scenario was formulated to serve as the basis for a worst case failure:

1) Due to malfunction, the electrical breakers and/or fuses which are directly associated with the particular transformer are inoperative. The sustained presence of a severe overload current therefore goes undetected, permitting the insulating oil to overheat significantly.
2) As a consequence of the overload, the average oil temperature in the tank reaches 135°C. (Note, the NEMA study accepted this temperature as "an arbitrary but reasonable level to assume under the circumstances.")
3) Insulating deterioration occurs, causing charring or tracking of the high voltage insulation. A fault to ground develops through the train of bubbles or along the insulation surface to ground. The tracking fault then rapidly flashes into a long, high current arc.
4) Once the long arc has developed the sequence involves the following:

    a. The long arc rapidly decomposes the surrounding oil into gases.
    b. A high pressure (several hundred psi) gas bubble develops around the arc under the oil.
    c. The pressurized bubble rapidly accelerates the oil upward towards the gas space and also subjects the tank walls below liquid level to a severe overpressure.
    d. Depending on the tank geometry, failure occurs either near the bottom, below the oil surface or at the top by ejection of the cover.

5) Ignition of the hot oil spray results from one or more of the following sources:

    a. The hot solid particles of insulation and conductor which are produced by the arcing fault.
    b. The arc drawn by the ejected bushing.
    c. The hot gas bubble of arc decomposed oil which, at a temperature of about 3000°C initially drove the oil out of the tank, is itself ejected.

The theory of pressure phenomena due to arcing in liquid filled transformers has been studied and relationships derived between arcing and tank pressure (EPRI project 325[88]). Tests were conducted on different types of arcs to corroborate the results. For example, current flowing through an expulsion fuse produces a higher arc-voltage gradient than does the same current in a free arc in oil. Higher voltage resulted in higher arc energy, and was accompanied by higher peak pressure in the tank. Faults contained within the coil's windings were found to produce less pressure on the transformer than either an open arc or an expulsion fuse. In particular, short length winding faults were found to be less severe than one inch arcs drawn directly in oil. Peak values of below-oil pressures were observed when the melting of a fuse wire initiated the arc, since this resulted in a high, near instantaneous rise in arc current. The maximum pressures developed under the oil for fuse initiated arcs were found to be very high, in the range of 20–30 atmospheres.

Electrical failure resulting in rupture of cylindrical 10 kVA tanks has been reported to exceed 100 kW-sec. but for rectangular tanks this value increases to 800–1400 kW-sec. Values for cylindrical tanks are found to grow rapidly with increasing tank size. Therefore, eventful failure is much more likely in small cylindrical tanks.

The real world test parameters discussed above have been incorporated into the fire hazard assessment model used by UL (Webber[89]) to determine the compliance of transformer fluids with section 450.23 of the National Electrical Code. In order to be complaint with the code, transformers insulated with less-flammable liquids are permitted to be installed without a vault in Type I and Type II building of approved noncombustible materials in areas in which no combustible materials are stored, provided there is a liquid confinement area, the liquid has a fire point of not less than 300°C and the installation complies with all restrictions provided for in the listing of the liquid. UL has identified the need for pressure relief devices and current limiting fusing to limit the effect of possible high current arcing faults.

The purpose of the tests was to determine the flammability of dielectric liquids after an explosion. Approximately ten galloons of each of the fluids under test were placed in separate transformer tanks and preheated to 150°C. Each container had internal electrodes designed to force the arc upward into the gas space. Although the temperature at the point of arcing was several thousand degrees, the duration was only for a few cycles and therefore the temperature of the fluid in a container the size of even a small distribution transformer cannot be significantly affected. The 150°C used in the test was probably higher than would reasonably by expected to occur and therefore the same tests were also run at 120°C. It was found that a fireball was not produced at the lower fluid temperature.

When the test conditions were made sufficiently severe to expel liquid and gas from the transformer at high temperature, a high molecular weight paraffinic hydrocarbon oil produced a smaller fireball, lower fireball temperature and shorter fire residence time than any of the other fluids tested, including silicone dielectric oils. Catastrophic failure tests methods such as that described above are difficult to standardize because of the many factors which contribute to the results. Nevertheless, the scenario used is substantially more realistic than many other fire tests applied to less flammable fluids. For example, ignitability tests to determine whether a fluid would contribute to a fire showed that, under all the test conditions applied, a high temperature hydrocarbon fluid exercised a considerable advantage over other fluids with respect to time of ignition. The advantage was lost in badly conceived experiments where gasoline was added to the fluid as a fire starter to the hydrocarbon oil. Tests to determine the ignitability of fluids are difficult to standardize and have been abandoned because of it. However, the ignition of a combustible material is the first step in any fire scenario and therefore is important to fire prevention.

Since arcing is the most common failure mode of devices incorporating insulating fluids, R. Hemstreet[90] made an attempt to assess spray flammability. His experimental arrangement starts a flow through a nozzle and the liquid temperature is gradually increased until the spray sustains ignition at a measured distance of the igniter flame from the nozzle.

The burning rate in a gas is proportional to the square of the gas pressure. Therefore, for two fluids of different viscosity, or the same fluid at different temperatures, the pressure in the aerosol cloud will be greater for the thinner fluid because more of it can be pumped through the nozzle in a given time. One would therefore expect that the minimum distance of the igniter flame to cause combustion would be larger in the case of a non-viscous fluid than in a viscous one or, equivalently, that the minimum igniter distance will increase for a given fluid as the temperature increases. This is shown to be the case in Hemstreet's results. Clearly, the experimental factors in an apparently simple test to rank fluids can be extremely complicated. For example, it was found that an air velocity of greater than about two feet per second had a considerable effect on the fluid temperature required for ignition as the nozzle distance was increased.

The overall shape of a temperature–distance curve in experiments of this sort is likely to be strongly influenced by the presence of decomposition products, either originally present in the fluid or induced by the igniter flame. The minimum temperature at which a liquid ignites depends on such factors as the degradation produced in the fluid by the flame and prior thermal and electrical stresses.

Heat release rate was measured in Hemstreet's work after removal of the ignition source and measurements are sometimes significantly different from results obtained using a sustained ignition source. Heat release raw data derived from measurements done on a quiescent pool of fluid provide information which has been used to calculate suggested clearances between the burning pool, presumably around the transformer, and the walls and ceilings of the room in which the unit is housed. The test does not consider the interference of silica crust produced over a quiescent pool of silicone oil. In the unlikely event that the conditions necessary to maintain a quiescent condition were to exist in a real building fire, the wicking action of the crust causes the silicone oil to burn longer than high temperature hydrocarbon oils under the same conditions (Webber[89]).

Thermal Decomposition of PCB Replacement Fluids

The suitability of fluids as replacement dielectrics for PCBs depends not only on the electrical properties of the liquids but also their decomposition products. Part of the reason is that the buildup of oxidation products in in-service fluids degrades its electrical properties and also because the decomposition products may be toxic. In particular, combustion products derived from PCB replacement fluids should not be toxic.

Several fluids have been promoted as replacement dielectric fluids in capacitors and transformers, of which, some are listed in Table 42.

The capacitor fluids are usually based on a single major constituent whereas the transformer oil are typically derived from oils or are polyalkylated siloxane oils. Decomposition products derived from single or, at least well characterized, compounds are more easily analyzed than those obtained from complex mixtures such as refined oils. Dielectric mineral oils have been in use in transformers for many years and are not usually considered to represent a high health risk even though they contain many aromatic compounds. Indeed, part of the usefulness of naphthenic oils which gives them an advantage over their paraffinic counterparts, is their ability to solubilize the sludge of oxidation products formed during in-service aging. Table 43 lists the percentage of different classes of hydrocarbons found in naphthenic transformer oils to illustrate the starting complexity with which oxidation studies are faced.

RTEmp transformer oil was developed as a non-toxic PCB replacement fluid and was derived from high molecular weight paraffinic material. The major thermal decomposition products at 600°C in air include benzene, which is known carcinogen, and traces of other compounds, as shown in Tables 44 and 45. Benzene is also formed, in greater yield, when RTEmp is heated[158] at 700°C under nitrogen. It is very difficult to assess the health risks associated with the presence of trace quantities of pyrolysis products. Since the formation of monoalkylated benzenes is two to three orders of magnitude less than the amount which occurs naturally in nontoxic naphthenic transformer oils, it is not likely that the decomposition products formed from high molecular weight paraffinic oils would cause much concern.

Thermal decomposition products derived from PCB replacement capacitor fluids are much more easily characterized and are shown in Table 46. While the thermal decomposition products of PCBs yield PCDF and PCDD products which are toxic at the μg/Kg level, the toxicities of products derived from the alkylaromatic capacitor fluids in the absence of chlorine sources, are toxic at the mg/Kg level. In the presence of chlorine sources, however, it would be possible to generate PCBs as well as polychlorinated methylanthracenes and, conceivably, PCDFs as well.

TABLE 42
PCB replacement fluids and their thermal decomposition products

PCB replacement fluid	Thermal decomposition products
RTEmp	benzene, toluene, cyclopentadiene, phenol, acetic acid
Polydimethylsiloxane	cyclic siloxanes, short-chain hydrocarbons, amorphous silica
Chlorinated benzenes	chlorophenols, PCBs, PCDDs, PCDFs, PCNs, PCDPEs, PCBPs
Tetrachloroethylene	HCl, phosgene
Phthalates	phthalic anhydride, alkenes, alcohols, aldehydes
Phenylxylylethane	styrene, phenylxylylmethane, phenol, methylanthracene
n-Propyl-biphenyl	dihydrophenanthrene
Iso-propyl-biphenyl	biphenyl, biphenylcarboxaldehyde

TABLE 43
Hydrocarbon class analysis of new naphthenic transformer oils

Hydrocarbon class	Gulf transcrest	Esso univolt 60	Esso univolt	Esso EEMAC paraffinic	Sun elect AH	Shell diala A	Gulf Canada hydrotreated
Paraffins iso-/normal	7.9	11.8	20.1	22.7	5.7	7.2	20.2
Monocyclo-paraffins	15.5	17.9	26.7	23.7	11.7	22.5	28.6
Condensed cycloparaffins	39.9	42.1	32.8	32.6	46.3	62.5	50.6
Total Saturated compounds	63.3	71.8	79.6	78.6	63.7	92.2	89.4
Monoalkyl benzenes	7.2	6.9	5.7	5.7	6.8	1.5	0.0
Indans/benzocyclo paraffins	8.9	6.9	5.2	5.2	8.4	2.0	0.2
Benzocyclo paraffins	7.8	6.3	4.0	4.2	8.3	2.4	0.4
Total Monoaromatics	23.9	20.1	14.9	15.1	23.5	5.9	0.6
Naphthalenes	6.0	4.4	2.8	3.2	6.5	1.4	—
Biphenyls/Naphthocyclo-paraffins	2.5	1.6	1.3	1.5	2.6	0.3	—
Fluorenes/Naphthobicyclo-paraffins	2.4	1.2	1.1	1.1	1.4	0.1	—
Total Diaromatics	10.9	7.2	5.2	5.8	11.0	1.8	—
Phenanthrenes/Anthracenes	0.7	0.3	0.1	0.2	0.5	—	—
Total Triaromatics	8.0	0.4	0.1	0.2	0.5	—	—
Total thiophenes	0.0	0.0	0.0	0.2	0.3	—	—
Furans/Benzofurans	0.8	0.5	0.0	0.0	0.7	—	—

TABLE 44
Thermal decomposition products from RTEmp

Decomposition product	% Yield
	700°C/N^2
Cyclopentadiene	0.11
Cyclohexadiene/methylcyclopentadiene	0.055
Benzene	0.28
Toluene	0.13
Ethylbenzene	0.05
Xylene	0.022
Styrene	0.061
C^3-alkylbenzene	0.028
Indene	0.042
Methylbenzofuran	0.012
Methylindene	0.013
Naphthalene	0.038
Methylnaphthalene	0.012
Acenaphthalene	0.0099

TABLE 45
Thermal decomposition products of RTEmp at 600°C in air

Decomposition product	% Yield
	700°C/N^2
Benzene	0.14
Acetic acid	0.060
Toluene	0.020
Styrene	0.0074
Quinone	0.0058
Benzaldehyde	0.016
Phenol	0.089
Benzofuran	0.0036
Naphthalene	0.0058
Phthalic anhydride	0.0063

THE ANALYSIS OF PCBS

The widespread applicability of PCBs since their introduction in 1930 up to 1976 resulted in worldwide environmental contamination. By 1968, PCBs had become the most abundant of the chlorinated aromatic pollutants, rivaling DDE (Risebrough *et al.,* 1968).

Jensen's paper in 1966 confirmed the presence of PCBs in fish after repeated encounters of gas chromatographic elution patterns while analyzing for DDT and other chlorinated pesticides. The earlier failures by researchers in this field to recognize the presence of PCBs must certainly have contributed to the overestimation of DDT in the environment. The shift in emphasis to the analysis of PCBs in environmental matrices after 1966 reversed the roles almost overnight and PCB residues were then described to have interferences from a wide variety of organochlorine pesticides.

Use of a halogen-sensitive detector such as an electron capture detector in conjunction with a nonpolar stationary phase such as OV-101 or SE-30 in a gas chromatographic system had resulted in interferences with DDT, DDD and DDE as well as early eluting pesticides such as BHC isomers, aldrin, and heptachlor epoxide because of similar retention times.

Dependent upon the matrix selected for analysis, three distinct analytical protocols emerged involving either no prior separation, separation via column chromatography or separation with destruction or conversion of interfering compounds.

Cleanup methods have been devised to separate DDT and its analogues from some of the PCBs (Amour and Burke, 1970).

The figure below shows the structural similarity between DDT and PCB structures. PCBs extracted from air have been analyzed by Jacobs (1949) before the advent of gas chromatography. The PCBs were quantitated by determining chlorine after combustion over heated platinum.

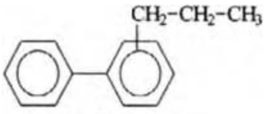
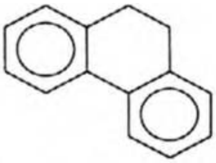
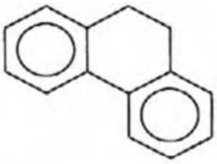
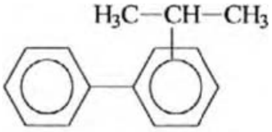
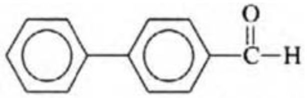
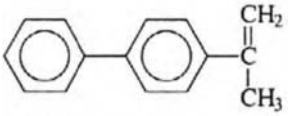
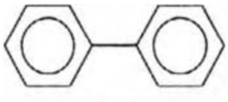
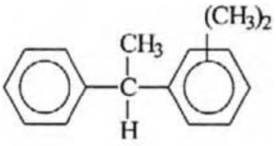
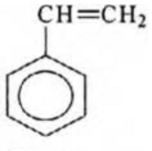
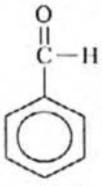
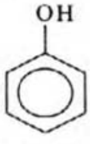
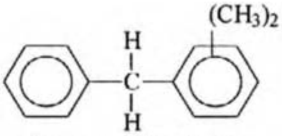
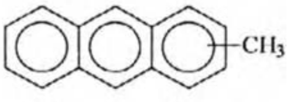
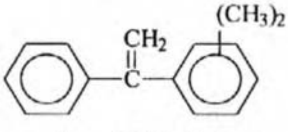
Several papers in the late 1960's and early 1970's discuss problems associated with the sampling of environmental matrices for PCB analysis. Harvey and Teal (1973) showed that nylon netting adsorbs PCBs and hydrocarbons and that they are desorbed if exposed to water containing PCBs and hydrocarbons in lower than equilibrium concentration.

The boat used to take water samples can be an important source of analytical contamination (Grice *et al.,* 1972). Polystyrene particles recovered from the ocean were found (Carpenter *et al.,* 1972) to contain 5 ppm Aroclor 1254. Jensen *et al.* (1972) found that plankton samples collected from the wake of the boat were contaminated by PCBs from the ship's paint and contained 4 to 14 times more PCBs than samples collected 7 ft. abeam of the boat.

Laboratory contamination is always of concern in trace analysis and several workers have found that significant interferences can be retained by reagent water unless precautions are taken (Bevenue *et al.,* 1971). Several workers (Giam and Wong, 1972; Bevenue *et al.,* 1971; Levi and Nowicki, 1972) have discussed cleanup treatments for solid materials used in PCB analyses such as silica gel, florisil, sodium sulfate, glass wool and aluminum foil. Lamont and Cromartie (1969) recommended a cleanup method for extraction thimbles used to hold solid samples during solvent extraction of PCBs.

Ahling and Jensen (1970) extracted PCBs from water by passing the water sample over a coated solid material. From 10 to 176 L of water were treated in this way and then the solid material was extracted with petroleum ether to recover the PCBs. Gesser *et al.* (1971) described the use of polyurethane foam to quantitatively extract PCBs from water.

TABLE 46
Combustion and Pyrolysis Products of PCB Replacement Fluids

Trade Names	Primary Components	Combustion Products	Pyrolysis Products
Wemcol Edisol II Edisol III	 n-Propylbiphenyl	 Dihydrophenanthrene	 Dihydrophenanthrene
	 Isopropylbiphenyl	 Biphenylcarboxaldehyde  C3-Alkylbiphenyl	 Biphenyl
Selectrol Dielektrol	 Phenylxylylethane	 Styrene  Benzaldehyde  Phenol	 Phenylxylylmethane  Methylanthracene  Phenylxylylethene

Aue *et al.* (1969, 1972) showed the applicability of surface-bonded silicones to adsorb PCBs from water. Extraction methods for PCBs from soil, sediment, sludge and paper have used more polar solvents or solvent mixtures than that used for the analysis of water. Holden (1970), for example, used a mixture of hexane and isopropanol to extract PCBs from sewage sludge.

PCBs from non-carbon copy paper were extracted with acetone in a Soxhlet apparatus for 3 days (Kuratsune and Masuda, 1972).

PCB extraction from biological samples tends to be complicated by problems of incomplete extraction, cleanup losses etc. Extraction techniques have most often used Soxhlet extraction to conserve the volume of solvent or a percolation system with large volumes of solvent to obtain satisfactory recovery efficiency. Soxhlet extraction was used by Zitko (1972) in the extraction of tissue samples using 100 mL of hexane in a 1 hour extraction. Acker (1970) used a combination of solvents passed through a tissue sample homogenized with sea sand and packed into a column. The method served to remove interferences from the PCBs contained in a pentane fraction. A procedure was developed for the PCB extraction of milk by Noren and Westöö (1968) using ethyl ether and petroleum ether, centrifuging and removing the separated phases.

The Yusho incident (1968) occurred shortly after Jensen's report on PCBs (1966). The incident prompted the US Food and Drug Administration (FDA) to initiate a national survey to determine the exact extent and levels to which PCBs might have made their way into the food chain by indirect use of PCB contaminated animal feed, industrial and environmental sources, and the use of PCB-containing paper food-packaging materials. The survey resulted in a 'Notice of Proposed Rule Making' to limit the levels of PCBs in foods in 1972.

The survey found that several accidents involving PCBs had contaminated animal feed and subsequently the poultry and eggs intended for human consumption. Other parts of the survey indicated that the use of PCB-containing coatings on the inner walls of grain silos had been responsible for PCB residues in milk derived from dairy cows which fed on the grain stored in such silos. The FDA concluded that it would be in the best interest to limit the ways in which PCBs might enter the food chain as well as limit the levels of PCBs in food.

Since part of the problem with PCBs was the vulnerability of food and feed commodities to direct contamination through accidental causes, the EPA issued rules governing the continued deployment of PCBs in certain industrial applications. These regulatory controls were made under the Toxic Substances Control Act of 1976 and proposed the discontinued use of PCBs in heat transfer systems in plants manufacturing or processing food, drugs and cosmetics. An interagency alert notice [1979] was then issued by the EPA to urge voluntary compliance in removal of equipment containing PCBs and replacement with non-PCB units to prevent food contamination.

On February 24, 1969 the San Francisco Chronicle carried a major feature on work reported by R. Risebrough (1968) and that of S. Jensen (1966). Jensen had reported the identification of PCBs while analyzing for DDE.

The early analytical method applied to the quantitation of PCBs in oils uses a calibration technique developed by Webb and McCall in 1973. The method is based upon the detector response comparison obtained for a commercial PCB mixture with that of the sample and uses packed column gas chromatography with, typically, an electron capture detector (see ASTM method D-4095). Slight variations in the Aroclor manufacturing process, or chemical alteration of the analyte PCB, causes problems with the quantitation. Reasonable results can be achieved, however, in the analysis of insulating oils where the PCB have not been degraded. More sophisticated analytical methods have been published by Albro *et al.*[22]; Erickson[23]; Leo *et al.*[24]; and Stalling *et al.*[25]

The USEPA methods manual, SW-846, of test procedures is commonly used to quantitate PCBs in environmental samples. The manual describes both the method of sample cleanup as well as of analysis. The protocols are given in SW-846 as method 8080 using packed columns or 8081 with capillary columns. Each sample is spiked with surrogate compounds and extracted using methods 3510 or 3520 for liquids or methods 3540, 3541 or 3550 for solids. Cleanup methods use silica gel adsorption (method 3630) sulphuric acid/potassium permanganate destruction of interfering compounds, including pesticides (method 3665) and/or sulfur cleanup by agitation with powdered copper, elemental mercury or tetrabutyl ammonium sulfite (method 3660).

The electron capture detector used in the gas chromatograph is extremely sensitive towards compounds which contain halogen atoms, such as the chlorine of PCBs. However, the detector is not totally insensitive to other elements such as oxygen and sulfur. Indeed, the sensitivity of the gas chromatographic quantitation is very dependent upon the level of interferences present.

The sample chromatograms derived from the electron capture detector in the gas chromatograph are interpreted according to the pattern of peaks and their retention times relative to known standard mixtures. In the case of PCB contamination, the Aroclor products, such as Aroclor 1016, 1242, etc., each has a characteristic "fingerprint".

Aroclor 1016 was derived from the vacuum distillation of Aroclor 1242 which removed the less volatile, more highly chlorinated congeners of the mixture. The PCB congener composition of Aroclors 1016 and 1242 are given in Table 47.

Packed columns for GC analysis do not have enough separating power for individual components of a PCB mixture to yield anything but a series of envelopes. The pattern formed by the envelopes tend to be characteristic of the PCB mixture and therefore the pattern, taken as a whole, is usually recognizable as PCB and may be quantitated as such by comparison with the envelopes produced by a standard Aroclor mixture.

The pattern produced on packed columns by some samples shows that the electron capture detector was responding to components which, in all probability, contained a highly electronegative i.e., electron attracting, element, such as chlorine. The pattern of envelopes in the chromatogram may not be distinguishable as any standard Aroclor product. Since there is often no reason to suspect that any other chlorinated compounds similar to PCBs may be present in a sample the origin of the chlorine is usually an sometimes erroneously attributed to PCB.

Attempts are sometimes made to remove interferences from the oil to allow an unequivocal identification of PCB by gas chromatography/mass spectroscopy (GC/MS). The mass selective detector in place of the electron capture detector is typically used as an identification tool for individual compounds and compound types. The response obtained in the machine is compared with a computer stored library of known compound responses. The closest matched response at the gas chromatographic retention time of the species is usually evidence enough to establish the identity of the component of interest.

The GC/MS technique applied to complex matrices such as oils tends to swamp the detector with naturally occurring

TABLE 47
PCB composition of Aroclors 1242 and 1016

	Mol. %			Mol. %			Mol. %	
PCB	1242	1016	PCB	1242	1016	PCB	1242	1016
BP	0.01	0.50	2,2′,6,6′	0.17	0.19	2,3,3′,4,5′	0.46	0.16
2	0.68	0.80	2,2′,3,6	0.90	1.00	2,2′,3,4,4′	0.40	—
3	0.04	0.10	2,2′,5,6′	0.97	1.07	2,2′,3,4,5′	0.09	—
4	0.22	1.00	2,2′,5,5′	4.08	4.35	2,3,3′,4,4′	0.25	—
2,6	0.13	0.20	2,4,4′,6	2.18	2.40	2,2′,3,5,5′	0.12	—
2,2′	3.99	4.36	2,2′,3,5	0.44	0.47	2,2′,4,4′,5	0.55	—
2,5	0.31	0.34	2,2′,3,6′	0.31	0.33	2,2′,4,5,5′	0.27	—
2,4	1.04	1.16	2,2′,4,5	1.33	1.41	2,3′,4,5,5′	0.31	—
2,3′	1.24	1.37	2,2′,4,5′	3.28	3.48	2,3′,4,4′,5′	0.36	—
2,4′	8.97	10.30	2,2′,3,5′	1.06	1.14	3,3′,4,5,5′	0.05	—
3,5	0.35	0.37	2,2′,3,4	1.67	2.00	3,3′,4,4′,5	0.03	—
3,4	0.09	0.11	2,2′,3,3′	0.15	0.18	2,2′,3,4,5,6′	0.07	—
3,4′	0.12	0.12	2,3,3′,4′	0.60	—	2,2′,3,4,4′,5′	0.08	—
4,4′	0.99	1.07	2,2′,4,4′	1.65	1.81	2,2′,4,4′,5,5′	0.02	—
2,2′,6	0.97	1.08	2,4,4′,5	2.02	1.35			
2,2′,5	9.36	10.87	2,3′,4′,5	1.11	—	Monochloro (%)	0.94	2.40
2,2′,4	2.92	3.14	2′,3,4,5	trace	—	Dichloro (%)	17.23	19.40
2,2′,3	3.25	3.50	3,3′,4,5	0.52	—	Trichloro (%)	51.67	56.31
2,3′,6	0.54	0.58	3,4,4′,5	0.28	—	Tetrachloro (%)	24.65	21.32
2,4′,6	2.15	2.31	3,3′,4,4′	0.34	—	Pentachloro (%)	5.24	0.40
2,3′,5	0.55	0.62	2,3′,5,5′	0.33	—	Hexachloro (%)	0.17	—
2,4′,5	4.53	4.72	2,3′,4,4′	0.81	0.14			
2,3′,4	1.68	1.79	3,3′,4,5′	0.24	—			
2,4,4′	13.30	14.48	2,3,4,4′	0.21	—			
2,3,3′	3.64	3.99	2,3′,4,5′,6	0.92	—			
2,3,4′	2.64	2.80	2,2′,3′,4,6	0.13	0.04			
2′,3,4	2.83	3.08	2,2′,3,5′,6	0.53	0.18			
3,3′,4	0.66	0.38	2,3,3′,5′,6	0.39	0.01			
3,4,4′	1.62	1.89	2,2′,3,4′,6	trace	—			
3,4′,5	1.03	1.08	2,2′,3,3′,6	0.38	0.01			

compounds in the oil matrix. The signal then combines to give a resultant which is unintelligible to the mass spectrometer's computer.

If, at first, a standard method of quantitation does not provide an unequivocal identification of PCBs, an unconventional method might be adopted. For example, extensive cleanup of the sample, starting with solvent extraction to concentrate the PCBs, if present, and followed by the recommended cleanup protocols would yield a sample which could be usefully analysed by a combination of coupled instruments known as gas chromatography/mass spectrometry/mass spectrometry (GC/MS/MS).

The coupled technique is typically used for ultra-trace analysis of dioxins and was designed to provide unequivocal identification of trace amounts of PCB-like compounds.

Quantitation of the gas chromatographic results obtained from the use of packed columns and an electron capture detector, i.e., GC/ECD, are calculated on the assumption that peak area is directly proportional to the amount of PCB present. The relationship between peak area and component concentration is usually linear and a sample peak which is half the size of a standard peak will result from a sample which contains half as much as the standard. If, however, some of the peak area is the result of a co-eluting material which is not the analyte of interest then it will appear that there is more present than is actually the case and the quantitative result will be biased on the high side.

PCBs are not highly volatile compounds and consequently the GC conditions have to be such that the column used for separation must be heated to about 200°C. The volatility of PCBs increases with temperature and as the amount of chlorine in the molecules decreases. Aroclor 1016, because it contains a high proportion of the less chlorinated congeners is more volatile than, say, Aroclor 1254 which has a higher average number of chlorines per molecule. The effect is that at the normal GC operating temperature, Aroclor 1016 is eluted much faster, in about 10 minutes, than Aroclor 1254, in about 20 minutes. The interfering components of the mixture tend to be more volatile than even the constituents of Aroclor 1016 and consequently these are eluted quickly and tend to spill over into the early part of the PCB "fingerprint". Since the amount of interfering substances can be very much higher than the amount of PCB, and even though the detector is less sensitive to the interfering substances, the co-elution of PCBs with other detector responsive materials causes a combinative effect on the peak area and a high quantitation bias is obtained. The relative effect of the interfering compounds on the quantitation of Aroclor 1254 is usually much less because the peaks which are used for quantitation elute much later than the interfering materials and the extent of overlap of the interference and PCB peak areas is usually much less.

Some of the problematic effects of interferences can be avoided by a combination of rigorous sample cleanup and subsequent analysis by capillary column GC. Capillary column technology allows for much greater resolution of the components of mixtures and consequently PCB components tend to become separated from interfering materials. There is still a very strong need for adequate sample cleanup but the presence of interfering non-PCB components becomes much more easily recognized.

The following analytical results illustrate the discussion:

- the correlation coefficient for the standard curve, which is a measure of the linearity of the detector response in the concentration range of interest, is sometimes better if one chooses peaks which appear later in the chromatogram than those normally used in the early part of the chromatogram;
- the correlation coefficient tends to become less than satisfactory for three early peaks usually chosen for quantitation when co-eluting interferences are present;
- the apparent concentration of PCBs in the samples can fall dramatically when the standard curve is calculated on the basis of peaks which appear later in the chromatogram than the early peaks which may be the ones routinely chosen for quantitation. This indicates that there is a substantial difference between the chromatograms of the standards and the samples.

The analytical data in Table 48 were predicated on the assumption that the PCB contamination in the samples were due to a single Aroclor, namely Aroclor 1016, and that the relative peak intensities in the chromatogram matched those of an available standard Aroclor 1016. This can clearly lead to a high bias in the result when the quantitation method uses peak areas which contain a significant contribution from interfering substances. Such a simplistic approach is not suitable for the quantitation of PCB contamination derived from partially decomposed Aroclors, mixtures of Aroclors, or PCBs which may not have originated from an Aroclor.

TABLE 48
The apparent PCB concentration of samples is drastically altered by the method of calculation

	Calculated PCB concentration (ppm Aroclor 1016)		
Sample #	Early peaks	Late peaks	Later peaks
A	61.7	51.7	8.7
B	16.7	<D.L.	<D.L.
C	19.0	<D.L.	<D.L.

The chromatograms of the samples used for the above discussion do contain a distribution of PCBs which match the pattern found in Aroclor 1016. However, the relative amounts of PCB congeners in the standard and the samples do not match. The samples may have been derived, for example, from a batch of Aroclor 1242 which had been distilled to yield only the most volatile fraction as Aroclor 1016. This might explain why there is relatively little of the less volatile, more highly chlorinated congeners. Chemical degradation, on the other hand, tends to dechlorinate the more highly chlorinated material first and then, as the chlorines are removed, the mixture becomes more concentrated in the less chlorinated material until it too eventually becomes dechlorinated.

Different isomers of PCB with the same number of chlorine substituents can cause substantially different responses from electron capture detectors. The accurate quantitation of PCBs therefore relies on the similarity between the sample chromatogram and the standard Aroclor.

A calibration technique was proposed by Webb and McCall which employed individual peak response factors.

A table was developed for each Aroclor in which the weight-percent composition of each peak in the chromatogram was identified by whole numbers which represented their retention times relative to a reference compound defined as 100. The weight-percent compositions of Aroclors were determined using GC/MS.

The column used was a packed column with poor chromatographic resolution compared with today's capillary columns. The individual peak calibrations were valid only for the specific lots of Aroclors tested.

Sawyer[26] has conducted a similar characterization to Webb and McCall and concluded that individual peak calibration is the most reliable approach for samples containing a non-Aroclor residue.

The method of peak by peak comparison is used in ASTM D-4059. The concentration of PCB in each peak in the chromatogram is calculated from the individual response factors of the detector. The total PCB content is then obtained by summing the concentrations associated with each peak.

An alternative, but more approximate, calculation, noted by ASTM, can be made using the responses of the larger, more cleanly separated peaks in both the standard and the sample. Again, this makes the assumption that the chromatograms of the standard and sample match one another. ASTM notes that the PCB concentration calculated in this way may be incorrect because the PCB content in an individual peak may be reduced or relatively enhanced by a non-standard isomer distribution or impurities. In order to minimize the effect of variations between Aroclors, ASTM suggests the use of a minimum of three peaks in the sample being analyzed but adds that the simplified calculation should not be used in circumstances where maximum accuracy is required. Indeed, EPA SW-846-8080 indicates in $7.6.5.3 that PCB to residues should be quantitated by comparing total area or height of residue peaks to total area or height of peaks from appropriate Aroclor reference materials.

The following calculations take the above considerations into account on a different set of samples. Based on three sets of calibration data for early, late and total areas of peaks, chromatograms were recalculated to yield significantly different data as shown in Table 49.

The area of the chromatogram attributable to Aroclor 1016 in a standard was approximately 59%. The remaining 41% is attributable to a region from the start of timing to 2.16 min. Of this total, 39% is the response of the detector to the solvent. This leaves only 2.5% which might be attributable to Aroclor 1016. Thus, the error induced in the calculation by leaving this portion of the chromatogram out entirely is very small relative to the error induced by including non-PCB detector response in the same retention time range. For example, for the sample marked as [‡] in Table 49, the area of the total chromatogram which occurred prior to 2.19 min. was 74%. This should be compared with the 41% in an Aroclor 1016 standard. In other words, in this case, about 70% of the total sample chromatogram area is attributable to non PCB components.

§ 7.6.5.3 of the USEPA SW-846/8080 method states the following: "Quantitate PCB residues by comparing total area or height of residue peaks to total area or height of peaks from appropriate Aroclor(s) reference materials. Measure total area or height response from a common baseline under all peaks. Use only those peaks from the sample that can be attributed to chlorobiphenyls. These peaks must also be present in a chromatogram of reference materials. Mixtures of Aroclors may be required to provide a best match of GC patterns of sample and reference."

The gas chromatographic trace, in many cases, does not conform with what would be expected from a standard Aroclor and, consequently, the quantitation of the PCB components must be done by comparing total area or height of residue peaks to total area or height of peaks from appropriate Aroclor(s). These requirements have been met in Table 50 and are contrasted with other methods of calculation.

It was not reasonable to attempt to quantitate the PCB residue contained in these samples by peak height because in part, the chromatogram derived from the instrumentation typically does not lend itself to that type of quantitation method. In order to illustrate the point, PCB concentrations have been calculated using the peak height method with peaks appearing at different retention times and the results listed in Table 51.

The peak height method is clearly not acceptable. Inconsistencies in the peak height ratios have also been calculated. The peak height ratios of peaks in a standard Aroclor pattern should remain the same from sample to sample. However, because the samples contained PCB residues and were not standard Aroclors the peak height ratios vary. Peak height ratios are presented in Table 52. For this fact alone, the method of peak height analysis is not valid for these samples and the analyst must revert to the only other alternative provided by the EPA SW-846 8080 method, that is, the comparison of total areas.

Similar inconsistencies are shown by determining relative area ratios. Table 53 shows the result of calculating area ratios relative to a peak at $R_t = 7.75$ min which is prominent in the standard samples.

Clearly, the experience of the analyst is critical to the analytical data derived. The precision and accuracy of an analytical protocol, as measured by standard methods, may indicate that the quantitation of ideal mixtures is within the expected limits. Without such a measure of control there is a much reduced likelihood of achieving reliable, scientifically defensible, results.

THE DISPOSAL OF POLYCHLORINATED BIPHENYLS: CONCENTRATIONS OF CONCERN

Methods for the dechlorination and destruction of polychlorinated biphenyls are reviewed in terms of the fundamental principles involved and the reaction products formed. This section is organized into three parts (1) the disposal of askarel liquids; (2) the disposal of PCB contaminated liquids and (3) the remediation of PCB contaminated soils. Figure 46 presents disposal methods for high concentration PCBs.

(1) PCBs are refractory organic compounds which require a large energy input for their decomposition. The methods in use for the destruction of fluids which contain high percentages of PCBs are therefore typically large energy sources. The most widely used method for the disposal of askarel fluids, and that preferred by the EPA is high temperature incineration. However, other methods discussed here include chemical methods, a DC arc and the use of a microwave plasma.

(2) The EPA defines a PCB fluid as one which contains more than 500 ppm of PCB. For this discussion, liquids are considered to be contaminated with

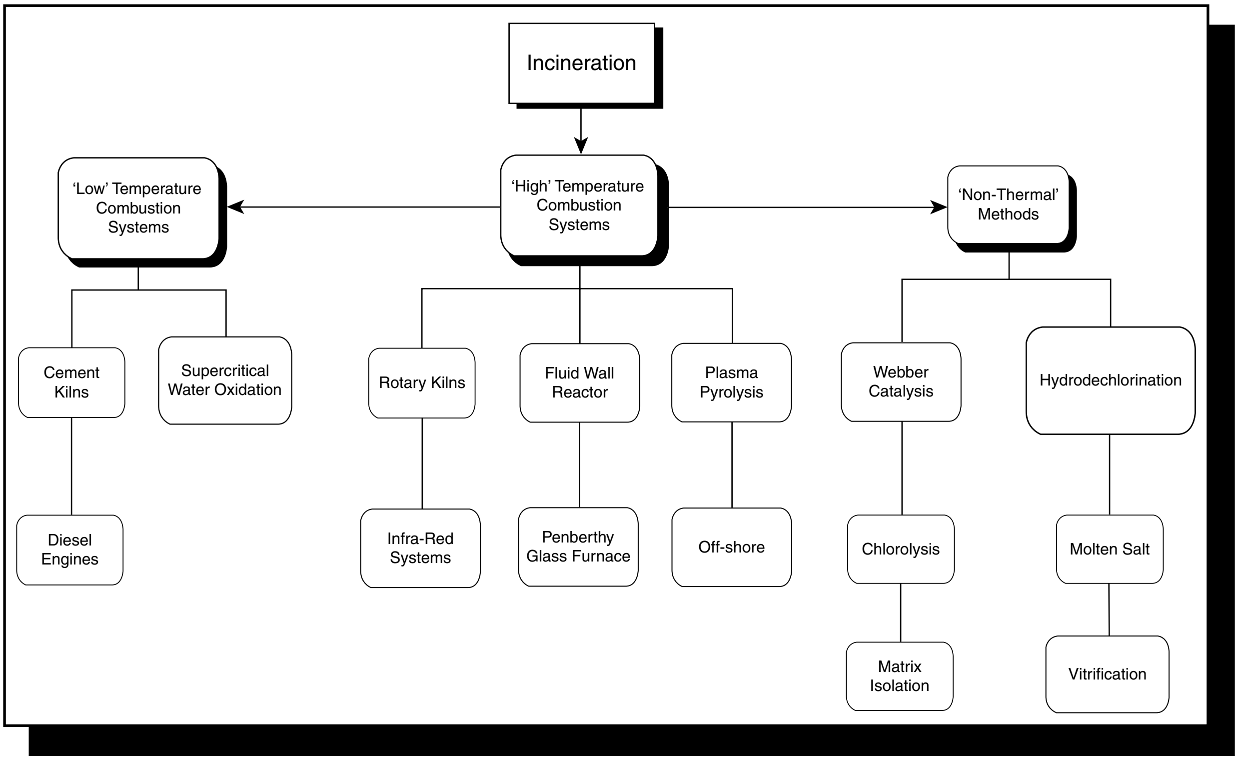


FIGURE 46 Disposal methods for the high concentration PCB fluids.

PCBs even if they are at the several hundred ppm concentration level. This is because the methods which are aimed primarily at the decontamination of low concentration liquids tend to remain effective up to levels as high as, perhaps, 1% (10,000 ppm) and, in any case, well into the range defined by the EPA as "PCB fluid".

(3) The remediation of PCB contaminated soils can present a wide range of PCB concentrations for destruction. This section reviews several different methods used on a large scale or being developed for field use and includes an outline of information on the application of a novel electrochemical technology.

It is sometimes possible to reclaim the liquid for re-use. Several of the numerous alternative chemical methods of PCB dechlorination as well as physical separation methods can be made to yield decontaminated liquids.

The stability and lack of reactivity of PCBs are benefits and at the same time, liabilities. Their thermal stability to oxidation and in the extreme case, combustion, and resistance to harsh chemicals such as strong acids, bases, oxidizing and reducing agents, has made the removal of PCBs from electrical and other uses extremely difficult.

The PCB decontamination methods presently available fall into four major classes and are either physical, physicochemical, chemical or biochemical. The literature concerning PCBs is expanding rapidly. This paper is not intended to be a complete review of every disposal option, but rather an overview of some of the more scientifically interesting systems.

(1) THE DISPOSAL OF PCB LIQUID

PCB liquid is taken to include those commercial PCB mixtures which were diluted with tri- and tetra-chlorobenzenes and which typically contain about 70% PCBs. There is an important distinction to be made between askarel fluids in which PCBs are the majority constituent and fluids which contain more than 500 ppm PCB but in which PCBs are not a majority component. The EPA requires that any fluid with a concentration of more than 500 ppm PCB must be considered as a "PCB fluid". The name "askarel" is a generic name used for PCB formulations. Other names found frequently in the literature are manufacturer's trade names such as "Aroclor". "Inerteen", "Chlorphen", "Safe-T-Kuhl", "Pyrochlor" etc. The most commonly used askarel in North America was Monsanto's "Aroclor" and these two names are often interchanged. Figure 57 shows some of the disposal options available for askarels.

INCINERATION

EPA research into the thermal degradation of PCBs has involved the use of commercial mixtures and pure isomers

TABLE 49
PCB concentrations derived from different correlation curves can produce drastically different result

Concn. of PCB as ppm Aroclor 1016		
Early peaks	Late peaks	Total area
55.7	19.0	8.6
57.3	31.5	6.9
46.8	6.3	40.1
48.9	4.8	33.4
53.4, 62.6‡	6.9	32.6, 19.5
63.5	26.5	46.7
62.2	27.0	51.4
82.9	58.2	85.8

TABLE 50
A high bias is obtained when residue peaks are inappropriately calculated

PCB concentration (ppm Aroclor 1016)		
Conventional	Total area	Peak height
30.7	18.7	12.5
59.6	13.7	0.9
51.6	56.8	19.7
60.6	22.0	N.D.
55.0	4.4	N.D.
147	37.7	33.0
73.0	44.1	11.6
65.5	23.2	N.D.
53.9	27.9	7.6
54.6	27.6	7.6
50.7	44.4	13.1
60.4	23.0	4.5
58.4	30.3	6.3
48.8	24.3	6.9
48.0	25.2	7.2

N.D.—not determinable.

TABLE 51
The apparent concentration of aroclor 1016 in the samples can vary by several hundreds of percent depending upon the peak chosen for quantitation

Retention time (min)/concentration		
4.39 min	6.47 min	8.71 min
21.5	25.0	12.5
58.8	22.6	0.9
37.8	37.0	19.7
73.0†	21.0	N.D.
9.2	N.D.	N.D.
38.3	<91.7	33.0
65.0	46.7	11.6
51.8	34.6	N.D.
53.1	33.0	7.6
50.7	36.1	7.6
50.0	34.0	13.1
66.3	28.0	4.5
60.2	30.3	6.3
55.5	33.7	6.9
46.3	30.2	7.2

TABLE 52
Peak height ratios for standard aroclors remain constant but vary for the sample under investigation

	Peak height ratios relative to R_t = 7.75 min				
Sample	4.27 min	6.28 min	7.52 min	8.04 min	11.4 min
EPA	0.64	0.76	1	0.57	0.53
Standard	0.54	0.59	1	0.54	0.43
Oil (1)	0.48	0.54	1	0.45	0.45
Oil (2)	7.60	3.30	1	N.D.	N.D.
Oil (3)	0.50	0.52	1	0.44	0.46
Oil (4)	9.8	3.36	1	N.D.	N.D.

in a precisely controlled thermal environment. Temperature and residence time relationships were established for the conditions necessary to produce a specified destruction efficiency. It was concluded that with a 1 sec. residence time, most of the pyrolysis occurred within a narrow temperature range between 640°C and 740°C (Mescher *et al.*[56]). The results were based on the disappearance of PCBs rather than the accumulation of degradation products. The lower molecular weight PCBs are less thermally stable than the higher molecular weight congeners. The residence time of the molecules in the high temperature zone is therefore a critical parameter in determining the overall efficiency of the degradation. However, the destruction and removal efficiency (DRE) of an incineration process for a given compound does not consider the toxicity of any pyrolysis products. The DRE is defined as

$$DRE = \frac{(W_{\text{feed}} - W_{\text{out}})}{W_{\text{feed}}} \times 100$$

where,

W_{feed} = mass feed rate of a compound to the incinerator,

TABLE 53
Area ratios for standard samples remain constant but vary for the samples in question

	Area rations relative to R_t = 7.75 min				
Sample	4.27 min	6.28 min	7.52 min	8.04 min	11.4 min
EPA	0.34	0.40	1	0.50	0.43
Standard	0.34	0.39	1	0.46	0.36
Oil (1)	0.33	0.40	1	0.44	0.34
Oil (2)	20.6	8.67	1	1.48	0.20
Oil (3)	0.39	0.41	1	0.43	0.36
Oil (4)	45.7	8.71	1	1.64	N.D.

W_{out} = Mass emission rate of the incoming compound past the incineration zone.

Water and carbon dioxide are the main products of combustion for organic material. However, incineration equipment does not always produce complete combustion and the effluent or residues can contain organic material of varying complexity which is derived from either incomplete combustion or thermal synthesis. A better understanding and acceptance of the pyrolysis data accumulated over the last few years has given rise to the concepts of Principal Organic Hazardous Constituents (POHCs) and Products of Incomplete Combustion (PlCs).

POHCs can be any incoming hazardous waste stream but in the particular case of refractory organic compounds such as PCBs, TCBzs and other halogenated compounds, the pyrolysis chemistry is complex and can lead to the generation of PlCs which are much more toxic than the original POHC. There is a wide variety of equipment available for the incineration of waste organic materials, but the suitability of the incinerator for refractory compounds such as PCBs is restricted to those which allow the use of high temperatures and a suitably long residence time to achieve a DRE of 99.9999%.

A rotary kiln is usually large and expensive to operate but does fulfill the required conditions. Such kilns are used by Energy Systems Corporation in El Dorado, Arkansas and SCA/Chemical Waste Management in Chicago, Illinois. The kiln consists of a refractory lined cylinder with its axis slightly inclined to produce a tumbling action which mixes the wastes. The advantage of the design lies in the wide range of wastes which can be accommodated because the residence time can be controlled to the extent necessary for complete combustion. Combustion temperatures range from 870°C to 1650°C.

Halogenerated wastes can be used to fire cement kilns. This has been done by the St. Lawrence Cement Co. in Mississauga, Ontario, and the Peerless Cement Co. in Detroit, Michigan. The calcium oxide of the cement apparently reacts with the chlorine in the PCB to form calcium chloride. The resulting cement is not contaminated by organic material because the kiln operates at about 1500°C. The process has the advantage that some amount of chloride imparts beneficial mechanical properties to the cement.

Hazardous waste has been used as a supplemental fuel in a lime kiln located near Manitowoc, Wisconsin in a research program run by Monsanto Research Corp. for the EPA. The DRE for each POHC has been measured together with the chlorine and trace metals pyrolysis products. The concentration of stack gas pollutants under baseline and waste fuel test burn conditions were found to be similar and the DREs exceeded RCRA regulations for incineration. Perhaps a disadvantage of both rotary kilns and cement kilns is that they are operated on fixed land sites so that the PCB wastes must be transported to them through the surrounding communities. The siting of hazardous waste disposal facilities is, in general, complicated by the "not in my back yard" (NIMBY) syndrome. The result has been that companies have attempted to mobilize their disposal technology. Offshore incineration has been proposed. Bayer AG in Germany and Solvay Cie in Belgium began using a converted tanker, the 984 ton Matthias I, in 1969. After five years, the ship was replaced with the 3,445 ton tanker the Matthias II. Bayer, Solvay and Westab GMBH are also currently using the Vesta, a 1,356 ton incinerator ship which was converted in 1979.

Oceanic Combustion Services, a chemical waste management company, has used a converted chemical tanker fitted with two large incinerators. This ship, the Vulcanus, was the first to be used in US coastal waters. It sailed from the Mobile-Chickasaw port facility in Alabama to an EPA designated incineration site in the Gulf of Mexico. At-Sea Incineration Co. of Greenwich, Connecticut, also presented a plan to burn PCBs on an incinerator ship operating out of Port Newark, New Jersey. The $74.5 million program was aimed at the construction of two ships, each of which was to be capable of burning 1.3 million gal. of hazardous wastes at each sailing. The disposal rate of liquid wastes through each of the two onboard 170 ton incinerators was expected to be 70 gal./min.

The construction of large scale facilities for the land-based support of ocean incineration was an unresolved issue which contributed to the demise of ocean incineration. Instead, mobile incinerators are being tested which can be used onsite. The EPA has built and tested a reactor built onto a series of four flatbed trailers. A 4.9 meter long, 1.2 meter diameter kiln is mounted on the first trailer and operates at 1000°C with a nominal solids retention time of up to 60 minutes. A secondary combustion chamber operating at 1300°C is housed on the second trailer and provides a secondary retention time of 2 to 3 seconds. The third trailer houses a water quenching apparatus, a wet electrostatic precipitator and an alkaline scrubber for off-gas treatment. The fourth trailer is used for monitoring instruments and controls. The unit was developed at a cost of $10 million and subsequent units are projected to cost $5 million each.

The operating parameters of such equipment must typically provide very high temperatures to make up for the restriction on size which limits residence time. The unit has been extensively tested on 900,000 Kg of dioxin contaminated soil and 80,000 Kg of liquid waste and found to have a destruction and removal efficiency which meets the EPA criterion of 99.99999% while operating at a feed rate of over 2000 Kg per hour.

ENSCO Corporation has developed a mobile rotary kiln incinerator which has been used to treat up to 100 tons per day of contaminated soil. The unit operates in the range of 500° to 1000°C. The off-gas from the rotary kiln has a residence time of over two seconds in a secondary combustion chamber operating at 1000° to 1300°C in an atmosphere with excess oxygen. Acid gases produced in the combustion air are removed with a countercurrent flow of water which is passed through activated carbon filters and then neutralized. The unit operates at a feed rate of about 6 tons per hour and has been demonstrated to achieve 99.9999% destruction and removal efficiency in an 11,000 ton filed trial.

The J.M. Huber Corp. in Borger, Texas began working with Thagard Research Corporation on high temperature fluid wall (HTFW) reactor technology in 1976 and acquired the patents from Thagard. The general operating principle of the process is to rapidly heat materials to temperatures in excess of 4,000°C with a residence time of tenths of a second. The energy input to the waste measured as the integral over temperature and time is then sufficient to obtain the DREs required by RCRA. The technical problem posed in the application of the principle is that the reaction vessel must be able to withstand the operating conditions without becoming chemically or physically degraded. The core is a porous refractory cylinder heated externally by electric elements so that wastes fed into it are heated by radiant energy. The core material allows the permeation of a gas such as nitrogen through the core wall into the interior. The gas is admitted radially to produce an annular envelope which blankets the walls and reduces the contact of the reactants with the reactor. This device tends to avoid the degradation of the reactor. Nitrogen is used partly for the reason that it is transparent to the radiant infrared energy and also because the production of partially oxygenated pyrolysis products, such as PCDFs and PCDDs, is avoided.

In Huber's "Advanced Electric Reactor" a graphite core is maintained at 2200°C with a fluid wall of gaseous nitrogen to avoid oxidation of the carbon. The burned material emerges into a sealed steel container and the offgases are vented through activated carbon cylinders. Fixed site and mobile reactors based on the HTFW technology have been found effective on 2,3,7,8-TCDD contaminated soils. The capacity of the pilot plant is 5,000–15,000 tons/yr with costs in the range $77 to $385/ton.

Plasma pyrolysis of PCBs has been investigated (Barton *et al.*[92]) by both Lockheed Aerospace and the Royal Military College in Kingston, Ontario. The principle is applied on a large scale for the recovery of metals from waste dust generated in steel mills. A plasma consists of both charged and neutral particles exhibiting collective behaviour with an overall charge of close to zero. A common method of plasma generation is electrical discharge through a gas. In the case of a plasma torch, the plasma is stabilized by collimation and resembles the heat source used for atomic emission spectroscopy. Arc temperatures can occur of up to 50,000°K and the material entering the plasma tends to become atomized and ionized.

The overall efficiency of the process depends upon the conversion of the plasma electrical energy into thermal energy. Large scale generators convert about 85–90% of the electrical energy into usable heat and are therefore useful for the rapid decomposition of materials which would otherwise require a long residence time for complete pyrolysis. Nevertheless, extensive work has been done to determine the conditions necessary to avoid the release of chlorinated compounds with molecular weights larger than tetrachlorobenzene.

Pyrolysis Systems Inc. in Welland, Ontario has built the prototype mobile unit for the Department of Environmental Conservation of New York State, which is based on the work done at the Royal Military College. The unit has a throughout rate of one gallon per minute and has shown destruction efficiencies for askarel of the order of 99.9999999%. Atomization of fluids has been estimated to take place in less than one third of a millisecond. The estimated cost is about $300,000 per unit with electrical consumption of about 1.2 kWh/kg of feed. This is about one-tenth the energy cost of incineration. The system is not designed to destroy solid material.

The SKF Steel engineering plant in Hofors, Sweden is constructing a 6 MW plasma generator designed to treat 70,000 tons/yr of waste dust from the steel mill. The cost of the facility is estimated at $23 million dollars. Thus, there is no doubt that the plasma technology can be scaled up but the size of the associated generators and equipment imposes a limitation on its mobility. In the case of the treatment of PCB wastes the siting of a fixed facility is often a difficult, if not impossible issue to resolve so that the limitation of the method in terms of throughput rate is governed by political factors rather than technical considerations.

The circulating bed combustor manufactured by GA Technologies Inc., is an application of fluidized bed technology to the disposal of hazardous wastes. In a fluidized bed unit the temperature is limited to the melting point of the inert support material. In the case of sand this is about 1100°C. The bed acts as a large radiative surface to produce high destruction efficiencies as well as several operating advantages over other incinerator designs. Ogden Environmental Services uses a design in which a high velocity air flow suspends the fluidized bed to give a turbulent combustion zone at about 800° to 1100°C. Solids have a very long residence time of 30 minutes and are discharged with the bed materials into a cyclone and recirculated through the furnace. Gases have a 2 second residence time and are vented through a convective gas cooler and flue gas filter. Ogden has received a National TSCA PCB permit for the use of its circulating bed combustors.

Waste-Tech Services use fluidized bed systems for the destruction of both solids and liquids. Aluminum silicate firebrick particles are used as the bed material in a conical combustion chamber. Air is used to fluidize the bed.

A secondary reaction chamber maintains gases and suspended particulates at about 1100°C for 2 seconds in order to ensure that the required DRE is obtained. A wet scrubber is used to neutralize the acidic off-gas.

The Penberthy glass furnace for hazardous waste was developed by Penberthy Electromelt Inc. in Seattle, Washington. The Unit is a tunnel incinerator with a pool of molten glass maintained electrically at 1200°C. The high thermal capacity bed acts as the heat transfer medium. The unit is insensitive to the form of the waste introduced and the ash is periodically removed with the molten glass. The product is a non-leachable solid matrix which can be landfilled.

Offgas scrubbing is accomplished using either calcium carbonate towers, which react to produce calcium chloride, or wet scrubbing towers using alkaline water followed by an electrostatic precipitator. Particulates are removed with a ceramic filter element. The capacity of the unit can be provided from 100 lb/h to 25,000 lb/h. Its application to the incineration of PCBs has not yet been documented and its usefulness will probably depend on the relative cost of alternative energy sources.

Shirco Infrared Systems have applied a large scale combustion system in which the waste was treated by infrared energy developed by electrically powdered silicon carbide rods in a primary chamber at 1000°C. The residence time of material could be adjusted between 10 and 90 minutes. Volatile materials were given a 2 second residence time in a secondary chamber at 1260°C.

Each mobile unit could be erected on-site in about a week and be capable of processing between 10 and 45 Kg per hour.

SUPERCRITICAL WATER OXIDATION

The oxidation of organics by supercritical water oxidation is being developed by Modar Inc. of Natick, Massachusetts. A mobile pilot plant is being developed and has shown destruction efficiencies for chlorinated organics of 99.99%. The equipment is mounted on a flatbed truck with a capacity of about 0.5 gal. of organics/min. A fullscale unit with a capacity of 25,000 gallons per day is being designed.

The operating principle utilizes the properties of water above 375°C and 3,200 psi where the liquid and vapor phases have the same density (about 0.05 to 0.30 g/mL). Under these conditions hydrogen bonding is much reduced and the dielectric constant is diminished to between 3 to 10 with the result that water becomes a much better solvent for organics and salts are only partially ionized.

High pressure air or oxygen injected into the system causes rapid oxidation of organic material to carbon dioxide and water with chlorine eliminated as chloride. Inorganic salts are precipitated and the supercritical water used for power generation or high pressure steam. The waste stream is slurried with water to give a mixture of 5 to 10% organics and is heated and pressurized to supercritical conditions without charring. The exothermic heat of reaction is able to sustain the process at about 550°C. At the present time the application of this technology is assessed on a site by site basis. Supercritical oxidation of 2,3,7,8-TCDD wastes has been shown to take place with a DRE of 99.9999%.

WET AIR OXIDATION

Wet air oxidation of PCBs and aqueous wastes has been developed by Zimpro Inc., Rothschild, Wisconsin. Zimpro is a subsidiary of ENSCO. The operating principle of the process is similar to that of supercritical water oxidation. The pressure is adjusted to the autogenous pressure of water at the operating temperature, that is, in the range 300 to 3,000 psi between 175°C and 320°C. The process is primarily aimed at the treatment of aqueous wastes which are either too dilute to incinerate or which are toxic to biological treatment systems. If the chemical oxygen demand (COD) is greater than 10 to 15 g/L the exotherm of the oxidation reactions is sufficient to sustain the reaction, whereas incineration does not usually achieve energy sufficiency until the COD of the waste reaches 300 to 500 g/L, depending upon the temperature.

Chlorinated aromatics containing non-halogen functional groups such as phenols and anilines are relatively easily oxidized by the process compared with chlorobenzenes or PCBs. Zimpro is therefore evaluating the use of homogeneous catalysts to increase the destruction efficiency of refractory organics.

The Wetox process is similar to Zimpro's technology. It has been developed by the Ontario Research Foundation and is being marketed by Wet Com Engineering Ltd., Toronto, Ontario. A 25 gpm full scale unit has been put into successful operation at a chemical plant.

The turbulence produced by mixers in the reactor vessel creates eddies and induces the formation of bubbles. The air or oxygen in the system can then react at a faster rate and this, in turn, allows operation of the system at temperatures which are considerably less than in the wet oxidation process (about 225°C/580 psi compared with 175°–320°C/300–3,000 psi). The Wetox process typically reduces the COD of the waste feed by 75 to 95% and produces low molecular weight species in addition to carbon dioxide. Its usefulness for the oxidation of PCBs has not been reported.

MOLTEN SALT

The molten salt process is an old technology which has been considered during the past 25 years as a method for coal degassification. It can, however, be applied (Barclay, 1980[93]) to the destruction of organic and inorganic hazardous wastes, including PCBS. Rockwell Energy Systems Group has developed a system to treat both liquid and solid PCB waste. In 1986 Hydro Quebec built a $4 million salt system housed in several forty foot trailers. The PCB material is introduced at the bottom of a pool of molten sodium carbonate which is maintained at a temperature of about 900°C and reacts as it migrates to the surface. The sodium carbonate is converted to sodium chloride. PCB destruction in a 6-inch depth of molten salt was determined in the Rockwell Energy Systems Group process as a function of the pool temperature using 53–65% excess air and the destruction efficiencies found to vary in the range 99.99917% at 745°C to 99.99971% at 912°C. As the depth of the molten salt pool is increased so the destruction efficiency increases because of the longer residence time and Rockwell points out that the deeper salt beds of commercial units would probably achieve higher destruction efficiencies than was observed with the laboratory scale system.

The product gases were analyzed for CO, CO_2, HCl, Cl_2 and $COCl_2$. The sodium carbonate tends to react with acid gases and reduces the need for downstream scrubbing systems. It was found that less than 2 ppm HCl was formed provided that 1–2%, Ca_2CO_3 remained in the melt. Less than 0.25 ppm of phosgene was detected and about 0.5 ppm of Cl_2 (or perhaps NO^x).

Inorganic, high melting, products such as sodium chloride build up in the melt until it eventually solidifies and the destruction efficiency is reduced. The continuous process provides for this by adding more sodium carbonate to maintain the fluidity of the pool. However, even when the pool had been almost completely converted to sodium chloride PCB destruction continued. Under these circumstances the process can no longer be classified as a high temperature organic/inorganic reaction but rather is a pyrolysis. The EPA requires a 1200°–1600°C temperature for incineration coupled with a sufficient residence time at the chosen temperature to avoid inadvertent generation of products of incomplete combustion (PlCs) such as PCDFs and PCDDs.

Mobile molten salt reactors have been constructed but are costly. A system with a capacity of 225 lb/h was estimated by Rockwell to cost approximately $1.5 million. The solids handling equipment increases the cost by a further $250,000. The operational costs are not high once the capital cost is overcome because the reaction exotherm provides heat and the only reagent expended is sodium carbonate.

DIESEL DISPOSAL

D. and D. Disposals Ltd., based in Smithville, Ontario have evaluated the use of diesel engines to incinerate PCBs. Three six cylinder units were mounted onto a flatbed truck to give a throughout of about 8 gal. of PCB per hour. The cost estimate for each engine was about $300,000. The Ontario Research Foundation has undertaken measurements of PCB destruction efficiencies and found that 99.998% can be achieved when an 80/20 blend of diesel fuel/PCBs is burned at about 590°C in the cylinder head.

OZONOLYSIS

The Royal Military College in Kingston, Ontario and the University of California, Riverside, have both investigated the use of ozone to oxidize PCBS. A six- to eightfold excess of ozone over the amount required for stoichiometric reaction was found to produce a 90–95% conversion of PCBs. The extent of formation of PCDFs and PCDDs from askarel fluids can be expected to depend upon the particular reaction conditions so that the potential problems which the operation of the method may have are intended to be alleviated by critical control. About 100 lb of PCBs require approximately 100 lb of ozone. About 1000 kWh of electricity are used to produce this quantity of reagent and therefore the method is not likely to be as cost competitive as some incineration processes although it may be proven useful for the destruction of PCBs diluted into trace quantities in the environment. The application of the ozonolysis reaction to PCB contaminated waters is discussed below in the discussion on radiation methods.

Palauschek and Scholz[94] have applied ozone destruction methods to PCDFs and PCDDs in water. They found that degradation takes place only under alkaline conditions of pH 10 as a result of the generation of hydroxyl radicals by reaction between ozone and water. Ozone does not directly attack PCDDs/PCDFs in aqueous solution. The method is therefore not likely to be useful for the treatment of oily wastes in which the chlorinated aromatics are confined to the oil.

CHLOROLYSIS

Hydrocarbon wastes have been reacted into useful products by adding rather than removing chlorine. The chlorolysis reaction has been used on a large scale by Hoechst A.G. (Jackson[95]). The process involves the reaction of chlorine gas at high temperature (about 640°C) and pressure (about 25 MPa) with the partly chlorinated waste stream to yield carbon tetrachloride and hydrogen chloride. The generalized reaction is

$$C_xH_yCl_z + (4x + y - z)/2Cl_2 \rightarrow X\ CCl_4 + y\ HCl$$

The method has been used primarily to treat aliphatic wastes rather than PCBs and is limited to feedstocks which contain approximately 5% aromatics unless considerable modifications are undertaken. The reaction conditions necessary to treat askarels may involve an increase in the process residence time or the use of a catalyst to promote the reaction. The reaction products are easily separated from the process by distillation and the mixture which remains is then recycled to the reactor.

Carbon tetrachloride and hydrogen chloride are marketable reaction products. However, the health effects associated with carbon tetrachloride have tended to curtail its use. The reaction is exothermic, which helps to keep operating costs low, but the disadvantage of the process lies in the need to use high purity nickel construction at a high capital cost in order to overcome the extremely corrosive nature of the chemicals involved.

HYDRODECHLORINATION

Ahonkhai *et al.*[96] have investigated the thermal hydrodechlorination of hexachlorobenzene by passing chlorobenzene vapor into a quartz tubular flow reactor at different temperatures. Above about 900°C a $> 99\%$ conversion to HCl was observed. When H_2 was replaced by CH_4 the same reaction took place but it occurred with the formation of copious quantities of soot.

The catalytic hydrodechlorination of PCBs was investigated by Lapiere and co-workers[97] using either 61% nickel on Kieselguhr or 10% palladium on charcoal. The method requires the handling of high pressure hydrogen at 30–50 atmospheres and the HCl produced is a source of catalyst poisoning and corrosion problems. Approximately 90% dechlorination is achieved in five hours at 100–200°C.

A nickel catalyzed zinc reduction of PCB has been reported by N.S. Chu and S.C. Vick[98]. The reaction is best carried out in an aprotic solvent, such as dimethylformamide (DMF), rather than a protic solvent such as water or alcohol. The major reaction product of PCBs is biphenyl. The reaction of trichlorobenzene in a mixture of MDF and methanol was found to be essentially complete after four hours at 50°C.

The mechanism of reaction was suggested to parallel the $Pd(OAc)_2$ catalyzed dehalogenation reaction reported by Okamoto and Oka[99]. These authors explain their observations by proposing that a low valent metal complex was generated in situ and reacts with aromatic halides to give oxidative addition products which are then reduced by the alkoxide ion. The nickel catalyzed reaction of Chu and Vick was therefore suggested to involve similar steps as follows.

- the reduction of Ni(II) to Ni(0) by Zn;
- oxidative addition of aryl halide to the lower valent Ni and;
- reaction of the aryl-Ni species with alcohol solvent to form a reduced product or reaction with another aryl halide to give a coupling product.

A similar mechanism may exist for similar catalytic reactions involving the Pd or Ru catalyzed reduction of aromatic halides by hydrogen sources, such as hydrazine,[100] amine,[101] formate,[102] alcohol or alkoxide.[99]

A uniquely effective catalytic dechlorination method currently under investigation by the author and which will be the subject of patents, has been shown to completely dechlorinate Arochlor 1260, a transformer dielectric fluid, *in 4 minutes at room temperature.* The reagent, packed into a column configuration in a mobile system, allows the reaction to be used on-site for the destruction of askarels under mild chemical conditions without the liability of transportation and interim storage at a high temperature incinerator facility.

MATRIX ISOLATION

Solid solutions which are held together by weak Van der Waals forces are probably inapplicable to the fixation of PCBs. A chemically fixed product, on the other hand, is able to hold the contaminant even though it may fragment under service conditions. A third type of fixation product is one which contains microencapsulates. Microencapsulates consist of microscopic and macroscopic contaminant particles encased by the fixative to yield a product which maintains its integrity under environmental stress (Lubowik *et al.*[103]). PCB fluid has been disposed of as a nonleachable solid waste by microencapsulating it in concrete. The effectiveness of the method is limited for large volumes of PCBs because of the amount of concrete which is generated and the possible risk associated with the long term leachability of the product.

Organic wastes are not usually amenable to fixation but Envirosafe Services Inc., in Horsham, Pennsylvania has used pozzolanic chemistry to chemically fix PCBs in a rock-like product which has good load-bearing properties and which is hydrolytically stable. Three PCDD contaminated sites in Missouri have been used to assess the feasibility of cementitious and asphaltic stabilization techniques. Asphalt alone was found not to be satisfactory but a combination of lime and asphalt was found to solubilize the 2,3,7,8-TCDD sufficiently to avoid its detection in leachate studies.

VITRIFICATION

It situ soil vitrification is being proposed by Battelle (des Rosiers *et al.*[104]) in which heat from the electric current passed into electrodes placed in the ground converts the soil and its contaminants into a non-leaching fused glasslike material. Temperatures exceeding 20,000°C are produced. Gases and volatile materials are collected and treated on the surface. The product obtained is a grasslike material which is more or less homogeneous depending upon the duration of heating and the extent of convective mixing which has occurred while the zone is molten.

TRANSFORMER RETROFILL

The retrofilling of askarel transformers to a safe and cost-effective level would practically eliminate the 99.8% of PCBs located in transformers. Work done by M. Erickson *et al.*[23] at Midwest Research Institute on the formation of toxic pyrolysis products in PCB contaminated transformer oils has shown that there is a narrow range of temperature and oxygen conditions which permit the formation of PCDFs and PCDDs from PCBs and chlorobenzenes from dilute solutions. The probability of formation of toxic products from low concentration askarel residuals in retrofill fluids at concentrations of concern under real world conditions, however, seems to be extremely small. For example, a fire which occurred in the basement transformer vault of the Electrical Engineering Laboratory of the University of Manitoba during March, 1982 involved six mineral oil transformers contaminated with 250 ppm PCB. An electrical failure had caused one of the transformers to arc and catch fire. All six transformers were damaged by the fire. During fire-fighting attempts, a transformer exploded and sprayed PCB contaminated soot onto the walls. No PCDFs or PCDDs were detected.

When a transformer is retrofilled to < 500 ppm PCB one would expect that the rate of conversion of PCBs would be very much less than in the case of an askarel fire. It becomes increasingly less probable that a large enough quantity of pyrolysis products would be produced as the concentration of PCBs and chlorobenzenes is reduced. Indeed, the contaminated mineral oil fire in the University of Manitoba resulted in no detectable concentration of PCDFs in the soot. PCDDs, which are typically formed at an order of magnitude less concentration than PCDFs, were not detected either.

In 1970 Vos *et al.*[12] showed a correlation between the toxic effects of European PCBs and the concentration levels of PCDFs. The major PCDF components contained in Yusho oil were the highly toxic 2,3,7,8-TCDF and 2,3,4,7,8-pentaCDF. The relative concentrations of the PCDF isomers present in Yusho oil and in samples of used heat exchanger PCBs (Kanechlor KC 400 and Mitsubishi-Monsanto T 1248) were found to be strikingly similar (Kuratsune *et al.*[29]). The overall toxicity of the fluid may then be attributable to the presence of small quantities of PCDF as degradation products of PCBs.

A 1981 EPRI study[105] of the equilibrium distribution of PCBs between transformer solid materials and the liquid dielectric showed that, for a typical 500 kVA transformer, about 97.5% of PCBs are dissolved in the oil while only 2.5% are unevenly distributed among the paper, core steel and Formvar wire. The majority of the PCB on the solids is contained either adsorbed or absorbed in the paper.

An oil temperature fluctuation between 50°C and 110°C causes a maximum shift of only 0.6% in the amount of dissolved PCBs when adsorption/desorption is considered alone. The effect is much larger when multiple layers of paper are involved because the migration of absorbed PCBs through the capillaries of the impregnated paper causes a gradual leaching which results in a final PCB concentration in the bulk oil which usually exceeds the EPA definition of a PCB fluid.

The study contains important implications regarding the efficacy of retrofilling transformers using *current* technology. The difficulties in obtaining a cost effective retrofill solution would be avoided if technology were proven to circumvent the problem of PCB migration into the bulk oil. Here, again, it is important to recognize clearly the difference between the retrofill of PCB contaminated transformers and askarel transformers which have been impregnated with PCBs.

In the case of a PCB contaminated mineral oil transformer it would be possible to reduce the PCB concentration of the working fluid by first draining the transformer and refilling it with non-contaminated oil. Since the large majority of PCBs are contained in solution, and are not trapped by the porous insulation, it is possible to retrofill the transformer to a required concentration level by repeated washings, as necessary. The drained, PCB contaminated oil can be disposed of by one of the several suitable methods discussed in this section. On the other hand, even if all the askarel fluid of a PCB transformer were to be washed out there would, according to the EPRI report, still be about 2.5% of the original askarel contained in the core/coil assembly. For a typical 235 gal. PCB transformer this amounts to about 5.9 gal. of askarel impregnating the porous insulation. A PCB concentration of 500 ppm in a 235 gal. transformer amounts only to about one coffee mug in volume.

The diffusion of PCBs under the action of a strong concentration gradient from the interior of the paper towards the very low concentration in the bulk retrofill fluid has been discussed earlier. If the trapped PCBs are able to leach out into the bulk retrofill fluid at a rate of as little as 1 ppm per day, it would need less than 1.25 years before the fluid concentration had climbed above 500 ppm and the unit would have to be regarded once again as a PCB transformer.

The molecular geometry of the PCB molecule makes it susceptible to adsorption on carbon. Therefore, when an askarel transformer is drained and filled with polymethylsiloxane (silicone) oil rather than an oil which contains molecules with the geometry or benzene rings, the only aromatic structures present are those of the contaminant to be removed. The capacity of the carbon for the adsorption of aromatic species can be determined from Freundlich's characteristic isotherm. This can then be used to calculate the number of filters of a given size required to reduce the residual askarel concentration in the bulk oil to a defined lower concentration. The higher the initial concentration of PCBs remaining after draining the transformer, the larger will be the number of filters needed to adsorb the residual PCBs.

As more filters are used so the overall cost of the retrofill approaches the cost of a new transformer. The rate of PCB reduction is partly governed by the recontamination of the oil and partly by the adsorption/desorption/leaching process itself.

(2) THE DISPOSAL OF PCB CONTAMINATED LIQUIDS

Disposal options for PCB contaminated liquids are shown in Figure 47. As stated earlier, PCB contaminated liquids are understood to mean that the PCB is a minority constituent of the fluid, usually at a concentration measured in parts per million, but not necessarily less than the 500 ppm definition of PCB fluid regulated by the EPA.

The most easily identified large volume of PCB contaminated fluid is that used by the electrical industry in its mineral oil transformers and the emphasis in this paper is therefore placed upon the disposal of contaminated dielectric oils. Contaminated fluids can be incinerated in the same way as askarels but, in addition, the EPA has granted approval for some companies to destroy their own low-level oils in high efficiency boilers. The disadvantage of the process is that it not only destroys the PCBs but also the transformer oil.

Environmental Science and Engineering Inc. in Gainesville Florida has conducted an analysis of the pyrolysis products of dilute PCBs in a utility "High Efficiency Boiler" and found that PCDFs and PCDDs were not detectable by the GC/MS method used. PCBs were destroyed with

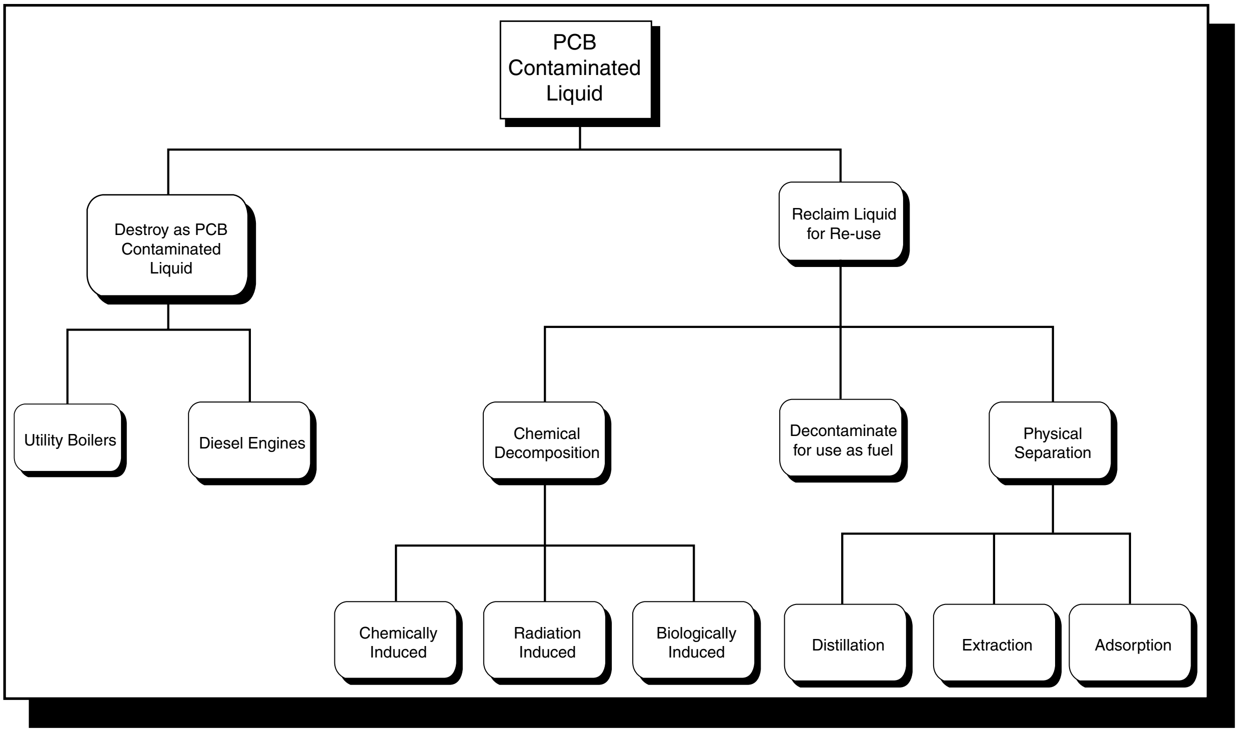


FIGURE 47 The disposal of PCB contaminated liquids.

99.99999% efficiency. Because of the increasing short supply of base stocks suitable for use as electrical transformer fluids it is becoming increasingly necessary to reclaim the oil. Also, the once empirical reclamation procedures practiced in the industry are on a sound scientific footing today so that there is no need for the concern that there once was regarding the performance of reclaimed product (Webber[106]).

Physical processing methods which do not chemically alter the components of the oil, and which do not leave residues of the process in the decontaminated fluid, are less likely to result in changes to the properties of the oil than are chemical destruction methods.

CHEMICAL EFFECTS OF DECONTAMINATION PROCESSES

The accepted properties of new mineral transformer oil and oil requiring reclamation are summarized in Table 54. It is appropriate at this point to describe the characteristics of transformer oil.

Power Factor The ratio of the power in watts dissipated in the oil, to the product of the effective sinusoidal voltage and current in volt-amperes. It is expressed numerically as the sine of the dielectric loss angle. An increase in power factor corresponds to an increase in the concentration of polar constituents in the oil and is used as a measure of the extent of oxidation.

Neutralization Number The total acid number is defined as the quantity of base, expressed in milligrams of potassium hydroxide, required to neutralize the acidic compounds in 1 g of oil. The acids present in service-aged oils are due to oxidative degradation.

Interfacial Tension The molecular attractive force between oil and water expressed in units of dynes per centimeter. As the concentration of polar contaminants in the oil increases the interfacial tension (IFT) value decreases. The test is therefore an indirect measure of the oxidation which has taken place in the oil.

An important question concerning reclamation is whether or not reclaimed oil is equivalent to new oil and, in particular, whether or not reclaimed PCB decontaminated oil is equivalent to new oil. The fuller's earth treatment of partially oxidized oils is primarily effective in reducing the neutralization number and power factor to acceptable levels but is relatively less effective in producing values of interfacial tension, color, etc., typical of new oils. Fuller's earth treatment does not remove PCBs or ditertiary-butyl-p-cresol (DBPC) oxidation inhibitor (Webber[106]). However, chemical processes for the PCB decontamination of oil involve highly reactive alkali metals or organometallic reagents which have a strong affinity for the most easily reduced components of the oil.

Polarographic reduction is a measure of the ease with which a compound will accept an electron and this therefore allows an assessment of the relative reactivities of radical

TABLE 54

Test	ASTM test method	New oil	Typical properties of oil requiring reclamation
Dielectric Breakdown	D–877	>30	22
(kV)	D–1816	>56	—
Power Factor 60 Hz, 25°C	D–924	<0.05	1.6
Neutralization Number (mg KOH/g oil)	D–974	<0.03	0.40
Interfacial Tension (dynes/cm)	D–971	40	18
Moisture Content (ppm)	D–1533	—	60

anions. Polarographic measurements are usually made in 75% aqueous dioxane solution of the flow of current through the solution as a function of the applied potential. When reduction takes place a wave of current is produced. The potential measured at the mid-point of the wave is referred to as the $e_{1/2}$, value. The $e_{1/2}$ values for some of the aromatic hydrocarbons which can be used for PCB decontamination agents, together with values for their electron affinities, are listed in Table 55.

TABLE 55

Compound	Electron affinity (eV)	$e_{1/2}$
Biphenyl	—	2.70
Naphthalene	0.152	2.50
Phenanthrene	0.308	2.46
Anthracene	0.552	1.9

As one goes down the list, the electron affinity of the larger aromatic molecules reflects the relative ease with which an electron added to the system can spread itself through the structure without causing high energy points of stress. The relative ease of reduction therefore follows the same order and anthracene is much more easily reduced than biphenyl. The reactivity of the radical ions is therefore expected to decrease down the list. Thus sodium biphenyl has been known for many years to quickly and quantitatively react with organochlorine compounds. The reaction is useful for use in the field as an inexpensive screening tool for potentially contaminated oils. Also, sodium naphthalenide is used in the Goodyear/(Smith and Bubbar) process (Smith and Bubbar[107]) for the PCB decontamination of fluids. The reactions of alkali metals and organometallic reagents with the classes of compounds typically found in insulating oils is perhaps best discussed in terms of compound classes. A typical distribution of compound types in mineral transformer oils is shown in Table 46.

ALKANES

Oils which are totally paraffinic are essentially inert to radical ions. It has been found for example, that when a sodium dispersion is made in a high molecular weight paraffinic oil direct from the manufacturer, and then subsequently filtered, that the oil properties remain unchanged.

ALKENES

This class of compounds is not expected to occur in significant quantities in transformer oils. If they were present, they would most probably undergo their most common reaction of polymerization and precipitate out of the oil into the sludge layer. An alternative possibility is that the olefin may add to the alkyl side chain of an alkyl aromatic compound. In either case it is most unlikely that the products of such reactions would deleteriously affect oil properties and indeed such reaction products may occur naturally in some oils.

ALDEHYDES AND KETONES

Aldehydes are a class of compound which, because of their reactivity, are not expected to be present in oxidized oil in significant quantities. Ketones, on the other hand, are known to be present as primary oxidation products (Webber[106]; Holy[108]).

The oxidation of hydrocarbons produces, successively, hydro-peroxides, alcohols and ketones, and finally acids. The fuller's earth reclamation process is effective in the removal of acids but does not remove their precursors. The presence of even small quantities of acid can cause a catalytic effect on the decomposition of ketones to give further acids. Oil which has been reclaimed by the conventional fuller's earth treatment is therefore likely to oxidize more rapidly than new oil with the same concentration of oxidation inhibitor and this is usually found in practice.

The reactions of aromatic ketones with radical anions lead to reduction of the ketone and yield an alcohol as the major product (Holy[108]). The importance of alcohols as oxidation inhibitors can be understood in terms of the mechanism of oil oxidation as follows:

$$
\begin{array}{lll}
\text{Initiator} + [\text{R}-\text{H}] & \rightarrow & \underset{\text{I}}{\text{R}\cdot} + \underset{\text{I}}{\text{R}'\cdot}\ (\text{Initiation}) \\
\text{R}\cdot + \text{O}_2 & \rightarrow & \underset{\text{II}}{\text{ROO}\cdot} \\
\text{ROO}\cdot + \text{R}-\text{H} & \rightarrow & \underset{\text{III}}{\text{ROOH}} + \text{R}\cdot\ (\text{Chain Propagation}) \\
\text{ROO}\cdot + \text{R}\cdot & \rightarrow & \text{ROOR}\ (\text{Termination}) \\
\text{R}\cdot + \text{R}'\cdot & \rightarrow & \text{R}-\text{R}'\ (\text{non-radical products}) \\
\text{ROO}\cdot + \text{R}'\text{OO}\cdot & \rightarrow & \underset{\text{IV}}{\text{ROH}} + \underset{\text{V}}{\text{RCOR}}
\end{array}
$$

The hydrocarbon in the initiation step splits into free radicals (I) which, with dissolved oxygen, forms peroxide radicals (II). The reactivity of a peroxide radicals depends upon its structure. If it is sufficiently reactive, it will be able to abstract a hydrogen atom from a nearby oil hydrocarbon molecule and partially stabilize itself as a hydroperoxide molecule and yield a further radical. This is the chain propagation step.

The peroxide radical, (II), can be terminated by reaction with the initial hydrocarbon radical, (I), to produce an ester. Alternatively, two peroxide radicals can react to give an alcohol and a ketone or two hydrocarbon radicals can react to form non-radical products.

The hydroperoxide can decompose as follows (degenerate branching):

$$
\begin{array}{l}
\text{ROOH} \rightarrow \text{RO}\cdot + \cdot\text{OH} \\
2\text{ROOH} \rightarrow \text{RO}\cdot + \text{H}_2\text{O} + \text{ROO}\cdot
\end{array}
$$

The reaction scheme illustrates why it is that a large variety of oxidation products is typically found in used oil. The rate of oxidation is determined by the concentration of peroxy radicals. When an oxidation inhibitor is added to the system, its primary function is to scavenge peroxy radicals and in doing so itself becomes a radical.

The important difference between the initial peroxy radical and the newly formed species is that the product radical is more stable than the reactant radical. The degree to which the product radical is more stable is a measure of the effectiveness of the compound as an oxidation inhibitor. Such compounds as phenol show an inhibitive effect towards oxidation in white mineral oil (Ingold[109]), but the most efficient inhibitors are found to be sterically hindered phenols such as di-tertiary-butyl-p-cresol, DBPC. From the above it would be excepted that the formation of aromatic alcohols from a chemical decontamination process in particular would give rise to "natural" oxidation inhibitors. It is probable that the inhibition produced in this way is less than would be achieved using DBPC but at least the oil should not be deleteriously affected.

The presence of oxidation inhibitors is important for the reasons that they are essential to the oxidation stability of the reclaimed oil and that oxidation inhibitor retards the free radical dechlorination reaction of reagents with PCBs. In some cases, for example reactions with metallic sodium, the dechlorination reaction shows a very marked "induction" period while the alkali metal reacts with oxidation inhibitor. The reaction products are highly colored because of their quinoidal structure and it is these compounds which are the cause of the yellow color normally attributed to oxidized oil. The quinoidal product is not sufficiently polar to be adsorbed by fuller's earth during the reclamation process and is not detrimental to the dielectric properties of the oil.

A characteristic of the reaction of alkali metals and organo-metallic reagents with oil components and PCBs is that a sludge is produced. Filtration of the fine precipitate, followed by washing and analysis by a variety of analytical procedures, shows (Webber *et al.*[61]) that the sludge contains predominantly non-chlorinated polyphenyls and is therefore not a toxic hazard.

Biphenyl is a reaction product together with a large range of high molecular weight polyphenyls. It is the production of partially soluble compounds in the intermediate molecular weight range which can sometimes cause problems. The reason is that long chain polyphenyls can exist as biradicals (Brown *et al.*[63]) and contribute strongly towards a high power factor oil. The effect is particularly noticeable when treatments are applied to mineral oils which contain more than a trace of high molecular weight aromatics. In these cases, if the oil is left standing in contact with organometallic reagents for a period of a few hours, the aromatics in the oil provide enough solubility for the polyphenyls to produce a power factor of more than 50%. The problem can usually be avoided by separating excess chemical reagent from the decontaminated oil immediately after the dechlorination reaction has taken place.

DISTILLATION AND EXTRACTION

The separation by distillation of transformer oil and askarels has been evaluated by Battelle-Columbus Laboratories and D.&D. Disposal Inc. Transformer oil is usually a high boiling fraction of crude oil which is distilled under vacuum to avoid degradation. The oil contains a large variety of low vapor pressure compounds. PCBs also have a low vapor pressure and the difficulty in separating a small quantity of contaminant from the oil lies in the overlap of the boiling ranges of the components. Since the PCBs are typically higher boiling than most mineral oil components, conventional distillation requires that most of the fluid be distilled before the PCBs can be separated as a residue. The process is both inefficient and requires a lot of energy.

Steam stripping is an alternative to vacuum distillation. Separations can be achieved at atmospheric pressure at temperatures which do not cause thermal degradation. The method shares the disadvantage of conventional distillation that many components exhibit overlapping boiling ranges. The addition to the PCB in oil system of an agent capable of forming a low boiling azeotrope has been investigated. In this case, separation can be accomplished by distilling a small quantity of azeotrope rather than a large quantity of transformer oil.

Solvent extraction of PCBs from transformer oil has been investigated by Oak Ridge National Laboratory using dimethyl formamide. A second stage solvent extraction with water reduces the volume of fluid for disposal. The process may be found particularly useful in the treatment of PCB contaminated lubricating oils since fluids of this type typically contain additives which tend to react with alkali metal organometallic dechlorinating reagents thereby making the process expensive in both reagent and additive replacement cost.

The Franklin Research Center has reported (laconianni *et al.*[110]) a study in which measurements were made of the partition coefficients of PCBs in a range of potential solvents while General Electric, under EPRI contract, has developed a solvent extraction process for separating PCB from mineral oil. Patents have been assigned to the USA for the extraction of PCBs from oils using organic solvents such as methanol or isopropanol. In these cases the extracted mixture can be distilled to recover solvent and the solvent recycled. The difficulties encountered with these methods may out-weight their potential advantages in the treatment of transformer oils. It does not seem likely that the solutions offered by physical methods will compete favourably with those offered by chemical destruction systems.

RADIATION

The dehalogenation of askarels involves an initial addition of an electron to the aromatic molecule. There is little chemical difference whether the electron is derived from metallic alkali metals or organometallic compounds or from electrons generated in solution by the absorption of radiation. The UV-photolysis of PCBs in organic solutions produces reductive dechlorination of the PCB as the main photochemical reaction while in aqueous or alcoholic solution replacement of the chlorine by a hydroxyl group also occurs (Plummer[111]). A review on the photochemical degradation of PCBs to yield less toxic and more toxic products has been written by Safe *et al.*[112]

The Atlantic Research Corporation in Alexandria, Virginia experimented with a UV/H_2 system to dechlorinate PCBs in 1983. More recently, G.A. Epling and coworkers at the University of Connecticut[113] have reported on a hydride-enhanced photoreaction of PCDDs in which the reaction system was found not to be overwhelmingly influenced by the presence of impurities. In general, impurities tend to quench target molecule excited states and competitive light absorption by impurities also causes a slowdown in the rate of photoreaction. In the PCDD isomer systems tested, however, the presence of borohydride caused an enhancement of photoreaction in the range of about 2 to 6 times the rate observed without borohydride. Also, the photoreaction in the presence of sodium borohydride proceeded to yield predominantly the direct dehalogenated product. Ultimately, the completely dechlorinated dibenzo-p-dioxin was formed.

BH_{4-} can be used as a strong nucleophile in the presence of catalytic amounts of tris(triphenylphosphine)-nickel(0). The reaction has been reported[159] by Shaw-Tao Lin and J.A. Roth (1979). The reaction uses conditions of about 70°C, a nitrogen blanket, and a reaction time of 16 hours to produce an 80% benzene yield.

Roth *et al.* (1994)[160] describes the hydrodechlorination of a series of PCBs. They found that $NaBH_4$ caused the dechlorination of mono- and di-chlorobiphenyls in DMF solution without catalysts, but the reaction was very slow (15–20 hours) and the extent of dechlorination was very small (1–5%). On the other hand, Ni_2B reactions in which a reactive form of hydrogen is on the surface of the solid was said to produce dechlorination. Roth also used a homogeneous dechlorination reaction with tetrakis-(triphenylphosphine)-nickel (0), which is an air sensitive complex soluble in dimethyl formamide. Small concentrations of around 2000 ppm PCB were dechlorinated between 60–100 minutes at room temperature.

The Westgate Research Corporation in Los Angeles, California has developed a process known as the Ultrox system which involves the UV-catalyzed ozone oxidation of chlorinated organics in industrial waters. Hydroxyl radicals are frequently proposed as the reactive species in the UV-ozonation of organic compounds. However, according to the experimental data obtained by Leitis *et al.*[114] the major oxidizing species in the UV/O_3 Ultrox system is ozone itself. Carbonhalogen bonds are broken before other structures are oxidized and the halogen atoms appear in solution as halide ions. A synergistic effect between ozone and UV light has been observed. The magnitude of the effect is greater at low (1 ppm) that at high (100 ppm) concentrations of PCB.

It is postulated that the mechanism of oxidation of chlorobenzene in the Ultrox process is as follows:

Chlorobenzene + UV/O_3 → chlorophenol
→ mixture of [(chloro)
catechol + (chloro)
resorcinol +
(chloro)hydroxyquinone → trihydroxyl(chloro)
benzenes
→ (chloro)muconic acids
→ maleic aldehyde
+ glyoxylic acid
→ oxalic acid
→ $CO_2 + H_2O$

A 40,000 gallon per day (gpd) Ultrox system consisting of a reactor and ozone generator, costs about \$125,000 with O&M costs per day of about \$175. For a 150,000 gpd system the equipment cost is about \$300,000. O&M costs for the large system are about \$330 per day so that the cost to process 1000 gal. of waste water, including monitoring labor, is about \$4.35 for the smaller unit and about \$2.21 for the larger system.

Virontec has applied the UV/O_3 technology in its RADINOX® process. A full-scale treatment system was applied to a PCP-contaminated wood preserving site in Oregon during 1988. Data from bench scale tests showed that PCP concentrations were reduced from 10 mg/L to 0.5 mg/L in a 30 gal./min. continuous flow reactor.

C.H. Langford and his co-workers[115] have studied the UV photodecomposition of several chloroaromatic compounds using a semiconductor catalyst. It was suggested that a photogenerated reactive species is desorbed from the TiO_2 surface by band-gap excitation. The species was not identified but it was pointed out that hydroxyl radicals have been considered as the reactive species in several studies and that the superoxide ion (O_2) shows nucleophilic and oxidative characteristics in a variety of systems (Matthews[116]). In particular, Sugimoto *et al.*[117] showed that polychlorinated aromatics were degraded by a superoxide ion initiated pathway.

R. Niessen *et al.*[118] have studied the phototransformation of phenols by excitation of nitrate ions. They pointed out that nitrates are probably one of the main sources of hydroxyl radicals in natural waters exposed to daylight. The gamma-ray radiolysis of halogenated benzenes was investigated by Choi and Willard" using a high activity Co^{60} source. It was found that many similarities existed between UV photolysis reactions and those initiated by gamma rays. The work was applied in 1971 to an environmentally persistent pesticide, DDT, and in the following year Arai *et al.*[58] applied the method to solutions of PCBs.

A high energy gamma-ray passing through matter typically leaves a track of ionized particles and electrons. The solvated electrons have a very high affinity for atoms such as chlorine and abstract them from the PCB molecule as chloride ion. The reaction leaves the original PCB in a highly reactive "free radical" state with one chlorine less than it had originally. A free radical chain reaction can then occur in which there is a cascade of progressively more dechlorinated free radical products. Eventually, a stable configuration is reached and biphenyl and polyphenyl products are observed.

Van de Graff generated electrons have been used to study the effects of high energy electron irradiation (3 MeV) on persistent toxic trace components in sludges and wastewater by Merrill *et al.*[119] at MIT. It was observed that in pure, deoxygenated solutions dechlorination did occur and that the reaction products became more water soluble than the starting compounds. Aromatic carbons and typical insulating oil compounds are resistant to breakdown by high energy radiations. A project was therefore undertaken for the Canadian Electrical Association (Webber[60]) to investigate the possibility of decontaminating electrical insulating oils using beta-particles (2 MeV electrons) from a radioactive Sr^{90} source.

Irradiations were conducted at Simon Fraser University using different askarels in transformer oil both with and without isopropanol, potassium hydroxide and air. The PCB mixtures were found to be dechlorinated by beta-particle radiolysis most efficiently in isopropanol/KOH solutions. The solution acts as an alkaline hydrogen donor and allows free radical chain propagation to occur in the absence of oxygen as a free radical scavenger. The reaction also occurs when commercial askarels are dissolved in transformer oils but very large radiation doses, and hence reaction times, are then required for complete dechlorination. The efficiency of the Sr^{90} induced beta-particle reaction measured in terms of its $G(Cl^-)$ value, was found to be only about 7×10^{-3} i.e., about 10^5 times smaller than the efficiency of the Co^{60} gamma-induced reaction in Willard and Choi's work[57] but about 400 times more efficient than the reaction induced by high energy, accelerated electrons. The work was stopped because of the lack of applicable reaction kinetics and because the method has the potential to produce the hydroxylated precursors of PCDFs and PCDDs. The reaction mechanism was discussed in an earlier section on the formation of compounds of concern and shown in Figure 18.

BACTERIOLOGICAL

The distribution and fate of PCBs in the environment depends on their dispersion and partition as well as their susceptibility to degradation by various routes.

A number of papers have been written on the subject[120–132] and this area of investigation may become increasingly important for the in-situ decontamination of soils and the cleanup of spills. Biological treatments for industrial waste systems have been in use for many years but it is only recently that new microbial strains have been developed which show promise for the treatment of environmentally persistent compounds.

In a study on the persistence of Aroclor 1254 in various California soils it was found by Iwata *et al.*[132] that some considerable changes had taken place after 1 yr. of incubation at 30°C and 40% moisture content. It was pointed out, however, that the loss of some of the components may have been due to evaporation. Bailey *et al.*[133] used C^{14}-labelled compounds to obtain material balances in their PCB degradation study in river water. They found that monochlorobiphenyl degraded rapidly ($t_{1/2}$ <2 days) to chlorobenzoic acid which, in turn, degraded to carbon dioxide and water. By way of contrast, the 2,2',4,4'-tetrachlorobiphenyl studied under similar conditions showed no measurable degradation during a period of 50 days. The degradability of PCBs in lake sediments has been found to decrease with an increasing degree of chlorination of the biphenyl. Half-lives for the different congeners range from a few days for dichlorobiphenyl to 200 days for pentachloro-compounds. Biodegradability is also affected by the position of chlorine substitution on the biphenyl ring. Neely[134] has developed a mathematical model to predict the lifetime of selected chlorobiphenyl congeners in the environment. His conclusions predict a half-life of about 3 days for monochlorobiphenyl, 150 days for di-, 579 days for tri-, 1044 days for tetra- and 3445 days for penta-.

Camoni *et al.*[135] evaluated the microbial degradation of TCDD in Seveso soil. Samples were analyzed in the presence and absence of organic compost. The results indicated a 25 No degradation of any of the% reduction in TCDD concentration in 480 days. Studies have also indicated that the average TCDD remaining in the soil after 1 year of weathering was approximately 50% at all the concentrations tested (1 to 10 ppm). The results are complicated by the uncertain contribution of photodegradative effects. On the other hand, research by Huetter *et al.*[156] and also by Young *et al.*[137] indicates that microbial degeneration of TCDD in soil is very slow under the most optimum conditions and therefore tends to support a conclusion that PCDDs are essentially non-biodegradable. Several strains of bacteria have been discovered which can metabolize PCBs. These represent five genera, Acetobacter, Acinetobacter, Alcaligenes, Klebsiella and Pseudomonas. The bacteria are found in river sediments.

A combination of photodegradation and bacterial degradation has been investigated by R.A. Clyde in Asheville, N.C.A. porous, high area, fiber webbing impregnated with bacteria was mounted inside a tubular reactor so that the webbing could be rotated on an axle and drawn through the solution flowing through the lower half of the tube. A subsequent chamber admitted UV light to degrade any reactive material remaining after the bacterial treatment. The process has been reported to be useful for the treatment of wastewaters containing metal ions and may be found useful for large scale application to PCB contaminated water. A patent was issued on the process (#4,446,236 May 1, 1984).

It is well known that biological degradation of oil occurs at an accelerated rate in the presence of emulsifiers. Petroleum Fermentations Inc. has several patents concerned with the production and use of such emulsifiers. Emulsification clearly plays an important role in promoting contact between PCB contaminated oil and bacteria in the aqueous phase. Researchers at the University of Georgia have found, for example, that Pseudomonas aeruginosa and Serratia liquefaciens, incubated in a medium containing Tween 80, were able to destroy 97% of the PCB present within 90 to 130 days.

The Polybac Corporation in Allentown, PA, have demonstrated the in-situ degradation of PCBs using a proprietary blend of bacteria. The higher chlorinated PCB isomers were found to be unaffected during the 5–6 month treatment program and were therefore chemically reduced prior to bacterial treatment.

Fundamental biochemical degradation mechanisms are not well understood. Chemical pathways are therefore difficult to predict and the potential exists for the release of hazardous compounds into the environment through incomplete degradation or system failure. Degradation rates in the field tend to be much slower than laboratory rates where pure cultures are tested on pure compounds. Also, the problem of PCB identification and quantitation can be very difficult in partially degraded mixtures and process streams.

The identification of chlorobiphenyl congeners has been the subject of extensive research. Safe *et al.*[138] at Texas A&M University has synthesized most of the 209 PCB congeners and has used these to identify the constituents of Aroclor mixtures using high resolution gas chromatography (HRGC). Similar work has been done by Pellizzari *et al.*[139] who identified 73 PCB congeners in a mixture of Aroclor 1016, 1254 and 1260 by HRGC/negative chemical ionization mass spectrometry (HRGC/NCI-MS). Publications by Stalling and his co-workers[140,141] describe a data base for the isomer specific determination of PCBs and a pattern recognition method for the classification and determination of PCBs in environmental samples. Current research is focusing on the application of "white rot" fungus, Phanerochacte Chrysosporium, to degrade toxic halogenated organics. The fungus enzyme is one of the strongest, nonspecific, oxidizing enzymes known. In labscale tests, pentachlorophenol in water was reduced from 250 mg/L to 5 mg/L in 24h (des Rosiers[142]).

The mechanism of action of P. chrysosporium involves an initial one electron oxidation to produce an aryl cation radical which then undergoes cleavage and further oxidation to produce quinones. The quinones are degraded further by a quinone reducing enzyme until finally all of the organically bound chlorine becomes chloride.

ELECTROCHEMICAL

Russling *et al.*,[143] at the University of Connecticut, has investigated the application of controlled potential electrolysis as a method to dechlorinate PCBs. Radical anions of such compounds as anthracene, 9,10-diphenylanthracene and phthalonitrile in dimethylformamide solution form radical anions at a Hg electrode surface. The radical anions react with PCBs in solution to produce chloride in much the same way as the mechanism of action of organometallic reagents. Stepwise reduction is observed with biphenyl as an end product.

Kinetics of the reactions were characterized by cyclic voltammetry. It was observed that the rate constant for the rate determining electron exchange between radical anion and chlorobiphenyl varied inversely with the difference in reduction potential between the radical anion and the chlorobiphenyl.

Hydrogen peroxide generated by an alternating current field has been used as an oxidant for the PCB decontamination of aqueous wastes. Westinghouse Electric Corporation in Pittsburgh, Pennsylvania, has patented an apparatus for the electrolysis of water which consists of a series of packed beds of alternating high and low electrical conductivity. Each bed has separately controlled pairs of electrodes. The low conductivity bed contains conductive particles such as carbon or nonconductive particles coated with oxides such as MnO_2 or PbO_2. The proportion of particles in the more conductive bed is about 30% oxide coated particles with the balance as conducting and/or absorptive particles. A series of beds was constructed in the effluent flow path and 29 volts/12 amps passed into each of the less conducting beds while 10 volts/15 amps was passed into each of the more conductive beds. The electrodes were cooled so that the process temperature could be minimized. The PCB concentration was reduced from 120 ppb to 1.5 ppb.

DECHLORINATION REACTIONS OF ALKALI METALS AND ORGANOMETALLIC REAGENTS

The chemical decontamination systems which have been developed all share a common factor based upon the transfer of electrons to the PCB molecule from an alkali metal such as sodium or potassium, or an organometallic reagent, such as an alkali metal-aromatic compound. The processes differ in details such as the air and moisture sensitivity of the reagent, side reactions to give partially dechlorinated toxic products, inhibition by radical scavengers such as oxidation inhibitors and the degree of processing required to produce a useful, decontaminated product. Public scrutiny has required that processes be researched very carefully to elucidate the details of reactions so that the PCB "problem" is removed rather than substituted.

The first large scale process for the PCB decontamination of fluids was developed (Parker[144]) by Goodyear Tire and Rubber Company in 1980 using the known chemistry of sodium naphthalenide (NaNp) as a dechlorinating agent for strongly bonded organic halogen compounds such as chlorobenzenes. The process was applied to the decontamination of heat transfer fluids containing about 120 ppm of PCBs. Earlier, in 1978, Oku *et al.*[145] reported the use of NaNp for the dechlorination of PCBs. Smith and Bubbar at the University of Waterloo, Ontario performed a research contract for the Canadian Electrical Association on the application of NaNp to PCB contaminated transformer oils in 1982. A patent issued on their process in 1982 (Smith and Bubbar[107]).

Sodium naphthalenide is prepared by reaction of an excess of dispersed sodium metal with naphthalene in a dry ether solvent, such as tetrahydrofuran, under a nitrogen atmosphere. The air unstable compound forms as dark green crystals and reacts with PCBs to form chloride ions, biphenyl and polyphenyls with a yield, based on chloride ion concentration, of 99.5%. Analysis of some of the non-chlorinated polymeric, aromatic residue by NMR spectroscopy indicated that one solvent tetrahydrofuran molecule was associated with each aromatic ring.

The reaction of NaNp with PCBs was found to be more efficient when the recovery of naphthalene was increased. This was interpreted as being due to the presence of an impurity produced from a side reaction which acts as an inhibitor to the dechlorination reaction. A similar inhibition of the dechlorination reaction was studied by Webber and Wilson[62] in the reaction between dispersed sodium and PCBs in oxidation inhibited transformer oils. The chemistry of this method is discussed later in this section.

Acurex Waste Technologies Inc. in Mountain View, California, have adapted the Goodyear/University of Waterloo process for use in the field to decontaminate transformer oils because naphthalene is a priority pollutant. The Acurex process utilizes a chemically similar, proprietary alternative. The large scale process developed by Goodyear was intended to yield a decontaminated fuel. The Acurex process yields a decontaminated oil which is filtered and returned to the customer for reuse or disposal although the extent of required further treatment is uncertain. Acurex does not recommend reuse of the oil. Smith and Bubbar are proposing a continuous, cyclic system which, it is claimed, can be built into a portable unit that could be carried on a trailer for on-site processing to yield a dielectrically acceptable oil.

Brown and Lynch at General Electric Company, Schenectady, New York, have obtained a patent (#4,377, 471) in March 1983, on a method for the dechlorination of PCBs in insulating oils by addition of a mixture of dispersed metallic sodium, an aprotic ion-complexing solvent such as tetrahydrofuran, and an oil-soluble electron carrier such as naphthalene. It was noted in Smith and Bubbar's patent on the action of NaNp that the presence of free sodium metal in the reaction medium, resulting from the use of an excess of sodium metal in the formation of sodium naphthalenide, appeared to enhance the effectiveness of the dechlorination reaction. The use of sodium dispersions for the removal of halogenated compounds from hydrocarbon oils is a well known reaction in the refining of oils. The application of the method to PCBs was described as early as 1973 by Japanese workers who used it to decontaminate kerosene used as a solvent for the extraction of PCBs from sewage sludge.

The B.C. Hydro process (Webber *et al.*[146]) for the decontamination of electrical insulating oils utilizes an alkali metal such as sodium dispersed directly into the contaminated oil. Rate studies of the reaction performed by Pilgrim and Webber[61] showed a trend which is characteristic of chemical treatments, namely that the more highly chlorinated PCBs react in a stepwise fashion to produce less chlorinated species which are ultimately completely dechlorinated. Samples withdrawn from the reacting dispersion were analyzed at periodic intervals by gas chromatography according to a method similar to the now standardized ASTM D4509. The later peaks in the chromatogram showed an initial decrease, declining sharply as the reaction proceeded. The intermediate peaks showed an initial increase, followed by a short plateau phase and a sharp decline. The least retained, least chlorinated congeners retain an almost constant peak height until the reaction is almost complete, at which time they quickly disappear from the chromatogram.

The suggested reaction scheme involves the formation of anion radical intermediates followed by a dechlorination step. Polyphenyls are produced from the reaction between biphenyl free radical species and a chlorobiphenyl free radical, followed by further dechlorination and ultimate precipitation as sludge. The reaction appears to be similar to the ENSR (SunOhio) PCBX process.

Further work by Pilgrim and Webber[61] on an accelerated PCB destruction process by metallic sodium in the presence of small quantities of isopropanol as a hydrogen donor showed that abstractable hydrogen altered the reaction kinetics and mechanism to such an extent that decontamination was completed in about 20 minutes at room temperature. Part of the accelerative effect is probably due to the regeneration of active sodium surface by direct reaction with isopropanol. The reaction with alcohol would tend to renew the sodium surface and avoid the passivating effect caused by the formation of sodium phenolate derived from 2,6-di-tertiary-butyl-p-cresol (DBPC) used as oxidation inhibitor. Evidence which suggests that an alteration of mechanism may have taken place is that relatively little polyphenyl sludge is produced in the accelerated reaction.

The inhibition of dechlorination by DBPC in the reaction of dispersed sodium with PCBs was investigated by Webber and Wilson[62]. The oxidation inhibitor interacts with peroxy radicals in solution to produce a phenoxy radical stabilized by resonance. The phenoxy radical undergoes second-order disproportionation to yield an unstable quinone methide. Subsequent dimerization occurs to produce an intensely yellow colored compound 3,3′,5,5′,-tetra-tert-butyl-4,4′-stilbenequinone and 1,2bis(3,5-di-t-butyl-4-hydroxyphenylethane).

To determine whether the stilbenequinone affected the PCB dechlorination reaction, a 1% Aroclor 1242 in oil solution was prepared with 0.3% of the stilbenequinone and heated with sodium dispersion. No reaction was observed in 10 hours where, in the absence of the stilbenequinone, complete dechlorination was observed in about 15 minutes. Exclusion of oxygen from the reaction, and the use of a dispersed form of sodium, was found to prevent the formation of the stilbenequinone and to overcome the inhibition of the PCB dechlorination reaction at concentrations of DBPC typical of transformer oil. When the sodium was dispersed into particles of about 10 μm and reacted under a nitrogen blanket, a 10,000 ppm Aroclor 1242 solution was completely dechlorinated in less than 15 min. at 130°C in the absence of DBPC. When the maximum recommended inhibitor concentration in dielectric oils of 0.3% DBPC was used, the reaction was complete in less than 30 min. under the same conditions. Exclusion of oxygen from the reaction mixture prevented the formation of stilbenequinone and thereby avoided the inhibition of dechlorination.

The interaction of metallic sodium with the oil produces a reducing environment which tends to react with both natural and added alcohol oxidation inhibitors to yield sodium chloride and, upon complete reaction, a dechlorinated polyphenyl sludge. When the reaction is conducted under a blanket of nitrogen there is little likelihood of producing toxic oxidation products such as PCDFs or PCDDs. Also the formation of the sodium salts of oxidation inhibitors and acids tend to precipitate and can be easily separated from the oil by filtration or washing. The product can then be easily reclaimed as a decontaminated dielectric oil suitable for further use by adding oxidation inhibitor.

Ontario Hydro has developed the laboratory scale reactions performed by the author at B.C. Hydro and has had success in cost effectively treating large quantities of oil to yield a reusable product. A disadvantage of the system is that excess alkali metal is very reactive and needs to be used with caution. Ontario Hydro has had built (1989) mobile, 40 foot trailers, containing processing equipment to apply the sodium technology.

ENSR's (SunOhio's) "PCBX" process is contained in a forty foot trailer accompanied by an auxiliary trailer containing reclamation and oil test equipment. The process has been engineered to be continuous, as opposed to a batch operation, and runs at about 10 gal./min. Oil is pumped from the transformer or storage tank, through a heater to a reaction chamber where dispersed sodium is metered into the process stream. The slurry is passed through a recovery system containing heat exchangers, filters, a centrifuge and vacuum degassing apparatus. PCBs are dechlorinated to less than 2 ppm PCB in oil.

Oxidation inhibitor is reported to be added to the product oil. No alteration of the oxidation stability of the product was observed when measured by ASTM D2440 and a low temperature modified version of ASTM D2112. The presence of sodium compounds in the oil returned to the transformer was shown to be less than 1 ppm. The cost of be "PCBX" trailers is high at about $500,000, but, like the B.C. Hydro process, it has been shown to be cost-effective in the treatment of a wide range of PCB concentrations.

Vertac Chemical Corporation in Memphis, Tennessee has developed and patented (Howard *et al.*[147]) a method for the destruction of PCDDs in agricultural chemicals. The PCDDs are formed in unwanted side reactions during the product synthesis. The process uses alkali metal alcoholates of short chain alkanols or polyalkoxyalkane glycols or alkali metal hydroxides to react with PCDDs. The extension of the process to the dechlorination of PCBs is obvious. The reaction takes place at about 140°–220°C at atmospheric pressure when sodium ethylene glycolate is used.

The alkali metal alcoholate can be produced by reacting alkali metal with a polyol. Alternatively, the reagent can be formed from an aqueous solution of sodium or potassium hydroxide with an alkane-polyol such as ethylene glycol in the presence of an azeotropic solvent under azeotropic distillation conditions.

The mechanism of reaction with TCDD is believed to occur in two steps:

$$\begin{aligned}\text{TCDD} + 4\text{NaOC}_2\text{H}_4\text{OH} &\rightarrow (\text{OC}_2\text{H}_4\text{OH}) \text{ — dibenzo-p-dioxin} + 4\text{NaCl}\\ &\rightarrow (\text{NaO})_4 \text{ — dibenzo-p-dioxin} + 4(\text{HOC}_2\text{H}_4)_2\text{O}\end{aligned}$$

The polyhydroxy aromatic species formed in the process are more readily decomposed by aerobic bacteriological action than the original toxic, water insoluble, stable compounds. The sodium phenolate product can be acidified and safely separated from the aqueous wash solution containing dissolved inorganic salts.

Brunelle and Singleton[148] have reported the reaction between potassium hydroxide, polyethylene glycol (PEG) and PCBs. It was found that the complete removal of up to 1% PCBs from transformer oil could be achieved in less than two hours at 100°C. The advantage of the system over process which involve alkali metal organometallic reagents is that it is not moisture or air sensitive.

The Franklin Institute in Philadelphia, Pennsylvania has patented the "NaPEG" process for the detoxification of transformer oils (Lee[149]). The reagent used is derived from the reaction of alkali metal with a polyglycol such as polyethyleneglycol with an average molecular weight of about 400. The preparation can be described in the following equations:

$$\text{Na} + \text{PEG} \rightarrow \text{NaO — (PEG)} + \text{H}_2$$

$$\text{NaO} \quad \text{(PEG)} + \text{O}_2 \rightarrow \text{NaPEG reagent}$$

The NaPEG agent, like the Vertac reagent and the GE reagent, is both air and water stable and therefore has a long shelf life without special handling precautions. Dechlorination of PCBs in transformer oils can be achieved by heating the reagent and oil at temperatures between 40°C and 180°C depending upon the particular reagent chosen and the degree of chlorination of the PCB. The reaction evolves hydrogen gas and produces polyhydroxylated biphenyls and sodium chloride. The effectiveness of the system has been proven on a large scale (Garland[150]) by the Philadelphia Electric Company (PECO) who have subsequently reused the decontaminated oil produced in a pilot plant. It was observed that air or oxygen was necessary to make the reagent active towards PCBs in insulating oils and that the use of pure oxygen enhanced the rate of dehalogenation by a factor of five. This is in contrast to Brunelle and Singleton's work described above since they found that the presence or absence of air (oxygen) had no effect on the rate or outcome of the reaction. Also, in order to preserve the oxidative stability of the oil the reactions were recommended[148] to be conducted under an inert atmosphere. Webber and Wilson's work[62] on the inhibition of PCB dechlorination in oxidation inhibited oils suggests that the cause of the NaPEG reagent inactivity may be that the DBPC oxidation inhibitor oxidatively dimerizes to produce the observed yellow solution of stilbenequinone which then acts as an efficient scavenger of the radicals produced in the dechlorination steps. The formation of phenoxy radicals in Webber and Wilson's work was indicated by ESR spectroscopy. The presence of peroxy species in the NaPEG process was also observed.

The structures of naphthalene, anthracene and biphenyl radical anions with alkali metal gagenions have been interpreted by the alkali metal spin density distributions in the aromatic systems (West[151]). Webber *et al.*[152] have utilized the chemistry of this organometallic series of complexes to produce a preferred process for the PCB decontamination of insulating oil which is unaffected by the presence of oxidation inhibitors and which can be used at moderately low temperatures in a continuous system to yield a dielectric quality, decontaminated product.

Tiernan *et al.*[153] at Wright State University have reported on the use of KPEG reagent on PCDD/PCDF contaminated soil at hazardous waste sites. A similar paper has also been given by DesRosiers[154]. A 40 foot mobile trailer containing a 2700 gallon stainless steel reactor was used with 600 gallons of KPEG to treat 1400–2000 gallons of waste material. The mixture was heated to 150°C for 1.5 h. Under optimal conditions the chemistry allows complete dechlorination. A practical destruction efficiency of $>99\%$ is observed at reaction temperatures above 70°C with reaction times of 15 minutes to 1 hour. Interestingly, bioassays of residues from the reaction were negative. The cost of the KPEG treatment has been estimated to be approximately $24.00/gallon i.e., about 10% of the cost of on-site incineration.

UNCONTROLLED REACTIONS

The extreme toxicity of PCDFs and PCDDs discussed earlier makes it imperative that uncontrolled side reactions in processes used to chlorinate PCBs should be investigated and steps taken to avoid the generation of concentrations of compounds of concern.

Uncontrolled reactions during the large scale preparation of trichlorophenol from tetrachlorobenzene and sodium hydroxide in ethylene glycol have resulted in runaway conditions on at least six occasions and resulted in the formation of PCDDs. The Seveso incident, which occurred in 1976, is the most recent and is probably the best documented. Ethylene glycol used as solvent in the reaction can undergo an exothermic base catalyzed polymerization above 180°C to produce substantial quantities of TCDD from the dimerization of chlorophenol.

If the reaction conditions in a process to destroy askarels can yield chlorophenols as a side reaction product then the potential exists for the generation of PCDFs and/or PCDDs. The reaction mechanisms which explain the formation of the cyclized products is described in an earlier section.

The dimerization of chlorophenates is a bimolecular reaction and therefore the products formed should be highly dependent upon the chlorophenate concentrations. The particular PCDD isomers formed in the system, and their quantities, will depend upon the relative kinetics of alternative reaction routes.

The pyrolysis of PCDPEs follows two competitive reaction pathways, viz., dechlorination or ring closure to PCDFs. The cyclization of pre-dioxins is a bimolecular reaction caused by heating. The final concentration of PCDDs in a heated chlorobenzene system should ultimately depend upon the concentration of polychlorinated phenols which are formed at an intermediate stage.

Louw, *et al.*[96,96a] have interpreted the pyrolysis of chlorobenzenes as a radical chain reaction involving $\cdot C_6H_4Cl$,Cl· and H· as carriers. The authors discuss a radical reaction sequence which explains the observed product pattern, including the formation of PCBs.

The formation of polychlorinated naphthalenes (PCNs) in Buser's experiments (Buser[55]) can be explained by either invoking the formation of benzyne intermediates or the rearrangement of intermediates formed between an ortho-chloro phenyl radical with a chlorobenzene.

The overall effect of radical reactions on the product distribution in a pyrolysis reaction will be affected by both temperature and the availability of oxygen. This is well illustrated in the product distributions found after accidental fires involving PCBz/PCB filled electrical equipment (see earlier section).

It would be ironic if a process aimed at the detoxification of PCBs, which have been shown to be for the most part nontoxic in humans, should inadvertently create a hazard as the result of uncontrolled side reactions to give products with a much greater apparent toxicity.

Analytical chemistry at the time of the "Yusho" incident in 1968 was at a stage of development in which the determination of analytes at the ppm concentration level was considered to be the forefront of the field. Vos showed in 1970 that PCDFs could be determined in the presence of a matrix of PCBs but it was not until about 10 years later in 1977/78 that analytical methodology had advanced to the point that PCDFs and PCDDs could be determined, albeit laboriously, with anything like a "standardized" method. At the present time, EPA method 613 for the target compound 2,3,7,8-TCDD, uses high resolution gas chromatography coupled to low resolution mass spectrometry and is capable of detection limits of about 10 ppb. Developments in tandem mass spectrometry promise to provide a very rapid screening method for the presence of target compounds and is sensitive to PCDDs in the ppt range.

A part of the problem in the interpretation of PCB disposal methods lies in the analytical requirements of uncontrolled reactions. For example, even though methodology has been developed which is able to determine compounds which were not considered at the time of the Yusho incident, it is currently not possible to determine some of the polychlorinated polyaromatic pyrolysis products of PCBs because of a lack of standard compounds. The key question which remains unanswered today is "What are the concentrations of compounds of concern?"

(3) THE PCB REMEDIATION OF SOILS

The cost-effective remediation of PCB contaminated soils is discussed in terms of the availability of current options and an assessment of the best available technology. Major categories of in-place treatment techniques are reviewed as well as a wide range of technologies for the treatment of excavated soils.

Treatment methods are used to reduce the toxicity, mobility or volume of PCB contaminated material and fall into two distinct groups. One group requires the use of large energy sources and requires a significant amount of setup time. The other group of technologies tend to be more mobile, use much less energy and apply a decontamination process to excavated soil.

This section focuses on the use of electrochemical methods for the treatment of soil extracts and, in particular, on a new electrochemical technology which has been demonstrated to be significantly superior.

The Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) mandates the Environmental Protection Agency (EPA) to select remedial actions which "utilize permanent solutions and alternative treatment technologies or resource recovery technologies to the maximum extent practicable". Also, a preferred remedial action is one which "permanently and significantly reduces the volume, toxicity, or mobility of hazardous substances, pollutants, and contaminants as a principal element."

The Toxic Substances Control Act (TSCA) requires that material contaminated with PCBs at concentrations above 50 ppm should be disposed of in an incinerator approved for the purpose or by an alternate method that achieves a level of destruction which is equivalent to incineration.

Other applicable or relevant and appropriate requirements (ARARs) which can apply to a PCB contaminated site might involve the Resource Conservation and Recovery Act (RCRA) as well as the Clean Water Act (CWA) and Safe Drinking Water Act (SDWA).

RCRA applies to PCBs when liquid waste that is hazardous under RCRA contains PCBs at concentrations greater than 50 ppm or non-liquid hazardous waste contains total hazardous organic constituents at concentrations greater than 1000 ppm. PCBs are specifically addressed under RCRA in 40 CFR 268, which describes the prohibitions on land disposal of various hazardous wastes. The land disposal restrictions require that prior to placing PCB contaminated material on the land, it must be incinerated unless a treatability variance is obtained.

The National Contingency Plan (NCP) established a general presumption that a treatability variance is warranted for CERCLA soil and debris because when the Act was put into effect there were no standards for disposal. Consequently, alternate treatment levels(161) are justified according to the treatability guidance levels set out by the EPA. To qualify for a treatability variance for PCBs, residuals after the alternate treatment should contain 0.1 to 10 ppm PCBs for initial concentrations up to 100 ppm, and for initial concentrations above 100 ppm, treatment should achieve 90 to 99% reduction in concentration.

Soil remediation techniques applied to hazardous waste sites are assessed in terms of the soil processes which affect them and the type of system to be used. Once a technology has been selected it must be monitored for treatment effectiveness until it is finally proven that decontamination has taken place to an acceptable level of risk.

In general, contaminated soil systems must be considered as four phases: (1) aqueous; (2) vapour or gas; (3) organic and (4) solid. The distribution of contamination between the phases is linked to the relative affinity of the components of concern for each phase and are represented as 'Distribution Coefficients'. A distribution coefficient is calculated as the ratio of the concentration of a chemical in one phase to its concentration in a second phase. PCBs, for example, are hydrophobic and so are largely retained by the non-aqueous phase liquid, usually referred to as *NAPL,* and the organic matter in the soil. Hence the distribution coefficient for PCBs in oil is very large.

The mobility of a component of concern through a soil bed is measured in terms of a retardation factor. The retardation factor describes the relative velocity of a constituent compared to the rate of movement of water through the subsurface. A retardation factor which is greater than unity indicates that a component is moving more slowly than water through a system. The retardation factor for PCBs is very large.

The combination of large distribution coefficients for NAPL and soil organic matter and a large retardation factor means that PCBs tend to remain at the surface of a spill site and migrate either by erosion or by dust emission. As a consequence, spill sites or landfills tend to be long term sources of PCBs.

Remedial response levels for PCBs are as follows:

soil	<10 mg/Kg (EPA/NYSDEC soil removal criteria)
air	<1.67 μg/m3 (NY State Ambient Air Level)
ground water	<1.00 μg/L (NYSDOH advisory level)
surface water	$<7.9 \times 10^{-5}$ μg/L (Clean Water Act)

The surface water criterion of $<7.9 \times 10^{-5}$ μg/L i.e., less than about 1.0 parts per trillion, for a lifetime cancer risk of 10^{-6} is below the method detection limit for EPA method 608 and even below the MDL for the method.

The extent and type of information obtained from modeling the migration and persistence of contamination is used to select treatment approaches for further evaluation with respect to their feasibility and cost effectiveness. Remediation is accomplished using one or more of three types of systems: (1) in-situ; (2) prepared bed, and (3) in-tank reactor.

There are five major categories of in-place treatment techniques:

- Extraction,
- Immobilization,
- Degradation,
- Attenuation, and
- Reduction of voltalization.

Extraction

Extraction involves washing contaminants from the soil with a suitable solvent. The solvent must be non-polluting and must not alter the soil properties in an unacceptable way. An efficient method is needed to capture the elutriate and while the method has shown application in the mining industry it is not likely to be suitable for PCBs.

Immobilization

Immobilization methods are intended to reduce the rate of release of contaminants from the soil so that the concentrations along pathways of exposure are restricted to within acceptable limits. Activated carbon is a strong adsorbent for PCBs and can be readily incorporated into soil. The amount of carbon which may be required is strongly affected by the amount of other organic material is the soil. Since the PCBs are retained at the site, re-application of carbon may be necessary as the original material erodes.

Degradation

Oxidation, reduction and polymerization reactions have been carried out in attempts to transform soil contaminants into less toxic or less mobile products.

Oxidation Chemical oxidation of compounds of concern with reagents such as ozone or hydrogen peroxide can lead to oxidation products which are more toxic than the original contaminants. For example, the formation of hydroxylated products would tend to increase the water solubility of PCBs and may even lead to the generation of polychlorinated dibenzofurans (PCDFs) or polychlorinated dibenzodioxins (PCDDs).

Reduction Reducing agents such as iron powder or sodium borohydride have been used to degrade toxic organics. An electrochemical method which uses iron/copper pellets has been applied to the treatment of PCBs in a tank reactor but has not been demonstrated as an in-place treatment method.

Reductive dehalogenation of PCBs can be applied using organometallic reagents and sodium based reagents similar to those developed by the author. The reagents are typically very reactive with water and were originally developed for the PCB decontamination of oils. A two-step procedure has been used for soil remediation in which the soil is first extracted and then solvent exchanged with a water immiscible solvent. The relatively dry, water immiscible solution is then reacted with a water sensitive, PCB reactive reagent.

Problems associated with water reactivity can be circumvented by using sodium polyethylene glycol. The alkali metal (APEG) reagent is water miscible and has been used with some limited success for the treatment of both PCB contaminated oils as well as soils(148,162).

Polymerization In-situ polymerization reactions have been initiated to react polymerizable organics into polymers which are less toxic and less mobile than the monomer compounds. This type of reaction is not applicable to PCBs.

Bioremediation In-situ biological remediation is an attractive technology because it represents a permanent cleanup remedy which, although slow compared with an in-tank reaction, can be allowed to proceed in-situ for as long as necessary. The disadvantage of the method is usually associated with the need to maintain environmental factors such as moisture, temperature, pH and microbial nutrients. A possible result of PCB degradation by microbial treatment is the formation of PCDFs.

Photolysis Photodecomposition of PCBs at soil surfaces has been reported in the presence of triethylamine as a proton donor. Methanol has been used as a proton donor for the photodecomposition of PCDDs. Some difficulty has been observed at the very low concentrations of concern in distinguishing photolysis from evaporation. Photodegradation is limited to the surface of contaminated soil.

Treatment Techniques Applicable to Excavated Soils

In contrast with in-situ soils a relatively wide range of technologies is applicable to the treatment of excavated soils:

- incineration,
- biological treatment,
- matrix isolation
- vitrification,
- chemical dechlorination,
- electrocoagulation
- solvent extraction,
- physicochemical adsorption.

Incineration Incineration methods are required to achieve a destruction and removal efficiency of 99.9999%. This is described as the six 9's rule in 40 CFR 761.70.

There is a wide variety of equipment available for the incineration of PCBs which ranges from large fixed site equipment to more mobile units. Rotary kilns, for example, are usually large and expensive to operate but can fulfill the six 9's destruction efficiency requirement.

A kiln consists of a refractory lined cylinder with its axis slightly inclined to produce a tumbling action which mixes the wastes. Residence times of PCB contaminated material in the hot zone can be controlled to the extent necessary for complete combustion. Combustion temperatures range from 870°C to 1650°C.

The following high energy destruction methods are either being used on a large scale or are in the development stages.

- Circulating Bed Incineration
- High temperature fluid wall reactor
- Rotary kiln
- Plasma arc
- Penberthy glass-melting furnace
- Molten salt
- Wet oxidation
- Supercritical water oxidation

Biological Treatment Biological treatments for industrial waste systems have been in use for many years but it is only recently that new microbial strains have been developed to degrade PCBs by either enhancing conditions for existing microbes or the application of proprietary strains.

Fundamental biochemical degradation mechanisms are not well understood. Chemical pathways are therefore difficult to predict and the potential exists for the release of hazardous compounds into the environment through incomplete degradation or the formation of more toxic products.

The EPA considers biological processes which involve the excavation of contaminated material for treatment to be eligible for a treatability variance when the process can be shown to achieve a residual concentration of 2 ppm PCB or less. The treatment residuals can only be re-deposited on-site without long-term management controls when the treatment by-products do not present a threat to the environment.

Matrix Isolation Solidification is a term which is applied to the encapsulation of a waste to prevent its leaching into the environment. For example, PCB contaminated soils have been encapsulated into concrete without the formation of chemical bonding between the PCB and the encapsulant. On the other hand, solidification implies that some degree of chemical interaction is taking place to inhibit leaching. For example, pozzolanic chemistry has been used to chemically fix PCBs in a rock-like product which has good load bearing properties and which is hydrolytically stable.

Vitrification Vitrification utilizes an electrical process which uses about 4 MW of power to produce an in-situ pyrolyzed mass between large electrodes placed in the soil. Volatilized organics have to be captured and treated. The process results in complete pyrolysis of a localised area. The perimeter of the treated area needs to be defined by an appropriate analytical procedure to determine the need for on-going management controls.

Chemical Dechlorination Once PCBs have been extracted from the soil into a non-aqueous solvent the concentrated extract can be treated with organometallic reagents such as those invented by the author. Alternatively, alkali metal polyethylene glycol reagents, (APEG), can be used which are not directly reactive with water.

Both sodium and potassium PEG derivatives have been used with limited success. The reagent efficiency varies with the type of Aroclor and the amount of dilution the reagent receives from water contained in the soil. Recovery of excess reagent by water washing tends to reduce the cost-effectiveness of the process although PCB reduction to below 2 ppm PCB has been achieved.

Electrocoagulation Coagulation is the process by which particles in water are destabilized so that they can be removed by solid/liquid separation processes such as sedimentation and filtration.

Electrocoagulation has been applied to the removal of contaminated soil washings, waste oils and the enhanced dewatering of sludges. The applicability of the technique to soils which have been solvent extracted to remove PCB contamination is not yet known.

Solvent Extraction Three extraction techniques are used to clean soil: water washing, solvent extraction and air stripping. The first two methods are applicable to the remediation of soils contaminated with PCBs.

Water washing alone is not effective in the removal of hydrophobic contaminants such as PCBs. However, when surfactants are added to the water washes laboratory studies have shown an extraction efficiency of more than 90%.

Water washes used without added surfactant were found to reduce the PCB concentration in a sample of soil by only about 13%. The action of the surfactant is to reduce the adsorptive bonding strength of the PCBs on the soil particles with the result than 90% PCB removal was achieved.

Approximately 95% of the surface area of a soil is represented by fine particulates. The fines represent about 15% of silt and 8% of clay. Therefore, a sieving process alone is able to provide a particulate mass which contains a large fraction of the total PCB contamination. The EPA Mobile Soil Washing System (MSWS) was designed[163] around the particle size concentration effect.

Solvent extraction with organic solvents tends to be more effective than water washing with surfactant solutions. Two general techniques are applied: (1) immersion extraction and (2) leaching.

An immersion extractor is simply a tank containing stirred solvent into which the soil is suspended and mixed. When the extraction equilibrium has been established, the solids are separated either by settling or centrifugation and the solvent is passed down the treatment train to be decontaminated and recycled.

Leaching extraction is a process in which solvent is allowed to percolate through the soil contained in a screened bottom tank. The solvent desorbs and dissolves the PCB contamination and is collected at the bottom of the tank. A series of beds of soil are treated in a countercurrent fashion to increase the extraction efficiency.

Critical fluid extraction is a method in which liquified hydrocarbon gases such as propane and butane are used as solvents for separating organic contaminants from soils, sludges and sediments. Feed material is typically screened to remove particles which are greater than 1/8″ diameter and then mixed with water to provide a pumpable slurry. The technology has been applied on a mobile 60-tpd treatment scale. A 98% PCB reduction(164) has been achieved from soil initially contaminated with 350 ppm PCB.

A basic extractive sludge treatment uses aliphatic amines, typically triethylamine, (TEA) as the solvent to remove contaminants, including PCBs. TEA has the interesting property that below 15°C it can simultaneously solvate oils and water. Above 15°C, water becomes immiscible and separates from the oil and solvent. The process has been designed to remove water and organics at low temperatures, separate the water from the organic phase at higher temperatures, and recover the solvent in high yield through distillation. A 70-tpd unit has been demonstrated[165] on PCB contaminated soils.

Physico-chemical Adsorption Activated carbon is a strong adsorbent for PCBs because of its structure. Powdered carbon which has a very high surface area to mass ratio is used. The long-term physical and chemical stability of activated carbon added to soil systems is unknown but the material tends to erode with the soil and also become distributed in the environment as particulate emissions.

Carbon beds have been used to adsorb PCBs from water as well as oils. The system acts as a PCB concentrator which is usually then sent for incineration.

The following Table 56 shows a number of innovative technologies which are being developed for the remediation of contaminated soils.

The wide variety of technologies in the above table have been applied to the PCB decontamination of soils with varying degrees of success. The methods can be separated into two groups. One group requires the use of large energy sources and need a significant amount of setup time but are relatively less selective with regard to the contaminants to be treated. The second group tend to be more mobile, use much less energy and apply a second process to the soil extract to treat a specific group of contaminants.

Electrochemical Treatment Processes for Excavated Soils

This section will discuss electrochemical methods which have been used for the dechlorination of PCBs and focus on data obtained with a method which appears to have widespread applicability. Other electrochemical methods will be discussed briefly.

A Westinghouse patent deals with an electrochemical method for the treatment of PCBs in which hydrogen peroxide is generated by an alternating current field passed through a series of packed beds of conductors. The high and low conductivity beds are placed adjacent to one another and each has separately controlled pairs of electrodes. The low conductivity bed contains conductive particles such as carbon or non-conductive particles coated with oxides such as MnO_2 or PbO_2. The proportion of particles in the more conductive bed is about 30% oxide coated particles with the balance as conducting and/or absorptive particles. A series of beds was constructed in the effluent flow path and 29 volts/12 amps passed into each of the less conducting beds while 10 volts/15 amps was passed into each of the more conductive beds. The electrodes were cooled so that the process temperature could be minimized. The PCB concentration was reduced from 120 ppb to 1.5 ppb.

Controlled potential electrolysis has been investigated as a method for the dechlorination of PCBs. Radical anions of such compounds as anthracene, 9,10-diphenylanthracene and phthalonitrile in dimethylformamide solution are formed at a mercury electrode surface. The radical anions react with PCBs in solution to produce chloride in much the same way as the mechanism of action of organometallic

TABLE 56

Developer	Technology	Applicable waste
In development	Electrochemical	PCBs, PCDFs, PCDDs, volatile & semivolatile organics, PAHs, PCPs, petroleum by-products, metals, synthetic organics
American Toxic Disposal Inc.	Vapor extraction	Volatile & semivolatile organics, PCBs, petroleum by-products
CF Systems Corp./ENSCO	Solvent extraction with liquified hydrocarbon gases	PCBs, volatile & semivolatile organics, petroleum by-products
Ozonics Recycling Corp.	Soil washing, catalytic ozone oxidation	Semivolatiles, pesticides, PCBs, PCP, PCDDs
Harmon Environmental Services	Soil washing	Heavy organic compounds
Western Research Institute	Contained Recovery of oily wastes	Coal tar derivatives, petroleum by-products
Resources Conservation Co.	Solvent extraction with triethylamine	High molecular weight organics
Chemical Waste Management	Low temperature thermal desorption	Volatile & semivolatile organics, PCBs
International Waste Technologies	In situ solidification/stabilization	PCBs
MoTec Inc.	Liquid/solid contact digestion	Halogenated & non-halogenated organics, pesticides
USEPA	KPEG process	PCBs
J.M. Huber	Mobile incineration	PCBs, organics
Penberthy Electromelt	Infra red vitrification	PCBs, organics
Westinghouse	Plasma pyrolysis	PCBs, organics
Battelle Memorial Institute	Radiofrequency heating	PCBs, organics

reagents. Step-wise reduction is observed with biphenyl as an end product. The results of a research program to demonstrate the effectiveness of an electrochemical process are described in detail in this section. Data are presented for the dechlorination of monochloro-benzene, hexachlorobenzene and PCBs.

Chlorobenzenes and PCBs are very insoluble in water and the initial solutions of these compounds were so concentrated that they entered the process as emulsions. The system broke the emulsion in five to ten minutes and produced a 98% reduction of contamination in the same time. Reactions were essentially complete in 30 minutes. PCBs, as Aroclor 1260, were reduced from an emulsion of 3,300 ppm PCB down to 0.2 ppm PCB in 30 minutes at room temperature in a safe, totally enclosed system.

Table 57 illustrates some of the results obtained with the process.

A solution of monochlorobenzene (MCBz) in methanol was added to water containing a supporting electrolyte. The resulting solution contained 39,400 ppb MCBz.

A low D.C. current was passed through the mixture at an initial pH of 7.4 and an ambient temperature of 24°C. As the reaction progressed the pH decreased to a plateau value of about 3.0 after 15 minutes and the temperature increased to a stable value of 47°C in the same time. Samples were withdrawn at regular intervals to be analysed by GC/ECD to determine residual chlorobenzene and GC/MS to identify reaction products. The final concentration of MCBz was 0.54 ppb.

The initial concentration of benzene in solution was 32 ppb. After 15 minutes this had decreased to less than 1 ppb. No other chlorinated species were found in the reacted mixture.

A similar experiment was conducted with hexachlorobenzene (HCBz) at a starting concentration of 11,500 ppb. The mixture was at first an emulsion of HCBz in water. The emulsion had completely disappeared within 5 minutes and the pH decreased from an initial 7.4 to about 4.0 in the same time. The temperature of the reaction mixture was controlled by a circulating water jacket which maintained the temperature at about 50°C.

The first sample analysed was taken after 30 minutes reaction. The HCBz concentration had decreased by a factor of 1000 from 11,500 ppb to 11.3 ppb.

A comparison of the chromatograms obtained after different reaction times clearly shows the formation of intermediate chlorobenzene congeners.

After 30 minutes reaction time there is a dramatic change in the concentration of HCBz while the concentrations of penta-, and tetra- and tri-chlorobenzenes were greater than in the starting solution, albeit at only 4–5 ppb or less. The concentration of lower chlorinated congeners decreased below the analytical detection limit with increasing reaction time. The treated solution after 90 minutes contained 10.1 ppb HCBz and sub-ppb levels of the tetra- and lower chlorinated congeners.

A trace of unquantitated Aroclor 1260 is evident in the starting solution. The chromatogram showed all of the charac-

TABLE 57

Contaminant of Concern (ppm)	Initial concentration (ppm)	Final concentration (ppm)
PCBs (Aroclor 1260)	3,300	0.2
PCBs (Aroclor 1260)	254	4.6
PCBs (Aroclor 1260)	22.5	0.007
Hexachloro-benzene	11.5	0.01
Monochloro-benzene	39.4	0.0005
2,3,7,8-TCDF	0.5 (ppt)	<0.03 (ppt)
AOX	43	0.97

teristic peaks of Aroclor 1260. After 30 minutes reaction time the peaks attributable to the more highly chlorinated congeners are either missing entirely or are much reduced. Peaks attributable to the constituents of Aroclor 1242 are missing from the original chromatogram but some are present in the sample after 30 minutes reaction time. The fractional reduction of PCBs in the presence of a relatively large quantity of HCBz is much smaller than the fractional reduction of HCBz in the same time.

A similar experiment was performed with a starting emulsion of Aroclor 1260 in water at 22,500 ppb PCB. Approximately 320 mg of Aroclor 1260 was dissolved in 100 mL of warm methanol and added to 7.9 L of water. The electrochemical cell arrangement was set up in the solution with a supporting electrolyte.

After about five minutes the emulsion had become a clear solution.

After thirty minutes the PCB concentration decreased from 22,500 ppb PCB to 27.9 ppb PCB. The last sample was taken after 90 minutes and the PCB concentration was found to be 7.1 ppb PCB. The chromatogram of the final solution was significantly different from that of the starting material and resembled a mixture of Aroclor 1260 and Aroclor 1242. The penta-, hexa- and septachlorinated biphenyls appeared to have completely reacted while trace quantities of di-, tri- and tetrachlorinated isomers remained.

The reaction was complete at 90 minutes. A sample was analysed to determine any residual PCB content as well as the complete range of chlorobenzene isomers and benzene. Trace quantities of 1,2,3-trichlorobenzene, tetraCBz, pentaCBz were found at sub ppb levels while hexaCBz was 4.3 ppb. Trichloro- and lower chlorinated benzenes were below the method detection limit of 0.3 ppb.

Other work by the author has shown that the action of sodium reagents on PCBs causes the sequential removal of chlorine atoms from the biphenyl molecule until a polymerized polyphenyl sludge is obtained. The change in the relative peak heights shown in the chromatograms of samples taken at different times was clearly indicated. The later peaks contain $C_{12}H_3Cl_7$ isomers which appear to lose chlorine and add to the earlier peaks with six, five and fewer chlorine atoms. The PCBs which elute at shorter retention times probably contain five and four chlorine atoms. Both of these peaks are relatively larger than in the starting material which indicates that the more highly chlorinated PCB congeners react faster than the less chlorinated congeners. The kinetics of the process were not established in detail because the dechlorination process takes place too rapidly under the conditions used to allow precise correlations to be developed.

The reduction in concentration by three orders of magnitude in 30 minutes reaction time is very similar to the reduction observed in the HCBz experiment.

Chemical reaction from a relatively concentrated PCB solution of Aroclor 1260 has been found to be reduced from an initial value of 254 ppm PCB to 4.6 ppm PCB in 30 minutes.

At this point in time it is uncertain whether the final product is polyphenyl hydrocarbons or whether the reaction continues to break the phenyl-phenyl bond. There is no observable precipitation of polyphenyls or other organic residue. The turbidity of the initial solution caused by the emulsion is completely eliminated and points to the probable elimination of organics from the water. This is supported by the observation that total organic carbon (TOC) is reduced simultaneously with PCB dechlorination.

A further experiment was conducted with a 3,300 ppm Aroclor 1260 emulsion in water. The process reduced the concentration to 0.2 ppm in 30 minutes.

The following table summarises the data (Table 58).

The electrochemical process by its nature is easily controlled and safe. The system is closed and has no emissions, it operates at ambient temperature.

In any chemical dechlorination system in which there is a possibility of the formation of hydroxyl radicals or hydroxylated species there is a possibility that incomplete reaction might produce hydroxylated species which are more water soluble than the original PCB. While the toxicity of the hydroxylated product might be less than that of the original PCB it is much more likely to be dispersed into the environment because of its increased water solubility. The electrochemical process yields products which do not contain chlorine as evidenced by the chromatograms of samples taken from the process at different reaction times.

No mechanistic studies have so far been done to determine the formation of intermediate reaction products.

Once an hydroxylated product has been formed in a dechlorination reaction, there is a possibility that, given the right conditions for reaction, PCDFs or PCDDs might be formed. Such reactions are known to occur under pyrolysis conditions and also when hydroxyl radicals are produced during the course of a reaction such as ozone oxidation and the interaction of solutions with high energy electrons generated by a Van der Graff machine or Sr^{90} β^--particles. These reactions have been discussed in an earlier section.

The dimerization of chlorophenate precursors to yield dibenzofurans is a bimolecular reaction and therefore the products formed should be highly dependent upon the chlorophenate concentrations. The particular PCDD isomers formed in the system, and their quantities, will depend upon the relative kinetics of alternative reaction routes.

The data in Table 59 indicate that aqueous samples which contained PCDFs and PCDDs can be dechlorinated to below the method detection limit of 0.03 parts per trillion. Work to determine the efficiency and optimal treatment parameters of the method towards the elimination of PCDFs and PCDDs has not been specifically addressed. The sample analyses shown in the table derive from real samples which contained PCDF and PCDD contamination together with other contaminants. All of the contaminants were successfully removed. During this experiment the system was not intended to specifically reduce PCDF and PCDD compounds but rather to remove other components of concern. Since these classes of compounds are chemically more reactive than chlorobenzenes it is expected that PCBs/PCDFs/PCDDs would also react to the observed low levels found in the current research.

Analytical chemistry at the time of the ‘Yusho’ incident in 1968 was at a stage of development in which the determination of compounds of concern at the parts per million level was considered to be the state-of-the-art. Vos showed in 1970 that PCDFs could be determined in the presence of a matrix of PCBs but it was not until about 10 years later that analytical methodology had advanced to the point that PCDFs and PCDDs could be determined with a standard, quality assured protocol.

The analysis protocol used to determine the PCDF data in Table 59 is designed to precisely and accurately quantitate the 2,3,7,8-TCDF and 2,3,7,8-TCDD isomers which are of greatest health concern. The remaining congeners are classified according to the number of chlorine atoms contained in the isomeric group. The concentration of the isomeric group is reported as the sum of the concentrations of the unseparated isomers.

The technology might be designed into a closed loop system. For example, the excavated soil is washed, the solvent mixed with supporting electrolyte and fed into the electrochemical cell. Compounds of concern are destroyed and the wash water and solvent regenerated and recirculated for further use.

The cost-efficiency of the process can be optimized by reducing the concentration of contaminants of concern to an intermediate level in the closed-loop system. The incompletely reacted mixture can be re-used to extract the excavated soil for a second pass treatment. The solvent is re-loaded with contaminants and treated in a second pass. Finally, the mixture is completely reacted to a prescribed level which will allow the clean soil to be returned to the site. The solvent need not be treated to extremely low levels because it is used in the closed-loop system to re-extract subsequent batches of contaminated soil.

TABLE 58
PCB emulsions in water are reduced to low parts per trillion

	Time (min.)/PCB Concentration (ppm)			
Initial (ppm PCB)	10 min.	20 min.	30 min.	90 min.
22.5			0.028	0.007
254	13.2	8.7	4.6	
3,300	25.5	2.9	0.2	

TABLE 59

	Polychlorinated dioxins (ppt)		Polychlorinated dibenzofurans (ppt)	
Component	Untreated	Treated	Untreated	Treated
2,3,7,8-TCDF	N.D.	N.D.	0.5	<0.03
Tetra-chloro-	1.2	0.15	1.5	0.21
Penta-chloro-	0.95	0.48	1.5	0.12
Hexa-chloro-	1.1	0.17	1.5	0.15
Hepta-chloro-	1.7	0.37	2.8	0.32
Octa-chloro-	1.3	0.22	1.6	N.D.

N.D. = not detected.

The process is operated at ambient temperature. There are no emissions from the closed-loop system and only clean water, solvent and soil leave the process. The vessel is not pressurised and does not require a high capital cost to construct. The operation is conducted with a low power requirement of 25 or fewer amps and 8 volts which can be generated in the field. If a power outage occurred, the reaction would stop instantaneously. All of these factors make for a safe working environment.

The electrochemical process has been used on effluent samples from various industry groups. Amongst these is the effluent from the pulp and paper industry which is known to contain large quantities of chlorinated organic materials, referred to as AOX components, which may include PCBs, PCDFs and PCDDs. The particular sample treated contained originally a total of 43,000 ppb AOX. The electrochemically treated solution showed a final concentration of 970 ppb which was below the required AOX industrial discharge limits. The process could have been continued to further reduce the AOX if that had been required.

The electrochemical process, whether intended for the breaking of emulsions, the removal of dissolved or suspended organic compounds, the destruction of chlorinated synthetic organic compounds including PCBs, PCDFs and PCDDs, operates on all of the contaminants simultaneously. The application of the process is mobile and safe.

The electrochemical cell is not pressurised and does not require expensive materials of construction to avoid chemical corrosion. Operating parameters are simple and safe—no hazardous by-products are formed since all of the chlorinated organic compounds are completely dechlorinated. The system is self-contained and has no emissions—the discharge of clean soil and the reuse of water and solvent are key products of the operation.

The capital cost of construction of this alternative technology is much lower than incineration. The extent of treatment, expressed in terms of electrolytic conditions and time, can be controlled to achieve a required level of destruction of organochlorine compounds of concern.

APPENDIX

COD	Chemical oxygen Demand
DBPC	Di-tertiary Butyl-P-Cresol
DRE	Destruction and Removal Efficiency
ESR	Electron Spin Resonance
HRGC	High Resolution Gas Chromatography
HTFW	High Temperature Fluid Wall
NaNp	Sodium nephthalenide
NAPEG	Sodium polyethylene glycol
NMR	Nuclear Magnetic Resonance
NCI-MS	Negative Chemical Ionization-Mass Spectrometry
PCB	Poly Chlorinated Biphenyl
PCBZ	Poly Chloro Benzene
PCBP	Poly Chloro Bi Phenylene
PCCY	Poly Chlorinated Chrysene
PCDD	Poly Chloro Diberizo-p-Dioxin
PCDF	Poly Chlorinated Dibenzo Furan
PCDPE	Poly Chlorinated Di Phenyl Ether
PCN	Poly Chlorinated Naphthalene
PCPY	Poly Chlorinated Pyrene
PCQ	Poly Chlorinated Quaterphenyl
PIC	Product of Incomplete Combustion
POHC	Principal Organic Hazardous Constituent
RCRA	Resource Conservation and Recovery Act
TSCA	Toxic Substances Control Act

REFERENCES

1. Hutzinger, O., S. Safe and V. Zitko: "The Chemistry of PCBs", The Chemical Rubber Co. Press, Cleveland, 1974.
2. Bowes G.W., M.J. Mulvihill, B.R.T. Simoneit, A.L. Burlingame, R.W. Risebrough; "Identification of Chlorinated Dibenzofurans in American Polychlorinated Biphenyls", *Nature* **256**, 305 (1975).
3. Nagayama, J., Y. Masuda, M. Kuratsune; "Chlorinated Dibenzohimns in Kaneclors and Rice Oils Used by Patients with Yusho", *Fukuoka Acta Med.* **66**, 593 (1975).
4. Kuratsune, M., T. Yoshimura, J. Matsuzaka and A. Yamaguchi; "Epidemiologic study on Yusho, a poisoning caused by ingestion of rice oil contaminated with a commercial brand of polychlorinated biphenyls"; *Environ. Health Perspect.* **1**, 119 (1972).
5. Masuda, Y. and H. Yoshimura; "PCBs and PCDFs in patients with Yusho and Their Toxicological Significance"; *Am. J. Ind. Yed.* **5**, 31 (1984).
6. Kunita, N., T. Kasimoto, H. Miyata, S. Fukushima, S; Hori and N. Obana; Causal agents of Yusho; *Am. J. Ind. Med.* **5**, 45 (1984).
7. Chen P.H.-S., M.-L. Luo, C.-K. Wong, C.-J. Chen; PCBS, PCDFs and PCQs in the toxic rice-bran oil and PCBs in the blood of patients with PCB poisoning in Taiwan; *Am. J. Ind. Med.* **5**, 133 (1984).
8. Hsu, S.-T., C.-I. Ma, S.K.-H. Hsu, S.-S. Wu, N.H.-M. Hsu and C-C. Yeh; Discovery and epidemiology of PCB poisoning in Taiwan; *Am. J. Ind. Med.* **5**, 71 (1984).
9. Bertazzi, P., C. Zocchetti, S. Guercilena, M. Foglia, A. Pesatori, and Riboldi, L. 1981 Int Symposium on Prevention of Occupational Cancer, Helsinki, Finland April 21–24, 1981. Mortality study of male and female workers exposed to PCBS.
10. Humphrey, H. 1983. D'ltri, F.M., and M.A. Kamrin, eds., PCBs: Human and Environmental Hazards, Butterworth Publishers, Boston.
11. Unger, M., J. Olsen, and J. Clausen, 1982. *Environ. Res.* **29**, 371: "Organochlorine compounds in the adipose tissue of deceased persons with and without cancer: a statistical survey of some potential confounders."
12. Vos, J.G., J.H. Koeman, H.L. van der Mass, M.C. Ten Noever de Braun, and R.H. de Vos 1970. *Fd. Cosmet. Toxicol.* **8**, 625: Identification and toxicological evaluation of PCDDs and PCNs in two commercial PCBS.
13. Allen, J., L. Abrahamson and D. Norback; Biological effects of polychlorinated biphenyls and terphenyls on the subhuman primate; *Environ. Res.* **6**, 344 (1973).
14. Goldstein, J.; Structure–activity relationships for the biochemical effects and the relationship to toxicity. In R.D. Kimbrough, ed., Halogenated Biphenyls, Terphenyls, Naphthalenes, Dibenzodioxins and Related Products. Elsevier/North Holland Biomedical Press, New York, p. 151 (1980).
15. Kunita N., T. Kasimoto, H. Miyata, S. Fukushima, S. Hori and N. Obana; Causal agents of Yusho; *Am. J. Ind. Med.* **5**, 45 (1984).
16. Kimbrough, R.; In Toxicology of Halogenated Hydrocarbons: Health and Ecological Effects; Khan, M.A.Q. and R.H. Stanton, eds.; Pergamon Press, New York (1981).

17. Poland, A. and E. Glover; Chlorinated biphenyl induction of hydrocarbon hydroxylase activity; A study of the structure-activity relationship. *Mol. Pharmacol.* **13**, 924 (1977).
18. Smith, R.M. *et al. 1982. Chemosphere,* **11**, 715; TCDF and TCDD in a soot sample from a transformer explosion in Binghamton, New York.
19. Rappe, C., S. Marklund, P. Berggvist, and M. Hansson, 1982. *Chem. Scripta.*, **20**, 56; PCDDS, PCDFs and other polynuclear aromatics formed during PCB fires.
20. Williams, C., C. Prescott, P. Stewart, and G. Choudhary, 1984. In Keith, L.H., ed. Chlorinated Dioxins and Dibenzofurans in the Total Environment. Butterworth Publishers, MA.; Formation of PCDFs and other potentially toxic chlorinated pyrolysis products in PCB fires.
21. Rappe C., S. Marklund, P.-A. Bergqvist and M. Hansson; "Polychlorinated Dibenzo-p-Dioxins, Dibenzofurans and Other Polynuclear Aromatics Formed During Incineration and Polychlorinated Biphenyl Fires"; in "Chlorinated Dioxins and Dibenzofurans in the Total Environment"; ed. G. Choudhary, L.H. Keith and C. Rappe; Butterworth Publishers 1983.
22. Albro, P.W. and C.E. Parker; Comparison of the compositions of Arocior 1242 and Aroclor 1016; *J. Chromatog.* **161** (1979).
23. Erickson, M.D., C.J. Cole, J.D. Flora, P.G. Gorman, C.L. Haile, G.D. Hinshaw, F.C. Hopkins, S.E. Swanson and D.T. Heggem; PCDF formation from PCBs under fire conditions; *Chemosphere,* **14**, 855 (1985).
24. Lao, R.C., R.S. Thomas, J.L. Monkman, 1976. *Dynamic Mass Spectrom.* **4**, 107: Application of computerized GC–MS to the analysis of PCBs.
25. Stalling, D.L., J.N. Huckings, J.D. Petty, J.J. Johnson, J.O. Sanders, 1979. *Ann. N. Y. Acad. Sci.* **320**, 48: An expanded approach to the study and measurement of PCBs and selected planar halogated aromatic environmental pollutants.
26. Sawyer, L.D.; *J. Assoc. Off. Anal. Chem.,* **61**, 272 (1978).
27. Kuratsune, M., Y. Morikawa, T. Hirohata, M. Nishizumi, S. Kohchi, 1. Yoshimura, J. Matsuzaka, A. Yamaguchi, N. Saruta, N. Ishinishi, E. Kunitake, 0. Shimono, K. Takigawa, K. Oki, M. Sonoda, T. Ueda, and M. Ogata; *Fukuoka Igaku Zasshi* **60**, 513 (1969); "An epidemiological study on "Yusho" or chlorobiphenyls poisoning."
28. Kuratsune, M., T. Yoshimura, J. Matsuzaka and A. Yamaguchi; "Epidemiologic study on Yusho, a poisoning caused by ingestion of rice oil contaminated with a commercial brand of polychlorinated biphenyls"; *Environ. Health Perspect.* **1**, 119 (1972).
29. Kuratsune, M.; *Environ. Health Perspect.* **1**, 129 (1972); Results of lab examinations of patients with Yusho symptoms.
30. Kuratsune, M., Y. Masuda and J. Nagayama; Proceedings of the National Conference on Polychlorinated Biphenyls, USEPA, Office of Toxic Substances, Washington DC, November 1975 EPA 560/6–75–004; pp. 14–29 (1976).
31. Kuratsune, M.; "Epidemiologic studies on Yusho"; in Higuchi, K., ed., "PCB Poisoning and Pollution", Kodansha Ltd./Academic Press, Tokyo/New York pp. 9 (1976)
32. Kuratsune, M.; "Halogenated Biphenyls, Terphenyls, Napthalenes, Dibenzodioxins and Related Products"; Elsevier/North Holland Biomedical Press, New York pp. 287–302, (1980).
33. Kuratsune, M. and R. Shapiro; "PCB Poisoning in Japan and Taiwan"; *Am. J. Ind. Med.* **5**, 1 (1984).
34. Cordle, F., P. Comeliussen, C. Jelinek, B. Hackley, R. Lehman, J. McLaughlin, R. Rhoden, and R. Shapiro; "Human Exposure to polychlorinated biphenyls and polybrominated biphenyls."; *Environ. Health Perspect.* **24**, 157 (1978).
35. Urabe, H., H. Koda and M. Asahi; "Present State of Yusho Patients"; *Ann. NY Acad. Sci.* **320**, 273 (1979).
36. Urabe, H. and M. Asahi; "Past and Current Dermatological Status of Yusho Patients"; *Am. J. Ind. Med.* **5**, 5 (1984).
37. Kikuchi, M.: Autopsy of Patients with Yusho"; *Am. J. Ind. Med.* **5**, 19 (1984).
38. Yoshimura T.; "Epidemiological analysis of Yusho patients with special reference to sex, age, clinical grades and oil consumption"; *Fukuoka Acta Med.* **62**, 104 (1971).
39. Nagayama J., M. Kuratsune and Y. Masuda; Determination of chlorinated dibenzofurans in Kanechlors and "Yusho oil"; *Bull. Environ. Contam. Toxicol.* **15**, 9 (1976).
40. Miyata, H., T. Kashimoto and N. Kunita; "Detection and determination of polychlorodibenzofurans in normal human tissues and Kanemi rice oils caused Kanemi Yusho"; *J. Food Hyg. Soc.* Japan; **18**, 260 (1977).
41. Miyata H. and T. Kashimoto; PCDFs in commercial PCBs; *Shokuhin Eiseigaku Zasshi* **17**, 434 (1976).
42. Nagayama, J., M. Kuratsune and Y. Masuda; Formation of PCDFs by heating PCBS. In Japanese, summary in English; *Fukuoka Igaku Zasshi* **72**, 136 (1981).
43. Chittim, B.G., B.S. Clegg, S.H. Safe and O. Hutzinger; PCDFs and PCDDs: Detection and quantitation in electrical equipment and their formation during the incineration of PCBs; Environment Canada report 05578–00067 (1979).
44. Cull M.R. and A.J. Dobbs; "Analysis of Polychlorinated Biphenyls from Used Transformers for PCDFS"; *Chemosphere,* **13**, 1085 (1984).
45. Brown J.F., J.C. Carnahan, S.B. Dorn, J.T. Groves, W.V. Ligon, R.J. May, R.E. Wagner and S.B. Hamilton; "Levels of Bioactive PCDF Congeners in PCB Dielectric Fluids from Capacitors and Transformers"; *Chemosphere,* **17**, 1697 (1988).
46. Rouse T.O.; "PCDF and PCDD in Utility Transformers and Capacitors: Volume 1: Formation of PCDF and PCDD in Askarels and Contaminated Mineral Oil Equipment. EPRI Report EL EA-4858, volume 1, Electric Power Research Institute, Palo Alto, CA, October 1986.
47. C. Rappe, H. Buser, H. Kuroki and Y. Masuda; Identification of PCDFs, retained by patients with Yusho; *Chemosphere,* **8**, 259 (1979).
48. Kashimoto, T., H. Miyata and N. Kunita; The presence of PCQs in the tissues of Yusho victims; *Food Cosmet. Toxicol.* **19**, 335 (1981).
49. Chen P.H.-S., M.-L. Luo, C.-K. Wong, C.-J. Chen; PCBS, PCDFs and PCQs in the toxic rice-bran oil and PCBs in the blood of patients with PCB poisoning in Taiwan; *Am. J. Ind. Med.* **5**, 133 (1984).
50. Masuda Y., H. Kuroki, T. Yamaryo, K. Haraguchi, M. Kuratsune and S. Hsu; Comparison of causal agents in Taiwan and Fukuoka PCB poisoning; *Chemosphere,* **11**, 199 (1982).
51. Nagayama J., M. Kuratsune and Y. Masuda; Determination of chlorinated dibenzofurans in Kanechlors and "Yusho oil"; Bull. *Environ. Contam. Toxicol.* **15**, 9 (1976).
52. Masuda, Y. and H. Yoshimura; PCBs and PCDFs in patients with Yusho and their toxicological significance; *Am. J. Ind. Med.* **5**, 31 (1984).
53. Chen, P.H.-S., M.-L. Luo, C.-K. Wong, C.-J. Chen; PCBs, PCDFs and PCQs in the toxic rice-bran oil and PCBs in the blood of patients with PCB poisoning in Taiwan; *Am. J. Ind. Med.* **5**, 133 (1984).
54. Rappe C., H. Buser, H. Kuroki and Y. Masuda; Identification of PCDFs retained by patients with Yusho; *Chemosphere,* **8**, 259 (1979).
55. Buser, H.R. and C. Rappe; Formation of PCDFs from the pyrolysis of individual PCB isomers; *Chemosphere;* **8**, 157 (1979).
56. Mescher, J.A., R.A. Carnes, D.S. Duvall, and W.A. Rubey, 1978. USEPA Research Brief; Thermal degradation of PCBs.
57. Choi, S.U., and J.E. Willard, 1962. *J.A.C.S.,* **66**, 1041; The radiolysis of chloro-, bromo- and iodobenzenes with Co-60 gamma-rays
58. Arai S., M. Matsui, J. Moriguchi and H. Imamura; *Reports Inst. Phys. Chem. Res.* **48**, 185 (1972).
59. Sawai T., T. Shimokawa and Y. Shinozaki; "The Radiolytic Chain Dechlorination of Polychlorinated Biphenyls in Alkaline 2-Propanol Solutions"; *Bull. Chem. Soc. Japan,* **47**, 1889 (1974).
60. Webber, I. 1983. CEA report 77–70 phase 11: The radiolytic degradation of PCBs in electrical insulating oils.
61. Pilgrim, D. and I. Webber, 1980. Simon Fraser University work term report: PCB dechlorination treatments for insulating oils.
62. Webber, I. and D.E. Wilson, 1981. University of Victoria work term report: Inhibition of PCB dechlorination by 2,6-ditertiarybutyl-p-cresol.
63. Brown, C.E. and I. Webber, 1982. Unpublished results and also, Brown, C.E., Bezoari, M.D. and Kovacic, P. 1982. *J. Polymer Science,* **20**, 1697: Characterization of aromatic polymers (benzenoid and heterocyclic) by cross-polarization, magic-angle C-13 NMR spectroscopy.
64. Rubey, W.A., J.L. Graham, D.L. Hall, B. Dellinger and R. Moumighan: The laboratory approach to studying the thermal destruction behavior of organic substances; 11th Annual Hazardous Waste Research Symposium (1985).

65. Safe S., T. Sawyer, G. Mason, S. Bandiera, B. Keys, M. Romkes, J. Piskorska-Pliszczynska, B. Zmudszka and L. Safe; *Chemosphere,* **14**, 675 (1985).
66. Bradlaw, J. and J. Casterline, Jr.; Induction of enzyme activity in cell culture: A rapid screen for detection of planar polychlorinated organic compounds; *J. Assoc. Off. AnaL Chem.* **62**, 904 (1979).
67. Nagayama J., C. Kiyohara, M. Nishizumi, A. Horie, S. Handa, S. Yamaguchi, and Y. Masuda; "Comparative Toxicologic Study with 2,3,7,8-Tetrachlorodibenzo-p-dioxin in Inbred Strains of *Mice";* *Chemosphere,* **14**, 949 (1985).
68. Nagayama J., C. Kiyohara, M. Nishizumi, A. Horie, S. Handa, S. Yamaguchi and Y. Masuda; "Comparative Toxicity of 2,3,4,7,8-Pentachlorodibenzofuran in Inbred Strains of Mice"; **14**, 953 09851.
69. Nagayama J., C. Kiyohara, Y. Nakamura, M. Ikeda, M. Asahi and T. Hirohata; "Aryl Hydrocarbon Hydroxylase in Yusho Patients"; *Chemosphere,* **16**, 2073 (1987).
70. Furst P., H.-A. Meemken, Chr. Kruger and W. Groebel; "Polychlorinated Dibenzodioxins and Dibenzofurans in Human Milk Samples from Western Germany"; *Chemosphere,* **16**, 1983 (1987).
71. Rappe C. WHO Consultation on Organohalogen Compounds in Human Milk and Related Hazards, Bilthoven, 1985.
72. Takizawa Y. and H. Muto; "PCDDs and PCDFs carried to the human body from the diet"; *Chemosphere,* **16**, 1971 (1987).
73. Eadon G., K. Aldous, G. Frankel, J. Gierthy, D. Hilker, L. Kaminsky, P. O'Keefe, R. Smith, J. Gierthy, J. Silkworth and R. Smith; "Comparison of Chemical and Biological Data on Soot Samples from the Binghamton State Office Building"; New York State Dept. of Health Report (1982).
74. U.S. Environmental Protection Agency; "Interim Procedures for Estimating Risks Associated with Exposures to Mixtures of Chlorinated dibenzo-p-dioxins and dibenzofurans"; EPA/625/3–87/012 (1987).
75. Olson L.R., J.S. Bellin and D.G. Bames; "Reexamination of Data Used for Establishing Toxicity Equivalence Factors (TEFS) for Chlorinated Dibenzo-Dioxins and Dibenzofurans"; *Chemosphere,* **18**, 371 (1989).
76. Schecter A., J.J. Ryan and J.D. Constable; "PCDD and PCDF levels in human breast milk from Vietnam compared with cow's milk and human breast milk from the North American Continent"; *Chemosphere,* **16**, 2003 (1987).
77. Ahleborg U.G., F. Waern and H. Hakansson; "Interactive Effects of PCDDs Occurring in Human Mother's Milk; *Chemosphere,* **16**, 1701 (1987).
78. Sielken R. L; "Statistical Evaluations Reflecting the Skewness in the Distribution of TCDD Levels in Human Adipose Tissue"; *Chemosphere,* **16**, 2135 (1987).
79. Schecter A. and T.A. Gasiewicz; "Health Hazard Assessment of Chlorinated Dioxins and Dibenzofurans Contained in Human Milk"; *Chemosphere,* **16**, 2147 (1987).
80. Ogaki J., K. Takayama, H. Miyata and T. Kashimoto; "Levels of PCDDs and PCDFs in Human Tissues and Various Foodstuffs in Japan"; *Chemosphere,* **16**, 2047 (1987).
81. USEPA; "Health Assessment Document for Polychlorinated Dibenzo-p-Dioxins (EPA/600/8–84/014F), App. B-20; 1985.
82. Schecter A., T. Tiernan, F. Schaffner, M. Taylor, G. Gitlitz, G.F. VanNess, J.H. Garrett, and D.J. Wagel; *Environmental Health Perspectives,* **60**, 241 (1985).
83. Schecter A., J.J. Ryan, R. Lizotte, W.F. Sun and L. Miller; *Chemospheres,* **14**, 929 (1985).
84. Gross M.L., J.O. Lay, P.A. Lyon, D. Lippstreu, N. Kangas, and R.L. Harless, S.E. Taylor and A.E. Dupuy; *Environmental Research,* **33**, 261 (1984).
85. Ryan J.J., R. Lizotte and B.P.-Y. Lau; "Chlorinated Dibenzo-p-Dioxins and Chlorinated Dibenzofurans in Canadian Human Adipose Tissue; *Chemosphere,* **14**, 697 (1985).
86. Schecter A.J., J.J. Ryan, M. Gross, N.C.A. Weerasingbe, and J. Constable, in "Chlorinated Dioxins and Dibenzofurans in Perspective"; C. Rappe, G. Choudhary and L. Keith, Eds., Lewis Publishers, Chelsea, Michigan, pp. 35–50, (1986).
87. Kominsky J.R. and C.D. Kwoka; Background Concentrations of PCDFs and PCDDs in Office Buildings in Boston, MA"; *Chemosphere,* **18**, 599 (1989).
88. EPRI Report #325,1976: "Distribution Transformer Tank Pressure Study".
89. Webber I.: "Assessment of Fire Hazard Safety for Askarel Substitute Fluids: RTE Research & Development:; in collaboration with Underwriters Laboratories.
90. Hemstreet R.A.; "Flammability Tests of Askarel Replacement Fluids"; FMRC report, Serial No. IA7R3.RC prepared for the National Electrical Manufacturer's Association (1978).
91. Ballard J.G. and S.H. Hawthorne: "Solvent Decontamination of PCB Electrical Equipment", Canadian Electrical Association Engbm&g and Operating Division: Part 1, 81-A-66 (1981).
92. Barton T.G. and J.A.G. Mordy; "Mobile Plasma Pyrolysis"; EPRI 1983 PCB Seminar, Atlanta, Georgia, December, 1983.
93. Barclay K.M.; "Destruction of Polychlorinated Biphenyls by the Molten Salt Combustion Process"; Rockwell International Technical Report N00ITR000023, 1980.
94. Palauschek N. and B. Scholz; "Destruction of PCDDs and PCDFs in Contaminated Water Samples Using Ozone"; *Chemosphere,* **16**, 1857 (1987).
95. Jackson A.W.; "Chemical Destruction of Concentrated PCB Containing Solutions"; Ontario Hydro Research Division Report, 1979.
96. Louw R., J.W. Rothuizen and R.C.C. Wegman; "Vapour Phase Chemistry of Arenes. Part 11. Thermolysis of Chlorobenzene and Reactions with Aryl Radicals and Chlorine and Hydrogen Atoms at 500°C"; *J. Chem. Soc. Perkin Trans.,* **2**, 1635 (1973).

96a. Ahonkhai S.I., R. Louw and J.G.P. Born; "Thermal HydroDechlorination of Hexachlorobenzene"; *Chemosphere,* **17**, 1693 (1988).

97. Acherman D.B., L.L. Scinto, P.S. Bakshi, R.G. Delumyea, R.J. Johnson, G. Richard, A.M. Takata and E.M. Sworzyn; "Destruction of PCBs by Thermal and Non-Thermal Methods"; Noyes Data Corp., Park Ridge, New Jersey; Ch. 3, p. 7 (1983).
98. Chu N.S. and S.C. Vick; "Nickel Catalyzed Reduction of Polychlorinated Biphenyls"; Montech '86 Conference on PCBs and Replacement Fluids, Montreal, Quebec, Canada p. 167.
99. Okamoto T. and S. Oka; Bull. *Chem. Soc. Japan,* **54**, 1265 (1981).
100. Mosby W.L.; Chemistry and Industry, 1349 (1959).
101. Imai H., T. Nishiguchi, M. Tanaka and K. Fukuzumi; *J. Org. Chem.,* **42**, (1977).
102. Corteze N.A. and R.F. Heck; *J. Org. Chem.,* **42**, 3491 (1977).
103. Lubowitz H.R., R.W. Telles, B.M. Eliash and S.L. Unger; "Contaminant Fixation: Practice and Theory"; EPA Tenth Annual Research Symposium on the Land Disposal of Hazardous Waste, Ft. Mitchell, KY, April 1984.
104. desRosiers P.E. and J.H. Skinner; "Advances in Dioxin Risk Management Control Technologies"; *Chemosphere,* **18**, 41 (1989).
105. Ro B. and M.A. Thompson; "Equilibrium Study of PCBs Between Transformer Oil and Transformer Solid Materials"; EPRI Project No. TPS81–788, 1981.
106. Webber, I., 1980: "A Study of the Factors Affecting the Reclamation of Degraded Electrical Insulating Oil"; Canadian Electrical Association Report 77–70, Phase 1 (1980).
107. Smith J.G. and G.L. Bubbar; "Destruction of PCB's"; US patent No. 4,326,090 April 20, 1982.
108. Holy N.L.; "Reaction of the Radical Anions and Dianions of Aromatic Hydrocarbons"; *Chemical Reviews,* **74**, 243 (1974).
109. Ingold K.U.; "Inhibition of Oil Oxidation by 2,6-Di-tertiarybutyl-4-Substituted Phenols"; *J. Phys. Chem.* **64**, 1635 (1960).
110. Iaconianni F.J., A.J. Saggiomo and S.W. Osborn; "PCB Removal from Transformer Oil"; EPRI PCB Seminar, Dallas. TX, Proceedings September 1982.
111. Plummer J.R.; *Residue Review,* **33**, 47 (1970).
112. Safe S., N.J. Bunce, B. Chittim, O. Hutzinger and L.O. Ruzo; in "Identification and Analysis of Organic Pollutants in Water"; 35, L.H. Keith (Ed), Ann Arbor Science Publishers, 1976.
113. Epling G.A., Q. Qiu and A. Kumar; "Hydride-Enhanced Photoreaction of Chlorodibenzodioxins"; *Chemosphere,* **18**, 329 (1989).
114. Leitis E., J.D. Zeff, M.M. Smith and D.G. Cosby; "The Chemistry of Ozone and Ozone/UV Light for Water Reuse"; Proceedings-Vol. 2, Water Reuse Symposium **11**, p. 1121, August 1981.
115. Xu Y.M., P.E. Menassa and C.H. Langford; "Photodecomposition of Several Chloroaromatics Using a Commercial Prototype Reactor"; *Chemosphere,* **17**, 1971 (1988).

116. Mathews R.W.; *J. Chem. Soc., Faraday Trans.* **1**, 80, 457 (1984).
117. H. Sugimoto, S. Matsumoto and D.T.S. Sawyer; *J. Am. Chem. Soc,* **109**, 1081 (1987).
118. Niessen R., D. Lenoir and P. Boule; "Phototransformation of Phenol Induced by Excitation of Nitrate Ions"; *Chemosphere,* **17**, 1977 (1988).
119. Merrill E.W., D.R. Mabry, R.B. Shulz, W.D. Coleman, J.G. Trump and K.A. Wright; "Destruction of Trace Toxic Compounds in Water and Sludge by Ionizing Radiation"; paper presented at E.I.Ch.E. National Meeting, November 1977.
120. Choi P.S.K., H. Nack and J.E. Flinn; Distribution of PCBs in an Aerated Biological Oxidation Wastewater Treatment System"; *Bull. Environ. Contam. Toxicol.,* **11**, 12 (1974).
121. Wong P.T.S. and K.L.E. Kaiser: "Bacterial Degradation of PCBs II. Rate Studies"; *Bull. Environ. Contam. Toxicol.,* **13**, 249 (1975).
122. Kaiser K.L.E. and P.T.S. Wong; "Bacterial Degradation of PCBs I. Identification of Some Metabolic Products from Arocior 1242."; *Bull. Environ. Contam. Toxicol.,* **11**, 291 (1974).
123. Ahmed M. and D.D. Focht; *Bull. Environ. Contam. Toxicol.* **10**, 70 (1973). A pure culture of Achromobacter, isolated from sewage, was able to degrade various PCBs.
124. Ahmed M. and D.D. Focht; *Can. J. Microbiol.* **19**, 47 (1973). Two species of Achromobacter were isolated from sewage. Washed cell suspensions degraded PCBs without dechlorination.
125. Furukawa K. and F. Matsumura; *J. Agric. Food Chem.* **24**, 251 (1976). An Alcaligenes species isolated from lake sediments degraded highly chlorinated PCBs less rapidly than those with less Cl. PCBs with no Cl on one ring were degraded faster than those in which Cl occurred on both rings.
126. Baxter R.A., P.E. Gilbert, R.A. Lidgett, J.H. Mainprize and H.A. Vodden; *Sci. Total Environ.,* **4**, 53 (1975). PCBs with up to six Cl atoms were degraded by microorganisms. Less chlorinated PCBs were degraded before those with a higher Cl content. Some PCBs were more readily degraded as mixtures than when pure.
127. Neu H.J. and K. Ballschmiter; *Chemosphere,* **6**, 419 (1977). A Study of the Metabolism of 4- Chlorobiphenyt by Soil Bacteria to Form Phenolic Derivatives.
128. Ballschmiter K., C. Unglert and H.J. Neu; *Chemosphere,* **6**, 51 (1977). A mixed culture of cell bacteria grown on benzene as sole C-source was incubated with PCBs. Chlorobenzoic acid derivatives accumulated.
129. Sayler G.S., M. Schon and R.R. Colwell; *Microb. Ecol.,* **3**, 241 (1977). A Pseudomonas species, isolated from Chesapeake Bay, degraded pure isomers and mixtures of PCBS. Up to 84% of Aroclor 1254 was degraded in 60 days.
130. Nissen T.V.; *Tiddskr. Planteavl.,* **77**, 533 (1973) Aroclor 1254 was not degraded after nine weeks in moist soil at 25°C.
131. Vockel D. and F. Korte; *Chemosphere,* **3**, 177 (1974). 2,2′Dichlorobiphenyl was not degraded by any microorganisms or soil samples tested under several conditions of incubation.
132. Iwata Y., W.E. Westlake and F.A. Gunther; "Study of the Persistence of Aroclor 1254 in California Soils"; *Bull. Environ. Contam. Toxicol.,* **9**, 204 (1973).
133. Bailey R.L., W.L. Rhinehart, S.J. Gonsior, T.L. Batchelder, C.G. Mendoza and W.B. Neely; "Mass Balance of "C-Labelled PCBs Biodegraded in River Water"; paper presented at the Society of Env. Toxicol. And Chem. symposium, Arlington, VA, November 1981.
134. Neely W.B.; "Mathematical Model to Predict the Lifetime of PCB Congeners in the Environment"; *Chemtech.*, **11**, 249 (1981).
135. Camoni I. *et al.,*; "Laboratory Investigation for the Microbiological Degradation of 2,3,7,8-TCDD in Soil by Addition of Organic Compost" in "Chlorinated Dioxins and Related Compounds, Impact on the Environment" O. Hutzinger *et al.* Eds. Pergamon Press, 1982 pp. 95–103.
136. Huetter R. and M. Philippi; "Studies on Microbial Metabolism of TCDD under Laboratory Conditions" in "Chlorinated Dioxins and Related Compounds, Impact on the Environment" O. Hutzinger *et al.* Eds. Pergamon Press pp. 87–93, 1982.
137. Young A.L., W.J. Caimey and R. Channell; "Movement and Decontamination Studies of TCDD in Storage Sites Massively Contaminated with Phenoxy Herbicides"; paper presented at the 3rd International Symposium on Chlorinated Dioxins and Related Compounds, Salzburg, Austria, October 1982.
138. Safe S., L. Robertson, T. Sawyer, S. Bandiera and L. Safe; "PCBS: Structure-Activity Relationships"; paper presented at the Advances in Exposure, Health and Environmental Effects Studies of PCBs Symposium, Bethesda, MD, May 1982.
139. Pellizzari E.D., M.A. Moseley, S.D. Cooper, J.V. Harry, B. Demian and M.D. Mullin; *ibid.*
140. Schwartz T.R., R.D. Campbell, D.L. Stalling, R.L. Little, J.D. Petty, J.W. Hogan and E.M. Kaiser; "Laboratory Data Base for Isomer-Specific Determination of Polychlorinated Biphenyls"; *Anal. Chem.* **56**, 1303 (1984).
141. Dunn W.J. III, D.L. Stalling, T.R. Schwartz, J.W. Hogan, J.D. Petty, E. Johansson and S. Wold; "Pattern Recognition for Classification and Determination of Polychlorinated Biphenyls in Environmental Samples"; *Anal. Chem.,* **56**, 1308 (1984).
142. desRosiers P.E.; "Evaluation of Technology for Wastes and Soils Contaminated with Dioxins, Furans and Related Substances"; J. *Hazardous Materials,* **14**, 119 (1987).
143. Connors T.F. and J.F. Rusling; "Ultrasonically-Assisted Electrocatalytic Dechlorination of Polychlorinated Biphenyls"; *Chemosphere,* **13**, 415 (1984).
144. Parker D.K.; "A Safe, Efficient Chemical Disposal Method for Polychlorinated Biphenyls"; Goodyear Tire and Rubber Co., Akron, Ohio 1980.
145. Oku A., K. Yasufuku and H. Kataoka; Chemistry and Industry, 841 (1978).
146. Arato G.J., D.R. Pugh and I. Webber; "Process for the Dechlorination of Polychlorinated Biphenyls"; B.C. Hydro R. & D.; patent March 10, 1983.
147. Howard K.J. and A.E. Sidwell; Vertac Chemical Corp.; "Chemical Detoxication of Toxic Chlorinated Aromatic Compounds"; patent **4**, 327,027 April, 1982.
148. Brunelle D.J. and D.A. Singleton; "Destruction of PCBs in Transformer Oil" ACS Conference, Las Vegas, March 1982.
149. Lee L.A.; "Removal of PCBs from Contaminated Oils with NAPEG Reagents"; 1981 Spring Convention of Southeastern Electric Exchange, April 23, 1981.
150. Garland J. J4 "Philadelphia Solution to a National Problem"; EPRI PCB Seminar, Dallas, Texas, December 1981.
151. West P.; "Organoalkali Metal Compounds"; Inorganic Chem. Series One Vol. 4, MTP International Review of Science, Organometallic Derivatives of the Main Group Elements.
152. Maday J., M.A. Thompson and I. Webber; "Process for the Destruction of PCBs by Chemical Reagent"; RTE patent pending.
153. Tiernan T.O., D.J. Wagel, J.H. Garrett, G.F. Van Ness, J.G. Solch, L.A. Harden and C. Rogers; "Laboratory and Field Tests to Demonstrate the Efficacy of KPEG Reagent for Detoxification of Hazardous Wastes Containing PCDDs and PCDFs and Soils Contaminated with Such Wastes"; *Chemosphere,* **18**, 835, (1989).
154. des Rosiers P.E.; "Chemical Detoxification of Dioxin Contaminated Wastes Using Potassium Polyethylene Glycolate"; *Chemosphere,* **18**, 343 (1989).
155. Larsen B., S. Bwadt and S. Facchetti; "Separation of Toxic Congeners from PCB Mixtures on two Series Coupled Narrow-Bore Columns (50 m SIL-8 and 25 m HT-5); Gordon and Breach Science Publishers; Current Topics in Environmental and Toxicological Chemistry, Vol.16, 1993.
156. Buser H.R.; "Formation of PCDFs and PCDDs from the pyrolysis of Chlorobenzenes"; *Chemosphere,* **8**, 415 (1979).
157. Cullis C.F. and J.E. Manton; Trans. Faraday Soc.; **54**, 381 (1958).
158. Mazer, Sueann L.; 'Products of Pyrolysis and Combustion of Liquid PCB Substitutes"; in "Hazards, Decontamination, and Replacement of PCB—a Comprehensive Guide" edited by Jean-Pierre Crine, Environmental Science Research Volume 37, 1986.
159. Shaw-Tao Lin and Roth, J.A.; *J. Org. Chem.,* **44**, 309 (1979).
160. Roth, J.A., S.R. Dakoji, R.C. Hughes and R.E. Carmody; *Environ. Sci. Technol.,* **24**, 80 (1994).

161. USEPA; "Superfund Ground Water Issue—Facilitated Transport"; EPA/540/4–89/003; Office of Research and Development, August 1989.
162. des Rosiers P.E.; "Chemical Detoxification of Dioxin Contaminated Wastes Using Potassium Polyethylene Glycolate"; Dioxin '87, 7th International Symposium on Dioxins and Related Compounds, Las Vegas, NV, October 1987.
163. Masters H., B. Rubin, R. Gaire and P. Cardenas; "EPA's Mobile Volume Reduction Unit for Soil Washing"; Presented at the Seventeenth Annual Hazardous Waste Research Symposium, Cincinnati, OH, April 1991.
164. U.S. EPA; Technology Evaluation Report. CF Systems Organics Extraction System, Volume 1. Report to be published.
165. U.S. EPA; Innovative Technology: B.E.S.T. Solvent Extraction Process, OSWER Directive 9200.5–253FS, 1989.

IAN WEBBER
Chief Advisor to the Indonesian Government
on Laboratory Planning, Certification and Training

PESTICIDES

INTRODUCTION

Man employs pesticides as purposeful environmental contaminants in order to improve environmental quality for himself and his domesticated animals and plants. In agriculture, pesticides are used to increase the cost/benefit ratio in favor of the farmer and of the ultimate consumer of food and fiber products, the citizen. It has been widely estimated that the U.S. farmer receives an average net return of about $4 for every $1 invested in pesticides (PSAC, 1965, Pimentel and Levitan, 1986). In our present era of managed ecology of monocultures, of farm mechanization, and of the complex system of food harvesting, processing, distribution and storage, the use of pesticides often represents the slender margin between crop production and crop failure, and between economic profit and economic loss. In the developing countries where food supplies are marginal, pesticide use may represent the margin between survival and starvation.

In public health, pesticides often provide the only feasible means for the control of the invertebrate vectors of human and animal diseases. It is difficult to place monetary values on human health, but for malaria in India, the World Health Organization has estimated that an investment of $200 million in malaria control by DDT residual house spraying during 1956–66 saved 179.5 million man days of labor or an estimated saving of $490 million. The cost/benefit ratio has thus been about $2.7 return for every $1 invested. In addition, during this period the annual number of cases of malaria has decreased from 75 million to 150,000 and deaths from about 750,000 to 1500 (World Health, 1968).

In surveying the role of pesticides in environmental quality it must be remembered that pests themselves generally affect adversely the quality of the environment. The spectrum ranges from a mosquito in the bedroom or a cockroach in the pantry to a plague of locusts or the tsetse flies (*Glossina* spp.) which as vectors of trypanosomiasis have effectively prevented the development of 4.5 million square miles of Central Africa. The presence of vicious biting black flies (*Simulium* spp.) or wide-spread defoliation of forest and shade trees by the gypsy moth (*Porthetria dispar*) or other defoliators are effective deterrents to the resort industry in many northland vacation sites. Who can place a realistic value on the loss to environmental quality from the chestnut blight or the Dutch elm disease which have destroyed millions of North America's finest shade trees? Therefore it must be recognized that the purposeful environmental contamination by pesticides generally provides environmental benefits substantially greater than the risk of environmental pollution. It is also necessary to distinguish carefully between *environmental contamination*, which may not pose any risk or hazard to the environment, and *environmental pollution* where the health or well being of man or other animals and plants may be severely threatened. Environmental contamination is often a matter of degree, as for example with selenium, which at very low levels is essential for the normal growth and development of vertebrates yet at higher levels is an extremely poisonous pollutant. What is wanted then in an exploration of "pesticides in the environment" is a scientific appraisal of all these elements, a judicious weighting of risk/benefit ratios, and where deleterious effects on environmental quality are detected, the prompt substitution of remedial measures and/or alternative pesticides which pose no environmental hazard (Brown, 1978, McEwen and Stevenson, 1979).

Another measure of the economic value of pesticides is the total extent to which they are used. Pesticide production in the United Sates over the period of 1962–1986 is shown in Table 1. The average annual increase in total production from 1962 to 1976 was about 5% per year but since that time growth has averaged only about 1% per year. Over this period, there have been major changes in use patterns with herbicide use doubling from 1962 to 1968 and doubling again by 1980. The use of insecticides grew slowly from 1962 to 1974 and since that time has decreased about 5% per year. Agricultural use represents about 68% of the market, industrial use 17%, home and garden use 8%, and governmental use 7% (Storck, 1984, 1987). In global terms, the United States uses about 26% of total production, W. Europe 25%, the Far East 22%, E. Europe and USSR 10%, Latin America 9%, and the other regions about 8% (Chem. Week, 1985).

The most recent major inventory of pesticide use in United States agriculture was made in 1976 when it was estimated that 295 million kilograms of pesticide active ingredients were applied to 84 million hectares of cropland or about 61% of the total crop acreage. Herbicides were applied to 56%, insecticides to 18%, and fungicides to 2% of cropland. Corn was the most heavily treated crop with about 36% of total farm use, followed by soybean 13%, and cotton 12%. These three crops accounted for 61% of the total farm use of pesticides (Eichers *et al.,* 1978).

TABLE 1
Production of synthetic organic pesticides in the United States[a]

Year	Millions of kilograms			
	Fungicides	Herbicides	Insecticides	Total
1962	70	85	216	370
1964	70	118	208	396
1966	76	124	256	455
1968	87	183	265	534
1970	64	184	223	410
1972	65	205	256	526
1974	74	275	295	644
1976	65	298	257	620
1978	67	302	275	644
1980	71	366	230	667
1982	50	285	172	506
1984	56	325	159	540
1986	51	330	155	536

[a] Data from International Trade Commission.

TABLE 2
Estimated farm use of pesticides in the United States[a]

Pesticide	Millions of kilograms			
	1966	1971	1976	1982
Alachlor	—	6.7	40.2	37.7
Atrazine	10.7	26.1	41.0	34.1
2,4-D	18.2	15.2	17.5	9.5
Trifluralin	2.4	5.2	12.9	15.9
Toxaphene	15.7	16.8	14.0	1.4
DDT	12.3	6.5	—	—
Aldrin	7.1	3.6	0.4	—
Carbaryl	5.6	8.1	4.2	—
Carbofuran	—	1.3	5.3	3.2
Methyl parathion	3.6	12.5	10.4	4.1
Parathion	3.8	4.3	3.0	—

[a] Data from USDA Agr. Econ. Repts. 254 (1974), 418 (1979), Council Environ. Qual. Rept. 15 (1984).

An estimate for 1982 indicated that 113 million hectares of cropland were treated with 337 million kilograms of pesticides, with herbicides applied to 59%, insecticides to 18%, and fungicides to 3% of cropland. It was estimated that 16% of the total area of the United States received some direct application of pesticide annually (Pimentel and Levitan, 1986).

Major changes have occurred in the types of pesticides applied in the United States during the last half of the 20th Century. Available data for the farm use of major pesticides during the last two decades is presented in Table 2. The organochlorine insecticides DDT, BHC, toxaphene, aldrin, dieldrin, endrin, heptachlor, and chlordane dominated the market for about 25 years following their introduction after World War II. These persistent and broad-spectrum insecticides proved to be environmentally uncontrollable, and legal restriction of the use of DDT in 1973 was followed by restrictions on the use of aldrin and dieldrin in 1974, chlordane and heptachlor in 1976, and toxaphene in 1983. As a result the use of these organochlorines, which comprised 46% of all farm insecticides in 1971, declined to 29% in 1976 and was nearly phased out by 1984 (Table 2). Similar regulation of use has occurred in Western Europe, USSR, China, and Japan, although extensive use of these insecticides still occurs in developing countries of Asia, Africa, and South America. The organochlorine insecticides were almost entirely replaced during the last decade with biodegradable organophosphate insecticides such as parathion, methyl parathion, diazinon, malathion, and chlorpyrifos; and with the carbamate insecticides carbaryl, carbofuran, methomyl, and aldicarb. During the past decade a new group of insecticides, the synthetic pyrethroids permethrin, deltamethrin, cypermethrin, fenvalerate, and flucythrinate have become used extensively.

The use of herbicides has increased four-fold since 1960 as the preplanting technology of weed control developed. There has been a decrease in the use of the chlorinated phenoxyacetic acids and chlorinated benzoic acid herbicides. Presently the market is dominated by the triazines such as atrazine and cyanazine, the acetamides such as alachlor, the nitroanilines such as trifluralin and pendimethalin, the carbamate butylate, and the urea linuron.

There are approximately 600 individual chemical compounds registered as pesticides in the United States including 80 fungicides, 200 insecticides and 300 herbicides. These are available in about 50,000 formulations. On a world basis more than 900 chemicals are in commercial use as pesticides and the number of different formulations is estimated to exceed 100,000 (Melnikov, 1971, Büchel, 1983). However, as shown in Table 2, about 10 individual pesticides comprise 50% of the total farm use in the United States and the herbicides alachlor and atrazine comprise about 36% of the total herbicide use.

The chemical properties, biological behavior, and environmental fate of this large array of pesticides are exceedingly complex and are beyond the scope of this discussion (Melnikov, 1971, White-Stevens, 1971, Büchel, 1983). The role of these chemicals in pest control and crop production has been studied intensively and their use has become virtually indispensable to modern agriculture. Nevertheless, for the majority of the individual pesticides thee is only superficial knowledge of the effects of their long term use on the quality of the environment (Brown, 1978, McEwen and Stevenson, 1979).

During the 1980s the introduction of several groups of pesticides with 10- to 100-fold greater activity than conventional pesticides that are normally applied at about 1 kg per ha, provided for major improvements in the effects of pesticides on environmental quality. The synthetic pyrethroids are now widely used to control cotton insects at dosages of 20 to 50 g per ha. The sulfonyl urea herbicides such as chlorsulfuron are effective in preplanting applications at doses as low as 4 to 8 g per ha. There is increasing use of microbiological insecticides that are highly specific and do not leave persisting residues. The *delta*-endotoxin of *Bacillus thuringiensis* (BT insecticide) and nuclear polyhedrosis viruses (NPV insecticides) are both highly specific and essentially nonpolluting. The avermectins from *Streptomyces avermitilis* are parasiticides and insecticides that control certain agricultural pests at doses of a few grams per ha.

PESTICIDES AS ENVIRONMENTAL CONTAMINANTS

Pesticides are microchemical environmental contaminants, and their rates of utilization are such that they contaminate soil, water and food in terms of parts per trillion (0.000001 ppm) to parts per million (ppm).* Thus a pesticide applied at 1 lb/acre (1.12 kg/ha) contaminates the top 1 foot (30 cm) of soil (approximately 4 million lb. or 1800 metric tons) to 0.25 ppm. Where such contamination is deleterious to environmental quality the pesticide becomes a microchemical pollutant, exhibiting such objectionable properties as (1) high physiological and ecological specificity, (2) resistance to biochemical degradation, (3) sequential concentration in organisms of the trophic web, and (4) capacity for delayed onset of intoxication (Warner, 1967).

Pesticides in Air

The widespread application of pesticides in particulate sprays and dusts insures that appreciable contamination of the air is a consequence of pesticide use. Much of the total application of pesticides is from aircraft (about 80% of pesticide application in California in 1963 was by aircraft; Mrak, 1969) where the propeller wash and wing vortices characteristically throw small particles high into the air and wind currents may drift them for miles away from the target site. When released 10 ft above the ground into a 3 mph wind, 2μ dia. particles drifted 21 miles, 10μ particles 1 mile, and 50μ particles 200 ft (Akesson and Yates, 1964). It is evident that air applications produce increased air pollution, and Wasserman *et al.* (1960) found air concentrations in forests after air application of DDT at 18.9–170.9 mg/m^3 as compared with 4.6–25.5 mg/m^3 for ground application. Comparative values for BHC were: air 4.1–53.7 mg/m^3 and ground 4.6–25.5 mg/m^3. It has been estimated that 50% of the pesticide released in aerial spraying of forest drifts away from the target site.

The use of more than 200 million individual aerosol spray dispensers annually in the United States discharges substantial amounts of such pesticides as pyrethrinspiperonyl butoxide, dichlorvos, malathion, methoxychlor, chlordane, diazinon, propoxur, and so on indoors in homes, stores, warehouses, and aircraft, in the $5-30\mu$ dia range. Organic insecticides have appreciable vapor pressures, ranging from 1.5×10^{-7} mm Hg for DDT, 1.2×10^{-2} for dichlorvos and 1420 mm for methyl bromide. Thus volatilization is a major factor in dispersal of pesticides into the air, and accounts for much of the dissipation of pesticides from treated plant surfaces and buildings and from soil (Harris and Lichtenstein, 1961). Wind erosion of dust from treated soil and so on is also a substantial factor in the dispersal of pesticides through the air.

Occupational Exposure The degree of air pollution resulting from a variety of occupation uses of pesticides is shown in Table 3 (Jegier, 1969). These values illustrate the maximum air-borne concentrations to which humans are likely to be exposed and demonstrate the substantially higher exposures experienced indoors as compared to outdoors. Spray operators during average spraying operations are exposed to minute fractions of the combined dermal and respiratory toxic dose: 0.29% for endrin, 0.43% for parathion, 0.72% for azinphos, and 1.43% for demeton; four highly toxic insecticides. The greatest hazards from air borne exposure to highly toxic materials are found during filling spray tanks with wettable powders.

Residential Exposure Agricultural spraying operations, especially those from aircraft produce considerable air pollution and traces of pesticides have been identified for many miles downwind. Where highly toxic pesticides such as tetraethyl pyrophosphate have been dusted on orchards by air, typical cholinergic symptoms of poisoning have been observed in inhabitants of neighbouring rural communities

TABLE 3

Occupation exposure to insecticides in air[a]

Insecticide	Use	Mean concentration μm/m^3
Azinphos	Orchard spraying	670
Carbaryl	Orchard spraying	600
Malathion	Orchard spraying	590
Parathion	Orchard spraying	150–360
Endrin	Vegetable spraying	50
Demeton	Greenhouse spraying	9150
Chlordane	Household spraying	440
Diazinon	Household spraying	2680
Methoxychlor	Barn and cattle spraying	7680
Lindane	Household vaporizer	100
Azinphos	Tank filling	2270
Parathion	Tank filling	530

[a] Data from Jegier (1969).

---

* Contamination of a substrate at 1 ppm represents 1 μg of contaminant per gram of substrate.

(Quinby and Doornick, 1965). Spraying of apple orchards from ground air blast equipment has resulted in concentrations in nearby residential areas of azinphos, carbaryl, malathion, and parathion as high as 0.5 mg/m^3 (Jegier, 1969). Studies of the effect of DDT spraying operations upon air pollution of rural and urban communities have produced values ranging from <0.1 ng/m^3 (0.000001 mg/m^3) to >8500 ng/m^3. Ranges observed in specific localities included: Fresno, California 0.3–19; Sacramento, California 0–2; Florida City Florida 0.1–7.6; Fort Valley, Georgia 0.3–9.9; Leland, Mississippi 0.4–22; and Lake Apopka, Florida 0.3–8500 (Jegier, 1969). Pesticides are apparently present in the air everywhere as Risebrough *et al.* (1968) measured the concentrations in the air over Barbados as ranging from 13×10^{-6} to 380×10^{-6} ng/m as compared to that of La Jolla, California where the average was 7.0×10^{-2} ng/m^3.

A comprehensive study by Stanley (1968) to determine the atmospheric contamination by pesticides in urban and rural sites of 9 U.S. cities showed that only DDT was present in all localities. The maximum concentrations found were DDT 1560 ng/m^3, toxaphene 2520 ng/m^3, and parathion 465 ng/m^3. The pesticides were mostly present in the atmosphere as particulates and the levels were generally correlated with spraying practices on particular crops.

In another study of pesticides in the air near 10 urban communities (Tabor, 1966), the maximum concentrations found near the center of town and at least a mile from agricultural operations were: DDT 22 ng/m^3, chlordane 6, aldrin 4, and toxaphene 15. Where communities were being fogged for control of pests DDT was found up to 8000 ng/m^3 and malathion up to 140.

Pesticides in Dust The windblown erosion of dust from agricultural lands treated by pesticides can become a substantial source of air pollution. Cohen and Pinkerton (1966) investigated the transport of pesticides to Cincinnati in a violent dust storm originating in the southern high plains of Texas in 1965. The major pesticide components of the dust and their concentrations in the dried dust particles after precipitation by rain in Cincinnati were: DT 0.6 ppm, chlordane 0.5 ppm, ronnel 0.2 ppm, DDE 0.2 ppm, heptachlor epoxide 0.04 ppm, 2,4,5-T 0.004 ppm, and dieldrin 0.003 ppm.

Pesticides in Rainwater Many analyses of rainwater have shown substantial content of various pesticides indicating their general distribution in the atmosphere. As an example, the following mean concentrations of pesticides were found in three locations in Ohio in ppt: DDT 0.07–0.34, DDE 0.005–0.03, and BHC 0.006–0.05. Chlordane, heptachlor, aldrin, dieldrin, and 2,4,5-T isoctyl ester were also found in rainwater samples (Cohen and Pinkerton, 1966).

Recent investigations have shown that persistent residues of toxaphene, a chlorinated camphene with at least 177 separate components, have permeated the Great Lakes ecosystem. More than 2.5×10^8 kg of toxaphene were applied to cotton in the southern United States between 1947 and 1977. The characteristic gas chromatographic "fingerprints" of the multiple components indicate that lake trout, *Salvelinus namaychush* from Lake Michigan contain 6–10 ppm of these toxaphene components and trout from remote Lake Siskowit on Isle Royale in Lake Superior, contain 1.7–4.5 ppm (Rice and Evans 1984). This contamination of the Great Lakes and its biota could only have resulted by airborne transport and precipitation in rain.

Study and the evaluation of pesticide pollution in air has been less intensive than comparable investigations of water and food. However, it is clear that pesticide residues are constantly being transported and redistributed from their sites of application through the atmosphere and are present in some degree in the air everywhere. The degree of human exposure is related to occupation and to geographic location and is highest for workers in pesticide plants, spray operators, and users of household aerosol sprays. Inhabitants of rural regions or those dwelling in houses where spraying operations are conducted for agricultural or public health purposes are obviously exposed to substantially higher concentrations of airborne pesticides than are typical urban dwellers.

Pesticides in Water

The water environment provides the ultimate sink for pesticide residues which enter it by direct contamination from rain precipitating pesticide aerosols or atmospheric codistillates, by direct application to surface waters, by runoff from treated plants and soils, by industrial and household sewage effluents, and by residues in human and animal excreta. An intrinsic property of most pesticidal molecules in high lipid solubility and low water solubility and this property strongly favors concentration from water to the lipids of living animals through partitioning through the animal cuticle and the gills. Such absorption and subsequent storage and concentration may result in aquatic animals accumulating pesticide residues hundreds and even thousands of times greater than that in the surrounding aquatic medium. This concentration process is dependent upon the initial pesticide residue in water, the length of animal exposure, i.e. lifetime, and the rate of metabolism or breakdown of the pesticide in the organism, i.e. biological half-life. Thus this subject has assumed great ecological importance.

Ground Waters Pesticide residues in ppt to ppb quantities eventually percolate into ground waters. This problem has become a major environmental concern because of greatly increased soil applications of pesticides (more than 100 million ha are treated annually), improper waste disposal, and the enhanced analytical capability provided by gas chromatography/mass spectrometry. The Safe Drinking water Act of 1974 (Public Law 95–523) requires EPA to promulgate and enforce nonpolluting drinking water standards by establishing maximum contamination levels at which no adverse health effects are observable. The Act requires EPA to publish such non-polluting drinking water standards for some 83 water contaminants by Jan. 1991.

An EPA survey of 1988 has disclosed the presence of 74 different pesticides in the ground waters of 38 states. Aldicarb insecticide has been detected as the sulfoxide and sulfone derivatives in ground waters of 15 states and 29%

of the wells in the potato growing area of Suffolk Co. New York had aldicarb residues of 7 ppb with maximum concentrations reaching 600 ppb. In Massachusetts 220 drinking water wells were closed because of aldicarb contamination from 1–50 ppb (Pesticide & Toxic Chemical News 1985). Ethylene dibromide soil fumigant has been detected in the ground waters of 8 states and in 11% of more than 1000 wells in Florida, with exposure of more than 50,000 people (Pesticide & Toxic Chemicals News 1985). Iowa ground water has been found to be contaminated with the herbicides atrazine, cyanazine, metalochlor, alachlor and the insecticides terbufos and sulprofos; all applied as preplanting applications to corn and soybeans. In California, more than 50 different pesticides were detected in water sampled from over 8,000 wells in 24 counties. In Ontario, Canada the herbicides alachlor, butylate, dalapon, dicamba, MCPA and simazine were detected in the waters of 159 of 237 wells analyzed (Frank *et al.,* 1979).

This widespread pollution of drinking water sources is one of the most important environmental problems of the 1990s. Many of the soil applied pesticides migrate very slowly through the soil and even if annual applications were discontinued, water pollution levels are expected to increase for another 5 to 10 years. Pesticides have long residence times in ground water because of the absence of light, air, and microorganisms that are primarily involved in degradation. The number of persons exposed for long periods of time is very large. The mandatory provisions of the Safe Drinking Water Act may change forever how pesticides are applied.

Surface Waters The Mrak Commission (Mrak, 1969) observed that the current U.S. annual production of pesticides, ca. 1×10^9 lb, applied to the annual U.S. runoff, ca. 1×10^6 gal/day, could result in a maximum concentration of 0.3 mg/l (0.3 ppm). Fortunately, most of the environmental contamination with pesticides is directly or indirectly to the soil where the various compounds are often tightly bound to soil colloids and/or degraded by soil microorganisms. The average runoff concentrations of 9 organochlorine insecticides, obtained from analysis of 6000 samples at 100 locations in all major U.S. river basins from 1958–65, is summarized in Table 4 (Breidenback *et al.*, 1967). The maximum amounts determined, dieldrin 0.122 ppb, endrin 0.214 ppb, DDT 0.144 ppb, aldrin 0.006 ppb, heptachlor 0.002 ppb heptachlor epoxide 0.008 ppb, and BHC 0.022 ppb were well below the suggested Federal Drinking Water Standards (USPHS, 1968) of dieldrin 17 ppb, endrin 1 ppb, DDT 42 ppb, heptachlor 19 ppb, heptachlor epoxide 18 ppb and lindane 56 ppb, except for endrin. The yearly analyses showed dieldrin to be the dominant contaminant but declining, DDT and congeners virtually constant, endrin reaching a peak in 1964 and declining.

In another study of 12 organochlorine pesticides in 11 streams in Western United States (Brown and Nishioka, 1967) the contaminants, positive samples, and range detected were: DDT—82, 0.01–0.12 ppb; DDE—49, 00–0.06; DDD—35, 0.0–0.04; 2,4-D–41, 0.01–0.35; 2,4,5-T—28, 0.01–0.07; heptachlor–27, 0.01–0.04; dieldrin—24, 0.01–0.07; silvex—14, 0.01–0.21; lindane—11, 0.01–0.04; aldrin—11, 0.01–0.07; endrin—4, 0.01–0.07; and heptachlor epoxide—2, 0.02–0.04.

A major source of pesticide pollution of ground water is from soil particles contaminating attached pesticide residues through erosion runoff, or flooding. Application of aldrin to rice fields at 415 g/ha by seeding with treated rice seeds resulted in 1.6 ppb aldrin plus dieldrin in the water after 2 days and 0.07 ppb 14 weeks after seeding. Draining of the fields after 14 weeks produced 0.027 ppb in the ditches, 0.44 ppb in the stream receiving the ditches, and concentrations as high as 0.023 ppb in the river into which the stream flowed (Sparr *et al.,* 1955). In a companion study runoff from a cotton field treated 7 days before with 450 g/ha of endrin, contained 0.66 ppb endrin after a 1.15 in. rain. Used irrigation water contained 0.11 ppb endrin 3 days after spraying.

Lakes and Reservoirs These bodies of water often represent the sites of the most serious environmental pollution problems resulting from the applications of pesticides to both land and water. The direct application of pesticides to water for the control of mosquito or black fly larvae, snails,

TABLE 4
Organochlorine pesticides in major river basins of the United States[a]

Pesticide	Positive samples of 537	Range ppb	Area of highest concentration
Dieldrin	495	0.008–0.122	lower Miss. 1964
Endrin	217	0.008–0.214	Lower Miss. 1963
DDT	145	0.008–0.144	W. Gulf Basin 1963
DDE	176	0.002–0.011	lower Miss. 1965
DDD	231	0.004–0.080	N. Atlantic basin 1963
Aldrin	31	<0.001–0.006	S.W. basin 1964
Heptachlor	6	0–0.002	lower Miss. 1965
Heptachlor eposice	26	<0.001–0.008	N. Atlantic basin 1963
BHC	44	0.003–0.022	S.E. basin 1960

[a] From Breidenback *et al.* (1967).

or water weeds is an obvious source of contamination. Rates of application commonly range from 0.1–1 kg/ha of water surface for insecticides to as much as 100 kg/ha for 2,4-D herbicide. In water 0.3 m deep these rates would range from 2 ppm to 2000 ppm.

One of the first examples of serious water pollution by pesticides resulted from the application of the larvicide DDD to Clear Lake, California for the control of Clear Lake gnat *Chaborus astictopus* which was a severe nuisance. In 1949, 14,000 gal of emulsive concentrate of DDD was applied to the lake at a rate of 14 ppb. The gnats were nearly exterminated and it appeared that no damage to fish occurred. Reinfestation from nearby lakes resulted in retreatment with DDD at 20 ppb in 1954 and 1957. Dying Western Grebes (*Aechmophorus occidentalis*) in areas around the lake were observed in 1954, 1955, and 1957 and they had tremors characteristic of DDD poisoning. Their body tissues showed as much as 1600 ppm DDD. Subsequent study showed DDD residues of up to 10 ppm in plankton, and as much as 2375 ppm in the body fat of the white catfish, *Ictalurus catus* (Hunt and Bischoff, 1960). Thus this episode provided the first well studied example of ecological magnification of a pesticide from water through a food chain of plankton→fish→birds, which died from chronic DDD poisoning.

Large fresh water lakes may have astonishingly long water retention times, which magnify pesticide contamination problems. In the Great Lakes system Lake Superior with an area of 82,366 km^2 and a volume of 12,221 km^3 has an average water retention time of 189 yrs, and Lake Michigan with an area of 58,016 km^2 has an average water retention time of 30.8 yrs. Rainey (1967) has pointed out that in such bodies of water contamination is a major disaster for which there is no apparent solution. Thus the times for 90% waste removal are $>$500 yrs for Lake Superior and 100 yrs for Lake Michigan, as compared to 20 yrs for Lake Ontario and 6 yrs for Lake Erie. Lake Michigan is exposed to pesticide contamination from intensive agriculture and from effluents from the densely populated urban areas within its 117,845 km^2 watershed. Analyses of Lake Michigan surface waters in 1968–1969 showed DDT $2.0–2.8 \times 10^{-6}$ ppm, DDE $0.8–1.4 \times 10^{-6}$ ppm, and DDD $0.3–0.5 \times 10^{-6}$ ppm; while grab samples at the Chicago filtration plant had DDT $0.034–0.058 \times 10^{-3}$ ppm, lindane $0.01–0.02 \times 10^{-3}$ ppm, aldrin 0.019×10^{-3} ppm, and heptachlor epoxide $0.019–0.049 \times 10^{-3}$ (Mrak, 1969). The ecological significance of these trace amounts is shown by studies reporting concentrations of DDT in the Lake Michigan ecosystem of 0.014 ppm in bottom muds, 0.410 ppm in amphipods, 3.22 ppm in yellow perch, 6.9 ppm in lake trout, 6.71 ppm in lake herring, and 99 ppm in herring gulls (Reinert, 1970; Harrison *et al.,* 1970). The overall concentration of DDT from water to fish-eating bird is $>10^7$. Dieldrin was present in lake trout and lake herring to 0.20 ppm (Reinert, 1970), for an overall concentration of about 2×10^5.

Estuaries The importance of estuarine waters to commercial and sports fishing makes these locations especially vulnerable targets to the runoff of pesticides in streams and rivers from agricultural practices and industrial operations. It has been estimated (PSAC, 1965) that more than 50% of the total harvest of sea foods from waters of the United States is composed of species whose existence of spawning grounds are in the estuarine zone, and this harvest includes some of the most valued sea foods—shrimp, lobster, crabs, oysters, salmon, menhaden, and game fish. Agricultural pesticides are more toxic to the marine life than any other group of chemicals, and lethal concentrations for the organochlorine insecticides aldrin, dieldrin, heptachlor, endrin, DDT, lindane, and toxaphene range from 0.0006–0.06 ppm. Mollusks in particular can concentrate extraordinary quantities of stable pesticides, and oysters have been found to accumulate DDT to 70,000 times the amount in the surrounding water. Thus these organisms are especially useful as biological indicators of pesticide pollution. Woodwell *et al.* (1967) estimated that the Carmans River estuary of Long Island contained about 0.00005 ppm DDT in the water, with concentrations of 0.04 ppm in plankton, 2–3 ppm in small fish, and up to 75.5 ppm in the ring-billed gull (*Larus delawarensis*).

Oceans Little information exists about pesticide residues in the oceans. As these compounds are leached from the land or precipitated by rains, they circulate initially in the mixed layer above the thermocline and may eventually be transferred slowly into the abyss which provides a reservoir of virtually infinite capacity (Woodwell *et al.,* 1971). The organochlorine compounds such as DDT with their high lipid solubility and very low water solubility must be largely absorbed into organic matter. Woodwell *et al.* (1971) estimate concentrations of DDT in algae of the oceans ranging from 0.1–1.0 ppm and a maximum accumulation in the mixed layer of the ocean of about 15 ppt. Scattered observations suggest that DDT is present in most marine animals, with levels in whales of 0.4–6 ppm, tuna up to 2 ppm, oysters 0–5.4 ppm, and sea birds up to 10 ppm (Butler, 1966; Wolman and Wilson, 1971). Virtually nothing is known of the presence of other pesticides in marine organisms. Trapping of pesticides in petroleum slicks, which may contain up to 10,000 ppm DDT, provides another facet to marine pollution.

Pesticides in Soils

Pesticides are most frequently applied directly to soil or to plant surfaces above it and concern for the persistence of pesticide residues in soils has existed since the first widespread use of lead and calcium arsenate. A study by Jones and Hatch (1937) reported that 3500 lb of lead arsenate was applied to a commercial apple orchard over a 25 yr period. Most of the lead and arsenic was confined to the upper 6–8 in of soil and did not harm the roots of the fruit trees. However, the residual levels were highly toxic to cover crops or to young newly planted trees.

Pesticides are applied to crops and soils in most of the agricultural areas of the United States. Estimates in 1982 suggest that of the 113 million ha of cropland, herbicides were applied to 59%, insecticides to 18%, and fungicides to 13% (Pimentel and Levitan, 1986).

Soil Residues of Pesticides Monitoring studies made of pesticide residues in soils of the heavily treated cotton growing areas of the Mississippi Delta (Gentry, 1966) have given a general picture of the extent to which agricultural soils may be contaminated. In an area where a cumulative total of 30 kg of DDT/ha had been applied over a 9 yr period 1955–63, the level in the top soil in 1964 was 1.3 ppm or about 1.3 kg/ha. In another area where 13 applications of endrin at 0.2 kg/ha were applied since 1956, the average level in the soil in 1964 was 0.05 ppm. Aldrin and dieldrin, although not used since about 1955, were found in the study area in 1964 at levels of up to 0.06 ppm dieldrin and 0.13 ppm aldrin, and benzene hexachloride was found at about 0.02 ppm. Toxaphene was a present in the soils at 0.8–3.7 ppm, and calcium arsenate, although not used for many years, had left an arsenic level of 2.18–12.8 ppm.

The occurrence of residues of chlorinated hydrocarbon insecticides in soils from 31 farms in Southern Ontario has been explored by Harris *et al.* (1966). Orchard soils contained the highest levels of DDT 19.9–118.9 ppm, DDE 3.4–15.7 ppm, DDD 0.2–3.5 ppm, and dicofol 2.4–6.9 ppm. A vegetable farm where radishes were grown on muck soil had the highest concentrations of dieldrin 1.6 ppm, aldrin 2.1 ppm, endrin 3.8 ppm, together with DDT 13.8 ppm, DDE 0.8 ppm, and DDD 0.4 ppm. The highest levels of other pesticides found were: heptachlor 0.2 ppm, chlordane 0.6 ppm, and endosulfan 1.4 ppm. The average levels found in soils from 16 farms in this area in 1966 were aldrin 0.47 ppm, dieldrin 0.78 ppm, endrin 0.12 ppm, and DDT 23.9 ppm (Harris, 1971).

The rate of degradation of any pesticide in the soil is a function of its chemical structure and the formulation applied whether emulsion, granular, or seed treatment. Most pesticide degradation in soil is accomplished by the wide range of soil microorganisms which can use the compound as an energy source, although hydrolysis and photochemical oxidations may also play a role. Soil factors which determine the actual rate of persistence are (1) soil type, (2) soil moisture, (3) temperature, (4) uptake by plants, (5) leaching by water, (6) wind erosion. Thus it is difficult to generalize about soil persistence of pesticides, which is, however, greater in muck soils heavy in organic matter than in light sandy soils. Some idea of the relative persistence of various pesticides in soils is given in Table 5, from data by Edwards (1964), Lichtenstein (1969), and Harris (1971).

Uptake of Pesticides by Plants Plants obviously absorb pesticides from the soil and translocate them throughout the leaves and fruits or pre-emergent herbicides would have little activity in killing weeds and systemic insecticides would fail to protect plants against insect attack. The actual amounts absorbed by plant roots are dependent upon the solubility of the pesticide in the lipids of the root cuticle, and the amounts translocated are a function of the water solubility of the pesticide in the translocation stream.

The organochlorine insecticides are absorbed in trace amounts by root crops such as carrots, radishes, and potatoes, and these crops grown in soil treated with 1 kg/ha of aldrin contained as much as 0.03–0.05 ppm of the pesticide (Lichtenstein, 1969). Residues of these pesticides also enter above ground portions of crops and Lichtenstein (1969) has calculated that alfalfa grown on soil treated with aldrin and heptachlor at 1 kg/ha would contain approximately 0.005 ppm aldrin and dieldrin and 0.015 ppm of heptachlor and heptachlor epoxide. In contrast, residue studies made of the results of soil and seed treatments with phorate and disulfoton showed residues in the leaves after 39 days ranging from 5–12 ppm in alfalfa, 32–137 ppm in cotton, and 7–46 ppm in sugar beets (Reynolds *et al.,* 1957). These residues dissipated rapidly as the plants grew older and in practical use conform to the residue tolerance levels in hay and cottonseed.

Pesticides in Foods

The most direct avenue for pesticide contamination of the human body is through ingestion of pesticide residues on food products. Perhaps 75% of all pesticides production is used for the production or protection of agricultural commodities and wide-spread contamination of food products is the inevitable result. However, much of the initial pesticide load is lost by "weathering", through action of rain and dew and by photochemical oxidations, by enzymatic destruction in the tissues of plants or animals and through losses in harvesting and food processing. The processes of degradation and persistence of pesticide residues usually follow first

TABLE 5
Persistence of pesticides in soils

Pesticide	Approximate time for 70–95% loss[a]
DDT	4–10 yrs
Toxaphene	2–10 yrs
Dieldrin	3–8 yrs
Lindane	3–6 yrs
Chlordane, heptachlor	3–5 yrs
Aldrin	2–3 yrs
Picloram	1–2 yrs
Simizine, atrazine	10–12 months
Monuron, fenuron, diuron	8–10 months
Trifluralin	6 months
Carbaryl, carbofuran	4–6 months
2,4,5-T	3–5 months
Parathion, chlorpyrifos, diazinon	3–6 months
Amiben, dicamba, MCPA	2–3 months
Dalapon, propham, CDAA, CDEC, EPTC	1–2 months
Phorate, disulfoton	2–6 weeks
2,4-D	4–6 weeks
Captan	3–6 weeks
Malathion	1–2 weeks

[a] Data from Edwards (1961), Kearney, Nash and Isensee (1969).

order chemical kinetics and can be plotted semilogarithmically as straight lines of log. residue in ppm vs. time elapsed since treatment. Such residue persistence curves give values for residue half-lives or RL_{50} (Gunther and Blinn, 1965) which are independent of initial concentration and thus represent a characteristic of each pesticide on or in a particular substrate such as leaf surface, fruit peel, plant wax or juice, and so on This concept then supports the empirically derived practice of observing a "safe period" after pesticide application and before harvest to permit the pesticide residue to attenuate to levels which long term animal feeding studies have shown to be innocuous to animal health. Based on these animal feeding studies and incorporating a safety factor, ideally 100-fold to allow for peculiar human metabolic idiosyncrasies and sensitivities, together with the potential daily intake in foods, the U.S. Food and Drug Administration has established pesticide residue tolerances for each registered pesticide used on a food commodity. The range of tolerances for commonly used pesticides is shown in Table 6.

The actual occurrence of pesticide residues in human foods has been studied by the FDA from 1964–1966, through the examination of 26,326 samples including raw agricultural products, milk and dairy products, processed animal feeds, shell eggs, fish and shell fish, meat, canned and frozen foods, vegetable oils, and special dietary products (Duggan, 1969). Residues of 81 different pesticide chemicals were detected, with 11 chemicals accounting for 95% of the residues found. Those most commonly detected in domestic foods were, in order of importance: DDT 25%, DDE 24%, dieldrin 17%, DDD 9%, heptachlor epoxide 7%, lindane 4.5%, BHC 2.6%, endrin 2.3%, aldrin 1.8%, toxaphene 1.4% and dicofol 0.7%. Animal tissues contained, in addition, methoxychlor and chlordane. Approximately 95% of the residues were below 0.5 ppm, 78% below 0.1 ppm and 58% below 0.03 ppm. Excessive residues, above tolerances, were found in 3.6% of the samples. It is apparent that the chlorinated organics provided the great majority of readily detectable residues in food products, because of their high stability and fat solubility.

The exposure of humans to pesticide residues in a well balanced diet has been studied by FDA since 1962. For 1964–1966 this was calculated to contain an average of 0.025 ppm chlorinated organics, 0.003 ppm organophosphates, 0.003 ppm chlorophenoxy acids, and 0.05 ppm carbamates. The average daily intake of 15 pesticides in 516 diet composites sampled by FDA is shown in Table 7 (Duggan, 1969).

Increasingly sophisticated evolution procedures have shown that a number of pesticides traditionally considered safe are in fact carcinogens when fed to laboratory animals over their lifetimes. There is much public concern about the long term hazards of eating processed foods containing trace residues of such pesticides. In the United States pesticide residues in food are regulated under the 1954 Food Drug and Cosmetic Act (FDCA) and the 1947 Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) as modified in 1978. Under these laws, the Environmental Protection Agency grants approval for the presence of certain levels of a particular pesticide on a specific agricultural commodity, that is, a "tolerance". Typical pesticide residue tolerances are given in Table 6. However, the Delaney Clause of the FDCA states that no tolerance may be set for any pesticide found to be oncogenic. Based on risk studies determined from carcinogenic effects in laboratory animals together with human dietary contamination levels, a

TABLE 6

Typical pesticide residue tolerances in the United States[a]

Pesticide	Crops, commodities	Tolerance ppm
Alachlor	Forage	0.2–0.75
Atrazine	Forage	0.1–0.25
Azinphos methyl	Fruits, vegetables, forage	0.3–5
Butylate	Fruits, vegetables	0.1
Captan	Fruits, vegetables	100
Carbaryl	Fruits, vegetables, forage	5–100
Carbofuran	Forage	0.1–1.0
Coumaphos	Milk, eggs, meat	0–1
Cryolite	Fruits, vegetables	7
2,4-D	Fruits	5
Diazinon	Fruits, vegetables, forage	0.75–40
Disulfoton	Vegetables, forage	0.75–12
Diuron	Fruits, vegetables, forage	0.2–7
Dormant oils	Fruits	exempt
Endosulfan	Fruits, vegetables	2
Endrin	Fruits, vegetables	0
Ferbam	Fruits, vegetables	0.1–7
Lindane	Fruits, vegetables	10
Linuron	Vegetables, forage	0.25–0.5
Malathion	Fruits, vegetables, forage	8–135
Maneb	Fruits, vegetables	0.1–10
Methomyl	Fruits, vegetables	0.1–5
Methoxychlor	Fruits, vegetables, forage	1–100
Methyl bromide	Fruits, vegetables, nuts	2–240
Nicotine	Fruits vegetables	2
Paraquat		0.5
Parathion, methyl parathion	Fruits, vegetables, forage	1
o-phenyiphenol	Fruits, vegetables	10–20
Pyrethrins	Fruits, vegetables	exempt
Ronnel	Meat	0
Rotenone	Fruits, vegetables	exempt
Simazine	Alfalfa, grass	15
Tetradifon	Fruits	1–5
Thiram	Fruits	1–7
Trifluralin	Fruits, vegetables	0.05
Zineb	Fruits, vegetables	7–60

[a] Environmental Protection Agency (1971).

TABLE 7
Average daily intake and incidence of pesticides in U.S. human diet composites[a]

Pesticide	Daily intake from total diet mg/kg	% positive samples	FAO-WHO acceptable intake mg/kg
DDT	0.0005	37.4	0.01
Dieldrin	0.00009	20.2	0.0001
Lindane	0.00006	13.6	0.0125
Heptachlor epoxide	0.00004	12.6	0.0006
Carbaryl	0.0012	4.7	0.02
Malathion	0.001	4.5	0.02
Aldrin	0.00003	4.5	—
2,4-D	0.00005	3.5	0.017
Diazinon	0.000014	3.0	—
Dicofol	0.00011	2.5	—
Pentachlorophenol	0.00006	2.5	—
Endrin	trace	2.3	—

[a] Data from Duggan (1969).

National Academy of Sciences study (1987) concluded that nearly 80% of the estimated oncogenic risk to humans is from residues to 10 pesticides. The study concluded that 98% of the oncogenic dietary risk could be eliminated by revoking the registrations of 28 carcinogenic pesticides.

PESTICIDES IN WILDLIFE

The widespread use of pesticides has occasioned much concern about adverse affects on wildlife. Significant reviews of this important subject include Hunt (1966), Dustman and Stickel (1966, 1969), and Cope (1971). Concern in this area has moved from studies of acute toxicity (Rudd and Genelly, 1956) to physiological, behavioral, and ecological implications of the toxicity to non-target organisms (Cope, 1966; Newsom, 1967; Johnson, 1968; Pimentel, 1971). Despite the implications of selective toxicity inherent in pesticide application it is evident that pesticide usage often has deleterious effects on non-target organisms. As Pimentel (1971) points out, pesticide usage in the United States is aimed at about 2000 pest species of plants and animals, but many of the remaining 200,000 non-target plants and animals, the great majority of which are necessary for human survival, are affected directly or indirectly. The anthropomorphic concepts of pest and beneficial species are not valid ecologically and it is obvious that both target and non-target organisms respond to pesticide contamination in a variety of essentially similar ways (Newsom, 1967).

Acute Toxicity of Pesticides

An important first step in assessing pesticide effects on organisms is the determination of the acute toxicity as LD_{50} in terms of dose applied, eaten, or injected; or LC_{50} in terms of concentration in the water, causing 50% mortality of an animal population. Knowledge of the comparative toxicities of a variety of pesticides to representative organisms is important in selecting pesticides for various uses and in assessing risks to non-target species. Unfortunately most of the large amount of data on quantitative toxicology deals with effects of a few compounds on a wide variety of species and there is little uniform data on a wide representation of species for the important pesticides. Table 8 presents a selection of toxicity values to a variety of organisms for the widely used pesticides for which adequate data is available. From the comparative viewpoint, the table shows the impossibility of profound generalizations about pesticide toxicity. Methoxychlor and malathion, two of the very safest materials to mammals and birds, are highly toxic to fish and invertebrates. Carbaryl, with low toxicity to mammals, birds, and fish is highly toxic to invertebrates. Zectran and carbofuran, among the most toxic compounds to mammals and birds, are of low toxicity to fish. Endrin, phorate, disulfoton, and parathion are general biocides, highly toxic to nearly all animals. Only the herbicides such as dalapon, dicamba, diquat, diuron, endothall, paraquat, 2,4-D, 2,4,5-T and simazine seem relatively safe to all the animals listed.

Ecological Magnification

This descriptive term applies to the ability of living organisms to concentrate stable, lipid soluble, water insoluble substances in their bodies either through successive accumulation in food chains or directly from water by partitioning. In this manner residues of substances such as DDT and dieldrin have been concentrated in fish $>1 \times 10^6$ and $>1 \times 10^5$ fold respectively over the concentration of the water. The first demonstration

TABLE 8
Comparative toxicity of pesticides to various organisms[a]

				24-hr LC_{50}			Topical LD_{50}, mg/g			
	Oral LD_{50} m/kg			Rainbow		Fairy			*Musca*	*Apis*
Pesticide	Rat	Mallard	Pheasant	Trout	Blue gill	shrimp[b]	Stonefly[c]	Water flea[d]	*dometica*	*mellifera*
Aldrin	36–60	520	16.8	0.0061	0.010	45	0.008(2d)	0.028	2.9	4.5
Atrazine	3080	>2000		12.6(2d)				3.6		>100
Carbaryl	850	>2179	>2000	2.0	2.5	0.040	0.030	0.0064	900	2.3
Carbofuran	5	0.40	4.2		0.24(4d)				4.6	
Chlordane	335–430	1200		0.022	0.095	0.160	0.170	0.029	6.0	
Diazinon	108–76	3.5	4.3	0.380	0.052	0.80	0.06(2d)	0.0009	2.95	
DDT	113–118	>2240	1296	0.012	0.005(2d)	0.0047	0.041	0.0036	1.9	20.0
Dicamba	2900		673–800	35(ed)	130(2d)	10				>100
Dieldrin	46–46	381	79	0.0031	0.015	1.4	0.006	0.240	0.95	2.2
Disulfoton	6.8–2.3	6.5			0.040(2d)	0.110	0.040			
Diuron	3400	>2000				0.70	3.6	1.4		>100
2,4-D acid	375	>1000	472		8	1.4–6.8 (ester)	8.5 (ester)	0.2		>100
Endrin	17.8–7.5	5.6	1.8	0.0028	0.0008	0.0064	0.004	0.020	3.15	20.8
Heptachlor	100–162	>2000		0.013	0.026	0.150	0.008	0.042	2.25	
Lindane	88–91		500–600	0.018(2d)	0.10	0.120	0.012	0.460	0.85	2.0
Malathion	1375–1000	1485		0.130	0.110	0.0038	0.035	0.0018	26.5	1.1
Methoxychlor	6000	>2000		0.074	0.083	0.0047	0.030	0.00078	9.0	
Parathion, ethyl	13–3.6	1.9–2.1	12.4	2.0	0.047	0.012	0.028	0.0004	0.9	3.5
Parathion, methyl	14–24	10.0	8.2	2.75(3d)	8.0(2d)				1.2	0.84
Phorate	2.3–1.1	0.62	7.1		0.0055(2d)	0.024				
Toxaphene	90–80	70.7	40	0.004(2d)	0.0066	0.180	0.018	0.015	11.0	274
Trifluralin	>10,000	>2000	>2000	0.098	0.130	8.8	13	0.240		
Zectran	14.1–19	3.0	4.5	8	11.2	0.086	0.032	0.01	65	0.6

[a] Values from Pimentel (1971), Hayes (1963).

[b] *Gammarus lacustris.*

[c] *Pteronarcys californicus.*

[d] *Daphnid pulex,* 2 day values.

of this remark-able phenomenon resulted from the treatment of Clear Lake, California with DDD at 0.014–0.02 ppm and the observance of DDD residues in Western grebes and predaceous fish at 1600–2500 ppm. The overall magnification was about 120,000-fold (Hunt and Bishoff, 1960). Woodwell *et al.* (1967) describe a Long Island, New York estuary where a DDT concentration of 0.00005 ppm became successively magnified in plankton (0.04 ppm), invertebrates (0.16–0.42 ppm), fish (0.17–2.07 ppm), and predatory birds (3.15–75.5 ppm). The DDT level was concentrated about 10-fold in each trophic level and appeared in the upper levels of the food web largely as DDE the most stable metabolite. The overall magnification was about 120,00 fold. The magnification of aldrin through a terrestrial ecosystem where Missouri cornfields were treated over a 15 yr period with a total of about 25 kg/ha was studied by Korschgen (1970). The soil contained an average residue of 0.06 ppm aldrin, earthworms averaged 1.49 ppm; seed-eating ground beetles *Harpalus* contained 1.1 ppm; the predaceous beetle *Poecilus* 9.67 ppm; the seed-eating mice *Peromyscus, Mus,* and *Reithrodontomys* averaged 0.98 ppm; toads, *Bufo americanus,* feeding on insects and other invertebrates, 4.60 ppm; and garter snakes *Thamnophis sirtailis* which eat salamanders, toads, earth-worms, and small birds and mammals, accumulated 10.3–14.4 ppm. Most of the aldrin was stored as the more stable metabolic oxidation product dieldrin and the overall magnification was about 200-fold.

The aquatic habit favors the concentration of trace residues in the stable organochlorine compounds in invertebrates and fish. Dustman and Stickel (1969) cite examples such as: fathead minnows concentrated endrin 10,000 ×; croakers concentrated DDT 20,000 ×; trout concentrated dieldrin 3300 × ; oysters concentrated lindane 60 ×; DDT 15,000 ×; and heptachlor 17,600 ×. In these cases, the water and the animal tissue, especially the lipids, seem to form a system of liquid-liquid partitioning between two immiscible solvents and Hamelink *et al.* (1971) suggest that in fish the pesticide passes through two partitions from water→ blood, and from blood→fat. They suggest that the partition coefficients for the various organo-chlorine pesticides are roughly proportioned to their solubility in fat (ca. 10 g/l)→ solubility in water (DDT 0.0012, γ-chlordane 0.009, aldrin 0.027, heptachlor 0.056, dieldrin 0.186, endrin 0.23, heptachlor epoxide 0.35, and lindane 10 ppm). Calculated partition coefficients for fat/water are: DDT ca. 10^6, dieldrin 5×10^4, heptachlor epoxide 3×10^4, and lindane 10^3 which are remarkably close to the actual levels of concentration observed in nature.

Pesticides and Fish

It will be noted from Table 8 that many of the organic insecticides are extremely toxic to fish in the ppb range, e.g. endrin, dieldrin, aldrin, heptachlor, toxaphene, chlordane and DDT. The synthetic pyrethroids are also toxic to fish in the ppb range. Disastrous fish kills have resulted from industrial effluents, accidental spillages and unwise disposal of spray tank residues into ponds and streams where fish and invertebrates may be killed for many miles downstream. Examples of such accidents include massive fish kills in the lower Mississippi River in 1963 attributed to effluents from the manufacture of endrin (Barthel *et al.,* 1969) and in the lower Rhine River from the sinking of a barge load of endsulfan.

DDT spraying of the Yellowstone River System in 1955 at 1 lb/acre by air for spruce budworm control was responsible for widespread mortality of mountain white-fish (*Proposium williamstoni*), brown trout (*Salmo trutta*) and longnose sucker (*Coatostomus catostomus*). After subsequent spraying in 1957, dead fish were analyzed for DDT and the trout contained from 0.186–14.0 ppm DDT and 0–6.53 ppm DDE, and the whitefish 0.06–6.9 ppm DDT and 0–5.6 ppm DDE (Cope, 1961). The prolonged and deleterious effects of such spraying on important invertebrate sources of fish food are shown in Table 9 (Cope, 1961). DDT aerial spraying of the Lake George region of New York for gypsy moth control was responsible for widespread contamination of the lake and total loss of lake trout fry in 1955–1958. The affected fry were observed to float upside down and then to sink and die. The symptoms always occurred after absorption of the yolk sac and subsequent study showed that DDT was present at 4.75 ppm or more in the fish eggs which displayed this syndrome and that as the yolk sac was absorbed the fish were poisoned and subsequently died (Burdick *et al.,* 1964).

Many other examples of the adverse effects of the organochlorine pesticides on fish could be cited, and the entire problem is reviewed by Johnson (1968).

Pesticides and Birds

The effects of pesticides on bird populations have been more intensively considered than any other aspect of the impact of pesticides on the environment. As with fish in the aquatic environment, most of the concern has been directed at the use of the persistent organochlorine insecticides (Moore, 1965; Rudd and Genelly, 1956). Massive spraying of elms with DDT to control the elm bark beetle vectors (Scolytidae) of Dutch elm disease has resulted in dying robins and other birds (Barker, 1958; Hickey and Hunt, 1960). Such spraying contaminated the soil up to 18 ppm and earthworms contain 53–204 ppm DDT and DDE. Robins feeding on the contaminated earthworms suffered as much as 80% mortality and their bodies contained a median residue of 3 mg DDT, an amount which could be obtained by eating less than 100 earthworms. The widespread use of cereal seeds treated with dieldrin, DDT and other organo-chlorine insecticides has been an important cause of bird mortalities. Sowing of DDT-treated rice seed in California resulted in residues in pheasants averaging 740 ppm in body fat and lowered survival rates (Hunt, 1966). Spring seeding with dieldrin, aldrin, and heptachlor treated seed in Great Britain produced substantial evidence of birth mortality and the use of seeds treated with the organochlorine insecticides has been restricted in both the U.S. and Great Britain (Moore, 1965).

Many of the pesticides listed in Table 8, especially aldicarb, carbofuran, diazinon, disulfoton, endrin, parathion, and

TABLE 9
Effects of DDT forest spraying on bottom insects of Yellowstone River[a]

	Number of organisms per 10 ft^2					
	Plecoptera	Ephemeroptera	Trichoptera	Diptera	Other	Total
Prespray	160	102	17	0	0	279
After 2 weeks	14	3	12	0	0	29
After 1 month	31	9	0	1	0	41
After 1 yr	35	91	103	39	15	283

[a] Data from Cope (1961).

phorate are acutely toxic to birds but their use patterns are such that damage to bird populations has not been extensive.

Thin eggshell syndrome Drastic reductions in the populations of North American and European raptorial birds such as the peregrine falcon, (*Falco peregrinus*), sparrowhawk (*Accipiter niger*), kestrel (*Falco tinnuncujlus*), osprey (*Pandion haliaetus*), bald eagle (*Haeliaetus leucocephalus*), and golden eagle (*Aquila chrysactos*) have been observed during the past 2 decades and these population declines coincided with the period of large scale use of DDT and other chlorinated organic insecticides. Although the complex reasons for these population declines are highly controversial, there is ample evidence that reproductive failure is a major factor (Pimentel, 1971). Egg shell fragility or thin egg shells is the most common cause and on Anacapa Island of California, where the complete reproductive failure of the brown pelican (*Pelecanus occidentalis*) was observed, egg shells were found to average 0.38 mm in thickness, or a decrease of 34% over normal (0.57 mm). Residues of DDT and DDE as high as 1200 ppm (85% DDE) were found and the fat of the adult birds, contained from 738–2603 ppm DDT and DDE (Pimentel 1971). A similar effect has been shown for the Scottish golden eagle where egg shell thickness has declined from an index of 3.146 over the period of 1848–1946, to 2.839 during 1951–1965, or a decrease of 9.9% (Ratcliffe, 1970). Bald eagle eggshells in Florida declined in weight from an average of 12.15 g in the pre-DDT era to 9.96g in 1947–1962 or an 18% decrease, and this has been associated with a decline in the eagle population in Florida (Pimentel, 1971) although other factors must also be considered.

Laboratory experiments summarized by Pimentel (1971) show that feeding DDE to mallard ducks at 40 ppm induced a 14% decrease in eggshell thickness and in cracking of eggs. Eggshell thinning occurred at DDE levels as low as 10 ppm and duckling production per hen was reduced as much as 75% when these levels of DDE were fed. DDT was less effective in producing the thin-eggshell effect than DDE although this occurred in coturnix fed DDT at 10 ppm and 25 ppm.

After the cancellation of registrations of DDT in 1973, the average body concentration in the lake trout of Lake Michigan declined from 19.2 ppm in 1970 to 2.2 ppm in 1984. However, the cancellation of aldrin and dieldrin registrations in 1974 had little effect on the average body concentration of dieldrin in lake trout, which was 0.27 ppm in 1970 and 0.38 in 1984. There has also been a decrease in the residues of DDT in herring gull eggs from 33.4 ppm in 1986 to 7.1 ppm in 1986, and in dieldrin residues from 0.82 ppm in 1976 to 0.28 in 1986 (Council Environ. Qual. 1987).

Methylmercury Fungicides These compounds, especially Panogen$^{®}$ or methylmercury dicyandiamide, have been very widely used as fungicidal treatments for small grains. In Sweden, as much as 80% of the spring wheat sown from 1940–1965 was treated with such alkylmercury fungicides. As a result, mercury became very widely distributed in wild birds. Borg *et al.* (1966) studied the mercury content of the livers of seed eating birds, pheasants, partridges, pigeons, finches, and so on, which were found dead. Of 253 birds investigated, 48% had Hg levels above 2 ppm and 20% above 10 ppm. There was a remarkable increase in the incidence of high levels of Hg in these bird livers during May–June and October–November, which related to the planting of seed dressed with alkylmercury fungicides during these months (Borg *et al.,* 1968). When pheasants were fed wheat treated with 20 ppm methylmercury dicyandiamide, a normal rate of agricultural use, the eggs had reduced hatchability with Hg residues of 1.3–2.0 ppm and the birds died in 29–61 days with liver residues of Hg ranging from 30 to 130 ppm. Wild pheasants and partridge eggs in Sweden contained about 30 ppm Hg (Borg *et al.,* 1966).

Predatory birds, hawks, falcons, eagles, and owls were also examined for Hg content of the liver. Of a total of 412 predators found dead, shot, or trapped for examination, 62% had Hg levels above 2 ppm and 19% above 10 ppm (Borg *et al.,* 1966).

The methylmercury readily entered food chains and it was found that hens in Sweden fed grain grown from wheat whose seed was treated with alkylmercury fungicides produced eggs abnormally contaminated with Hg. Swedish eggs on the open market, 1964–1966, averaged 0.029 ppm Hg as compared with 0.007 ppm Hg in eggs on the market in Continental Europe (Westöö, 1969). Thus a daily consumption of two Swedish eggs could exceed the FAO/WHO recommended safe level for Hg intake of 0.00005 mg per kg body weight. As a result of the wildlife contamination and the dietary egg problem, alkylmercury fungicides were banned in Sweden on February 1, 1966. Subsequently, the Hg contents of Swedish hens' eggs decreased to an average of 0.009 ppm by September 1967 (Westöö, 1969).

Pesticides and Bees

The honeybee *Apis mellifera* is not only an important domesticated animal producing honey and bees wax valued at $50 million annually in the U.S., but also is the primary agent responsible for the pollination of fruit, vegetable, and field crops valued in excess of $12 billion annually. As representative insects, bees are highly susceptible to insecticides and are a prime non-target victim of pesticides applied during blossom time when the bees' foraging patterns take them to sprayed orchards and field; or from pesticides carefully applied near apiaries. As an example of the damage that insect control programs can do to bee colonies, the repeated application of carbaryl to cotton in the Imperial Valley of California as part of an eradication program for the pink boolworm *Pectinophora gossypiella,* is estimated to have killed 30,000 colonies of bees in 1967.

Anderson and Atkins (1968) have classified the toxicity of pesticides to the honeybee *Apis mellifera* in groups as follows:

Group I *highly toxic* LD_{50} 0.001–1.99 μg per bee: includes the organophosphorus esters parathion, methyl parathion, azinphos, malathion, dimethoate, dichlorvos, diazinon, chlorpyrifos, fenthion, mevinphos, phosphamidon, naled, and fenitrothion; the carbamates carbaryl, carbofuran, aldicarb, propoxur, and Zectran®; and the organochlorine compounds aldrin, dieldrin, heptachlor, lindane; and the arsenicals. The newer synthetic pyrethroids such as permethrin, fenvalerate, deltamethrin, and flucythrinate are extremely toxic to bees with LD_{50} values <0.01 μg per bee.

Group II *moderately toxic* LD_{50} 2.0–10.99 μg per bee includes DDT, endrin, endosulfan, mirex, chlordane, disulfoton, phorate, demeton, ronnel, coumaphos, temephos, and carbophenothon.

Group III *relatively non-toxic* LD_{50} greater than 11.0 μg per bee includes all the common acaricides, trichlorfon, cryolite, DDD, methoxychlor, rotenone, ryania, toxaphene, the fungicides, the herbicides, and the defoliants.

Simple precautions such as avoiding insecticide applications during bloom, or applying toxic insecticides only as soil and seed treatments can greatly minimize bee losses.

PESTICIDES AND HUMANS

Humans, in an ecological sense, are consumers at the top of the food pyramid and are exposed to pesticide residues contained nearly everywhere in cereals, fruits and vegetables, and in meat, fish and poultry. Additionally humans are exposed directly to pesticides from occupational use of aerosol products in the home, from home and garden sprays, and from effluents from agricultural operations. Therefore it is not surprising that traces of the stable organochlorine insecticides are found in the body fats of humans everywhere.

DDT in Humans

Elaborate monitoring studies of pesticide residues in humans have been reported by Hayes (1966), Durham (1969) and Mrak (1969). DDT was first found in human fat in 1948 and at least 9 general surveys of the DDT content of human fat have been made in various areas of the United States. These show that DDT is stored in the body fat of everyone in the United States, with a mean storage level in the 1960's of about 8 ppm combined DDT and its principal metabolite DDE. With the revoking of the registrations of DDT in 1973 and aldrin and dieldrin in 1974, the levels of these insecticide residues in human body fat have declined slowly over the next decade as shown in Table 10. However, detectable tissue levels of oxychlordane, the principal persistent metabolite of chlordane, very widely used for household pest control have persisted unchanged (Table 10).

TABLE 10
Insecticide residues in human body fat—United States[a]

Year	DDT-T	dieldrin	oxychlordane
		ppm in fat	
1972	6.97	0.18	0.10
1974	5.15	0.14	0.12
1976	4.35	0.09	0.11
1978	3.52	0.09	0.11
1980	2.82	0.10	0.12
1983	1.67	0.06	0.10

[a] Council on Environmental Quality, 17th Report, 1987.

The storage levels of DDT and DDE in people in other countries have also been investigated (Durham, 1969). The levels were lowest in countries where little DDT is used, for example Denmark with DDT 0.6 and DDE 2.7 ppm; and highest in countries such as India, DDT 16 and DDE 10 ppm, where large amounts of DDT are used in malaria eradication.

The storage of DDT in human fat is a direct function of level of intake and is lowest in meat abstainers, DDT 2.3 and DDE 3.6 ppm; and Eskimos, DDT 0.8 and DDE 2.2 ppm, whose diets contained the lowest amounts of DDT. Volunteers given 35 mg per day orally stored DDT 281 and DDE 40 ppm, and a formulator in a DDT factory had DDT 648 and DDE 483 ppm; yet these men remained in good health (Hayes, 1966).

DDT and its metabolites DDE and DDD are also found in other human tissues: blood DDT 0.0068 and DDE 0.0114 ppm, milk 0.08–0.13 ppm, adrenal glands 0.7 ppm (Durham, 1969). DDT ingested by humans is excreted slowly as 4,4-dichlorodiphenyl acetic acid (DDA) which is present in the urine of the general U.S. population at <0.02–0.18 ppm (Hayes, 1966). The levels of DDA in human blood and urine roughly parallel the rate of intake and can be used to monitor the exposure (Mrak, 1969).

The significance of levels of DDT and its metabolites DDE and DDD in the general population is complex.

Food is generally believed to contribute as much as 89% of the total intake of DDT but recent studies showing substantial fat storage of DDT in Eskimos and in institutional patients where DDT intake is very low suggest that non-dietary sources such as house dust may contribute up to 50% of the total body burden (Mrak, 1969). The non-white population of the U.S. has significantly higher fat storage of DDT than the white population suggesting the influence of dietary factors and socioeconomic factors resulting in greater use of house-hold insecticides. Fat storage of DDE increases with the age of the population, especially in non-whites where the level increased from 4.06 ppm in the 0–5 yrs category to 8.61 ppm in 41–50 years, and 15.50 ppm at 90+ yrs. Accumulation in states with cooler climates averaged 4.85 ppm as compared with 9.21 ppm in warmer States. There is no clear cut evidence of a sex difference in fat storage of DDT or in positive association with specific disease conditions (Mrak, 1969).

Other Pesticides in Humans

Other stable organochlorine insecticides or their stable metabolites are also found in human tissues. Surveys of the U.S. population have shown levels of these materials in fat and blood: BHC isomers—fat 0.20–0.60 ppm, blood 0.0031–0.0019 ppm; heptachlor epoxide—fat 0.10–0.24 ppm, blood 0.0008–0.0011 ppm (Hayes, 1966; Durham, 1969). In addition to these pesticides demonstrated in the blood of the general U.S. population, trace levels of chlordane, toxaphene, endrin, aldrin, pentachlorophenol, and 2,4-D have been found in tissues of humans with unusual occupational or accidental exposure.

Most of the other types of pesticides widely used, the organophosphorus esters, carbamates, phenoxy- and benzoic-acids, triazines, ureas are less persistent in the environment than the organochlorine compounds and are rapidly metabolized by enzymes in the human liver and converted to predominantly water-soluble products eliminated in the urine. Therefore there is little evidence of fat storage of these compounds. Monitoring of the urinary content of water-soluble metabolities is a very sensitive measure of intake of such pesticides and has been studied for parathion, malathion and related organophorus esters, the carbamates carbaryl and propoxur, dinitro-*o*-cresol, lead, arsenic, and mercury (Hayes, 1966).

Acute Toxicity in Humans

The substitution of the more hazardous organophosphate and carbamate insecticides for the organochlorines has greatly increased the incidence of acute human poisoning. WHO has estimated that worldwide there are as many as 500,000 accidental poisonings annually with 20,000 deaths (Copplestone, 1977). In the United States, EPA estimates suggest that there are 45,000 accidental poisonings with 3000 hospitalizations and 200 deaths annually (Pimentel and Levitan, 1986).

GLOSSARY

alachlor, 2-chloro-2′,6′-diethyl-*N*(methoxymethyl) acetanilide
aldicarb, 2-methyl-2-methylthiopropionaldoxime *0-N* methylcarbamate
aldrin, 1,2,3,4,10,10-hexachloro-1,4,4a,5,8,8a-hexachloro-1,4-*endo,exo*-5,8-dimethanonaphthalene
amiben, 3-amino-2,5-dichlorobenzoic acid
atrazine, 2-chloro-4-ethylamino-6-isopropylamino-*s*-trianzine
azinphos, methyl, *0,0*-dimethyl S-[4-oxo-1,2,3-benzotriazine-3(4H)-ylemthyl] phosphorodithioate
bendiocarb, 2,3-(isopropylidenedioxy)- phenyl *N*-methylcarbamate
BHC, mixture of stereoisomeric 1,2,3,4,5,6-hexachloro-cyclohexane
butylate, *S*-ethyl diisobutylthiocarbamate
captan, *N*-trichloromethylmercapto-4-cyclohexane 1,2-dicarboximide carbaryl, 1-naphthyl *N*-methylcarbamate
carbofuran, 2,2-dimethyl-2,3-dihydro-7-benzofuranyl *N*-methylcarbamate
carbophenothion, *S*- [*p*-chlorophenyl] thiomethyl] *0,0*'-diethyl phosphorodithioate
CDAA, 2-chloro-*N,N*-diallyl acetamide
CDED, 2-chloroallyl-*N,N*-diethyldithiocarbamate
chlorbenzilate, ethyl 4,4′-dichlorobenzilate
chlordane, 2,3,4,5,6,7,8,8-octochloro-2,3,3a,4,7,7a-hexa-hydro-4,7-methanoindene
chlordimeform, *N*′-(4-chloro-2-methyl)-*N,N*-dimethylmethanimidamide
chlorpyrifos *0,0*-diethyl *0*-(3,5,6-trichloro-2-pyridylphosphorothionate
chlorsulfuron, 2-chloro-*N*-[(4-methoxy-6-methyl-1,3,5-triazin-2-yl)-aminocarbonyl]-benzenesulfonamide
Ciodrin, α-methylbenzyl-3-hydroxy-*cis*-crotonate dimethylphosphate coumaphos, *0,0*-diethyl *0*-(3-chloro 4-methyl-7-coumarinyl)
Phosphorothionate
cryolite, sodium fluoaluminate
cypermethrin, cyano-(3-phenoxyphenyl)-methyl-3-(2,2-dichloroethenyl)-2,2-dimethyl-cyclopropanecarboxylate
cyanazine, 2-[[4-chloro-6-(ethylamino)-*s* triazine-2-y]-amino]-2-methylpropionitrile
dalapon, 2,2-dichloropropionic acid
2,4-D, 2,4-dichlorophenoxyacetic acid
DDA, 4,4′- dichlorodiphenylacetic acid
DDD, 1,1-dichloro-2,2-bis- (*p*-chlorophenyl)-ethane
DDE, 1,1-dichloro-2,2-bis- (*p*-chlorophenyl) ethylene
DDT, 1,1,1-trichloro-2,2-bis- (*p*-chlorophenyl) ethane
deltamethrin, cyano-(3-phenoxyphenyl)-methyl-3-(2,2-dichloroeth-enyl)-2,2-dimethyl-cyclopropanecarboxylate
demeton, 40:60 mixture of *0,0*-diethyl *S*-(2-ethylthio—ethyl phophorothiolate with *0,0*-diethyl *0*-(2-ethylthio)ethyl phosphorothionate
diazionon, *0,0*-diethyl-*0*-(2-isopropyl-4-methyl-6-pyrimidinyl) phosphorothionate
dibromochloropropane, 1,2-dibromo-3-chloropropane
dicamba, 2-methoxy-3,6-dichlorobenzoic acid
dichloran, 2,6-dichloro-4-nitroaniline
dichlorvos, 2,2-dichlorvinyl dimethylphosphate
dicofol, 4,4′-dichloro-α-(trichloromethyl) benzhydrol
dieldrin, 1,2,3,4,10,10-hexachloro-*exo*-6,7-epoxy-1,4,4a,5,6,7,8,8a-octahydro-1,4-*endo, exo*-5,8-dimethanonaphtha-lene
Dilan, mixture of one part 1,1-bis(*p*-chlorophenyl)-2-nitro-propane and two parts 1,1-bis(*p*-chlorophenyl)- 2-nitrobutane
dimethoate, *0,0*-dimethyl *S*-(*N*-methylcarbamoylmethyl) phosphorodithioate
dinitro-*o*-cresol (DNOC), 4,6-dinotro-*o*-cresol
diquat, 1,1′-ethylene-2,2′-dipyridylium dibromide
disulfoton, *0,0*-diethyl *S*-2(ethylthio)ethyl phosphorodithioate
diuron (DCMU)*, N*-3,4-dichlorophenyl-*N*′, *N*″-dimethyl urea
endosulfan, 6,7,8,9,10,10-hexachloro-1,5,5a,6,9,9a-hexahydro-6,9-methano-2,4,3-benzodioxathiepin 3-oxide
Endothall, 7-oxabicyclo (2,2,1)heptane-2,3-discarboxylic acid
endrin, 1,2,3,4,10,10-hexachloro-*exo*-6,7-epoxy-1,4,4a,5,6,7,8,8a-octahydro-1,4,5,8-*endo, endo*-dimethanonaphthalene
EPTC, *S*- ethyl,*N,N*- dipropylthiocarbamate
fenitrothion, *0,0*-dimethyl *0*-(4-nitro-*m*-tolyl) phosphorothionate
fenthion, *0,0*-dimethyl *0*[4-(methylthio)-*m*-tolyl] phosphorothionate
ferbam, ferric dimethyl dithiocarbamate
fenvalerate, cyano-(3-phenoxyphenyl-methyl-4-chloro-*α* (1-methylethyl)-benzeneacetate
flucythrinate, cyano-(3-phenoxyphenyl)-methyl-4difluoromethoxy)-*α*-(1-methylethyl)-benzeneacetate
heptachlor, 1,4,5,6,7,8,8-heptachloro-3a,4,5,5a-tetrahydro-4,7-*endo*-methanoindene
heptachlor expoxide, 1,4,5,6,7,8,8-heptachloro-3a,4,5,5a-tetrahydro-2,3-epoxy-4,7-*endo*- methanoindene
lindane, *gamma*-isomer of 1,2,3,4,5,6-hexachlorocyclohexane
linuron, 3-(3,4- dichlorophenyl)-1-methoxy-1-methylurea
malathion, *0,0*-dimethyl *S*-(1,2-dicarbethoxyethyl) phosphorodithioate
maneb, manganese ethylene bis(dithiocarbamate)
MCPA, 4-chloro-2-methylphenoxyacetic acid
methomyl, S-methyl-*N*-[(methylcarbamoyl)-oxy]-thioacetimidate
methoxychlor, 1,1,1-trichloro-2,2-bis(*p*-methoxyphenyl) ethane
methyl parathion, *0,0-dimethyl 0-p*-notrophenyl phosphorothionate
mevinphos, 2-methoxycarbonyl-1-methylvinyl dimethyl-phosphate
mirex, dodecachlorooctahydro-1,2,3-metheno-2H-cyclo-buta (cd) pentalene
monuron, 3- (*p*-chlorophenyl)-1,1-dimethylurea
naled, 1,2-dibromo-*2,2*-dichloroethyl dimethylphosphate
nicotine, 1–1-methyl-2-(3*α*-pyridyl)-pyrrolidine
Panogen, methylmercury dicyandiamide
Paraquat, 1,1′-dimethyl-4,4′-bipyridynium ion

parathion, *0,0*-diethyl *0-p*-nitrophenyl phosphorothionate
pendimethalin, *N*-(1-ethylpropyl)-3,4-dimethyl-2,6-dimitrobenzeamine
permethrin, (3-phenoxyphenyl)-methyl-3-(2,2-dichloroethenyl)-2,2-dimethyl-cyclopropane carboxylate
o-phenylphenol
phorate, *0,0*-diethyl *S*-(ethylthio)-methyl phosphorodithioate
Phosphamidon, dimethyl diethylamido-1-chlorocrontonyl-(2) phosphate
picloram, 4-amino-3,5,6-trichloropicolinic acid
propachlor, 2-chloro-*N*-isopropyl-*N*-phenylacetamide
propham, isopropyl carbanilate
propoxur, *o*-isopropoxyphenyl *N*-methylcarbamate
pyrethrins, insecticidal compounds in pyrethrum flowers, *Chrysanthemum cineraniaefolium*
ronnel, *0,0*-dimethyl *0*-2,4,5-trichlorophenyl phosphorothionate
rotenone, extracts of *Derris* or *Lonchocarpus* plants
ryania, roots and stems of *Ryania speciosa*
silvex, 2-(2,4,5-trichlorophenoxy) propionic acid
simazine, 2-chloro-4,6-bis-(ethylamino)-3-triazine
2,4,5-T, 2,4,5-trichlorophenoxyacetic acid
temephos, *0,0,0′,0′*-tetramethyl *0,0′*-thiodi-*p*-phenyclene phosphorothionate
terbufos, *0,0*-diethyl *S*- 2-(*tert*-butylthio)-methyl phosphorodithionate
tetradifon, *S-p*-chlorophenyl 2,4,5-trichlorophenyl sulfone
tetraethyl pyrophosphate (TEPP)
thiram, tetramethylthiuram disulfide
toxaphene, mixture of isomers of octachlorocamphence (67–69% chlorine)
trichlorfon, dimethyl (2,2,2-trichloro-1-hydroxyethyl) phosphonate
trifluralin, *α,α,α*-trifluoro-2,6-dinitro-*N,N*-dipropyl-*p*-toluidine
warfarin, 3-(1′-phenyl-2′-acetylethy)-4 hydroxycoumarin
Zectran, 4-dimethylamino-3,5-xylyl *N*-methylcarbamate
zineb, zinc ethylene-1,2-bisdithiocarbamate

REFERENCES

Akesson, N.B. and W.E. Yates (1964) *Ann. Rev. Entomol.,* **9**, 285.
Anderson, L.D. and E.L. Atkins (1968) *Ann. Rev. Entomol.,* **13**, 213.
Andrilenas, P. (1974) USDA Agr. Econ. Rept. 252.
Barker, R.J. (1958) *J. Wildlife Manag.,* **22**, 269.
Barthel, W.F., J.C. Hawthorne, J.H. Ford, G.C. Bolton, L.L. McDowell, E.H. Cressinger and D.A. Parsons (1969) *Pesticide Monitoring J.,* **3**, 8.
Borg, K., H. Wanntorp, K. Erne and E. Hanko (1966) *J. App. Ecol.,* **3**, Supp., 171.
Borg, K., H. Wanntorp, K. Erne and E. Hanko (1968) *Viltrevy,* **5**, (5)1.
Breidenback, A.W., C.G. Gunnerson, F.K. Kawahara, J.J. Lichtenberg and R.S. Green (1967) *Publ. Hlth. Repts.,* **82**, 139.
Brown, A.W.A. (1978) Ecology of Pesticides, Wiley, N.Y.
Brown, E. and Y.A. Nishioka (1967) *Pesticide Monitoring J.,* **1**, 38.
Büchel, K. ed. (1983) Chemistry of Pesticides, Wiley, N.Y.
Burdick, G.E., E.J. Harris, J.H. Dean, T.M. Walker, J. Skea and D. Colby (1964) *Trans. Am. Fish. Soc.,* **93**, 127.
Butler, P.A. (1966) *J. Appl. Ecol.,* **3**, Supp., 253.
Butler, P.A. (1969) The significance of DDT residues in estuarine fauna. *Chemical Fallout,* Ed. by Miller, M.W. and G.G. Berg, C.C. Thomas, Springfield, Illinois.
Chemical Week (1985) May 4, p. 22
Cohen, J.M. and C. Pinkerton (1966) Organic pesticides application of carbaryl rate to cotton in the environment, *Adv. Chem. Ser.,* 163.
Cope, O.B. (1961) *Trans. Am. Fish. Soc.,* **90**, 249.
Cope, O.B. (1966) *J. Appl. Ecol.,* **3**, Supp., 33.
Cope, O.B. (1971) *Ann. Rev. Entomol.,* **16**, 325.
Copplestone, J. (1977) p.147 in D. Watson and A. Brown eds. Pesticide resistance and insecticide resistance, Academic Press, N.Y.
Council Environment Quality (1984) Rept. 15; (1986) Rept. 17.
Duggan, R.E. (1969) *Ann. N.Y. Acad. Sci.,* **160**, 173.
Durham, W.F. (1969) *Ann. N.Y. Scad. Sci.,* **160**, 183.
Dustman, E.H. and L.F. Stickel (1966) Pesticide residues in the ecosystem. *Pesticides and Their Effects on Soils and Water,* Amer. Soc. Aronomy, Spec. Publ. 8.
Dustman, E.H. and L.F. Stickel (1969) *Ann. N.Y. Acad. Sci.,* **160**, 162.
Edwards, C.A. (1964) *Soils and Fertility,* **27**, 451.
Eichers, T., P. Andrilenas, and T. Anderson (1978) USDA Agr. Econ. Rept. 418.
Environmental Protection Agency (1970–1971) Summary of Registered Agricultural Pesticide Chemical Uses, Washington, D.C.
Frank, R., G. Sirous, and B. Ripley (1979) *Pestic. Monit. J.,* **13**, 120.
Gentry, J.W. (1966) Nat. Acad. Sci., Nat. Res. Coun., Publ. 1402, 303.
Gunther, F.A. and R.C. Blinn (1965) *Analysis of Insecticides and Acaricides,* Interscience Publishers, N.Y.
Hamelink, J.L., R.C. Waybrant and R.C. Ball (1971) *Trans. Am. Fish. Soc.,* **100**, 207.
Harris, C.R. (1971) Pesticides in soil, Proc. Internat. Symposium, Michigan State Univ. p. 58.
Harris, C.R. and E.R. Lichtenstein (1961) *J. Econ. Entomol.,* **54**, 1038.
Harris, C.R., W.W. Sans and J.R.W. Miles (1966) *J. Agr. Food Chem.,* **14**, 398.
Harrison, H.L., O.L. Loucks, J.W. Mitchell, D.F. Parkhurst, C.R. Tracy, D.G. Watts and V.J. Yannocone, Jr. (1970) *Science,* **170**, 503.
Hayes, W.J. Jr. (1963) Clinical Handbook Economic Poisons, U.S. Public Health Service, Communicable Disease Center, Atlanta, Ga.
Hayes, W.J. Jr. (1966) Monitoring food and people for pesticide content, Scientific Aspects of Pest Control, Nat. Acad. Sci., Publ. 1402.
Hickey, J.J. and L.B. Hunt (1960) *J. Wildlife Manag.,* **24**, 259.
Hunt, E.G. (1966) *J. Appl. Ecol.,* **3**, Suppl., 113.
Hunt, G. and A.I. Bishoff (1960) *Calif. Fish Game,* **46**, 91.
Jegier, Z. (1969), *Ann. N.Y. Acad. Sci.,* **160**, 143.
Johnson, D.W. (1968) *Trans. Am. Fish. Soc.,* **97**, 398.
Jones, J.S. and M.B. Hatch (1937) *Soil Sci.,* **44**, 37.
Kearney, P.C., R.G. Nash and A.R. Isensee (1969) Persistence of pesticide residues in soils, *Chemical Fallout,* Ed. by Miller, M.W. and G.G. Berg. C.C. Thomas, Springfield, Illinois.
Korschgen, L.F. (1970) *J. Wildlife Manag.,* **34**, 186.
Kraybill, H.F., Ed. (1969) Biological effects of pesticides in mammalian systems, *Ann. N.Y. Acad. Sci.,* **160**, Art. 1.
Lichtenstein, E.P. (1969) Persistence and degradation of pesticides in the environmental, *Scientific Aspects of Pest Control,* Nat. Acad. Sci., Publ. 1402.
Lichtenstein, E.P., G.R. Myrdal and K.R. Schultz (1965) *J. Agr. Food Chem.,* **3**, 126.
McEwen, F.L. and G.R. Stevenson (1979) The Use and Significance of Pesticides in the Environment, Plenum Press, N.Y.
Melnikov, N.N. (1971) *Chemistry of Pesticides,* Springer-Verlag, N.Y. Metcalf, R.L. (1972) *J. Environ. Qual.,* 1, 10.
Moore, N.W. (1965) Environmental contamination by pesticides, *Ecology and the Industrial Society,* Ed. by Goodman, G.T., R.W. Edwards and J.M. Lambert. John Wiley and Sons, N.Y.
Mrak, E.M. (1969) Report of the Secretary's Commission on Pesticides and Their Relationship to Environmental Health, U.S. Dept. Health Educ. Welfare, Washington, D.C.
National Academy Sciences (1987) Nat. Acad. Press ISBN-309.
Newsom, L.D. (1967) *Ann. Rev. Entomol.,* **12**, 257.
Pimentel, D. (1971) Ecological Effects of Pesticides on Non-Target Species, President's Office Science Technology, Washington, D.C.
Pimentel, D. and L. Levitan (1986) *Bioscience* **36**, 86.
President's Science Advisory Committee (PSAC) (1965) Restoring the Quality of our Environment, The White House, Washington, D.C.
Quinby, G.E. and G.M. Doornick (1965), *J. Amer. Med. Assoc.,* **191**, (1)95.
Rainey, R.H. (1967) *Science,* **155**, 1242.
Ratcliffe, D.A. (1970) *Nature,* **215**, 208.
Reinert, R.E. (1970) *Pesticide Monitoring J.,* **3**, 233.
Reynolds, H.T., T.R. Fukuto, R.L. Metcalf and R.B. March (1957) *J. Econ. Entomol.,* **56**, 527.
Rice, C. and M. Evans (1984) *Adv. Environ. Sci. Technol.,* **14**, 163.
Risebrough, R.W., R.I. Huggett, E.D. Griffin and E.D. Goldbert (1968) *Science* **159**, 1233.
Rudd, L.L. and R.E. Genelly (1966) *Calif. Fish Game Bull,* **7**, 309.
Rudd, R.L., and R.E. Genelly (1956) *Auk,* **73**, 529.
Stanley, C.W. (1968) Study to Determine Atmospheric Contamination by Pesticides, Final Rept. Midwest Research Inst.
Storck, W. (1984) *Chem. Eng. News* April 9, p. 35.
Storck W. (1987) *Chem. Eng. News* Nov. 16, p. 35.
Tabor, E.G. (1964) *Trans. N.Y. Acad. Sci.,* **28**, 569.

U.S. Dept. Interior (1968) Water Quality Criteria, Reports of National Technical Advisory Committee to Secretary of Interior, Fed. Water Pollution Control Administration, Washington, D.C.

Warner, R.E. (1967) *Bull World Health Org.,* **36**, 181.

Wasserman, M.I.S., G. Mandric and P. Horvath (1960) Amer. Med. Assoc., *Arch. Ind. Health,* **21**, 503.

Westöö, G., Methylmercury compounds in animal food (1969) Chemical Fallout, Ed. by Miller, M.W. and G.G. Berg, p. 75. C.C. Thomas, Springfield, Illinois.

White-Stevens, R. Ed. (1971) *Pesticides in the Environment,* **1**, Pt. 1, Marcel Dekker, Inc., N.Y.

Wolman, A.A. and A.J. Wilson (1971) *Pesticides Monitoring J.,* **4.**

Woodwell, G.M., C.F. Wurster and P.A. Isaacson (1967) *Science,* **156**, 821.

Woodwell, G.M., P.P. Craig and H.A. Johnson (1967) *Science,* **156**, 821.

Woodwell, G.M., P.P. Craig and H.A. Johnson (1971) *Science,* **174**, 1101.

World Health (1968) p. 10, April.

ROBERT L. METCALF
University of Illinois, Urbana

PHYSICAL AND CHEMICAL TREATMENT OF WASTEWATERS

INTRODUCTION

The substances in domestic and industrial wastewater having significance in water-pollution control, disposal, and reuse are (1) dissolved decomposable organic substances resulting in dissolved oxygen depletion in streams and estuaries and/or causing taste and odor; (2) suspended organic solids resulting in dissolved oxygen depletion; (3) inert suspended solids (SS) causing turbidity and resulting in bottom sediment deposits; (4) toxic synthetic organic substances and heavy metals; (5) oil, grease, and floating materials; (6) acids and alkalis; and (7) dissolved salts, including nutrients like phosphorus and nitrogen.

Conventional wastewater-treatment practices have been oriented to the removal of grit and floating matter followed by the removal of suspended and dissolved organic matter. The removal of suspended matter has been achieved by sedimentation, and the bulk of the soluble organic matter is removed by biological oxidation and flocculation. These processes, when carried out in combination, have proved to be economical and effective means for removing organic matter from wastewaters. However, there are certain disadvantages associated with them. These include the following:

1. Biological process require considerable operating control and often generate operating problems of a complex nature.
2. Biological processes are easily upset by shock loads and require time to regain efficient operation.
3. Biological processes are unable to remove certain nutrients, heavy metals, and inorganic salts, whenever there is a requirement for their removal.
4. Many waste streams contain certain compounds that do not respond to biological treatment or, alternatively, require extensive pretreatment.

In the last decades, physical-chemical treatment of wastewater has been studied both on laboratory and pilot-plant scales with important industrial and municipal wastewater-treatment applications. This type of treatment is used either as a pretreatment, tertiary treatment, or advanced treatment given to the effluent from secondary treatment, or as a substitute for conventional biological treatment. In the latter case, it is found to produce effluent of a quality at least equal to that produced by conventional biological treatment.

The first study on the treatability of raw wastewater by physical-chemical processes was reported by Rudolfs and Trubnick in 1935. In this study, solids were removed by chemical coagulation with ferric chloride followed by absorption of dissolved impurities with activated carbon.

Stander and Van Vuuren (1969) investigated the treatment of raw wastewater in a pilot plant where solids removal was achieved by primary sedimentation and chemical coagulation with lime, and adsorption with activated carbon. Rizzo and Schade (1969) have also reported results on the pilot-plant treatment of raw wastewater with chemical coagulation and anionic polymer and adsorption with activated carbon.

Zuckerman and Molof (1970) studied the efficiency of a treatment system in which raw wastewater was lime-clarified at high pH and then activated-carbon-treated: their results showed that the chemical oxygen demand (COD) values of the final effluent were significantly lower than those associated with good conventional treatment. Moreover, they concluded that the removal of soluble organics with activated carbon was enhanced because of the hydrolytic breakdown of high-molecular-weight organic compounds, at a higher pH value, which are absorbed more readily by activated carbon.

Weber et al. (1970) investigated the chemical clarification of primary effluent with ferric chloride followed by activated-carbon adsorption. Their results showed that 65% of the organic matter present in primary effluent was removed by chemical treatment with ferric chloride. Overall removal of biochemical oxygen demand (BOD) was reported as being consistently in the range of 95 to 97%. Final effluent from the system contained approximately 5 mg/l BOD as compared to 30 mg/l for the same wastewater treated conventionally.

In another study, Villiers et al. (1971) showed that the treatment of primary effluent by lime clarification and activated carbon, in a steady flow system, produced an effluent with total organic carbon (TOC) averaging 11 mg/l and turbidity averaging less than 2. Phosphates and SS removal were consistently 90% or better. These product characteristics are comparable to those associated with products from well-operated conventional treatment plants.

Shuckrow (1971) had developed a sewage-treatment process involving chemical coagulation for SS removal, followed by adsorption of soluble organics on powdered carbon. The advantages cited for this process were (1) a total treatment time of less than one hour, (2) a high-quality effluent,

(3) lower initial plant cost, (4) ability to remove nitrogen and phosphorus, and (5) a final sludge reduction to sterile ash in a centrifuge-incineration process combined with a chemical regeneration step to recover both the coagulant and the carbon. While the estimated operational costs were high, the overall costs during a 20-year plant life were considered to be significantly less than costs for comparable biological facilities.

Ecodyne Corporation's first complete physical-chemical treatment plant in Rosemount, Minnesota, with 0.04 to 0.08 m^3/sec peak capacity, consisted of bar screening, phosphate removal with sludge recirculation, dual media filtration, carbon absorption to remove dissolved organics, secondary filtration, and ammonia removal by ion exchange with zeolite. The plant included facilities for regenerating the carbon, recovering ammonia, and regenerating brine from the ion-exchange system (Ecodyne, 1972).

Examples of more recent research include an exhaustive review on the treatment of pulp- and paper-mill wastewater published by Pokhrel and Viraraghavan in 2004. This includes the different processes involved with their effluents, the different methods for treatment of these effluents, the integration of biological and physico-chemical processes, a comparison of them, and conclusions from this review. An article by Van Hulle and Vanrolleghem (2004) presents the development, calibration, and application of a model for the simulation and optimization of a wastewater-treatment plant. The constructed model proved to be able to predict large variations in influent composition. This could be an important tool for production scheduling when applied to industrial wastewater-treatment plants. Recent research in specific areas is included in for each process.

The number of water-treatment facilities in the United States by treatment capacities is presented in Table 1. Many of these facilities include some sort of physical-chemical treatment technology. A diagram of alternative technologies for wastewater treatment is shown in Figure 1; it includes most of the processes to be discussed in the next section. As environmental regulations, space availability, and cost factors affect the treatment of waste streams, more and more physical-chemical treatment will be needed to meet these constraints. This is an important research area that will continue to grow in the next years.

TABLE 1

Number of wastewater-treatment facilities in the United States (1996)

Flow ranges m^3/s	Number of facilities	Total existing flowrate m^3/s
0–0.00438	6444	12.57
0.0044–0.0438	6476	101.78
0.044–0.438	2573	340.87
0.44–4.38	446	511.12
>4.38	47	443.34
Other	38	—
Total	16204	1409.68

Source: Adapted from Tchobanoglous et al., 2003.

PHYSICAL AND CHEMICAL PROCESSES USED IN WASTEWATER TREATMENT

The following important unit operations and unit processes involved in the physical and chemical treatment of wastewater are discussed in detail:

Flow equalization and neutralization
Chemical coagulation, flocculation, and sedimentation
Filtration
Gas stripping
Ion exchange
Adsorption
Flotation
Chemical processes
Oxidative, photochemical, and electron-beam processes

Flow Equalization and Neutralization

Both domestic and industrial wastewater flows show considerable diurnal variation, and it is considered necessary to significantly dampen these variations in inflow to relieve hydraulic overload on both biological and physical-chemical plants. This process will also smooth out variations in influent characteristics.

Flow-equalization basins are basically flow-through or side-line holding tanks, and their capacity is determined by plotting inflow and outflow mass curves. These tanks are generally located after preliminary treatment and should be designed as completely mixed basins, using either diffused air or mechanical surface aerators, to prevent settling of suspended impurities. If decomposable organic matter is present in the wastewater, aeration will prevent septicity. The preaeration can also reduce the BOD on subsequent treatment units.

Flow-equalization basins can also be used to neutralize the acidity or alkalinity in incoming wastewater. The neutralizing chemicals are added to the inflow wastewater stream before entering the flow-equalization basins, and the retention period in these basins provides sufficient time for reaction. Any precipitates produced during neutralization are separated in subsequent sedimentation basins.

Chemical Coagulation, Flocculation, and Sedimentation

The use of chemical treatment appeared early in the development of sewage- and wastewater-treatment technology. Aluminum sulfate, lime, and ferrous sulfate, when used in the manner usually adopted for water clarification, were successful in producing an effluent of quality better than that obtained by plain sedimentation. An effluent that is generally fairly clear, with only very fine suspended or colloidal solids but with practically all of the dissolved solids remaining, can be produced. Under most favorable conditions and with skilled operation, SS may be reduced in an amount of up to 90% and BOD up to 85%. However, the

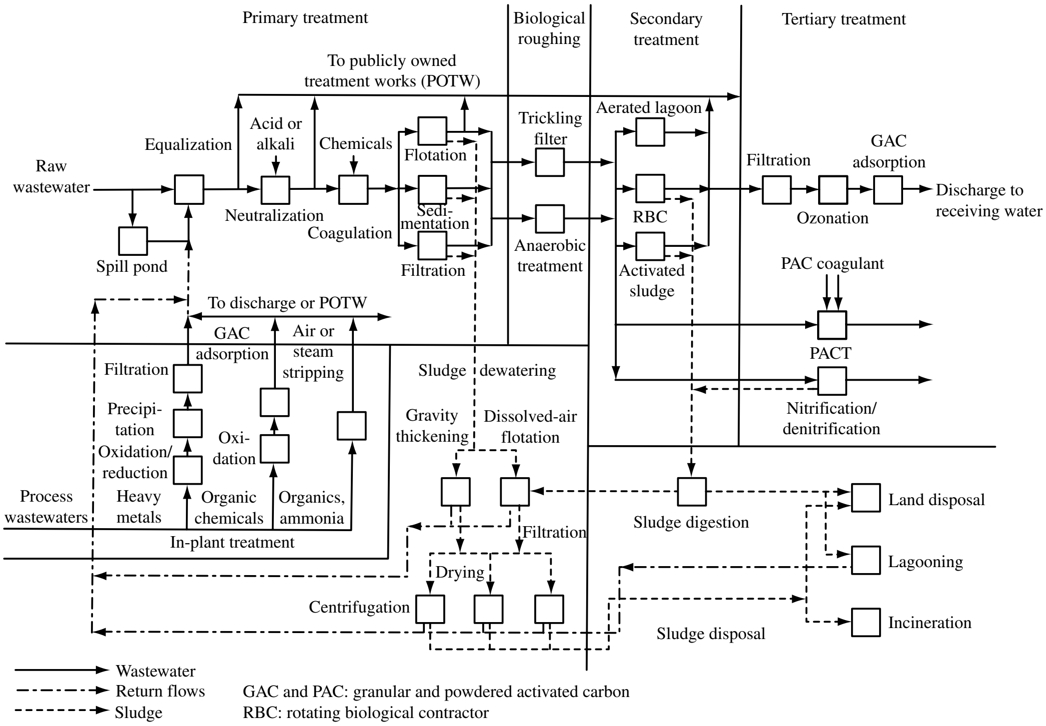


FIGURE 1 Alternative technologies for wastewater treatment (From Eckenfelder, 2000. Reprinted with permission from McGraw-Hill.)

cost of the coagulants and the difficulty of disposing of the larger amount of the sludge produced by this process caused it to be abandoned. The revival of chemical treatment can be attributed to a number of factors that accumulated as a result of continuous investigations and reevaluation of the process. These are (1) the decrease in cost of chemicals; (2) better understanding of floc formation and the factors affecting it; (3) the development of methods of sludge filtration and processing that overcome, in part, the difficulty of greater sludge bulk; and (4) the establishment of the relationship between eutrophication in streams and nutrients, particularly phosphorus, nitrogen oxides, and organic matter. This relation establishes the need of final effluent wastes free of such pollutants regardless of the cost of additional treatment.

The settling velocities of finely divided and colloidal particles in wastewaters are so small that removing them in a settling tank under ordinary conditions is impossible unless very long detention periods are provided. Therefore, it has been necessary to devise means to coagulate these very small particles into larger ones that will have higher settling velocities. The aggregation of dispersed particles in wastewater is induced by addition of chemical coagulants to decrease the effects of stabilizing factors such as hydration and zeta potential, and by agitation of the medium to encourage collisions between particles.

Because of the greater amount of suspended matter in sewage, the doses for chemical coagulants are generally considerably greater. Therefore, in order to keep costs down, it is important that the chemical reaction involved with each coagulant should be known and enhanced and that optimum pH values be obtained by adjustment with acid or base to get more efficient coagulation and clarification with least sludge production.

Coagulation Coagulation is a process in which chemicals are added to an aqueous system for the purpose of creating rapid-settling aggregates out of finely divided, dispersed matter with slow or negligible setting velocities. The potential applications of this process in treating wastewater are: (1) direct coagulation of organic matter present mostly as colloidal particles in wastewater; (2) the removal of colloidal substances prior to such tertiary treatment processes as ion exchange, carbon adsorption, and sand filtration; (3) the removal of colloidal precipitates formed in phosphate precipitation processes; and (4) the removal of dispersed microorganisms after a brief biooxidation process.

The majority of colloids in domestic wastewater or in organic wastes are of a hydrophilic nature; that is, they have an affinity for water. The affinity of hydrophilic particles for water results from the presence of certain polar groups such as $-COOH$ and $-NH_2$ on the surface of the particles. These groups are water-soluble and, as such, attract and hold a sheath of water firmly around the particle. The primary charge on hydrophilic colloidal particles may arise from ionization of the chemical groups present at the surface of the particles, e.g., carboxyl, amino, sulfate, and hydroxyl. This charge is dependent upon the extent to which these surface groups ionize, and thus the particle charge depends upon the pH.

The precise zeta potential that yields optimum coagulation must be determined for a given wastewater by actual correlation with jar test or plant performance. The control point has been reported to be in the range of 0 to −10 mV when raw sewage is coagulated by alum. It is important that coagulants contribute polyvalent ions of charge opposite to the zeta potential of the dispersion. On a molar basis, bivalent ions seem to be about 10 to 50 times and trivalent ions about 300 to 700 times as effective as monovalent ions for destabilization of dispersion in wastewater (Rich, 1963). The zeta potential is unaffected by pH in the range of 5.5 to 9.5 (Eckenfelder, 2000).

Since most dispersions encountered in wastewaters are stabilized by negative charges, coagulants required are polyvalent cations such as aluminum, ferric, ferrous, or calcium. Organic polyelectrolytes are also effective coagulants. Dispersions stabilized principally by electrostatic force are in general amenable to coagulation inasmuch as addition of small doses of suitable electrolytes may effect a significant change in zeta potential of the particles. The most widely used chemicals for coagulation of wastewater are the salts of aluminum and iron. Lime alone has also been used for precipitation of phosphates.

Aluminum Sulfate In order to form flocs, aluminum sulfate requires the presence of alkalinity, which, if naturally present in wastewater in the form of bicarbonate, would lead to the following reaction:

$$Al_2(SO_4)_3.xH_2O + 3Ca(HCO_3)_2 \rightarrow \\ 2Al(OH)_3 \downarrow + 3CaSO_4 + xH_2O + 6CO_2 \qquad (1)$$

In case of insufficient alkalinity in the wastewater, lime is generally added, and the reaction with alum becomes:

$$Al_2(SO_4)_3.xH_2O + 3Ca(OH)_2 \rightarrow \\ 2Al(OH)_3 \downarrow + 3CaSO_4 + xH_2O \qquad (2)$$

In the presence of phosphate, the following reaction also occurs:

$$Al_2(SO_4)_3.xH_2O + 2PO_4^{-2} \rightarrow \\ 2AlPO_4 \downarrow + 3SO_4^{-2} + xH_2O \qquad (3)$$

Aluminum hydroxide flocs are least soluble at a pH of approximately 7.0. The floc charge is positive below pH 7.6 and negative above pH 8.2 (Eckenfelder, 1966). The solubility of $AlPO_4$ is related to the pH and the equilibrium constant for the salt. Stumm and Morgan (1970) state that the solubility of aluminum phosphate is pH-dependent, and the optimum pH for phosphorus removal lies in the range of 5.5 to 6.5. Generally, at pH above 6.3, the phosphate removal occurs either by incorporation in a complex with aluminum or adsorption on the aluminum hydroxide flocs. According to Yuan and Hsu (1970), the reaction mechanism for precipitation of phosphates by aluminum hydroxide is very complex. They have proposed that the positively charged hydroxy-aluminum polymers are the species that accounts for the precipitation of phosphates and that effective phosphate precipitation can occur only when the positive charges on the polymers are completely neutralized. It is also reported that the effectiveness of aluminum is related to the nature and concentration of the foreign components present and to the ratio of phosphate to aluminum.

Alum has been used extensively for phosphate removal in raw wastewaters. Bench-scale tests of alum addition were conducted at Springfield, Ohio, and Two Rivers, Wisconsin (Harriger and Hoffman, 1971 and 1970, respectively). Raw wastewater at Springfield required an average Al:P mass ratio of 1.9:1 to achieve 80% removal, while at Two Rivers the average mass ratio was 0.93:1 to obtain phosphate removal of 85%. The stoichiometric equation (3) indicates that each kilogram of phosphorus requires 0.87 kg of aluminum for complete precipitation.

Ferrous Sulfate and Lime If ferrous salts are used for wastewater coagulation, addition of a small amount of base, usually sodium hydroxide or lime, is essential. The required dosage is related to the alkalinity of water. Ferrous sulfate reacts with calcium bicarbonate in water, but this reaction is much delayed and therefore cannot be relied on (Steel, 1960). Caustic alkalinity, due to the addition of lime to the wastewater, produces a speedy reaction. The lime is added first, and the following reaction takes place:

$$FeSO_4.7H_2O + Ca(OH)_2 \rightarrow Fe(OH)_2 + CaSO_4 + 7H_2O \qquad (4)$$

The ferrous hydroxide is not an efficient floc, but it can soon be oxidized by the dissolved oxygen in wastewater as ferric hydroxide:

$$4Fe(OH)_2 + O_2 + 2H_2O \rightarrow 4Fe(OH)_3 \downarrow \qquad (5)$$

An insoluble hydrous ferric oxide is produced over a pH range of 3 to 13. The floc charge is positive in the acid range and negative in the alkaline range, with a mixed charge over a pH range of 6.5 to 8.0. This process is usually cheaper than the use of alum but needs greater skill to dose with the two chemicals.

Ferric Chloride Ferric chloride has been used successfully for wastewater coagulation because it works well in a wide pH range (Steel, 1960; Wuhrmann, 1968). The reactions of ferric chloride with bicarbonate alkalinity and lime are, respectively:

$$2FeCl_3 + 3Ca(HCO_3)_2 \rightarrow 2Fe(OH)_3 \downarrow + 3CaCl_2 \\ + 6CO_2 \qquad (6)$$

$$2FeCl_3 + 3Ca(OH)_2 \rightarrow 2Fe(OH)_3 \downarrow + 3CaCl_2 \qquad (7)$$

Wuhrmann (1968) was successful in removing phosphates from sewage effluent by precipitation with a mixture of ferric salt and lime. The ferric dosages varied between 10 and 20 mg/l and the lime dosages from 300 to 350 mg/l in order to raise the pH to values between 8 and 8.3. The actual lime dosage required is related to the alkalinity of the water. According to Wuhrmann, the dominant reaction product between the phosphate ion and the ferric ion at pH above 7 is believed to be $FePO_4$, with a solubility product of about 10^{23} at 25°C. The colloidal particle size of the $FePO_4$ requires a sufficient excess of ferric ion for the formation of a well-flocculating hydroxide precipitate, which includes the $FePO_4$ particles and acts as an efficient adsorbent for other phosphorous compounds.

It has been reported that for efficient phosphorus removal (85 to 95%), the stoichiometric amount of 1.8 mg/l Fe required per mg/l P should be supplemented by at least 10 mg/l of iron for hydroxide formation. Also, the use of anionic polymer is considered desirable in order to produce a clear supernatant (Wukash, 1968).

Lime Lime reacts with the bicarbonate alkalinity of wastewater to form calcium carbonate, which precipitates, under normal conditions:

$$Ca(OH)_2 + Ca(HCO_3)_2 \rightarrow 2CaCO_3 \downarrow + 2H_2O \qquad (8)$$

Normally, 70 to 90% of the phosphorus in domestic sewage is in the form of orthophosphates or polyphosphates that may hydrolyze orthophosphates. The remaining phosphorus is present in the form of organic-bound phosphorus. The removal of phosphorus can be achieved by direct adsorption on the surface of calcium carbonate particles. Orthophosphates can also be precipitated in the alkaline range by reaction with calcium salts to form hydroxyapatite, according to the following reaction:

$$10Ca(OH)_2 + 6H_3PO_4 \rightarrow 10Ca + (PO_4)_6(OH)_2 \downarrow + 18H_2O \qquad (9)$$

Schmid and McKinney (1969) observed that hydroxyapatite was present in soluble form at a pH value above 9.5. They also found that at pH values of 9.5 or less, phosphorus was adsorbed onto the growing faces of calcium-carbonate particles, thereby inhibiting their growth. Buzzell and Sawyer (1967) have shown that at pH levels of 10 to 11 in the primary sedimentation tanks, BOD removal of 55 to 70%, nitrogen removal of 25%, phosphate removal of 80 to 90%, and coliform removal of 99% can be expected. Bishop et al. (1972) have reported that precipitation of domestic wastewater with lime removed approximately 80% of the TOC, BOD, and COD; 91% of the SS; 97% of the total phosphorus; and 31% of the total nitrogen. Phosphates from secondary effluent have been removed successfully at Lake Tahoe by precipitation with lime (Slechta and Culp, 1967). Albertson and Sherwood (1967) found that by recirculating calcium-phosphate solids, previously formed due to the addition of lime, it was possible to reduce the lime dosage by about 50%.

Galarneau and Gehr (1997) present experimental results of their studies on phosphorous using aluminum hydroxide. de-Bashan and Bashan (2004) present an extensive review of recent advances in phosphorous removal from wastewaters and its separation for use as a fertilizer or as an ingredient in other products.

In wastewater-treatment practices, it is detrimental to form large floc particles immediately in the flocculation step because it reduces the available floc surface area for adsorption of phosphorus. Therefore, it is essential to maintain fine pinpoint flocs in order to get a maximum phosphate removal by surface adsorption, and this can be achieved by minimizing the time of their flocculation. This is not the case if the goal is one of colloidal-solids removal, as is often the case in water treatment.

The process of coagulation and flocculation in wastewater treatment can be summarized in the following three steps:

1. As the coagulant dissolves, positive aluminum and ferric ions become available to neutralize the negative charges on the colloidal particles including organic matter. These ions may also react with constituents in solution such as hydroxides, carbonates, phosphates, sulfides, or organic matter to form complex gelatinous precipitates of colloidal dimensions that are termed "microflocs." This is the first stage of coagulation, and for greatest efficiency a rapid and intimate mixing is necessary before a second reaction takes place.
2. After the positively charged ions have neutralized a large part of the colloidal particles and the zeta potential has been reduced, the resulting flocs are still too small to be seen or to settle by gravity. The treatment, therefore, should be flocculation, slow stirring so that very small flocs may agglomerate and grow in size until they are in proper condition for sedimentation. Some evidence suggests that aggregation of microflocs with dispersed waste constituents is the most important mechanism affecting coagulation in water treatment (Riddick, 1961).
3. During the third phase, surface adsorption of particles takes place on the large surface area provided by the floc particles. Some of the bacteria present will also become entangled in the floc and carried to the bottom of the tank.

Electrocoagulation Electrocoagulation is a process in which the coagulating ions are produced by electrolytic oxidation of sacrificial electrodes. This technique has been successfully used in the removal of metals, suspended particles, colloids, organic dyes, and oils. An interesting review of this technique is presented by Mollah et al. (2001). In it the advantages and disadvantages of electrocoagulation are presented as well as a description and comparison with chemical coagulation. His group studied its use in the treatment of a synthetic-dye solution with a removal of 99% under optimal conditions (Mollah, Morkovsky, et al., 2004). Another publication (Mollah, Pathak, et al., 2004) presents the fundamentals of electrocoagulation and the outlook for the use of this process in wastewater treatment. Lai and Lin (2003) studied the use of electrocoagulation for the treatment of chemical mechanical polishing wastewater, obtaining a 99% copper removal and 96.5% turbidity reduction in less than 100 minutes.

Sedimentation Sedimentation basins are important components in water- and wastewater-treatment systems, and their performance greatly depends upon proper design. In chemical treatment of wastewater, the separation of chemically coagulated floc depends on the characteristics of the floc in addition to the factors normally considered in the design of conventional primary and secondary clarifiers. Field experience indicates that the usual values for surface overflow rates used in separating chemical floc in water-treatment plants must be reduced in order to obtain a good efficiency in the removal of floc from chemically coagulated wastewater (Weber et al., 1970; Convery, 1968; Rose, 1968; Kalinske and Shell, 1968). In wastewater-treatment practices, the recommended overflow rate for removal of alum floc is 30 m/day, while with use of lime or iron salts, it can be increased up to 40 m/day.

Filtration

Filtration of wastewater can be accomplished by the use of (1) microscreens, (2) diatomaceous earth filters, (3) sand filters, (4) mixed media filters, or (5) membranes. The filtration of sludges, on the other hand, is achieved by sand beds or vacuum filters.

The filtration characteristics of the solids found in a biological treatment plant effluent are greatly different from those of the floc formed during chemical coagulation for the removal of organic matter and phosphates. Tchobanoglous and Eliassen (1970) have noted that the strength of the biological floc is much greater than that of the flocs resulting from chemical coagulation.

Accordingly, biological flocs can be removed with a coarser filter medium at higher filtration rates than can the weaker chemical flocs, which may shear and penetrate through the filter more readily. Lynam et al. (1969) had observed that the chemical floc strength can be controlled, to some degree, with the use of polymers as coagulant aids. Their experiments yielded higher SS removal by filtration when 1 mg/l of anionic polymer A-21 was used along with alum.

The filterability of solids in a conventional biological plant effluent is dependent upon the degree of flocculation achieved in the biological process. For example, filtration of the effluent from a trickling-filter plant normally cannot yield more than 50% removal of the SS due to the poor degree of biological flocculation in trickling filters. On the other hand, the activated sludge process is capable of a much higher degree of biological flocculation than the trickling-filter process. The degree of biological flocculation achieved in an activated sludge plant was found to be directly proportional to the aeration time and inversely proportional to the ratio of the amount of organic material added per day to the amount of SS present in the aeration chamber (Culp and Hansen, 1967a). It has also been reported that up to 98% of the SS found in the effluent from a domestic sewage-treatment plant after a 24-hour aeration time could be removed by filtration without the use of coagulants (Culp and Hansen, 1967b).

Microscreening Microscreens are mechanical filters in which flow is passed through a special metallic filter fabric placed around a drum. The filter traps the solids and rotates with the drum to bring the fabric under backwash water sprays fitted to the top of the machine, in order to wash the solids to a hopper for gravity removal to disposal. The rate of flow through the microscreen is determined by the applied head, normally limited to about 150 mm or less, and the concentration and nature of the SS in the effluent.

Extensive tests at the Chicago Sanitary District showed that microscreens with a 23 μm aperture could reduce the SS and BOD of a good-quality activated sludge effluent, 20–35 mg/l SS and 15–20 mg/l BOD, to 6–8 mg/l and 3.5–5 mg/l, respectively (Lynam et al., 1969). It was noted that the microscreens were more responsive to SS loading than to hydraulic loading and that the maximum capacity of the microscreens was reached at the loading of 4.3 kg/m^2/day at 0.27 m/min.

Diatomaceous Earth Filtration Diatomite filters found their widest application in the production of potable waters, where the raw water supply was already of a relatively good quality, i.e., of low turbidity. Operating characteristics of diatomite filters can now be predicted under a wide range of operating conditions by utilizing several mathematical models (Dillingham et al., 1966, 1967). Several investigators have studied the filtration of secondary effluent by diatomite filters whose ability to produce an excellent-quality effluent is well established (Shatto, 1960; R. Eliassen and Bennett, 1967; Baumann and Oulman, 1970). However, the extremely high cost and their inability to tolerate significant variations in SS concentration limit the usage of diatomite filters in sewage-treatment practices.

Sand Filtration Sand filters have been operated as slow sand filters or as rapid sand filters made with one or more media. A slow sand filter consists of a 150- to 400-mm-thick layer of 0.4-mm sand supported on a layer of a coarser material of approximately the same thickness. The underdrainage system under the coarser material collects the filtrate. The rate of flow through the filter is controlled at about 3 m/day. This rate is continued until the head loss through the bed becomes excessive. Then the filter is thrown out of service and allowed to partially dry, and 25 to 50 mm of the sand layer, which includes the surface layer of sludge, is manually scraped from the top for washing. Disadvantages of slow sand filtration system are: (1) the filters may become inoperative during the cold winter weather, unless properly housed; (2) slow sand filters may not be effective due to the rapid clogging of filters (the normal frequency of cleaning filters varies from once to twice a month; Truesdale and Birkbeck, 1996); (3) the cost of slow sand filtration is three times the cost of rapid sand filters and twice the cost of microscreens (anonymous, 1967); and (4) the large space requirement.

Rapid sand filters consist of about a 400-mm-thick layer of 0.5- to 0.65-mm sand supported on coarser gravel. The rate of filtration ranges between 80 and 120 mm/min. At this high filtration rate, the filter beds need backwashing when the head loss becomes excessive.

Lynam et al. (1969) reported results of detailed tests conducted on filtration of secondary effluent from an activated sludge plant of the Chicago Sanitary District. They used a filter bed of 0.85-mm-effective-size sand in a 280-mm depth and a filtration rate of 120 mm/min, and analyzed the data in terms of both hydraulic and SS loadings. Poor correlations were obtained between effluent quality and hydraulic loading, effluent quality and solids loading, and solid removal and hydraulic loading. However, an excellent correlation existed between SS loading and SS removals. It was also observed that the sand filtration of alum-coagulated solids was no better than that of uncoagulated solids, and the optimum SS removal was obtained by alum and polymer coagulation in combination with sand filtration.

A review of the retention of pathogenic bacteria in porous media is presented by Stevik et al. (2004). The review includes the factor affecting bacteria retention and the factors that effect elimination of bacteria from porous media. The authors also suggest priority areas of research in this field.

Multimedia Filtration The limitation of the single medium rapid sand filter follows from its behavior as a surface filtration device. During filter backwashing, the sand is graded hydraulically, with the finest particles rising to the top of the bed. As a result, most of the material removed by the filter is retained at or very near the surface of the bed. When the secondary effluent contains relatively high solids concentrations, the head loss increases very rapidly, and SS clog the surface in only a few minutes. One approach to increase the effective filter depth is to use dual-media beds consisting of a discrete layer of coarse coal placed above a layer of fine sand.

More recently, the concept of mixed-media filters has been introduced in order to achieve a filter performance that very closely approaches an ideal one. In this case a third layer of a very heavy and fine material, garnet (with specific gravity of 4.2) or illmenite (with specific gravity of 4.5), is placed beneath the coal and sand. Conley and Hsiung (1965) have suggested the optimum design values for these filters. The selection of media for any filtration application should be based on the floc characteristics. An example of a typical dual-media filter is shown in Figure 2.

Moving-Bed Filters These types of filters were put on the market by the Johns-Manville Corporation in the late 1960s. It is a continuous sand filter in which influent wastewater passes through the bed and becomes product water. Solids trapped on the filter face and within the bed move with the filter media, countercurrent to the liquid. Solids and small amounts of filter media regularly removed from the filter face are educted to the filter media tower without stopping operations. Solids are scrubbed from the media and discharged as a waste sludge, while the washed media is fed back into the bed.

The filter medium usually used is 0.6- to 0.8-mm sand with a maximum sand-feed rate of 5 mm/min and maximum filtration rate of 85 m/day (2100 U.S. gal/day/ft^2). The advantages claimed for this system are (1) automatic and continuous operation, (2) that the filter allows much higher and variable solids loadings than is permissible with a sand bed, and (3) that through an efficient use of coagulant chemicals, the system has the flexibility to reduce turbidity, phosphorus, SS, and BOD to the desired level (Johns-Manville Corporation, 1972).

Membrane Filtration Membrane filtration is being applied more extensively as membrane materials are becoming more resistant and affordable. Fane (1996) presents a description of membrane technology and its possible applications in water and wastewater treatment. An extensive study on microfiltration performance of membranes with constant flux for the treatment of secondary effluent was published by his research group in 2001 (Parameshwaran et al., 2001). Kentish and Stevens (2001) present a review of technologies for the recycling and reuse of valuable chemicals from wastewater, particularly from solvent-extraction processes.

A feasibility study on the use of a physico-chemical treatment that includes nanofiltration for water reuse from printing, dyeing, and finishing textile industries was performed by Bes-Pia et al. (2003). In this work jar tests were conducted for flocculation using commercial polymers followed by nanofiltration. Their results show that the combination reduces COD from 700 to 100 mg/l. Another treatment approach by the same authors (2004) uses ozonation as a pretreatment for a biological reactor with nanofiltration as a final step. A combined approach is presented by Wyffels et al. (2003). In this case a membrane-assisted bioreactor for the treatment of ammonium-rich wastewater was used, showing this to be a reliable technology for these effluents.

Galambos et al. (2004) studied the use of nanofiltration and reverse osmosis for the treatment of two different wastewaters. For their particular case the use of reverse osmosis was more convenient due to the high quality of the effluent, but the permeate of the nanofiltration can only be released into a sewer line or would have to be treated, resulting in an economic compromise. A comparison between a membrane bioreactor and hybrid conventional wastewater-treatment systems at the pilot-plant level is presented by Yoon et al. (2004).

The removal of volatile organic compounds (VOCs) using a stripper-membrane system was studied by Roizard et al. (2004). Their results show that this hybrid system can be used for the removal of toluene or chloromethane with a global efficiency of about 85%.

Vildiz et al. (2005) investigated the use of a coupled jet loop reactor and a membrane for the treatment of high-organic-matter-content wastewater. The main function of the membrane is the filtration of the effluent and the recycle of the biomass to the reactor. One advantage of the system is its reduced size as compared with traditional treatment systems, as well as a better-quality effluent.

A comprehensive review on the use of nanofiltration membranes in water and wastewater use, fouling of these membranes, mechanisms of separation, modeling, and the use of atomic force microscopy for the study of surface morphology is presented by Hilal et al. (2004). The future of

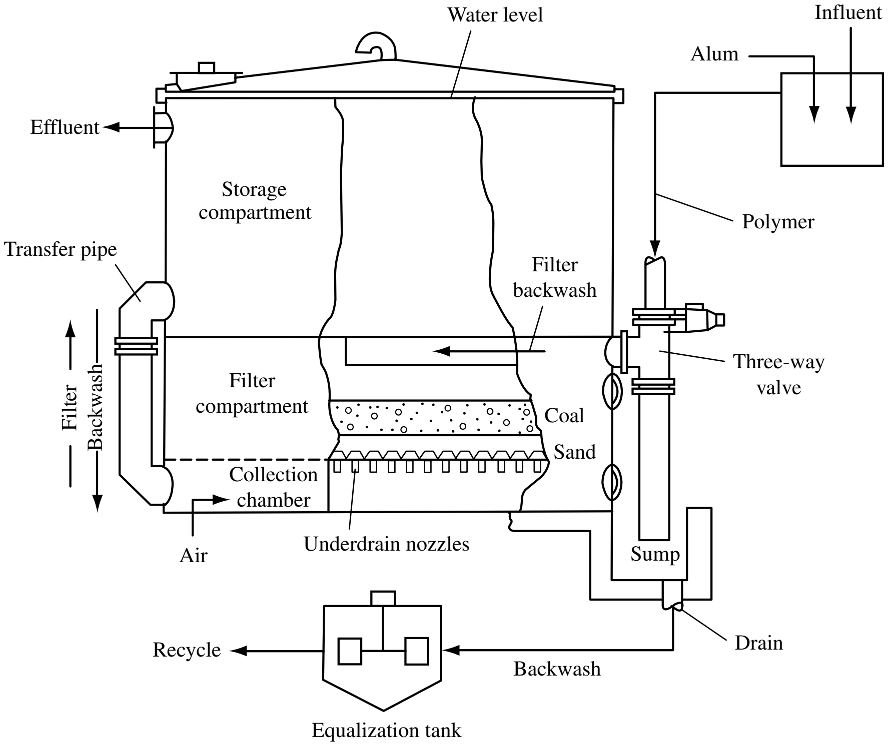


FIGURE 2 Typical dual-media filter (From Eckenfelder, 2000. Reprinted with permission from McGraw-Hill.)

membranes and membrane reactors in green technology and water reuse was published by Howell (2004). In it, water problems in different regions of the world are discussed, different membrane systems are presented, and different approaches for new research are introduced.

Reverse Osmosis There are several reverse-osmosis units currently in use to produce freshwater from seawater. With recent improvements in membranes, this process is also being used for purification of wastewater. Substantial removal of BOD, COD, total dissolved solids, phosphate, and ammonia by this process has been reported (Robinson and Maltson, 1967).

In a reverse-osmosis process, wastewater containing dissolved materials is placed in contact with a suitable semipermeable membrane in one of the two compartments of the tank. The pressure on this compartment is increased to exceed the osmotic pressure for that particular waste in order to cause the water to penetrate the membrane, carrying with it only a small amount of dissolved materials. Therefore, the dissolved material in the wastewater gets concentrated continuously, while highly purified water collects in the other compartment.

The performance of the reverse-osmosis process depends mainly on (1) the membrane semipermeability or its efficiency to separate dissolved material from the wastewater, and (2) the membrane permeability or the total amount of water that can be produced with appropriate efficiency for the removal of dissolved materials. It has been reported that the conventional cellulose-acetate membranes give adequate separation efficiency, but the flow rate of water is too small to be of practical interest. However, cellulose-acetate membranes allow a much higher flow rate of product water, at the same separation efficiency, which makes it applicable in wastewater-treatment practices (Goff and Gloyne, 1970).

The operating pressure, as well as the rejection performance of the membrane, is dependent on the membrane porosity. Rejection performances of three graded membranes with secondary sewage effluent were investigated by Bray et al. (1969).

Merten et al. (1968) evaluated the performance of an 18.9 m^3/day pilot reverse-osmosis unit in removing small amounts of organic material found in the effluent of carbon columns treating secondary effluent. With a feed pressure of 2760 kPa and water recovery of 80 to 85%, 84% removal of COD present in the carbon column effluent, averaging 10.8 mg/l, was achieved. Problems of clogging have occurred when operating with waters containing high concentrations of bicarbonate, and as such, adjustment of pH to prevent calcium-carbonate precipitation is normally required.

Sadr Ghayeni et al. (1996) discuss issues such as flux control and transmission in microfiltration membranes and biofouling in reverse-osmosis membranes in their use for the reclamation of secondary effluents. The process used for the study consisted of a microfiltration membrane followed by a reverse-osmosis membrane. The performance of this combined system was evaluated by Sadr Ghayeni et al. (1998b), as was the study of the adhesion of bacteria to reverse-osmosis membranes (1998a).

A study on the processing of composite industrial effluent by reverse osmosis was published by Sridhar et al. in 2003. The effluent used in the study was from combined bulk drug and pharmaceutical companies, obtaining a removal of 88% of dissolved solids, COD, and BOD, with reasonable water recovery. They also present a comparison between aerobic and reverse-osmosis treatment for this effluent.

A physical-chemical process for the treatment of chemical mechanical polishing process wastewater is presented by Lin and Yang (2004). In it the authors used chemical coagulation using different coagulants followed by reverse osmosis, obtaining water capable of being reused in the process due to its characteristics.

Electrodialysis Electrodialysis involves the removal of inorganic ions from water by creating an electrical potential across two electrodes dipped in water. One of the two strips serves as a cathode and the other as an anode. The treatments that can be achieved by electrodialysis include:

1. Removal of inorganic ions: Under the effect of applied potential, cations and anions migrate to the cathode and anode, respectively. By alternating membranes, a series of concentrating and diluting compartments can be created. For a long run and better efficiency, it is essential that turbidity, SS, colloids, and trace organics are removed from the wastewater before it enters the electrodialysis unit.
2. Effective bacteria reduction in wastewater: Most of the municipal wastewaters contain a high concentration of chloride ions. Oxidation of chloride at the anode produces chlorine, hypochlorite, or chloramines, depending on the nature of the wastewater. Chlorine in these forms is a good disinfectant and also provides an effective means of reducing soluble BOD.

In order to reduce the operating cost of the electrodialysis process, the eroding anodes made of aluminum or iron are now being replaced by nonconsumable noble anodes, which appear to have more potential in wastewater treatment (Culp and Culp, 1971). The cost of disinfection by electrodialysis is reported to be 0.053 \$/m^3 of wastewater as compared to 0.095 \$/m^3 for the conventional chlorination (unpublished proposal, 1970). However, some other sources have reported that the cost of electrolytic treatment of wastewater was too high for the removal of a large percentage of secondary effluent COD.

Grimm et al. (1998) present a review of electro-assisted methods for water purification, including electrodialysis. Fukumoto and Haga (2004) applied this technique for the treatment of swine wastewater with removal rates for NO_3^- and PO_4^{-3} ions of 99% and an average color reduction of 58%.

Gas Stripping

In domestic wastewaters, most of the nitrogen that gets converted to ammonia during biological degradation is present either as ammonia or in organic form. When the carbon concentration in wastewater becomes low and the nitrifying bacteria are populous, this ammonia can be oxidized by bacteria to nitrites and nitrates in the presence of dissolved oxygen. The stripping process can be employed either before or after secondary treatment for removing high levels of nitrogen that is present as ammonia. If it is to be used as pretreatment prior to a biological system, enough nitrogen, N:BOD = 5:150, must be left in the effluent to satisfy the nutritional requirement (Eckenfelder and Barnhart, 1963).

In wastewater, ammonium ions exist in equilibrium with ammonia and hydrogen ions:

$$NH_4^+ \leftrightarrow NH_3 \uparrow + H^+ \tag{10}$$

At pH levels of 6 to 8, ammonia nitrogen is mostly present in the ionized form NH_4^+. Increasing the pH to above 10 changes all the nitrogen to ammonia gas, which is removable by agitation. The stripping of ammonia from wastewater is carried out with air. In this operation, wastewater is agitated vigorously in a forced-draft countercurrent air-stripping tower when the ammonia is driven out from the solution and leaves with the air exhausted from the tower. The efficiency of ammonia removal in the stripping process depends upon the pH, airflow rate, tower depth, and hydraulic loading to the tower.

Slechta and Culp (1967) have shown experimentally that the efficiency of the ammonia-stripping process is dependent on the pH of the wastewater for pH values up to 10.8. However, no significant increase in ammonia removals was achieved by elevating the pH above 10. Kuhn (1956) had come to the same conclusion. It has also been reported that the efficiency of the ammonia-stripping process depends on maximizing the air–water contact within the stripping tower. Higher ammonia removals and lower air requirements were obtained with a 40 × 50-mm packing than with a 100 × 100-mm packing. Increased tower depth, which provides additional air–water contact, results in greater ammonia removals and lower air requirements. Ammonia removals of 90%, 95%, and 98% were obtained at airflow rates of 1875, 3000, and 6000 m^3 per cubic meter of wastewater, respectively.

Gas stripping is also used for removal of H_2S and VOCs from wastewater.

Ion Exchange

The ion-exchange process has been adopted successfully in wastewater-treatment practice for removing most of the inorganic dissolved salts. However, the cost of this method for wastewater treatment cannot be justified unless the effluent water is required for multiple industrial municipal reuse. One of the major applications of this technique is the treatment of plating-industry wastewater, where the recovery of chrome and the reuse of water make it an attractive choice (Eckenfelder, 2000).

Gaffney et al. (1970) have reported that the modified DESAL process, developed for treating acid mine drainage waters, can be applied successfully to the treatment of secondary sewage effluent. This process consists of passing secondary sewage-plant effluent upflow through an ion-exchange unit filled with a weak base anion exchange resin, Amberlite IRA-68, operated on a bicarbonate cycle. The effluent with a pH of 6.0 is then treated with a small quantity of bentonite and cationic flocculant, Prima floc C-7, followed first by aeration to drive out carbon dioxide, and then lime softening in proportion to its hardness concentration. A dosage of 30 mg/l of bentonite, 3 to 5 mg/l of polyelectrolyte, and normal lime levels are required. The effluent that is partially desalinated and essentially free of nitrates, phosphates, chlorides, alkyl benzene sulfonate (ABS), and COD can be produced. If the salinity is too high, it may be reduced further by passing a portion of the effluent through a weak acid cation, Amberlite IRC-84. It has been observed that IRA-68 can remove much of the organic contents and COD, thereby eliminating or markedly reducing the need for carbon treatment.

Slechta and Culp (1967) tested a cationic resin, Duolite C-25, for the removal of ammonia nitrogen from the carbon column effluent that was containing ammonia nitrogen in the range of 18 to 28 mg/l as nitrogen. A 100-mm-diameter Plexiglas cylinder filled to a depth of 700 mm with the resin served as the pilot ion-exchange column. The rate of application of influent waste to the ion-exchange column was 0.4 m^3/min per cubic meter of resin. Following breakthrough of the ammonia nitrogen to 1 mg/l, the bed was backwashed and the resin was regenerated. On the average, about 400 bed volumes of carbon column effluent had been passed through the ion-exchange resin prior to a breakthrough to 1 mg/l ammonia nitrogen. However, considering the operating and capital costs, they concluded that the ammonia-stripping process was more efficient.

Nitrate nitrogen, present in the effluent from the activated sludge process, has been removed by anion exchange regenerated with brine by R. Eliassen and Bennett (1967). This ion-exchange process also removes phosphates and some other ions; however, pretreatment by filtration is essential. The resin is restored by treatment with acid and methanol.

The removal of heavy metals with Mexican clinoptilolite was studied by Vaca Mier et al. (2001). In this study the interactions of lead, cadmium, and chromium competed for the ion-exchange sites in the zeolite. The authors also studied the influence of such factors as the presence of phenol and the pH of the solution to be treated.

Adsorption

Application of adsorption on granular active carbon, in columns of counterflow fluidized beds, for the removal of traces or organic pollutants, detergents, pesticides, and other substances in wastewater that are resistant to biological degradation has become firmly established as a practical, reliable, and economical treatment (Slechta and Culp, 1967; Weber, 1967; Parkhurst et al., 1967; Stevens and Peters, 1966; Presecan et al., 1972).

Adsorption can also be accomplished with powdered carbon (Davies and Kaplan, 1964; Beebe and Stevens, 1967), which is mixed in wastewater, flocculated, and ultimately settled. However, there are certain problems associated with the use of powdered carbon. These are: (1) that large quantities of activated carbon are needed in wastewater treatment, because it is used only on a once-through basis, and handling of such large quantities of carbon also creates a dust problem, and (2) problems in disposal of precipitated carbon unless it is incinerated along with the sewage sludge.

Carbon-Adsorption Theory Less polar molecules, including soluble organic pollutants, are removed by adsorption on a large surface area provided by the activated carbon. Smaller carbon particles enhance the rate of pollutant removal by providing more total surface area for adsorption, partial deposition of colloidal pollutants, and filtration of larger particles. However, it is almost always necessary to remove finely divided suspended matter from wastewater by pretreatment prior to its application on a carbon bed.

Depending on the direction of flow, the granular carbon beds are either of the downflow-bed type or upflow-bed type. Downflow carbon beds provide the removal of suspended and flocculated materials by filtration beside the absorption of organic pollutants. As the wastewater passes through the bed, the carbon nearest the feed point eventually becomes saturated and must be replaced with fresh or reactivated carbon. A countercurrent flow using multiple columns in series is considered more efficient. The first column is replaced when exhausted, and the direction of flow is changed to make that column the last in the series. Full countercurrent operation can best be obtained in upflow beds (Culp and Culp, 1971).

Upflow carbon columns for full countercurrent operations may be either of the packed-bed type or expanded-bed type. Packed beds are well suited to treatment of wastes that contain little or no SS, i.e., turbidity less than 2.5 JTU. However, the SS invariably present in municipal and industrial wastewaters lead to progressive clogging of the carbon beds. Therefore, expanded-bed upflow columns have certain potential advantages in operation of packed-bed adsorbers for treating wastes that contain SS. In expanded-bed-type adsorbers, water must be passed with a velocity sufficient to expand the bed by about 10%, so that the bed will be self-cleaned.

Experiments conducted by Weber et al. (1970) have shown that expanded-bed and packed-bed adsorption systems have nearly the same efficiency with regard to the removal of soluble organic materials from trickling-filter effluent, under otherwise similar conditions. The packed-bed system was found to be more effective for removal of SS, but the clogging that resulted from these solids required higher pumping pressure and more frequent cleaning of the carbon beds. Because of the time elapsed in cleaning, the expanded-bed production was about 9% more than the packed-bed production.

The Lake Tahoe Water Reclamation Facilities, described by Slechta and Culp (1967), included pretreatment of secondary effluent by chemical clarification and filtration, thereby providing a highly clarified feed and permitting an extended operation of carbon beds. Carbon beds were of the upflow full-countercurrent columns and were usually operated as packed types since the turbidity of the applied water was always less than 0.3 JTU. The columns were also operated at times as expanded beds by drawing 10% of the carbon.

The efficiency of carbon bed treatment depends on:

1. Contact time between the carbon and the wastewater
2. The pH of the wastewater (below pH of 9.0, the rate of adsorption of organics in wastewater increases with decreasing pH)
3. Temperature of the wastewater
4. Quality of the influent.

Carbon-Bed Design The following are the important design parameters used in the design of activated-carbon beds:

1. Properties of filtering material: The two most common sizes of granular carbon normally used for wastewater treatment are 8 × 30 mesh and 12 × 14 mesh; 8 × 30 mesh carbon is preferable, although its surface area is less, because it reduces losses during regeneration, head loss is less, and the operation of the filter becomes easier. Culp and Culp (1971) have reported the desirable physical properties for granular activated carbon for use in wastewater treatment.
2. Depth of carbon bed: 3 to 10 m.
3. Flow rates: 140 to 700 m^3/min per square meter of column cross-sectional area. For optimum performance of the bed, the actual depth and rate of flow should be determined by a dynamic pilot-plant test in the laboratory.
4. Activated carbon contact time: 15 to 35 mm, depending on the objective of treatment and the impurities to be removed from the wastewater.

A diagram of a carbon-gas-adsorption process is presented in Figure 3.

In the 1970s, over 30 granular activated-carbon plants were designed in the United States for use at municipal wastewater-treatment plants (DeJohn and Edwards, 1981). Thirteen of the plants are classed as physical-chemical and 15 as tertiary treatment plants, and 4 use carbon to dechlorinate. According to the authors, there have been some problems encountered at certain physical-chemical and tertiary treatment plants, but in their opinion, granular activated carbon is a viable treatment alternative when applied under proper conditions (Grieves et al., 1964).

One of the changes in conventional treatment technology is the use of the PACT (Powdered Activated Carbon Treatment) process using powdered carbon in the aeration tank of an activated sludge system (Meidl, 1981). The development

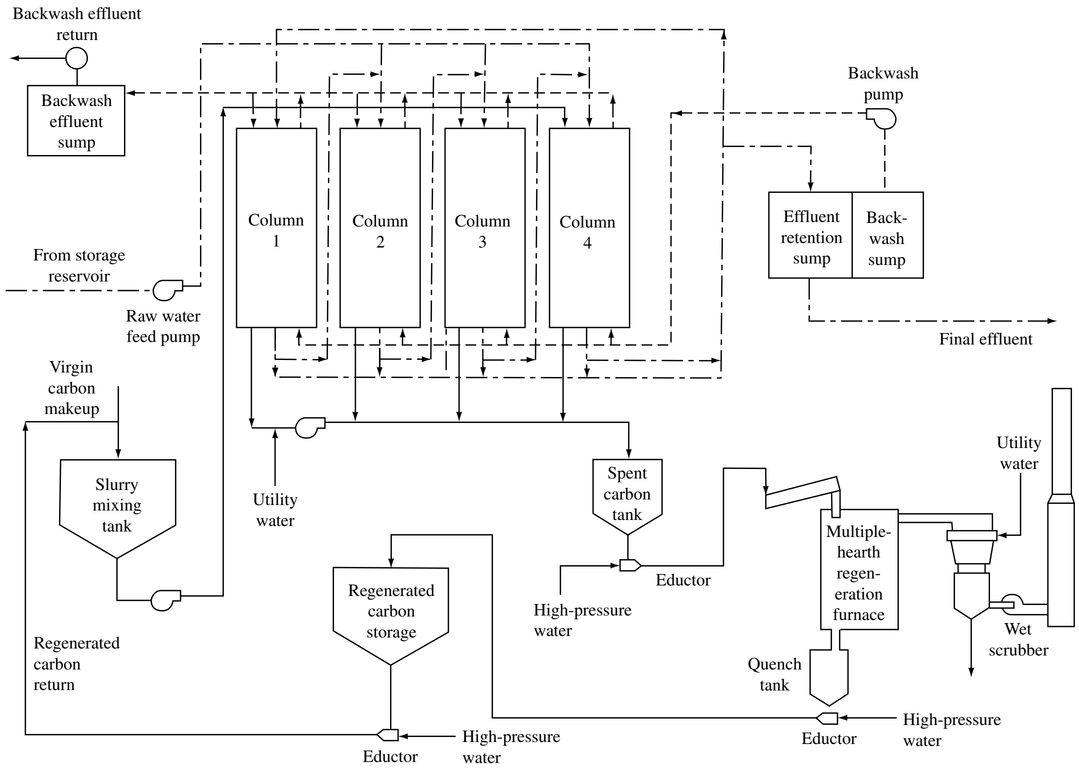


FIGURE 3 Carbon-adsorption process diagram (From Eckenfelder, 2000. Reprinted with permission from McGraw-Hill.)

of PACT in municipal wastewater treatment resulted from the inability of the physical-chemical treatment process to adequately treat wastewater. This process has been used successfully to treat domestic wastewater from a residential population of 30,000 in Vernon, Connecticut, and many other plants have been under construction. Similar applications of combined powdered-carbon and activated-sludge treatment to various industrial wastewaters, particularly coal-gasification wastewaters, are shown to be successful.

Miyake et al. (2003) studied the adsorption equilibrium isotherms of trichloroethylene (TCE) vapor stripped from TCE-polluted waters. These results can be used for wastewater treatment of waters with the same pollutant. A combined Al(III) coagulation/carbon adsorption process for the treatment of reactive dyes in synthetic wastewater was proposed by Papic et al. (2004). This process achieved 99.9% reduction of the dyes in the wastewater as well as 95.7 and 91.3% COD reduction for the two waters used. They conclude that the proposed process has many advantages, such as high efficiency, low use of coagulant, minimal sludge production, and high-quality product water with reuse potential.

Adsorption of Pollutants on Biomaterials A review of potentially low-cost adsorbents for heavy metals was presented by Bailey et al. (1999). Minamisawa et al. (2004) investigated the adsorption of cadmium and lead ions on different biomaterials, concluding that this is a promising alternative for the removal of these ions. Another review was published in 2004 by Gardea-Torresdey et al., concluding that Cd, Cu, Ni, Cr, and other ions have been successfully removed from solutions using different biomaterials. Gong et al. (2005) studied the use of peanut hull as a biosorbent for the removal of anionic dyes in a solution, obtaining sorption capacities of between 13.99 and 15.60 mg per gram of biosorbent for three different dyes.

Flotation

Surface-active contaminants, if present in wastewater, will produce foam upon aeration. This foam rises to the surface of the wastewater and can be separated and concentrated. Moreover, as the foam generates and rises, certain suspended impurities also get removed by entrainment. Thus, the foam-separation process can be developed to provide a selective removal of soluble and colloidal pollutants in various concentrations from water or wastewater. The process may either utilize the surface-active impurities present in the wastewater or may require the addition of a specific surface-active agent prior to aeration; soluble or colloidal impurities of interest may be precipitated to improve their removal.

Factors affecting the efficiency of the process of foam separation are:

1. Airflow rate/surface-active agent or airflow-rate-to-waste-flow-rate ratio (the removal efficiency increases when increasing the airflow rate, but yields wetter foam)
2. Air-bubble size (fine-bubble aeration improves the efficiency of foam separation, whereas coarse bubbles or deck aeration are not efficient)
3. Nature and concentration of surface-active agents
4. Foam stability
5. pH of the wastewater
6. Detention period (a short aeration time, 5 minutes or so, is considered sufficient)
7. Surface-to-volume ratio (a large surface-to-volume area is conducive to higher efficiency)

The process of foam separation has been used successfully in a number of waste-treatment applications. These include (1) removal of surface-active materials such as ABS from secondary effluent sewage (Klein and McGauhey, 1963; Grieves et al., 1964; Brunner and Lemlich, 1963; Fldib, 1963); (2) removal of radioactive ions from dilute aqueous solutions by the addition of anionic surface agents (Schoen and Mazella, 1961; Schoen et al., 1962; Schnepf et al., 1959); (3) removal of specific surfactants from refining and petrochemical wastewaters (Grieves and Wood, 1963); (4) foam fractionation of phenol (phenolate) from aqueous solutions using a cationic surfactant, ethyl hexadecyl dimethylammonium bromide, as the foaming agent (Grieves and Aronica, 1966); (5) the treatment of cyanide and acid chromate wastes (cyanide is precipitated with ferrous ions, and the ferrocyanide precipitates are floated with a cationic surfactant in a pH range of 5 to 7; acid chromate is reduced and precipitated as $Cr(OH)_3$, and these precipitates are floated with an anionic surfactant [Grieves, 1972]); (6) the treatment of liquid wastes from the scrubbing of phosphate and fluoride from air-pollution emissions of the phosphoric-acid-manufacturing industry (these include precipitation by lanthanum (III), followed by flotation with an anionic surfactant [Grieves, 1972]); (7) clarification of turbid raw water supplies by microflotation-activated carbon process (Grieves, 1972); and (8) physical separation of SS and coagulated dissolved organic impurities from wastewaters. ABS serves as activator for bubble attachment in flotation treatment of sewage (Klein and McGauhey, 1963). Other applications include the treatment of mine-drainage wastes combined with secondary sewage effluents and removal of H_2S from sour wastes.

An overview on flotation in wastewater treatment was published by Rubio et al. in 2002. The authors explain the process, present conventional and emerging flotation techniques and processes, and discuss the applications of the technique to different compounds. An example of a dissolved air flotation (DAF) process is presented in Figure 4.

Chemical Processes

An extensive review on catalytic abatement of water pollutants was published by Matatov-Meytal and Sheintuch in 1998. In it they present the oxidation and reduction processes involved as well as a discussion on catalyst support and deactivation. A discussion on the use of iron for wastewater treatment was published by Waite (2002). This article presents the research challenges and the new possibilities in

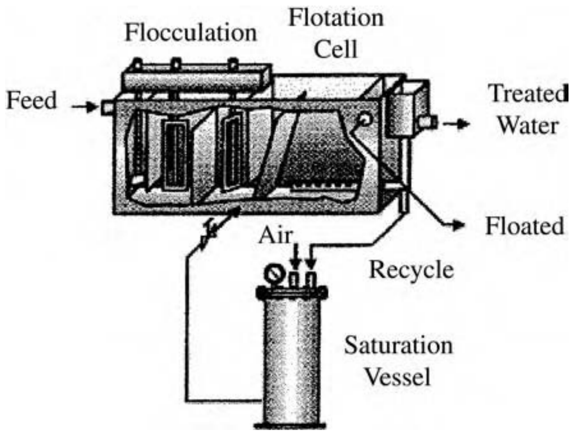


FIGURE 4 Conventional DAF unit with water recycle to the saturator (From Rubio et al., 2002, with permission from Elsevier).

this area. A review on the use of ferrate (VI) salt as a coagulant and oxidant for water and wastewater treatment is presented by Jiang and Lloyd (2002). In this paper the authors demonstrate the advantages of this approach in the treatment of microorganisms, heavy metals, and SS, as well as present the difficulties associated with this technique.

Ghoreishi and Haghighi (2003) propose a combined bisulfite catalyzed sodium borohydride reduction followed by activated sludge for the treatment of nonbiodegradable textile effluent, reposting reductions of BOD, COD, and TSS of 74–88%, 76–83%, and 92–97% respectively. Lin and Wang (2004) investigated the performance of a proposed granulated activated carbon bed combined with ozonation and preceded by chemical coagulation, obtaining interesting results. Akcil (2003) discusses the advantages of using biological methods for the treatment of cyanide in gold-mill effluents as compared to well-established chemical methods. Chung et al. (2003) evaluated the performance of a combined chemical absorption–biological oxidation system for the removal of H_2S from gaseous streams such as those produced by livestock wastewater-treatment plants. They demonstrated the potential of this technique and obtained reductions on the order of 85%. Another combined approach is presented by Libra and Sosath (2003). In this case they used ozonation followed by a biological process for the treatment of textile wastewater, comparing two configurations of treatment processes, concluding that the most attractive alternative is the simplest: ozonation followed by aerobic treatment as compared to anaerobic, aerobic, ozonation, and aerobic treatments for their effluent and conditions.

A comprehensive review on the properties of alumina, the chemical reactions involved in its use in water treatment, and adsorption and catalysis using alumina was published by Kasprzyk-Hordern in 2004. An interesting review on the enhancement of biodegradability of industrial wastes by chemical-oxidation pretreatment is presented by Mantzavinos and Psillakis in 2004. This exhaustive review includes many different pollutants, trends, and process schemes used as segregation by effluent type; evaluation of the pretreatment steps; and modeling of the individual steps involved.

Aiyuk et al. (2004) propose a chemically enhanced primary treatment followed by a UASB reactor for the treatment of domestic wastewater. The chemical pretreatment consisted of the addition of $FeCl_3$ or $Al_2(SO_4)_3$ and polymers in a mixing tank, and a posttreatment of the biologically treated water was achieved by zeolite adsorption for NH_4^+ removal. Another combined solution was proposed by Bressan et al. in 2004. This approach includes the use of an enhanced Fenton process followed by biological treatment of olive-mill wastewater.

Oxidative, Photochemical, and Electron-Beam Processes

Oxidizing agents like chlorine, ozone, hydrogen peroxide, potassium permanganate, and ultraviolet irradiation have been used successfully in oxidizing and stabilizing certain impurities like hydrogen sulfide, phenol, cyanide, and selected refractory organic substances. The reaction of chlorine with certain organic matter present in wastewater to form a more harmful persistent chemical has raised a serious question of continuing the use of chlorine for disinfection of wastewater effluents. Ozone is gradually becoming more popular, because it has strong oxidizing and disinfecting capacity and leaves no harmful residuals. It has been shown that ozone in combination with photochemical oxidation economically removed inorganic, organic, and toxic refractory species and effectively destroyed phenol in a bubble column or a stirred cell (Wall, 1980; Otake, 1979). Ganes and Staubach (1980) have reported the kinetics of the reaction between ozone and nitrilotriacetate in water and the effects of pH, metal content, and natural organics on the rate and extent of degradation.

In 1993, Legrini et al. published an extensive review on photochemical processes for water treatment, including UV, H_2O_2/UV, ozone/UV, O_3/H_2O_2/UV, TiO_2/UV, vacuum UV, photochemical electron-transfer processes, and energy-transfer processes. In 1999, Andreozzi et al. published studies on the use of O_3/UV and O_3/H_2O_2 for the treatment of mineral-oil-polluted wastewaters, performing different experiments and achieving an 80 to 90% reduction of COD for the O_3/UV treatment (Andreozzi et al., 1999). A discussion of the different advanced oxidation processes (AOP), as well as experimental apparatus and working procedures for the study of the application of this technique, is presented in Andreozzi et al. (2000).

An extensive review on photocatalytic degradation was presented by Bhatkhande et al. (2001). In it they include a description of the process, the mechanisms implied, the compounds that can be degraded with it, and the variables that can affect this technique. Another review was published by Kabra et al (2004), concluding that this process can be used to treat industrial wastewater for the removal of metal ions and nonbiodegradable organics. Another conclusion is that the costs of this method are slightly higher than for conventional methods, but that future research can make a competitive option. An interesting article on the evaluation of the use of peracetic acid for the disinfection of the effluent of Montreal's wastewater-treatment plant was published by Gehr et al. in 2003. In this study the authors presented and evaluated the different alternatives (with the exception of chlorine, due to environmental regulations) arriving at the conclusion that using the current process, the economically viable alternative is UV, while if some changes are made upstream, the use of peracetic acid could be a viable alternative.

A study on the application of different AOPs to the treatment of textiles, Kraft bleaching, photoprocessing, and pharmaceutical wastewaters was published by Balcoglu et al. in 2003. They conclude that the efficiency of the AOP process used for a particular application depends upon the pretreatment used, and concentrations and types of pollutants present in the influent.

Adesina (2004) investigated the use of photocatalysis for the treatment of spent industrial Bayer liquor and detoxification of paper-mill effluents, among others. The use of sunlight as an energy source is also discussed. A comparison of different AOPs and chemical-treatment options was conducted by Azbar et al. (2004) for the reduction of color and COD from an acetate- and polyester-dyeing process. They conclude that AOPs have better performances than chemical-coagulation methods for the parameters studied. They also found that UV/H_2O_2 achieved 99% and 96% COD and color removal, respectively. Their choice from the economic point of view was the Fenton's reagent process. The removal of the drug diclofenac by means of UV/H_2O_2 was studied by Vogna et al. (2004), showing that the proposed treatment was effective for the degradation of the drug. The behavior of the process Fe(III)/Air/UV was studied by Andreozzi and Marotta (2004) by using benzoic acid as the molecule to be treated, with the goal of developing kinetic models for this process. A study on the treatment of cork-processing wastewater was published by Acero et al. in 2004. They evaluated the use of different combinations of UV, H_2O_2, and O_3 and Fenton's reagent and photo-Fenton processes for the effluent under study and concluded that the best options that produce reusable water are those involving ozone.

Broad reviews on oxidation technologies at ambient conditions and on hybrid methods for wastewater treatment were published by Gogate and Pandit (2004a, 2004b). Tabrizi and Mehrvar (2004) present an interesting article on the integration of AOPs and biological processes, including recent developments, trends, and advances in this field. A review on the degradation of chlorophenols via AOP was published by Pera-Titus et al. in 2004. Among their conclusions is that although photocatalytic processes show higher half reaction times, they do not require oxidants or further separation of byproducts after the reaction.

Ding et al. (1996) present a review of catalytic oxidation in supercritical water, including the reactions involved, the processes available, a comparison with subcritical water oxidation, and an extensive review of the catalysts available. Kritzer and Dinjus (2001) published an interesting evaluation of the problems of supercritical water oxidation, with discussion and suggestions for its improvement.

Ultrasound Ultrasound is being studied as an alternative solution for environmental problems. This process works by generation of highly reactive oxidizing species, such as hydroxyl, hydrogen, and hydroperoxyl radicals as well as hydrogen peroxide by means of ultrasound waves (Vajnhandl and Majcen Le Marechal, 2005). In their review they include the use of ultrasound in the textile industry and its wastewaters.

Electrochemical Oxidation Panizza and Cerisola (2004) published results on a series of experiments using an electrochemical cell for the treatment of synthetic-tannery wastewater using two different electrodes under various experimental conditions, concluding that electrochemical methods can be effectively applied for the final treatment of these effluents, achieving total COD, tannin, and ammonium removals.

Electron-Beam Wastewater Treatment Water irradiation with ionizing radiation generates several very reactive ions and molecules. Getoff presents a review and the state of the art for radiation-induced degradation of water pollutants (1996). His group has studied the use of radiation for disinfection and decomposition of pollutants in water and wastewater for years. Among the experimental factors proposed to affect the efficiency of this process in pollutant degradation are: form of radiation, energy, absorbed dose and dose rate, pollutant concentration, pH, temperature, effect of oxygen and ozone, presence of ozone and TiO_2, and molecular structure of the pollutants (2002).

Pikaev et al. (2001) applied electron beams followed by coagulation for the treatment of mixed distillery and municipal wastewater; they conclude that the proposed scheme can treat the effluent at lower cost than the biological and sedimentation processes. In 2002 his group published pilot-plant experiments using electron-beam and biological oxidation for the treatment of dyeing wastewater. In this research they conclude that the proposed method reached the desired effluent characteristics in about 8 hours of treatment, as compared to 17 hours for biological treatment alone (Pikaev, 2002). In collaboration with researchers from Korea, he also used electron beams combined with coagulation, flocculation, and biological processes for the treatment of paper-mill effluents. This research concluded with the design of a commercial plant for the treatment of 15,000 m^3/day of wastewater with 80% of water reuse (Shin et al., 2002). In a final publication, Pikaev (2002) presents data on the treatment of different pollutants, including carboxylic acids, distillery slops, and petroleum products. In this work the author presents an economic evaluation of a commercial plant, concluding that the treatment costs for the proposed technology are about half of conventional methods.

REFERENCES

Acero, J.L., Javier Benitez, J. F., Heredia, J.B.D., and Leal, A.I. (2004). Chemical treatment of cork-processing wastewaters for potential reuse. *Journal of Chemical Technology and Biotechnology,* 79(10), 1065–1072.

Adesina, A.A. (2004). Industrial exploitation of photocatalysis: progress, perspectives and prospects. *Catalysis Surveys from Asia,* 8(4), 265–273.

Aiyuk, S., Amoako, J., Raskin, L., van Haandel, A., and Verstraete, W. (2004). Removal of carbon and nutrients from domestic wastewater using a low investment, integrated treatment concept. *Water Research,* 38(13), 3031–3042.

Akcil, A. (2003). Destruction of cyanide in gold mill effluents: biological versus chemical treatments. *Biotechnology Advances,* 21(6), 501–511.

Albertson, O.E. and Sherwood, R.J. (1967). Phosphate extraction process. Paper presented at the Water Pollution Control Federation meeting, Yakima, Washington, October.

Andreozzi, R. and Marotta, R. (2004). Removal of benzoic acid in aqueous solution by Fe(III) homogeneous photocatalysis. *Water Research,* 38(5), 1225–1236.

Andreozzi, R., Caprio, V., Insola, A., and Marotta, R. (1999). Advanced oxidation processes (AOP) for water purification and recovery. *Catalysis Today,* 53(1), 51–59.

Andreozzi, R., Caprio, V., Insola, A., Marotta, R., and Sanchirico, R. (2000). Advanced oxidation processes for the treatment of mineral oil-contaminated wastewaters. *Water Research,* 34(2), 620–628.

Anonymous (1967). *Renovation and reuse of wastewaters in Britain,* published by Water Pollution Research Laboratory.

Azbar, N., Yonar, T., and Kestioglu, K. (2004). Comparison of various advanced oxidation process and chemical treatment methods for COD and color removal from a polyester and acetate fiber dyeing effluent. *Chemosphere,* 55(1), 35–43.

Babbit, H.E., Donald, J.J., and Cleasby, J.L. (1962). *Water Supply Engineering.* McGraw-Hill Book Company, New York.

Bailey, S.E., Olin, T.J., Bricka, R.M., and Adrian, D.D. (1998). A review of potentially low-cost sorbents for heavy metals. *Water Research,* 33(11), 2469–2479.

Balcoglu, I.A., Alaton, I.A., Ötker, M., Bahar, R., Bakar, N., and Ikiz, M. (2003). Application of advanced oxidation processes to different industrial wastewaters. *Journal of Environmental Sciences and Health,* Part A—Toxic/Hazardous Substances and Environmental Engineering, A38(8), 1587–1596.

Baumann, E.R. and Oulman, C.S. (1970). Sand and diatomite filtration practices. *Water Quality Improvement by Physical and Chemical Processes,* Water Resources Symposium No. 3, University of Texas Press, Austin and London, 104.

Beebe, R.L. and Stevens, J.I. (1967). Activated carbon system for wastewater renovation. *Water and Waste Engineering,* 4(1), 43.

Bes-Pia, A., Iborra-Clar, A., Mendoza-Roca, J.A., Iborra-Clar, M.I., and Alcaina-Miranda, M.I. (2004). Nanofiltration of biologically treated textile effluents using ozone as a pre-treatment. *Desalination,* 167, 387–392.

Bes-Pia, A., Mendoza-Roca, J.A., Alcaina-Miranda, M.I., Iborra-Clar, A., and Iborra-Clar, M.I. (2003). Combination of physico-chemical treatment and nanofiltration to reuse wastewater of a printing, dyeing and finishing textile industry. *Desalination,* 157(1–3), 73–80.

Bhatkhande, D.S., Pangarkar, V.G., and Beenackers, A.A.C.M. (2001). Photocatalytic degradation for environmental applications—a review. *Journal of Chemical Technology and Biotechnology,* 77(1), 102–116.

Bishop, D.F., O'Farrell, T.P., and Stanberg, J.B. (1972). Physical-chemical treatment of municipal wastewater. *Journal of the Water Pollution Control Federation,* 44, (3), 361.

Bray, D.T., Merten, U., and Augustus, M. (1969). *Chemical Engineering,* 76(4), 60.

Bressan, M., Liberatore, L., d'Alessandro, N., Tonucci, L., Belli, C., and Ranalli, G. (2004). Improved combined chemical and biological treatments of olive oil mill wastewaters. *Journal of Agricultural and Food Chemistry,* 52(5), 1228–1233.

Brunner, C.A. and Lemlich, R. (1963). Foam fractionation: standard separator and refluxing columns. *Industrial Engineering Chemistry,* 2, 297–300.

Buzzell, J.C. and Sawyer, C.H. (1967). Removal of algal nutrients from raw wastewater with lime. *Journal of the Water Pollution Control Federation,* 39(10), R16.

Chung, Y.-C., Ho, K.-L., and Tseng, C.-P. (2003). Hydrogen sulfide gas treatment by a chemical-biological process: chemical absorption and biological oxidation steps. *Journal of Environmental Science and Health,* Part B—Pesticides, Food Contaminants, and Agricultural Wastes, 38(5), 663–679.

Conley, W.R., Jr. and Hsiung, K. (1965). Design and application of multimedia filters. *Journal of the American Water Works Association,* 57(10), 1333.

Convery, J.J. (1968). Phosphorus removal by tertiary treatment with lime and alum. Paper presented at the FWPCA Symposium on Nutrient Removal and Advanced Waste Treatment, Tampa, Florida, November 15.

Culp, R.L. and Culp, G.L. (1971). *Advanced Wastewater Treatment.* Van Nostrand Reinhold Company, New York.

Culp, G.L. and Hansen, S.P. (1967a). Extended aeration effluent polishing by mixed media filtration. *Water and Sewage Works,* 114(2), 46.

Culp, G.L. and Hansen, S.P. (1967b). How to reclaim wastewater for reuse. *American City,* June, 96.

Davies, D.S. and Kaplan, R.A. (1964). Activated carbon eliminates organics. *Chemical Engineering Progress,* 60(12), 46.

de-Bashan, L.E. and Bashan, Y. (2004). Recent advances in removing phosphorus from wastewater and its future use as fertilizer (1997–2003). *Water Research,* 38(19), 4222–4246.

DeJohn, P.B. and Edwards, R.W. (1981). Application of GAC to municipal wastewaters. Proceedings of the International Conference on the Application of Adsorption to Wastewater Treatment, Vanderbilt University, Nashville, Tennessee, 109–127.

Dillingham, J.H., Cleasby, J.L., and Baumann, E.R. (1966). Optimum design and operation of diatomite filtration plants. *Journal of the American Water Works Association,* 58(6), 657.

Dillingham, J.H., Cleasby, J.L., and Baumann, E.R. (1967). Diatomite filtration equations for flat and cylindrical septa. *Journal of the Sanitary Engineering Division, American Society of Civil Engineers,* 93(2) (SA1), 41.

Ding, Z.Y., Frisch, M.A., Li, L., and Gloyna, E.F. (1996). Catalytic oxidation in supercritical water. *Industrial and Engineering Chemistry Research,* 35(10), 3257–3279.

Eckenfelder, W.W. (1966). *Industrial Water Pollution Control.* McGraw-Hill Book Company, New York.

Eckenfelder, W.W. (2000). *Industrial Water Pollution Control.* McGraw-Hill Book Company, New York.

Eckenfelder, W.W. and Barnhart, E.L. (1963). Performance of high rate trickling filter using selected media. *Journal of the Water Pollution Control Federation,* 35(12), 1535.

Ecodyne (1972). Physical-chemical sewage treatment plant for the village of Rosemount. Environmental Currents News, *Environmental Science and Technology,* 6(4), 306.

Eliassen, K. and Bennett, G. (1967). Anion exchange and filtration techniques for wastewater renovation. *Journal of the Water Pollution Control Federation,* 39(10), R81.

Eliassen, R. and Bennett, G.E. (1967). Renovation of domestic and industrial waste water. Paper presented at the International Conference on Water for Peace, Washington, DC, May.

Fane, A.G. (1996). Membranes for water production and wastewater reuse. *Desalination,* 106(1–3), 1–9.

Fldib, L.A. (1963). Renovation of wastewater by foam fractionation. Paper presented before the Division of Water and Waste Chemistry, American Chemical Society, 143rd National Meeting, Cincinnati, Ohio, January 15.

Fukumoto, Y. and Haga, K. (2004). Advanced treatment of swine wastewater by electrodialysis with a tubular ion exchange membrane. *Animal Science Journal,* 75(5), 479–485.

Gaffney, J.G., Junin, R., and Downing, D.G. (1970). Ion exchange for industrial water. Water Quality Improvement by Physical and Chemical Processes, Water Resources Symposium No. 3, University of Texas Press, Austin and London, 258.

Galambos, I., Mora Molina, J., Jaray, P., Vatai, G., and Bekassy-Molnar E. (2004). High organic content industrial wastewater treatment by membrane filtration. *Desalination,* 162, 117–120.

Galarneau, E. and Gehr, R. (1996). Phosphorus removal from wastewaters: experimental and theoretical support for alternative mechanisms. *Water Research,* 31(2), 328–338.

Ganes, L.M. and Staubach, J.A. (1980). Reaction of nitrilotiacetate with ozone in model and natural waters. *Environmental Science and Technology,* 14, 571.

Gardea-Torresdey, J.L., dela Rosa, G., and Peralta-Videa, J.R. (2004). Use of phytofiltration technologies in the removal of heavy metals: a review. *Pure and Applied Chemistry,* 76(4), 801–813.

Gehr, R., Wagner, M., Veerasubramanian, P., and Payment, P. (2003). Disinfection efficiency of peracetic acid, UV and ozone after enhanced primary treatment of municipal wastewater. *Water Research,* 37(19), 4573–4586.

Getoff, N. (1996). Radiation-induced degradation of water pollutants—state of the art. *Radiation Physics and Chemistry,* 47(4), 581–593.

Getoff, N. (2002). Factors influencing the efficiency of radiation-induced degradation of water pollutants. *Radiation Physics and Chemistry,* 65(4–5), 437–446.

Ghoreishi, S.M. and Haghighi, R. (2003). Chemical catalytic reaction and biological oxidation for treatment of non-biodegradable textile effluent. *Chemical Engineering Journal,* 95(1–3), 163–169.

Goff, D.L. and Gloyne, E.F. (1970). Reverse osmosis development. Water Quality Improvement by Physical and Chemical Processes, Water Resources Symposium No. 3, University of Texas Press, Austin and London, 312.

Gogate, P.R. and Pandit, A.B. (2004a). A review of imperative technologies for wastewater treatment I: oxidation technologies at ambient conditions. *Advances in Environmental Research,* 8(3–4), 501–551.

Gogate, P.R. and Pandit, A.B. (2004b). A review of imperative technologies for wastewater treatment II: hybrid methods. *Advances in Environmental Research,* 8(3–4), 553–597.

Gong, R., Ding, Y., Li, M., Yang, C., Liu, H., and Sun, Y. (2005). Utilization of powdered peanut hull as biosorbent for removal of anionic dyes from aqueous solution. *Dyes and Pigments,* 64(3), 187–192.

Grieves, R.B. (1972). Water and wastewater treatment applications of the foam separation process. Paper presented at the Second Annual Environmental Engineering and Science Conference, University of Louisville, Louisville, Kentucky, April 20–21.

Grieves, R.B. and Aronica, R.C. (1966). Foam separation of phenol with a cationic surfactant. *Air and Water Pollution International Journal,* 10(1), 31–40.

Grieves, R.B., Crandall, C.J., and Wood, R.K. (1964). Foam separation of ABS and other surfactants. *Air and Water Pollution International Journal,* 8(8–9), 501–513.

Grieves, R.B. and Wood, R.K. (1963). Continuous accumulation in foam for the treatment of refining and petrochemical wastes. *Chem. In Can.,* 15, 13–15.

Grimm, J., Bessarabov, D., and Sanderson, R. (1998). Review of electro-assisted methods for water purification. *Desalination,* 115(3), 285–294.

Han, B., Ko, J., Kim, J., Kim, Y., Chung, W., Makarov, I.E., Ponomarev, A.V., and Pikaev, A.K. (2002). Combined electron-beam and biological treatment of dyeing complex wastewater: pilot plant experiments. *Radiation Physics and Chemistry,* 64(1), 53–59.

Harriger, R.D. and Hoffman, F.L. (1970*). Phosphorous Removal Tests—Two Rivers, Wisconsin.* for Black and Veatch Consulting Engineers. Technical Service Department, Allied Chemical Corporation, Industrial Chemical Division, Syracuse, New York.

Harriger, R.D. and Hoffman, F.L. (1971). *Phosphate Removal Tests—Springfield, Ohio.* for Black and Veatch Consulting Engineers, Technical Service Department, Allied Chemical Corporation, Industrial Chemicals Division, Syracuse, New York.

Hilal, N., Al-Zoubi, H., Darwish, N.A., Mohammad, A.W., and Abu Arabi, M. (2004). A comprehensive review of nanofiltration membranes: treatment, pretreatment, modelling, and atomic force microscopy. *Desalination,* 170(3), 281–308.

Howell, J.A. (2004). Future of membranes and membrane reactors in green technologies and for water reuse. *Desalination,* 162, 1–11.

Jiang, J.-Q. and Lloyd, B. (2002). Progress in the development and use of ferrate(VI) salt as an oxidant and coagulant for water and wastewater treatment. *Water Research,* 36(6), 1397–1408.

Johns-Manville Corporation (1972). Moving bed filter system. *Water and Sewage Digest,* May–June, 47.

Kabra, K., Chaudhary, R., and Sawhney, R.L. (2004). Treatment of hazardous organic and inorganic compounds through aqueous-phase photocatalysis: a review. *Industrial and Engineering Chemistry Research,* 43(24), 7683–7696.

Kalinske, A.A. and Shell, G.L. (1968). Phosphate removal from waste effluent and raw wastes using chemical treatment. Paper presented at the Phosphorus Removal Conference sponsored by the FWPCA, Chicago, Illinois.

Kasprzyk-Hordern, B. (2004). Chemistry of alumina, reactions in aqueous solution and its application in water treatment. *Advances in Colloid and Interface Science,* 110(1–2), 19–48.

Kentish, S.E. and Stevens, G.W. (2001). Innovations in separations technology for the recycling and re-use of liquid waste streams. *Chemical Engineering Journal,* 84(2), 149–159.

Klein, S.A. and McGauhey, P.H. (1963). Detergent removal by surface stripping. *Journal of the Water Pollution Control Federation,* 35(1), 100.

Kritzer, P. and Dinjus, E. (2001). An assessment of supercritical water oxidation (SCWO): existing problems, possible solutions and new reactor concepts. *Chemical Engineering Journal,* 83(3), 207–214.

Kuhn, P.A. (1956). Removal of ammonium nitrogen from sewage effluent. MS Thesis, University of Wisconsin, Madison.

Lai, C.L. and Lin, S.H. (2003). Electrocoagulation of chemical mechanical polishing (CMP) wastewater from semiconductor fabrication. *Chemical Engineering Journal,* 95(1–3), 205–211.

Legrini, O., Oliveros, E., and Braun, A.M. (1993). Photochemical processes for water treatment. *Chemical Reviews,* 93(2), 671–698.

Libra, J.A. and Sosath, F. (2003). Combination of biological and chemical processes for the treatment of textile wastewater containing reactive dyes. *Journal of Chemical Technology and Biotechnology,* 78(11), 1149–1156.

Lin, S.H. and Yang, C.R. (2004). Chemical and physical treatments of chemical mechanical polishing wastewater from semiconductor fabrication. *Journal of Hazardous Materials,* 108(1–2), 103–109.

Lynam, B., Ettelt, G., and McAloon, T. (1969). Tertiary treatment at Metro Chicago by means of rapid sand filtration and microstrainers. *Journal of the Water Pollution Control Federation,* 41(2), 247.

Mantzavinos, D. and Psillakis, E. (2004). Enhancement of biodegradability of industrial wastewaters by chemical oxidation pre-treatment. *Journal of Chemical Technology and Biotechnology,* 79(5), 431–454.

Matatov-Meytal, Y.I. and Sheintuch, M. (1998). Catalytic abatement of water pollutants. *Industrial and Engineering Chemistry Research,* 37(2), 309–326.

Meidl, J.A. (1981). Application of PACT to municipal wastewater. Proceedings of the International Conference on the Applications of Adsorption to Wastewater Treatment, Nashville, Tennessee, 357–383.

Merten, U., Nusbaum, I., and Miele, R. (1968). Organic removal by reverse osmosis. Paper presented at the American Chemical Society Symposium on Organic Residue Removal from Wastewaters, Atlantic City, New Jersey, September.

Minamisawa, M., Minamisawa, H., Yoshida, S., and Takai, N. (2004). Adsorption behavior of heavy metals on biomaterials. *Journal of Agricultural and Food Chemistry,* 52(18), 5606–5611.

Miyake, Y., Sakoda, A., Yamanashi, H., Kaneda, H., and Suzuki, M. (2003). Activated carbon adsorption of trichloroethylene (TCE) vapor stripped from TCE-contaminated water. *Water Research,* 37(8), 1852–1858.

Mollah, M.Y.A., Morkovsky, P., Gomes, J.A.G., Kesmez, M., Parga, J., and Cocke, D.L. (2004). Fundamentals, present and future perspectives of electrocoagulation. *Journal of Hazardous Materials,* 114(1–3), 199–210.

Mollah, M.Y.A., Pathak, S.R., Patil, P.K., Vayuvegula, M., Agrawal, T.S., Gomes, J.A.G., Kesmez, M., and Cocke, D.L. (2004). Treatment of orange II azo-dye by electrocoagulation (EC) technique in a continuous flow cell using sacrificial iron electrodes. *Journal of Hazardous Materials,* 109(1–3), 165–171.

Mollah, M.Y.A., Schennach, R., Parga, J.R., and Cocke, D.L. (2001). Electrocoagulation (EC)—science and applications. *Journal of Hazardous Materials,* 84(1), 29–41.

Otake, T. (1979). Photo-oxidation with ozone. *Journal of Chemical Engineering of Japan,* 12, 289.

Panizza, M. and Cerisola, G. (2004). Electrochemical oxidation as a final treatment of synthetic tannery wastewater. *Environmental Science and Technology,* 38(20), 5470–5475.

Papic, S., Koprivanac, N., Loncaric Bozic, A., and Metes, A. (2004). Removal of some reactive dyes from synthetic wastewater by combined Al(III) coagulation/carbon adsorption process. *Dyes and Pigments,* 62(3), 291–298.

Parameshwaran, K., Fane, A.G., Cho, B.D., and Kim, K.J. (2001). Analysis of microfiltration performance with constant flux processing of secondary effluent. *Water Research,* 35(18), 4349–4358.

Parkhurst, J.D., Dryden, F.D., McDermott, G.N., and English, J. (1967). Pomona activated carbon pilot plant. *Journal of the Water Pollution Control Federation,* 39(10), R70.

Pera-Titus, M., García-Molina, V., Banos, M.A., Gimenez, J., and Esplugas, S. (2004). Degradation of chlorophenols by means of advanced oxidation processes: a general review. Applied Catalysis B: *Environmental,* 47(4), 219–256.

Pikaev, A.K. (2002). New data on electron-beam purification of wastewater. *Radiation Physics and Chemistry,* 65(4–5), 515–526.

Pikaev, A.K., Ponomarev, A.V., Bludenko, A.V., Minin, V.N., and Elizareva, L.M. (2001). Combined electron-beam and coagulation purification of molasses distillery slops: features of the method, technical and economic evaluation of large-scale facility. *Radiation Physics and Chemistry,* 61(1), 81–87.

Pokhrel, D. and Viraraghavan, T. (2004). Treatment of pulp and paper mill wastewater—a review. *Science of the Total Environment,* 333(1–3), 37–58.

Presecan, N.L., Touhill, J.C., and Shuckrow, A.J. (1972). Cleveland westerly physical chemical wastewater treatment plant. Paper presented at the Second Annual Environmental Engineering and Science Conference, University of Louisville, Kentucky, April 20–21.

Rich, L.G. (1963). *Unit Processes of Sanitary Engineering.* John Wiley and Sons, New York.

Riddick, T.M. (1961). Zeta potential and its application to difficult waters. *Journal of the American Water Works Association,* 53(8), 1007.

Rizzo, J.L. and Schade, R.E. (1969). Secondary treatment with granular activated carbon. *Water and Sewage Works,* 116(8), 307.

Robinson, W.T. and Maltson, R.J. (1967). A new product concept for improvement of industrial water. Paper presented at Annual Meeting, Water Pollution Control Federation, New York.

Roizard, D., Teplyakov, V., Favre, E., Fefilatiev, L., Lagunstsov, N., and Khotimsky, V. (2004). VOC's removal from water with a hybrid system coupling a PTMSP membrane module with a stripper. *Desalination,* 162, 41–46.

Rose, J.L. (1968). Removal of phosphorus by alum. Paper presented at the Phosphorus Removal Conference sponsored by the FWPCA, Chicago, Illinois.

Rubio, J., Souza, M.L., and Smith, R.W. (2002). Overview of flotation as a wastewater treatment technique. *Minerals Engineering,* 15(3), 139–155.

Rudolfs, W. and Trubnick, E.H. (1935). Activated carbon in sewage treatment. *Sewage Works Journal,* 7(5), 852.

Sadr Ghayeni, S.B., Beatson, P.J., Schneider, R.P., and Fane, A.G. (1998a). Adhesion of waste water bacteria to reverse osmosis membranes. *Journal of Membrane Science,* 138(1), 29–42.

Sadr Ghayeni, S.B., Beatson, P.J., Schneider, R.P., and Fane, A.G. (1998b). Water reclamation from municipal wastewater using combined microfiltration-reverse osmosis (ME-RO): preliminary performance data and microbiological aspects of system operation. *Desalination,* 116(1), 65–80.

Sadr Ghayeni, S.B., Madaeni, S.S., Fane, A.G., and Schneider, R.P. (1996). Aspects of microfiltration and reverse osmosis in municipal wastewater reuse. *Desalination,* 106(1–3), 25–29.

Schmid, L.A. and McKinney, R.E. (1969). Phosphate removal by a lime-biological treatment scheme. *Journal of the Water Pollution Control Federation,* 41(7), 1259.

Schnepf, R.W., Gaden, E.L., Mirocznik, E.Y., and Schonfield, E. (1959). Foam fractionation: metals. *Chemical Engineering Progress,* 55, 42–46.

Schoen, H.M. and Mazella, G. (1961). Foam separation. *Industrial Water Wastes,* 6, 71–74.

Schoen, H.M., Rubid, E., and Ghosh, D. (1962). Radium removal from uranium mill wastewater. *Journal of the Water Pollution Control Federation,* 34(10), 1026.

Shatto, J. (1960). Aerobic digestion and diatomite filter. *Public Works,* December, 82.

Shin, H.-S., Kim, Y.-R., Han, B., Makarov, I.E., Ponomarev, A.V., and Pikaev, A.K. (2002). Application of electron beam to treatment of wastewater from papermill. *Radiation Physics and Chemistry,* 65(4–5), 539–547.

Shuckrow, A. (1971). Battelle's northwest laboratories. *American Society of Civil Engineers Newsletter,* August, 4.

Slechta, A.F. and Culp, G.L. (1967). Water reclamation studies at the South Tahoe Public Utility District. *Journal of the Water Pollution Control Federation,* 39(5), 787.

Sridhar, S., Prasad, K.K., Murthy, G.S., Rao, A.G., and Khan, A.A. (2003). Processing of composite industrial effluent by reverse osmosis. *Journal of Chemical Technology and Biotechnology,* 78(10), 1061–1067.

Stander, G.J. and Van Vuuren, L.R.J. (1969). The reclamation of potable water from wastewater. *Journal of the Water Pollution Control Federation,* 41(3), 355.

Steel, E.W. (1960). *Water Supply and Sewerage.* McGraw-Hill Book Company, New York.

Stevens, D.B. and Peters, J. (1966). Long Island recharge studies. *Journal of the Water Pollution Control Federation,* 38(12), 2009.

Stevik, T.K., Kari, A., Ausland, G., and Hanssen, J.F. (2004). Retention and removal of pathogenic bacteria in wastewater percolating through porous media: a review. *Water Research,* 38(6), 1355–1367.

Stumm, W. and Morgan, J.J. (1970). *Aquatic Chemistry.* John Wiley and Sons, New York.

Tabrizi, G.B. and Mehrvar, M. (2004). Integration of advanced oxidation technologies and biological processes: recent developments, trends, and advances. *Journal of Environmental Science and Health,* Part A—Toxic/Hazardous Substances and Environmental Engineering, A39(11–12), 3029–3081.

Tchobanoglous, G., Burton, F.L., and Stensel, H.D. (2003). *Wastewater Engineering, Treatment and Reuse* (4th ed.). McGraw-Hill Book Company, New York.

Tchobanoglous, G. and Eliassen, R. (1970). Filtration of treated sewage effluent. *Journal of the Sanitary Engineering Division,* American Society of Civil Engineers, 96(4) (SA2), 243.

Truesdale, G.A. and Birkbeck, A.E. (1966). Tertiary treatment processes for sewage works effluents. Paper presented to the Scottish Center of the Institution of Public Health Engineers, March 28.

Unpublished proposal to Clark County Sanitation District, prepared by Pacific Engineering and Production Co., Henderson, Nevada, 1970.

Vaca Mier, M., Lopez Callejas, R., Gehr, R., Jiménez Cisneros, B.E., and Álvarez, P.J.J. (2001). Heavy metal removal with Mexican clinoptilolite: multi-component ionic exchange. *Water Research,* 35(2), 373–378.

Vajnhandl, S. and Majcen Le Marechal, A. (2005). Ultrasound in textile dyeing and the decolouration/mineralization of textile dyes. *Dyes and Pigments,* 65(2), 89–101.

Van Hulle, S.W.H. and Vanrolleghem, P.A. (2004). Modelling and optimisation of a chemical industry wastewater treatment plant subjected to varying production schedules. *Journal of Chemical Technology and Biotechnology,* 79(10), 1084–1091.

Villiers, R.V., Berg, E.L., Brunner, C.A., and Masse, A.N. (1971). Municipal wastewater treatment by physical and chemical methods. *Water and Sewage Works,* 118(8), R62.

Vogna, D., Marotta, R., Napolitano, A., Andreozzi, R., and d'Ischia, M. (2004). Advanced oxidation of the pharmaceutical drug diclofenac with UV/H_2O_2 and ozone. *Water Research,* 38(2), 414–422.

Waite, T.D. (2002). Challenges and opportunities in the use of iron in water and wastewater treatment. *Reviews in Environmental Science and Biotechnology,* 1(1), 9–15.

Wall, A.E. (1980). Economically removed toxics. *Water and Wastes Engineering,* 17(2), 43.

Weber, W.J., Hopkins, C.B., and Bloom, R. (1970). Physico-chemical treatment of wastewater. *Journal of the Water Pollution Control Federation,* 42(1), 83.

Weber, W.J., Jr. (1967). Fluid-carbon columns for sorption of persistent organic pollutants. Advances in Water Pollution Control Research. Proceedings of the Third International Conference on Water Pollution Research, Washington, DC, Water Pollution Control Federation, 1, 253.

Wuhrmann, K. (1968). Objectives technology and results of nitrogen and phosphorus removal processes. Advances in Water Quality Improvement, University of Texas Press, Austin.

Wukash, R.F. (1968). New phosphate removal process. *Water and Wastes Engineering,* 5(9), 58.

Wyffels, S., Boeckx, P., Pynaert, K., Verstraete, W., and Cleemput, O.V. (2003). Sustained nitrite accumulation in a membrane-assisted bioreactor (MBR) for the treatment of ammonium-rich wastewater. *Journal of Chemical Technology and Biotechnology,* 78(4), 412–419.

Yildiz, E., Keskinler, B., Pekdemir, T., Akay, G., and Nuhoglu, A. (2005). High strength wastewater treatment in a jet loop membrane bioreactor: kinetics and performance evaluation. *Chemical Engineering Science,* 60(4), 1103–1116.

Yoon, T.I., Lee, H.S., and Kim, C.G. (2004). Comparison of pilot scale performances between membrane bioreactor and hybrid conventional wastewater treatment systems. *Journal of Membrane Science,* 242(1–2), 5–12.

Yuan, W.L. and Hsu, P.H. (1970). Effect of foreign components of the precipitation of phosphate by aluminum. Paper presented at the 5th International Water Pollution Conference, July–August.

Zuckerman, M.M. and Molof, A.H. (1970). High quality reuse water by chemical-physical wastewater treatment. *Journal of the Water Pollution Control Federation,* 42(3), 437.

ALESSANDRO ANZALONE
Polytechnic University of Puerto Rico

J.K. BEWTRA
University of Windsor

HAMBDY I. ALI
Ain Shams University

PLANNING

Planning is a field of study which encompasses a number of related physical development and social and scientific functions including land use analysis, transportation planning, housing policy, economic analysis, environmental planning, urban and rural development and redevelopment. The broad goal of planning is to provide thoughtful examination of physical development and related public policy initiatives.

Planning has evolved from the early social concerns of 19th Century urban life—crowding, squalor and unhealthful living conditions focuses attention on such concepts as the public interest and on laws which protect the public health and safety. Controls on the location of unhealthful industrial uses such as slaughterhouses have evolved into land use planning and zoning controls. Concern for over crowding and provision of adequate light and air are now measured by housing analysis and population surveys.

Planning is an inter-disciplinary field which brings an understanding of public health, legal and social issues, and architectural design principles to develop theories of the history and future of development patterns. From principles of urban design appropriate street layout, open space and urban densities are derived which provide light and air. Traffic transportation planners identify the need for mass transit systems as well as traffic improvements. Standards in environmental planning provide a background to evaluate costs, benefits and impacts of new developments and initiatives.

Planners are a diverse and loosely defined group who identify themselves by the branch of planning they engage in, thus land use planners, public health planners, economic planners, housing planners and transportation planners working side by side and within their own particular area of expertise in government and private industry.

The advancement of the public interest and the protection of the public health, safety and welfare are two principles which continue to support a myriad of planning initiatives. The goals can be seen within the studies of environmental planning: wetland management and other conservation efforts are viewed as controls which protect the public interest; air quality controls protect the public health, safety and welfare.

Planners utilize academia as a home base, where the lessons of urban planning are taught, together with analytical methods for the determination of social and scientific standards and criteria urban design and aesthetic principles. Professional planning societies also provide information and forums for discussion on planning issues. The American Institute of Certified Planners (AICP) offers a membership admittance test which serves to provide a roster of qualified professionals. Ethics standards promulgated by the AICP attempt to self-police the profession. Only two states, New Jersey and Michigan, license professional planners.

The key analytical methods and models of planners include survey and sampling methods, ranking strategies, program evaluation, location planning, population forecasting and models for measuring impact of landuse actions including air quality analysis, transportation capacity, and employment, economic and fiscal impact.

Land use planners utilize planning principles to determine appropriate locations for land uses within a specified area. The land use planner may be employed by a town or city to conduct long range planning and maintain and advance the master plan through day to day land use decision-making. Comprehensive plans provide forecasting tools for a variety of resources, population trends and social service needs, resource projections and utility and infrastructure investments as well as transportation planning, roadway improvements, and conservation lands and sensitive environmental lands.

The role of government in planning in the United States is generally regarding as a local activity, for example, defining the appropriate use for the abandoned industrial plant in your home town is a local decision-making process. However, planners are engaged at each level of government; that is, the federal, State and local level. From these broad perspectives the role of a planner shifts to reflects the changing role of government.

The federal government provides planning through public policy initiatives and laws and regulation. The two functions adhere to the two pronged principles of planning; the public interest and the public health safety and welfare. The public interest is served through programs which promote a particular public policy initiative which is encouraged, such as open space or local park funding. Federal programs which administer funding for state and local actions, such as Community Development Block Grants for urban redevelopment promote the public interest. Promulgation and enforcement of regulations which protect the public health safety and welfare include the Clean Air Act, the Clean Water Act, and the Safe Drinking Water Act.

The National Environmental Policy Act of 1969 is a key milestone within the federal government, aimed at promoting the general welfare, in that it requires federal actions to be subject to both inter-governmental and public review. The intent of this law is to provide for a careful analysis of the likely effects of an action prior to use of federal funds. This law serves to open up the process of government decision-making to the public. In addition, the NEPA legislation provides a model for similar environmental scrutiny at the state and local level, thus providing a network of environmental planning at all levels of government.

In terms of land use decision-making, there is no central planning. The federal government role is confined to public policy, grants in aid and regulatory control (the environmental controls placed on filling of wetland areas is an issue which most closely approximates central planning in that the effect of these regulations can often nullify development proposals which may be encouraged at the local level).

Information gathering, such as census data provided by the Department of Commerce, is the key central planning function used universally by planners in all disciplines and sectors: statistics and forecasts are used by public agencies, private consultants and commerce and industry to understand and plan for communities with regard to population, transportation, housing, the economy and employment and a host of demographic and industrial data.

At the State level, planning functions are both assumed from the federal government and conveyed to local governments. State governments serve to administer federal programs and assume responsibility for enforcement of some key planning and environmental responsibilities such as the Clean Air Act. Often, compliance with federal goals in one area forms the basis for government aid or funding of related public initiatives, for example, efforts to comply with air quality goals are often related to transportation funding. The State works closely with federal agencies as a conduit for policy and funding of government programs, often creating mirror agencies for administration of transportation, historic preservation and environmental protection. Also parallel and in association with the federal government, States provide for the collection and dissemination of statistical information such as the census and economic forecasting.

The state government assigns and provides to local governments the authority to plan and zone. The power to plan in its narrow sense, land use control, is based upon the police power. Planning is seen to protect the health safety and general welfare of the public. The majority of planning work is done at the municipal level (a certain amount of master planning and regional coordination is provided at the country and state level). The first element in municipal planning is the formulation of the Master Plan. The Master Plan or comprehensive plan is an amalgam of public policy objectives and practical information for future planning such as existing demographic information and trends with projections of the future to the benefit of landuse plans, transportation plans, utilities, and education. The master plan should be a unique reflection of both the physical elements of the city and the people who live in it. Thus the master plan, which culminates in a landuse map becomes the guide to the development regulations of a city.

The zone plant sets forth the physical controls on uses and densities within the jurisdictional area. The physical layout of the city is assigned uses, densities and standards for development in a land use plan. A zone plan presents the regulations which carry out a given land use plan. The integrity of a zone plan is as strong as the development review boards. Planning Boards and Zoning Boards are charged with maintaining the integrity of the plan with the power to review subdivision, site plan and various requests.

The site plan review process illustrates the typical planning exercise and scope and range of municipal review powers. The site plan review, and variance requests, for example, a use variance, are subject to review and scrutiny from a municipal board with regard to the merits of the site plan, the need for the use, how the use is consistent with the master plan, how it advances planning goals of the community and what hardships or unique physical characteristics would justify a departure from set development standards and proscribed uses.

Many municipalities have established environmental commissions, architectural or historical resource review boards all which serve to advise in the site plan review pursuant to the general welfare. The environmental review at the municipal level should encompass the full range of relevant issues to asses the scope of the proposed activity and its likely impact on site and surrounding resources.

ELIZABETH McLOUGHLIN
PS & S Key Span

PLANNING FOR NEW PROCESSES: ENVIRONMENTAL ASPECTS

INTRODUCTION

Planning for a major new facility must address the environmental impact of both the construction and operational phases of the project. It is essential to optimize alternatives, while evaluating performance relative to regulated emissions and ambient standards and to develop a cost effective permitting strategy.

For large scale projects, Quig (1980) recommends a highly integrated project approach for environmental compliance early in the planning stage based upon historical siting, licensing, engineering and construction experience with similar sized plants.

Strong emphasis on early process work is necessary to understand environmental impacts. This and other front-end engineering and planning should be executed in very close coordination with the staff charged with documenting the licensing effort. Extensive use of specialists is generally required.

The major federal acts to be addressed are:

1) National Environmental Policy Act, (NEPA), 1969.
2) Clear Air Act Amendments, revised 1990.
3) Federal Water Pollution Control Act Amendments, 1972 (FWPCA).
4) Resources Conservation and Recovery Act (RCRA), 1976.
5) National Historic Preservation Act of 1966.
6) Historical and Archaeological Preservation Act of 1974.
7) Endangered Species Act, 1973.

These federal and selected state environmental acts essentially address the following: Land Use Aspects (fuel storage, exclusion or buffer zones, waste disposal, zoning, and demography); Water Resources (availability and competitive uses, wastewater complexities and water quality, hazardous wastes, and waste heat); Air Quality/Meteorology (attainment/nonattainment areas, in terms of offset policy and lowest achievable emission rate; newsource performance standard for particulates; NO_x and SO_2; prevention of significant deterioration in Class I, II, III; stack height credit; hazardous wastes; minor meteorologic changes); and Regulatory (multiple lead agency involvement, licensing strategy, feasibility of concept, permit requirements, and federal/state implementation).

The environmental, health safety, and socio-economic impacts discussed above highlight the areas of concern which must be considered in the site characterization studies and subsequent reporting of the project compatibility with the proposed location. Baseline conditions must be identified in the areas of potential impact. The characterization of the environment, the definition of the process operations and the identification of the potential impacts are the elements required for input to a comprehensive program of facility design for impact mitigation. As such, the development of an environmental statement of the project serves as feedback to the design effort with the result being a facility licensable from the environmental viewpoint.

To illustrate the procedure we shall present a typical example, namely planning a new coal gasification plant. Technical details of gasification are discussed elsewhere in this Encyclopedia under Coal Gasification Processes. The example will focus on regulatory requirements and siting considerations.

REGULATORY REQUIREMENTS

The first step in any program of this nature is to define the regulatory requirements associated with the construction and operation of the proposed facility. This will define specific limitations and establish generally the study requirements for the program as they relate to the environmental, safety, health, and socioeconomic aspects of the development.

AIR QUALITY RELATED REGULATORY REQUIREMENTS

Federal Requirements

At the Federal level, this project will be required to comply with the following air quality regulations and requirements. *Primary and Secondary National Ambient Air Quality Standards* (NAAQS) A demonstration showing compliance with NAAQS must be made to EPA for approval to commence construction. This would involve modeling the anticipated plant emissions and imposing the resultant concentration increases on representative ambient air quality conditions and comparing these with NAAQS. Information necessary for this demonstration would include the facility emissions as discussed earlier and the ambient air quality developed from a monitoring program or from representative data as available.

New Source Performance Standards (NSPS) The proposed air quality control system (AQCS) for the facility must be designed to comply with existing NSPS for the coal preparation facilities (e.g., particulates), the gas turbine component (e.g., NO_x) and the auxiliary boiler (e.g., SO_2) of the plant. Since NSPS do not exist for the coal gasification component, appropriate AQCS Best Available Control Technology (BACT) evaluations will be performed to select the control system. In addition the AQCS design will have to be reviewed with the EPA for approval to construct.

Prevention of Significant Deterioration (PSD) No construction can commence until the PSD permit has been obtained. The report and application for the permit would have to consider the following: the emissions from the total facility; a BACT review for any regulated pollutant (NAAQS, allowable increments, NESHAP), which the plant emties above "de minimis" values, an air quality review for all pollutants emitted, after controls are applied, over the "de minimis" emission rates unless it were demonstrated that the air impacts of those emissions would not exceed the air quality impact "de minimis" values. As part of these demonstrations ambient air quality monitoring would have to be conducted for the same pollutants for which BACT demonstrations would be required unless representative monitoring data are available.

National Emission Standards for Hazardous Air Pollutants (NESHAP) The discharge to the atmosphere of pollutants regulated under NESHAP is not anticipated for this type of facility. However, tracking of EPA's continued development of NESHAP should be carried out to ensure compliance with the regulations as they develop.

New stationary sources and modifications to major stationary sources are required by the Clean Air Act to obtain permits prior to construction of a new process facility. The stringency of permit requirements depends on the regional status of its compliance with ambient standards for particular pollutants. For example, in zones having acceptable air quality referred to as "attainment areas" for a specific pollutant, the permits are of the prevention of significant deterioration (PFD) type. The code of federal regulations, US EPA Title 40 CFR, 51.166, specifies the set of minimum PSD air quality permit requirements to warrant approval by the US EPA. The primary objective of PSD is to insure new major sources and modifications of existing sources comply with NAAQS. Specific public notice requirements and subsequent hearings allow for public comment to be part of the PSD review process. On the other hand, in a non-attainment area, NAA permits are required. NAA permits address area improvement of pollutant levels and falls under the state's supervision, through a State Implementation Plan (SIP) enforced by the US EPA and DOJ. Either type of permit is subject to New Source Review (NSR). The physical change triggering regulation of pollutants is usually 100 or 250 tons per year, depending on the industrial source category. As of early 2005 the definition of major modification for coal fired power plants has come under dispute in the courts (see the discussion at the end of this article for further information).

State Requirements

The state may have air quality related requirements which will affect the proposed project.

State requirements may include a permit to construct a facility if the construction of operation of the facility will release air contaminants into the atmosphere. The applicant must submit a completed application for Approval of Emissions and an Emission Inventory Questionnaire (along with a copy of the PSD Application) which show compliance with state air quality standards, toxic substance limitations and emission control requirements.

WATER QUALITY RELATED REGULATORY REQUIREMENTS

Federal Requirements

At the Federal level, the major laws affecting the discharge of liquid effluents from the proposed facility are as follows:

Clean Water Act (CWA) Under the CWA, the proposed project will require a National Pollutant Discharge Elimination System (NPDES) permit before commencing construction and operation. The application to the EPA for these permits would be based on the conceptual design of the wastewater control systems which could ensure compliance with effluent limitations and water quality standards. Where effluent limitations are not specified for discharges from certain facilities, limitations on discharges from similar operations would be used as a guideline for design of the wastewater control systems. These designs would be used to support the application for an NPDES permit. The NPDES permit and the work effort necessary for its preparation will also address the discharge of toxic pollutants listed on the Section 307(a) toxic pollutants list and any other toxics discharged from the plant.

A section 404 permit is required by the Corps of Engineers for the discharge of dredge or fill material in the navigable waters of the United States. This permit is required for the river structures associated with the facility and would be prepared and obtained concurrently with the Section 10 permit required under Rivers and Harbors Act (see discussion below) and the NPDES permit.

Application for either an NPDES permit from the EPA or a Section 404 permit from the Corps will trigger the NEPA review process and is the basis upon which the preparation of an environmental report (ER), in support of a Federal EIS, is considered necessary.

Rivers and Harbors Act of 1899 (RHA) Under Section 10 of the RHA, any construction activity in a navigable waterway requires a permit from the Corps of Engineers. This permit will be required for the construction of the intact and discharge structures and the barge loading/unloading facilities. It would be submitted jointly as a common permit application with the Section 404 Dredge and Fill Permit application.

State Requirements

The state often has water quality related regulations and requirements which must be complied with before approval to commence construction and/or operation of the proposed project can be obtained. They are typically as follows:

A state regulatory agency may require that a certificate of approval be obtained prior to construction of a treatment facility for handling industrial wastes. A report containing detailed information about the operation of the treatment facility must be developed and submitted prior to construction.

Regulations may also require the submission of a permit application prior to discharge from an industrial source. The State may also issue a certification in accordance with the Clean Water Act which confirms that discharges from the facility will comply with effluent limitations and water quality standards.

SOLID WASTE RELATED REGULATORY ACTIVITIES

Federal Requirements

The major Federal law governing the handling and disposal of solid waste is the Resource Conservation and Recovery Act of 1976 (RCRA). The most significant sections of RCRA are Subtitle C, which deals with Hazardous Waste Management and Subtitle D, which deals with Non-hazardous Waste Management. Regulations pursuant to Subtitle C of RCRA address identification and listing of hazardous waste, standards applicable to generators, transporters, and owners and operators of hazardous waste treatment, storage and disposal facilities and permit requirements for treatment, storage or disposal of hazardous waste. The project will require a permit for disposal of any solid wastes determined to be hazardous by the criteria in Section 3001 regulations. Operation practices of the solid waste management facility are also regulated. In this regard the work necessary to determine the nature of the solid waste generated by this facility must be carried out. If the wastes are determined to be hazardous (Section 3001 Criteria) the applicable requirements of Subtitle C or RCRA must be incorporated into the facility design.

Regulations promulgated under Subtitle D or RCRA establish criteria for the development of State plans for management of solid waste. No requirements are directly imposed at the Federal level.

State Requirements

State plans for the management of solid waste (Hazardous and Non-hazardous) may be at varying stages of development. An application for a permit to operate a hazardous waste management facility may be filed with the state's DNR if any solid wastes to be generated at the proposed facility can be classified as hazardous.

NATIONAL ENVIRONMENTAL POLICY ACT (NEPA)

The major provision of NEPA which significantly impacts the planning and scheduling for major industrial facilities is the need for Federal agencies contemplating major actions, such as issuing permits, to prepare an environmental impact statement (EIS).

In the case of this coal gasification facility, the requirement for a Federal EIS would be triggered by the application for an NPDES permit from EPA and/or a Section 404 or Section 10 permit from the Corps of Engineers for anticipated river structures. Upon designation of the lead agency based on discussions with the various Federal agencies and submittal of applications for permits, the EIS would be prepared according to CEQ final regulations.

SITING THE PROJECT

Geology, Topography, and Soils

Geology studies should be performed to describe the soils, geologic and topographic setting of the site, particularly with respect to structural and topographic control of the local and regional groundwater flow systems. A secondary, albeit very important, purpose is the identification of potential geological hazards within the site area.

Information sought includes physical and chemical soil characteristics, general topography, paleontology, and geological framework. Descriptions are sought for aquifer systems and characteristics including their name, thickness, depth, stratigraphy, and areal extent. Mineral production and unique geologic/geomorphic features will be documented. Pertinent data is summarized in tabular and/or graphic format.

The results of the geology studies primarily define the soils, topographic, and geologic setting of the site. Potential impacts references these descriptive settings to evaluate impact magnitudes. The impact of plant site preparations and construction or localized site topography, soils and erosion characteristics, and site physical and economic geology are assessed. Geological hazards discussed include excessive slopes, unstable soils and fault zones.

Groundwater Hydrology and Water Use

The purpose of the groundwater studies is to understand the physical and chemical characteristics of the groundwater regime. This allows for an accurate assessment of groundwater impacts resulting from the proposed action in addition to formulation of mitigative measures to help alleviate these impacts. In addition, information necessary for the design of solid waste handling facilities as prescribed under RCRA is developed.

Information sought includes general topography and geological framework, description of aquifer systems and characteristics including their name, thickness, depth, stratigraphy, and areal extent; seasonal groundwater levels, rate, and direction of flow; aquifer hydraulic properties including permeability, transmissivity, and storativity; surface water/groundwater inter-relationships; location of aquifer recharge and discharge areas; ground water quality; and domestic, industrial, and municipal groundwater well distribution and characteristics. Long and short term regional and site specific (within 5 miles of the site) data is sought. Special efforts are made to document the location of contaminated areas.

Groundwater sampling can be conducted in conjunction with surface water sampling. Samples are taken quarterly from monitoring wells.

This information can then be evaluated in light of the projected facility emissions and demands on the area's resources. The issues and concerns to be addressed include:

1) Use of groundwater by the plant, and the effects of lowering water levels or pressures for this reason or for construction purposes.
2) Changes in water quality, or effects on rocks/deposits, caused by accidental leaks or spills, effluent discharge, slag or scrubber sludge pits, surface water and the like. The potential impacts identified are evaluated in light of their magnitude and importance. Extra attention is paid to those judged significantly high in either value.

By early identification of stresses that might affect the natural systems, steps can be taken to minimize the impacts or alleviate them to an acceptable degree. Mitigative measures that can be taken during plant design, construction or operation, such as adding clay liners, for example, can be evaluated and described.

Using the results of plant design, the initial impact evaluation and the adopted mitigative measures, a final evaluation is made of the effects of the construction and operation of the proposed plant and ancillary facilities on the natural environmental systems.

This activity evaluates expected effects of the proposed plant both during and after construction of the groundwater hydrologic environment. Each effect is evaluated as to its unavoidable adverse effects and favorable effects.

Surface Water Hydrology and Water Use

The purpose of the surface water studies is to determine, develop and present the surface water quantity and quality characteristics of the site and its surrounding environs. These data and information are analyzed and evaluated in recognition of the proposed facility's operation and construction related characteristics to determine and project potential effects and impacts on the surrounding surface water. Specifically, the objectives are:

1) To provide a quantitative description of the hydrologic setting of the site and its vicinity including any stream flow characteristics (i.e., flood and low flow frequencies, seasonal ranges, averages, and historical extremes), and the physical and chemical water quality characteristics of source and receiving waters. Annual and seasonal ranges and averages are developed.
2) To identify the other water uses (withdrawals as well as discharges) and users including the location and quantities involved;
3) To identify the existing water quality criteria and regulations affecting plant discharges, and
4) To evaluate the impact of construction and operation of the proposed plant on adjacent surface waters, with regard to the applicable water quality criteria, and related permit requirements.

The data and information needed for the description of the hydrologic setting of the surface waters of the site and evaluation of the plant's impact include the following:

1) Geographic and topographic maps of the site area containing varying degrees of local and regional details to delineate the drainage basin and its drainage patterns.
2) Watershed characteristics such as geometry, slope, vegetation types and density, and soil types to derive rainfall-runoff relationships (empirical runoff coefficient).
3) Records of rainfall events to estimate overland flow.
4) Records of stream flows from gaging stations on local water courses. These data are used in defining statistical stream flow characteristics.
5) Meteorological data including air temperature, relative humidity, solar radiation, wind speed and evaporation data, and thermal plume calculations (as needed).
6) Records of various water users, locations of withdrawal, quality, and quantities involved.
7) Proposed plant site location map, grade elevation, drainage pattern, character of soil types, and cover.
8) The physical and chemical water quality characteristics of the surrounding surface waters.
9) The facility description and operational characteristics relating to the discharge quantity and quality. In addition, construction procedures, methods, schedules, and erosion control features are needed.

In addition to a water quality characterization program, the required data is collected through existing sources. This would involve a thorough search, review, and compilation of the existing hydrological data base. Appropriate Federal, State, and local agencies are contacted and interviewed and published regulatory materials is reviewed to gain information regarding other water users and water laws affecting the plant construction and operation.

A field monitoring program is carried out to obtain water quality characteristics of intake and discharge waters. Water quality samples are taken quarterly from selected stations.

A hydrological assessment of the construction phase is undertaken to:

1) identify changes in drainage patterns and possible effects on flooding potential,
2) identify changes in riparian terrestrial habitat areas,
3) identify the potential for erosion and local soil losses.

These impact areas are addressed, and mitigating measures are specified for their control.

For the operational phase, the various aqueous discharges from the plant are inventoried and evaluated. Results of plume analyses, if appropriate, are critiqued with respect to compliance with applicable water quality criteria and standards. Recommendations concerning the potential optimization of the plant water management plan are made to reduce or eliminate environmentally objectionable discharges. Consumptive water use for the plant is identified to determine the effects of plant operation on intake waters and downstream users. Recommendations concerning optimization of the plant water management plan and the use of alternate or supplemental sources of water are made, if warranted.

Ecology

The ecological studies are designed to generate and assemble pertinent data to determine the status of threatened and endangered species, and commercially or recreationally important wildlife species, and to identify and locate sensitive, unique, and critical aquatic, riparian, and terrestrial habitat areas in the site area. Additionally, the status of commercially or recreationally important fish in any site intake and discharge waters is determined. The biological setting is then analyzed in light of the proposed plant construction and operational characteristics to arrive at assessments of impact potential.

Terrestrial Ecology

In evaluating the impact of the project on the terrestrial environment, the work objective is to assess both construction and operation of the facility utilizing "baseline" data developed and secured from field programs, literature, and agency contacts. Animal species, occurrence, abundance, distribution, and preferred habitat associations and principal ecological interactions are determined. Habitats are identified and described as natural plant communities within the site. Additionally, discussions and data gathering activities focus on vertebrates and prominent otherwise important plant community components.

Construction-related effects largely result from vegetation and habitat removal, which often constitutes the major impact of a major industrial facility on terrestrial communities.

Assessment of vegetation loss due to land clearing is based on previous identification and mapping of regionally productive rare, or otherwise important vegetation types. In this regard, the role of plants in soil stability warrants detailed consideration. Effects of facility construction on wildlife is also evaluated in terms of important habitat areas. Attention is focused on those species which appear sensitive to habitat loss (e.g., species already limited by factors relating to habitat availability), which function as critical components of a community, or which are considered "important." The latter category refers to wildlife designated uncommon, threatened or endangered, or wildlife of recreational or economic value.

An assessment of project operation including existing and proposed effects on vegetation must consider stack and cooling tower emissions. Predicted ground-level concentrations of stack emissions and cooling tower salt are compared to exposure levels considered thresholds for possible injury or damage, and to exposure levels documented as injurious under filed conditions.

Potential effects of facility operation (existing and proposed units) on wildlife from stack emissions, dust, increased human activities, and noise are evaluated. Additionally, the potential for bird collisions with plant components is evaluated. The magnitude of a potential bird-collision problem is evaluated from data compiled during field studies. Included in the impact assessment analysis is the use of the plant site by wildlife during station operation. Collecting ponds and other waste bodies provide habitat for waterfowl and amphibians, while areas cleared during construction and allowed to revegetate (or which are replanted) potentially provide habitat for a variety of species.

Aquatic Ecology

The aquatic ecology of the site intake and discharge waters as well as habitat removal associated with barge facilities is addressed, habitat and food web relationships of the system characterized and potential impacts to the system estimated. Data requirements are met by literature review, interviews, and discussions with local fishermen and scientists and field collections. While data gathering focuses on fish species, particularly the commercially or recreationally important species, other biotic elements of the lotic and lentic environments are identified.

Evaluation of potential impacts of the construction and operation of the proposed facility consist of projecting the effects of the various activities on the description of existing environmental conditions developed as a result of the field program, literature review and agency contacts described above.

The primary construction impacts likely to affect the aquatic habitat are those associated with the construction of the intake facility and secondarily increased erosion due to construction. Impact assessment of construction activities centers primarily on habitat lost or denied due to actual physical placement of structures and habitat degradation. Attention is focused on those species which appear sensitive to habitat loss; which function as critical components of the aquatic community; or are considered "important" (rate, threatened or endangered, or of commercial or recreational value).

Assessment of operational impacts centers on the effects of water withdrawal and the associated losses to the fish community due to entrainment and impingement. Potential changes in the population structure all addressed. Losses are estimated from population densities and from the field sampling program. Entrainment losses are expressed as "adult-equivalents" if warranted for important species. Potential discharge effects (thermal and chemical) are based on information developed from the literature review and input of engineering parameters.

Land and Waterway Use

The purpose of the land and waterway use—demographics effort is to:

1) determine the existing land use of the site and existing and future land and waterway use patterns in the surrounding area in order to assess any conflicts which may exist and to evaluate any impacts on land use that may occur from the construction and operation of the plant; and
2) determine the population growth patterns of the area in order to assess the impact the plant will have on nearby towns and communities in the area.

Based on the existing land uses, and the analysis performed by other disciplines such as terrestrial and aquatic ecology and air quality, impacts upon adjacent land uses caused by construction and operation are estimated. This impact assessment includes the impact of storage pond construction, noise, dust, plant appearance, stack emissions, cooling tower fogging and salt depositions, and construction stage traffic activity on residential, recreational, agricultural, and other adjacent land uses in the area, as well as the compatibility of the proposed plant with local land use plans, aesthetics, and regulations. Specific attention is given to the type and relative value of the land uses to be preempted or adversely affected by plant construction and operation.

The impacts of the increase in activities on a river if applicable is estimated including additional barge traffic staging in the area and the impact of these activities on existing movements and facilities in the vicinity of the area.

The demographic impact assessment consists of comparing the population projections for the study area to the expected population influx to be caused by plant construction and operation. The comparison is done by taking the estimated plant-related population influx as a percentage of the total projected population of the area to be affected by the incoming workers and families.

Socioeconomics

The purpose of the socioeconomics studies is to determine and describe the existing socioeconomic base for the plant region and surrounding major towns and to assess the changes, either positive or negative, which would occur as a result of the construction and operation at the proposed site.

The existing socioeconomic based is described for those areas likely to be impacted by the influx of construction and operational employment for the plant. Information required to describe the socioeconomic base of the area includes the following:

1) peak number of construction workers by craft during each year of plant construction;
2) estimate of the number of immigrant construction workers expected during construction of the plant;
3) existing and future capacity of the schools, hospitals, fire, sewer, etc., facilities in site area;
4) local government fiscal capabilities and local tax structure and tax bases;
5) employment and income statistics; and
6) economic base studies.

Socioeconomic impacts can result from the influx of immigrant construction workers to the area around the plant. This occurs when the construction force required to build the plant is fairly large and there are a number of large construction projects competing for the labor supply in the area.

The socioeconomic demand analysis qualitatively compares the demand for service facilities, and employment during the construction and operation of the plant with the baseline socioeconomic projections. Any perceived increases in demand for local facilities is qualitatively evaluated with respect to the cost of the facilities and the ability of local units to finance them.

Impacts associated with plant operation to be assessed include an evaluation of the change in local tax structure as a result of a large influx of new tax revenues to the local governments and the impact associated with the relocation of plant operating personnel into the area.

Noise

The purpose of this effort is to sample the existing ambient noise levels surrounding the proposed site, and to estimate the environmental noise impact produced by the plant operation.

In order to properly assess the noise environmental impact, plant noise emissions should be evaluated in terms of any State or local noise regulations. Consequently, various State and local regulatory agencies are contacted to determine the status of the regulatory constraints that might be imposed on the plant operation and construction noise emissions. In the absence of any such constraints, US EPA's guidelines for the protection of "Public Health and Welfare," as indicated in the "Levels Document" (550/9-74-004) are followed.

A literature review is conducted to assist in identifying the major sources of noise of the plant and in quantifying them. The search includes various professional journals, other environmental reports, and manufacturers' publications. Construction schedule, equipment list, general arrangement drawings, project description manual, and operational parameters of major plant equipment (Forced Draft Fans, Turbine Generator, Pumps and Motors) are obtained from the appropriate sources.

The facility noise levels are then evaluated and assessed in terms of existing regulations or guidelines and any potential restrictive conditions in either the working environment or the general site environment are identified. In addition, potential limitations to equipment are identified as are appropriate mitigative measures.

Cultural Resources

Cultural resource studies involve a review of appropriate records, and site-related literature to identify sensitive archaeological, historical, recreational, and aesthetic resources in the project area.

Most of the information required for cultural resource studies is available from State and Federal cultural resource agencies and societies. An on-site survey is conducted to locate any cultural resources eligible for the National Register of Historic Places.

Project components which affect significant cultural resources are identified and the magnitude of the impact evaluated. Mitigation alternatives are addressed. If serious impacts are discovered they should be brought to the attention of the developers promptly so that policy decisions can be made at the earliest opportunity to rectify the situation.

Air Quality/Meteorology

The purpose of this program is to obtain and analyze Air Quality/Meteorology data so that: the site can be characterized; the air quality implications of the facility can be evaluated; mitigative and control measures can be developed; and an Environmental Report and PSD application can be prepared.

The data required for the Air Quality/Meteorological program work efforts relate to: (1) The air quality/meteorological characterization of the existing site and region; and (2) The facility's atmospheric emissions and operating characteristics. Specifically, the existing site and region must be characterized in terms of the regulated pollutant such as SO_2, Particulates, NO_x, CO, Photochemical Oxidants, and the local meteorology, including winds, stability, and other physical characteristics. In addition, the facility's emission characteristics including their quantity and quality must be developed so that their impacts can be established.

To establish the required data base it is necessary to gather and update existing emission inventory information, collate meteorological and air quality data, review present and proposed PSD Class I and nonattainment areas, evaluate topographic influences and monitor the region's air quality characteristics.

The task of establishing adequate meteorological and air quality data bases includes evaluating any existing local meteorological and air quality data. The validity of the data and its representativeness with respect to the proposed site must also be assessed. As required, data from other sources is evaluated as a basis for comparison with local data, or as a supplement to local data where necessary. The objective is to establish meteorological and air quality data bases which are most representative of the proposed site.

A report should be prepared to provide technical support for a construction permit application under the PSD provisions of the Clean Air Act of 1977. Described in the report are the data bases, methodologies and models utilized in the analyses. The PSD report also includes appropriate maps, summary tables and figures necessary to display relevant information such as locations of plant sites, PSD Class I and nonattainment areas, and resultant pollutant concentrations.

An atmospheric impact assessment of the proposed and alternative cooling tower types is included in the ER. Operational impacts of the tower to be considered include elevated visible plumes and deposition of cooling tower drift. Computer modeling is utilized to predict the impact of these occurrences.

Ground level fogging/icing is also addressed. Computer modeling is utilized to predict the frequency and duration of ground level fogging and icing for the alternative cooling towers. The potential for interaction of the cooling tower and stack plumes must also be addressed.

National Weather Service (NWS) data can be supplemented by any available meteorological data to the fullest extent possible to develop the estimates of cooling frequencies, and salt deposition rates are given on an area basis and include more detailed information, as necessary, for any sensitive receptors.

Health Implications

The purpose of this work effort is to identify and evaluate the potential health concerns, including estimates of offsite exposures that may result from facility operation at the site. Once the concerns are identified, the need for controls and the feasibility of the gasification plant at a particular site from a public health perspective can be evaluated. Changes in coal type, process or waste treatment systems can be addressed if needed to mitigate a potential health concern. This allows and insures that alleviation of potential health problems is an integral part of the facility planning.

The health implications to the offsite population are identified and assessed. Specific analytical measurements from pilot plant studies, available information from similar industrial plants, and other existing studies regarding the health implications from coal conversion processes are used. The overall approach involves scaling the results of specific pilot plant runs and other study results to approximate a commercial size facility; applying standard air dispersion models and waste dilution criteria in order to predict exposure concentration; and evaluating the predicted concentration against known information on the toxicity of each contaminant.

The identification of potential public health concerns requires not only the estimation of exposure levels but also the evaluation of the relative toxicity of each chemical species. Consequently, a review of the toxic properties of each identified chemical substance or chemical group is required. At the conclusion of this evaluation each contaminant or chemical group is categorized into one of four groups—potentially significant health problems, potentially minor health problem, no expected health problem, and those for which insufficient information is available.

The results of this evaluation is then utilized to develop control systems and/or mitigative actions for the facility. In addition, the results are presented in the ER and discussed in terms of a cost/benefit framework.

With regard to occupational health, worker exposure to toxic substances is a potentially serious problem which could significantly lessen and limit the benefits of alternative fuel technology. As such, its implications must be carefully evaluated in the planning and design phases of this project. The Occupational Safety and Health Act (OSHA) contains basic worker protection guidelines and specific regulations which establish industry procedures for the protection of workers from exposure to potentially toxic or health impairing substances.

However, the current regulations do not specifically address a coal gasification process and only limited operating experience is available from existing gasification plants. Consequently, identification of potential occupational health problems must be done in an indirect manner through comparison with other industries.

This occupational hazard analysis yields identification of potential hazards, definition of possible control measures for as many of those hazards as possible, and identification of areas of concern where insufficient knowledge or control methodologies exist. In addition it provides input to:

1) design of worker protection programs to be implemented at the plant,
2) design of engineering controls to minimize worker exposure to hazardous substances, for example, isolation of process steps, ventilation changes, pressure control, etc.

Most of the procedural information, repeated from the article by Quig and Granger (1983) remains valid today. For a more quantitative treatment of the effluent emissions observed during plant operation the reader is referred to the study of Holt (1988) on the Cool Water plant and to the current Encyclopedia article, Coal Gasification Processes.

PERMITTING FOR LANDFILL GAS ENERGY RECOVERY

Purpose

New York State Air Guide 41 (1996) provides guidance on the permitting of emissions from municipal solid waste landfills, including the use of landfill gas for energy recovery, flares and, also, passive venting, as per the following:

Background

Landfill gas (LFG) is generated by the decomposition of wastes in all municipal solid waste landfills, regardless of age or size. The total volume of gas generated is a direct function of the quantity of wet, decomposable refuse available; however, the rate of gas generation can vary greatly over time, depending on numerous factors (such as the volume of waste, the depth of the landfill and the amount of rainfall the landfill receives), most of which are uncontrollable. Landfills the accept waste water treatment plant sludge for disposal tend to generate more LFG than those that do not. LFG is not generated until the available oxygen supply has been consumed and the decomposition process becomes anaerobic. The typical composition of LFG is essentially the same at all landfills and at all points within the landfill. The typical composition of LFG is:

Methane	50–58%
Carbon Dioxide	35–45%
Hydrogen	1–2%
Oxygen	1–2%
Nitrogen	2–5%
Non Methane Organic Compounds (NMOCs)[1]	3–5%

LFG can, and should, be used for energy recovery. The energy content of LFG comes entirely from the methane component, which has a basic heating value of 1,000 Btu/standard cubic foot (scf). Since the nominal concentration of methane in LFG is approximately 55%, the heating value of raw LFG is approximately 550 Btru/scf, although this figure can, and will, vary somewhat. By comparison, natural gas is composed of 95% methane, giving it a basic heating value of 950 Btu/scf.

At the majority of landfills in New York State, LFG is currently uncontrolled or passively vented to the atmosphere. Recovering and combusting such gas into useful energy will virtually eliminate harmful emissions from a fuel that is otherwise wasted. This also prevents the pollution associated with the use of fossil fuels (i.e., SO_2). If LFG is not combusted, it will still escape to the atmosphere through the path of least resistance (diffused from landfill, vented or flared).

Federal Regulations

Air In accordance with the Clean Air Act, the U.S. Environmental Protection Agency (EPA) has proposed New Source Performance Standards (NSPS) under 40 CFR 60 Subpart WWW for municipal solid waste landfills. These NSPS will affect landfills that began construction or modification after the standard was proposed (5/30/91) or existing landfills that have accepted waste since November 8, 1987. It must be noted that this proposed rule is currently being developed. The rule is subject to change and it is possible that it will not be released. However, the guidelines contained in this proposed rule should be used in developing a permit for the use of landfill gas. Additional information regarding this proposed rule is included in Appendix A.

In a recent Federal court case in Pennsylvania (Ogden Products Inc. vs. New Morgan Landfill Co.), the court ruled that the landfill in question is subject to New Source Review since it has the potential to emit more than 50 tons per year of volatile organic compounds. This decision, combined with the NSPS for landfills proposed by the EPA, will make all new landfills subject to the requirements of the CAA, particularly if the landfill has the potential to emit volatile organic compounds at levels exceeding air quality standards.

Hazardous Waste When LFG is recovered, it tends to cool, and some condensate is formed. It is stated in Section

124 of the Superfund Amendments and Reauthorization Act (SARA) of 1986 that if the aqueous or hydrocarbon phase of the condensate removed from the gas recovered from a landfill meets any of the characteristics of a hazardous waste (i.e., it fails the TCLP test), the condensate shall be considered hazardous waste and regulated accordingly. This section is an amendment to the Resource Conservation and Recovery Act (RCRA), but is not part of RCRA. Since this provision of SARA is not actually part of RCRA and there are no implementing regulations in 40 CFR, it may be binding upon EPA, but RCRA-authorized states (such as New York) are not obligated to enforce its requirements. This issue could arise if LFG is to be recovered from a municipal landfill that meets the size and NMOC criteria cited in Appendix A and is included on the Superfund priority list of inactive hazardous waste sites.

State Regulations

Solid Waste 6 NYCRR Part 360 has requirements for the control of LFG during both the active life of the landfill and after the landfill is closed. While the landfill is in operation, the owner must periodically (i.e., quarterly) monitor for the presence of LFG at or above 25% of the lower explosive limit (LEL) at on-site structures and any off-site areas. When the landfill is closed, an LFG control system must be included in the closure plans to prevent the migration of concentrated LFG away from the site and to prevent damage to a landfill cap. LFG is lighter than air and will tend to rise causing the overlying cap to rise also. Generally the LFG is allowed to vent to the atmosphere through a porous gas vent layer that leads to gas vent risers spaced at approximately one vent per acre.

Part 360-2.16 contains the regulations regarding LFG recovery facilities. These regulations require that anyone proposing an LFG facility obtain a permit to construct and operate the facility. The application for a permit must contain an engineering plan, engineering report and an operation and maintenance plan.

Hazardous Waste As cited above, some LFG condensate may exhibit hazardous waste characteristics. In an October 20, 1992 declaratory ruling applying to the Freshkills Landfill, the Department excluded landfill gas condensate from being regulated as a hazardous waste. This ruling was based on the grounds that the LFG was derived from a household waste and therefore excluded from hazardous waste regulation under New York State law. However, if the landfill received both municipal and industrial or hazardous waste, the condensate may be hazardous. The condensate would need to be analyzed using the TCLP method to determine if it is a hazardous waste.

Air NYSDEC's proposed Part 201 operating permit program (proposed to comply with the federal Clean Air Act Amendments of 1990) contains an exemption for LFG emissions vented directly (i.e., without a flare or energy recovery device) to the atmosphere that fall beneath major source thresholds as long as the facility is operating in compliance with 6 NYCRR Part 360. Such an exemption will not apply to landfills subject to NSPS or National Emissions Standards for Hazardous Air Pollutants.

A number of landfills in New York State currently use flares or energy recovery for control of their LFG. These emission sources must have a permit from the Division of Air Resources. These permits are issued under 6 NYCRR Part 201. All energy recovery projects produce NO_x in the combustion of the LFG. These projects must control NO_x emissions as required under Part 227-2. For example, if a lean burn internal combustion engine running on LFG is used for energy recovery, the emission limit for NO_x is 9.0 grams/brake horse power-hour (Part 227-2.4(f)).

LFG recovery projects would be affected by either Prevention of Significant Deterioration (PSD) or New Source Review in Nonattainment Areas (Part 231) regulations depending on the location of the project. Projects in nonattainment areas are likely to be affected by NO_x and CO requirements (Note: the entire state is nonattainment for VOC and NO_x because the state is in the ozone transport region). This is because recovering energy with a combustion unit will create NO_x and CO that often require emission offsets to be obtain and the installation of Lowest Achievable Emission Rate (LAER) technology.

The EPA has issued interim guidance stating that sources may be exempt from New Source Review (NSR) provided that the project is environmentally beneficial and there are no adverse air quality impacts. This exemption from NSR is referred to as a pollution control project. The EPA presently expects to complete rulemaking on an exclusion from major NSR for pollution control projects by mid 1996. However, in the case of nonattainment areas, EPA believes that the state or the source must provide offsetting emission reductions for any significant increase in a non-attainment pollutant from a pollution control project.

Presently, 6NYCRR Part 231 allows a Pollution control project exemption only at existing electric utility steam generating units (>25 megawatts of electrical output). Consequently, LFG projects which exceed the applicability thresholds, would have to obtain NO_x offsets at a ratio of 1.3 or 1.15 to 1, depending on the location of the project (i.e., in a severe non-attainment area or in a moderate non-attainment area), and would be required to install LAER technology. Please note that this may change if the EPA determines that this type of project is eligible to become a pollution control project.

Approach to Permitting—DAR

When the economics of an energy recovery project using LFG are favorable, these projects are to be encouraged. The following is the hierarchy of preferred LFG uses:

1. Gas cleaning and upgrade to pipeline quality gas;
2. Energy recovery with gas pretreatment or conversion to a reusable chemical product;
3. Energy recovery without gas pretreatment;
4. Flares with high combustion efficiency (i.e., 98% or greater);
5. Vents, if no economically feasible use for the gas is available.

The following procedure will be used for permitting of energy recovery facilities that utilize LFG:

If the LFG Is Pretreated (i.e., if the constituents other than methane are removed from the gas), then permit as a combustion source with no further emission testing or ambient modeling necessary to satisfy toxic concerns. The permit should address traditional combustion contaminants such as NO_x and CO. However, the permit application for this type of option must include a detailed description of the method(s) to be used for gas pretreatment. LFG can contain up to 50%, by volume, CO_2 (35–45%) and air toxics (1–2%). The pretreatment employed must remove these compounds before the LFG can be permitted for just the traditional combustion contaminants. Note that gas pretreatment will minimize toxic products of incomplete combustion and minimize system corrosion.

If the LFG Is Not Pretreated, then permit as a combustion source and use the total concentration of NMOC emitted to address toxic issues. Note that if the gas is burned, either by a flare or energy recovery process, generally the air toxics will be destroyed. It will be easier and more efficient to regulate the NMOC (or total VOCs) than trying to identify and regulate all contaminants of LFG emissions, since they can vary greatly depending on the waste disposed at the landfill. The permit should address traditional combustion contaminants such as NO_x and CO. The EPA proposed standard of 20 ppmvd NMOC should be used as BACT for the control of untreated LFG used as a combustion source. Periodic stack testing of the emissions is recommended at the discretion of the permit writer.

With regard to compliance with Part 231, the LFG facility may need to obtain NO_x and CO offsets at the ratio applicable to its location (i.e., 1.3 or 1.15 to 1). This requirement may change if the EPA decides that LFG-type facilities are eligible for a pollution prevention exclusion. The permit reviewer will need to exercise judgment to determine if the LFG facility is required to obtain these offsets. As stated above, an LFG facility can be used for energy recovery. While a combustion turbine or internal combustion engine is not normally considered add-on pollution control devises, they do serve the same function as a flare, namely to reduce VOC emissions at the landfill with the incidental benefit of producing useful energy (energy that would otherwise be produced using higher polluting fossil fuels). For an LFG facility the reviewer should proceed as follows:

1. Verify that the NO_x increase has been minimized to the extent practicable;
2. Confirm (through modeling or other appropriate means) that the actual significant increase in NO_x emissions will not violate the applicable NAAQS, PSD increment or adversely impact any air quality related value;
3. Apply all otherwise applicable SIP and minor source and permitting requirements and ensure that NO_x offsets are provided in an area in which nonattainment review applies to NO_x emissions increases.

Coordination within the Department

The use of LFG will require coordination of efforts between the Divisions of Air Resources (DAR) and Solid and Hazardous Materials (DSHM). If a landfill meets the criteria cited above and the emissions from the site must be controlled, the proposed plan for this control should be submitted to both Divisions. DSHM should focus their review of the proposal, based on the requirements of Part 360-2.16. DAR should focus their review on evaluating and permitting the combustion sources that utilize the LFG, as outlined in the previous section. Both Divisions must keep in mind that LFG can be a valuable resource for energy generation and that using this resource will conserve the use of other fossil fuels and permit the re-use of material otherwise considered waste. Further, the respective project managers handling a particular facility's permit application should maintain communication to ensure that there are no unnecessary delays on developing a permit for an LFG facility.

APPENDIX A OF AIRGUIDE 41

NSPS for Municipal Solid Waste Landfills

In accordance with the Clean Air Act, the U.S. Environmental Protection Agency (EPA) has proposed New Source Performance Standards (NSPS) under 40 CFR 60 Subpart WWW for municipal solid waste landfills. These proposed NSPS will affect landfills that began construction or modification after the standard was proposed (5/30/91) or existing landfills that have accepted waste since November 8, 1987. It must be noted that this proposed rule is currently being developed. This proposed rule would require landfills to install active gas collection and control systems if they exceed both of the following criteria:

- design capacity in excess of 2.500,000 Mg (2,700,000 tons); and
- NMOC emission rate in excess of 50 Mg per year (50.05 tpy).

Landfills closed prior to November 8, 1987 or having design capacities less than 2.5 million metric tons will be exempt once this rule is finalized.

The NMOC emission rate is determined by the following equation:

$M_{NMOC} = 2L_0R(1-e^{-kt})\,C_{NMOC}\,(3.595 \times 10^{-9})$ where,
M_{NMOC} = mass emission rate for NMOC, Mg/yr
L_0 = refuse methane generation potential, m^3/Mg refuse (default value = 170 m^3/Mg)
R = average annual acceptance rate, Mg/yr
k = methane generation rate constant, 1/yr (default value = 0.05/yr)
t = age of the land fill in years

C_{NMOC} = concentration of NOMOC, ppmv as hexane (default value= 4,000 ppm)
3.595×10^{-9}= conversion factor

(Note: In the absence of site specific data, use the given default values to determine NMOC emission rate.)

Each landfill of design capacity greater than or equal to 2,500,000 Mg must determine its NMOC emission rate, using the above equation. If the NMOC mass emission rate is less than 50 Mg/yr, the landfill owner would submit this information to the NYSDEC as part of the Part 360 permit. No further action regarding control of the landfill emission would be required. The emission rate will be recalculated annually, with a report submitted to the Department. This report must be submitted to the Department within 90 days of the issuance of a construction or operating permit or initial acceptance of refuse, whichever is earlier.

If the NMOC mass emission rate is greater than 50 Mg/yr, the landfill owner must submit a design for, and install a collection and control system at the landfill. This control system must be designed to reduce, in accordance with 40 CFR 60 Subpart WWW, the emission of NMOCs by 98% by weight.

Wehland and Earl (2004) present the inconsistencies in legal and enforcement interpretations of the PSD requirement for NSR major modifications and hence the criteria by which to determine the need for a PSD permit and the installation of BACT. There is no need to install equipment if the modifications are of the routine maintenance, repair or replacement without increase in unit capacity. For each of the four cases involving coal fired power plants in which decisions were reached prior to 2004, the court's conclusions were different, and in some cases significantly different. In *U.S. vs. Southern Indiana Gas & Electric Co.* (S.D. Ind. 2003), the district court ruled that the utility had fair warning about EPA's decision to review the records of the affected unit, only, in determining if the maintenance was routine, but deferred its final determination to the more general NSR applicability. In *U.S. vs. Ohio Edison* (S.D. Ohio 2003), the court concluded that fair warning was given to the utility in the plain language of the CAA.

The court gave weight to establishing the routine nature of the maintenance by analyzing the all of the factors below for the projects:

(1) budgeting and accounting,
(2) purposes and costs, and the
(3) net emissions increase

In *U.S. vs. Duke Energy* (M.D.N.C. 2003), the district court ruled that EPA's past practices in other cases require that industry-wide rather than the individual unit's past practices were the criteria to be followed. Also the court favored an increase in emissions analysis to be on a "Projected actual", rather than on a "future potential" basis. They also indicated that when determining the applicability of a modification, the NSPS rule of not increasing the maximum hourly emissions rate should be followed.

In the case, *Tennessee Valley Authority vs. U.S.* (11th Cir. 2003), the decision by the circuit court side-stepped the primary differences and ruled against the EPA because of its use of an administrative procedures that deprived TVA of its rights to a hearing. In October 2003 the EPA defined a routine project as one that would cost less than 20% of the total cost of replacing an emitting unit. The new rules and lack of enforcement were being challenged as of 2004. An alternative to the current regulations offered by the authors, "is to eliminate the NSR and incorporate plant-wide emission limits in operating permits." They also recommend modeling and employing maximum achievable control technology (MACT) when modeling results determine controls are necessary to protect the air quality.

REFERENCES

Holt, N.A., EPRI Report AP-5467, Feb. (1988).
Quig, R.H., *Chem. Eng. Prog.,* **76,** 47–54, March (1980).
Quig, R.H. and Granger, T., *Encyc. Of Env. Sci. & Eng. Vol. 1,* 103–113 (1983).
NYS Department of Environmental Conservation, Albany, New York, Air Guide 41, p. 1–5, March (1996).
USEPA TITLE 40 CFR 52.01–52.2, P. 9–55 US Government Printing Office, Washington DC: July 1, 2000.
Ibid Title 40 CFR 51.166, p. 167–186
Wehland C. and L. Earl, *Clearing the Air,* Environ. Protection, p. 20–22, 66, June 2004. *http://www.epa.gov/ttn/nsr/ gen/wkshpman.pdf*

ROBERT H. QUIG
THOMAS GRANGER
Ogden Products, Inc.

EDWARD N. ZIEGLER
Polytechnic University

PLANNING WATER SUPPLY AND SANITATION PROJECTS IN DEVELOPING NATIONS

INTRODUCTION

It is estimated that over 1.5 billion people in the world are without adequate water supply and waste disposal facilities.[1] Waterborne diseases kill an average of 25,000 people every day, and millions suffer the debilitating effects of the diseases.[2] Concerned about the need for safe water supply and sanitation, the United Nations (UN) declared 1981–1990 as the *International Water Supply and Sanitation Decade* with the professed goal of supplying potable water for all the people of the world by 1990, and also of providing all people with the means to safe disposal of human excreta. This goal was highly optimistic in view of the fact the nearly 80 percent of the population in the developing countries does not have access to piped water supply, and even a large percentage lacks public sewers and household waste disposal systems.[3,4]

At the end of the International Water and Sanitation Decade, it was estimated that almost 31 and 44 percent of the world population was lacking safe water supply and adequate human waste disposal, respectively. The global population being 5.28 billions, total population currently having inadequate water supply and sanitation is therefore staggering.[5–7] Based on an assessment of the successes and failures of the Decade, it is believed that much has been accomplished, but progress is needed on all fronts: rehabilitation and operation of existing systems, training of personnel, financing, and new construction to achieve the goals of the Decade within a reasonable time in the future. In this article many factors that contribute to successes and failures of the Water Decade goals are reviewed. The discussion is divided into (1) understanding of the needs, (2) appropriate technology for water supply and waste disposal, (3) commitment, (4) financial resources, (5) training of people to plan, design, build, operate and maintain water and sanitary projects, and (6) role of developed nations.[8–10]

UNDERSTANDING THE NEEDS

Adequate supply of safe water and basic sanitation are the foundation of health. Water pollution and poor sanitation are probably responsible for 80 percent of the morbidity and mortality in the developing countries. It is estimated that over 900 million cases of diarrhea related illness occur each year, resulting in the death of over 3 million children. Common diseases associated with polluted water are grouped in Table 1.[11] Prevention of such diseases depends on the improvement of the quality of the water supply, personal hygiene, food handling and preparation; as well as the provision of adequate sanitation, including sanitary facilities for human waste disposal.

The importance of personal hygiene specially on the part of mothers and children is the key to beneficial results in disease prevention. Therefore, more intensive activities directed to improve personal hygiene should be emphasized in the developing countries. Furthermore, a meaningful program of providing safe water supply in developing countries must also include safe disposal of human excreta, basic education, and improvement of sanitary conditions.

The major constraints on the progress of the International Drinking Water Supply and Sanitation Decade have been operation and maintenance and rehabilitation of the systems that were built earlier. Broken-down and poorly functioning facilities waste money, are a threat to health, and discourage future investments. Also, inappropriate technology is at the root of many water supply and waste disposal problems in the developing countries.[12,14]

APPROPRIATE TECHNOLOGY

Water Supply

The goal of the UN Water Decade was to provide "clean water and adequate sanitation for all by 1990". It should be noted that this did not mean a tap and a flush toilet in every house. Reasonable access to safe water is usually understood to be within 500 m of the household; for many developing countries, adequate sanitation probably refers to some technology intermediate between the water flush toilet and the simple pit latrine. The scope of the task of the UN Water Decade was truly enormous considering that the two priority areas are the rural population (71%), who often have to travel long distances for water, and the crowded urban poor (25%), for whom water supply often is grossly inadequate; 87% of rural population and 47% of urban fringe areas lack adequate sanitation.[12] Naturally, achieving universal coverage would mean not only supplying almost one billion people currently without safe water supplies and the 1.7 billion people without sanitation, but matching population growth as well. Annual investment is currently in the range of 15 to 20 billion dollars per year—about 2 to 3 percent of gross investment in developing countries. If the share

TABLE 1
Common disease associated with contaminated water and poor sanitation

Disease	Common Vehicle
Water-borne diseases	
Amoebic dysentery	Diseases transmitted by ingesting contaminated water and food
Bacillary dysentery	
Cholera	
Criptospridiosis	
Gastroenteritis	
Giardiasis	
Hepatitis	
Leptospirosis	
Paratyphoid fever	
Salmonellosis	
Typhoid	
Water-washed diseases	
Conjunctivitis	Lack of adequate quantity of uncontaminated water, and poor personal hygiene create conditions favorable for their spread
Hookworm (Ankylostoma)	
Leprosy	
Scabies	
Skin sepsis and ulcers	
Trihcuriasis	
Whipworm (Enterobius)	
Yaws	
Water-based diseases	
Bilharziosis	Diseases caused by infecting agents by contact with or without ingestion of water. An essential part of the life cycle of the infection agent takes place in an aquatic animal
Dracunculosis	
Oncholersosis	
Philariosis	
Schistosomiasis	
Treadworm	
Fecal-disposal diseases	
Clonorchiasis	Diseases caused by infecting agent mostly contracted by eating uncooked fish and other food
Diphyllobothriasis	
Fasciolopsiasis	
Paragonimiasis	
Water-related vectors	
Arbovirus	Diseases transmitted by insects which live and breed close to water. Infections are spread by mosquitoes, flies, and insect bite
Bancroftian	
Dengue fever	
Encephalitis	
Filariasis	
Hemorrhagic fever	
Malaria	

Adapted in part from Ref. 11.

remains constant over the next fifteen to twenty years, economic growth will allow investment to double in real terms, to 30–40 billion a year.[13] Therefore, selection of appropriate technology and service levels must be established in bringing down the capital costs of water supply and sanitation projects. It was also demonstrated in the first half of the Decade that the biggest priority is correcting the inadequate training of the operating personnel.[14]

Selection of appropriate technology for use by the developing countries requires (1) understanding of the cultural background, regional environmental conditions, and local needs; (2) selection of appropriate systems that will be within the financial (minimizing initial investment cost as well as maintenance cost) and technical (minimizing operational failures) resources; (3) training paraprofessional rural water technicians (using available labor where possible instead of expensive imported equipment); (4) promoting complementary activities to help people to obtain the most benefits from the systems (so that users will receive long-term benefits of these basic services); and (5) selection of the most appropriate water source and the most appropriate energy source for conveying water to its users.

The different sources of small community water supply systems in developing countries can be categorized as follows: (1) groundwater, (2) rain water, (3) springwater, and (4) surface water. Depending on the sources of water supply, many techniques have been found effective and applicable to the water and sanitation programs in the developing countries. Some of these techniques are presented below.

Groundwater In areas where *groundwater* is readily available at moderate depth, constructing a number of wells fitted with hand pumps is by far the cheapest means of providing a good water supply.[14] Although, community water systems piped under pressure to households and public standposts may be an ultimate goal, many areas will realistically have to seek hand pumps as an interim if not an ultimate measure. The trouble with most hand operated water pumps used in developing countries is that they are not hardy enough and require frequent maintenance. Pump parts are usually expensive to buy and are difficult to make locally to fit the pump. Development work started at the Consumer Research Laboratory (CRL) in England has led to the *snappy* plastic pump: a simpler, cheaper and hardier device that has many beneficial features. A complete description of such hand pumps is given by Sattaur.[15] Additional information on other types of hand pump manufactured in developing countries may be found in Ref. 15. Hofkes provides some other techniques and methods of groundwater withdrawal used by small water supply systems in developing countries.[11]

Rainwater In developing countries rainwater is sometimes used to supplement the other water supply sources. In some tropical islands rainwater is the only source of domestic water supply. Rainwater harvesting requires adequate provision for the interception, collection and storage of the water. Generally, cisterns are built to collect the runoff from the roofs. Water quality preservation is very important and some basic measures should be followed to exclude bird droppings, insects and dirt from the stored water. Also, storage in cool conditions, exclusion of light, and regular cleaning is essential. Simple disinfection devices may be very useful.

Rainwater catchments are relatively simple to construct and maintain. It is expected that these systems will be widely used in the future. In Kenya, concrete jars used as storage tanks are said to be the most popular appropriate technology. Their popularity is growing among the villagers in Thailand where the construction and maintenance of these units is undertaken by technicians of the Sanitation Division of the Department of Health. The technician directs the voluntary labor of villagers in constructing concrete storage tanks reinforced with bamboo. The Villagers then repay the costs of the tanks in 12 monthly installments. The owners of these tanks, having contributed so much of their time and money into their construction are usually very keen to operate and maintain it properly.[16] Design details and economics of many cisterns, storage tanks, and other rainwater harvesting systems may be found in Refs.[17–21]

Spring and Surface Water There are many situations in the developing world where water is available from a nearby stream or spring. Various devices have been constructed to utilize the energy available in flowing streams to pump the water to the point of use.

The more successful devices, that will represent substantial savings in pumping water are the hydraulic ram pumps (hydram) and low-head turbine pumps. One converts pressure energy to mechanical work and the other converts kinetic energy to mechanical work. These two practical devices have been described and compared in detail by Schiller.[22] He observed that the hydram is easier to operate and maintain, but installation of the turbine pump is easier and simpler.

Before the Water Decade program started the water quality control was given least importance in developing countries. Perhaps, the major reason was inappropriate operation and maintenance of the equipment. The primary goal of any water project started in the Drinking Water Decade was to provide a facility and system that can be operated, maintained and managed at local levels giving self sufficiency with proper personnel and financial needs.[23] Local capabilities for operation and maintenance must be developed through training local maintenance personnel, establishing a locally managed maintenance fund, selecting an easily maintained technology, and requiring capital contributions from the community to increase local sense of ownership and responsibility for the systems.[24] Finally, a successful drinking water program requires extensive local participation (materials and labor) and minimize dependence on overseas materials and equipment for construction and operation of facilities.[25]

Human Waste Disposal

Rapid growth of the urban populations (mainly because of migration from rural areas) has led to severe problems in providing human waste disposal systems. Only 32 percent of the population in the developing countries is directly connected to sewer systems. Therefore, the collection and disposal of human wastes constitute a serious environmental and health problem.[26] Several alternatives are available; each has its own limitations and constraints. Certainly there is no single option appropriate to all situations. The range of options described below is sufficient to cover the vast majority of the situations within the low income communities of the developing countries.

One of the fundamental principles of community sanitation is to remove all putrescible matter, particularly human wastes. A satisfactory excreta disposal method must satisfy the following requirements:[26]

1) Should not be accessible to flies or animals
2) Should not cause odors or unsightly conditions

3) Should be simple and inexpensive to construct and operate
4) Should not cause ground water or surface water contamination
5) Should not cause soil contamination.

The appropriate technology for human waste disposal in developing nations fall into two major applications (1) rural, and urban areas without sewers, and (2) urban areas having municipal sewers. Appropriate technology for each of the above applications are discussed below:

Rural and Urban Areas without Municipal Sewers The environmentally acceptable methods of human waste disposal in rural areas include various types of sewerless composting toilets. Common examples of such toilets are bored-hole latrine, pit privy, vault privy, septic privy, chemical toilet, soil or composting toilet, methane forming toilet, and box-and-can privy.[26,27] These devices are simple and inexpensive to build and operate. Each of these devices is briefly discussed below.

Bored-hole Latrine: Bored-hole latrines consist of a hole in the ground 25 to 60 cm in diameter, and 2 to 3 meters deep. The hole may be braced to prevent it from caving. A concrete slab with a hole cut in it may be placed over it, and the entire assembly is made into an outhouse. When the hole is filled, the structure is moved to another hole. Fly and odor problems are controlled by keeping the hole covered, and dropping dry soil in the hole on a regular basis. Deep bored-hole latrines present danger of ground-water pollution.[26]

Pit or Vault Privy: Pit or vault privies can serve the needs of homes, schools or groups of homes. A concrete pit or vault is constructed in the ground and a toilet seat is located on the top. The seat cover must be kept closed to prevent flies from entering. Odor is a problem. A vent pipe raised over the roof has been helpful in odor control. In 3 to 4 months when the pit or vault becomes full, there is the unpleasant job of cleaning it. Some vaults have two compartments. When one compartment is full it is kept closed for 2 to 3 months while the other side is used. During this time, the excreta is decomposed, and cleaning becomes less objectionable. Pit or vault privies with separate back covers have also been used. These back covers help in the cleaning operation.[26]

Septic Privy: Septic privies utilize a liquefying tank similar to a septic tank. The waste is digested and overflow may be discharged into a percolation field. The toilet has a shallow trap. After use the toilet must be flushed by pouring 2 to 3 liters of water.[26]

Chemical Toilet: Chemical toilets consist of a tank in which waste accumulates. The toilet seat is provided directly over the tank. Caustic soda (sodium hydroxide (NaOH)) is most commonly used to kill the bacteria and to liquefy the wastes. The caustic soda dosage is 1 to 2 kg of sodium hydroxide per month for a family of four. The tank is emptied periodically. If caustic soda is used the tank contents can be carefully applied over farming land. Many types of chemical toilets are available from various manufacturers.[26,28,29]

Soil or Composting Toilet: This is dry toilet capable of being used indoors. The toilet consists of a wooden frame with toilet seat, and a bucket lined with a plastic bag. The toilet is started with a layer of dry soil 10 to 15 cm deep. Users must sprinkle several scoops of soil after each use. Once or twice a week, the bag is removed, tied and stored. After a decomposition of several weeks the contents are spread over farm land.[30]

A more involved is the *Clivus Multrum* composting toilet. This toilet was developed and has been successfully used for some time in Sweden and is now manufactured in the United States.[31] The toilet is capable of composting toilet waste, kitchen wastes, and even leaves and grass. It has three sloping chambers and is started with a layer of peat moss four to five inches thick. Design and operational details may be obtained in Ref. 31. Another toilet similar to *Clivus Multrum* toilet is the *Toa-Throne* compost toilet also developed in Sweden.

Another dry composting toilet that has been proven highly successful in Vietnam, offers a suitable system for all developing nations.[32] The toilet consists of two-holer outhouse in which only one hole is used at any one time. There is no raised seat, instead each hole is placed in a squatting plate at floor level. The urine flows out and soaks into the ground outside the outhouse. The fecal wastes go into the other hole. The user must drop dry soil or ash after use. When the toilet is two-thirds full, the rest is filled with soil, and the vault is tightly sealed. The other side is started. After a composting period of approximately 45 days the contents are taken out and used over the farm as rich compost.[33,34]

Methane Farming Toilet: In India small biogas plants are extensively built in rural areas. These biogas plants anaerobically digest animal manure and other organic wastes, and the methane is used to light, heat, and power the farms. Singh gave the basic design and construction details of such toilet.[34] Other designs of anaerobic digesters utilizing human, animal and other organic wastes are used in Kenya and Brazil.[35–37] Morris developed a package unit for methane generation from human wastes.[37] Khandelwal provided design and operational details on dome-shaped biogas plants used in India.[38] Eusebio and Rabino provided design of large biogas plants used in Philippines, India and Africa.[39]

Box-and-can Privy: The toilet system consists of a wooden box, the lid and toilet seat. The can or pail is removable through the top or side of the box. The cans are removed manually under regular scavenger service, emptied, cleaned, and replaced. This system is most commonly used in cities that do not have a sewerage system.

This type of system is associated with odors and serious fly problems if the lids are not tightly closed after use. For satisfactory operation there should be a regular scavenger service, preferably under a governmental supervision with ordinances covering types and size of box and can, method of emptying and cleaning of the cans, and ultimate disposal of the human waste. A speedy and convenient way is to use identical toilet boxes and cans for the entire community. The scavengers remover the can and place it in his vehicle without emptying, and replacing a clean can. The vehicle when full is taken to a central disposal facility for emptying, and replacing with clean cans. The disposal facility should be located in a remote area and should utilize anaerobic digesters whereby the recovery and sale of methane gas, and digested sludge as soil conditioner can be achieved. The central disposal facilities should be equipped with running water, hose, and chemical solution for washing and cleaning the cans.

Experience has shown that such systems based on payments by the residents to private scavengers have been unsatisfactory. The service is poor, irregular and wastes may be dumped uncontrolled upon land and in waters.

The box-and-can system should be used only as a temporary means of waste disposal with continued effort to replace it by other more acceptable methods. Furthermore, it should be carried out only under the supervision of trained personnel and under strict governmental control.

Urban Areas with Municipal Sewers The wastewater collected through municipal sewers contains large volumes of water. Therefore, treatment technology is much more complex than discussed above. Treatment schemes to achieve secondary level of treatment (90% organics and total suspended solids removal) include screening, grit removal, primary sedimentation, biological treatment (activated sludge or trickling filter), final clarification, and disinfection. Other physical, chemical and biological treatment processes may be added to remove phosphorus, nitrogen, and additional organics and suspended solids. The solids fraction (sludge) is digested aerobically or anaerobically.

Although, such treatment technology may be desirable for the developed nations, their use in developing nations may not be appropriate. Such treatment processes are costly to build, and complex to operate. Desirable treatment methods for developing nations may include the following:

1) stabilization pond followed by effluent reuse on farming lands,
2) Imhoff tanks followed by stabilization pond and effluent reuse on farming land, and
3) upflow anaerobic sludge blanket process.

Imhoff tanks provide primary sedimentation, and anaerobic digestion of settled sludge. The methane is collected for energy source and digested sludge is used as compost. The stabilization pond (oxidation pond or lagoon) is an earthen basin that retains the wastewater for sufficient time to stabilize the organic matter, and destroy large percentage of pathogens. The effluent is used for irrigation, or discharged into natural waters.[40–43] The *upflow anaerobic sludge blanket* process (UASB) was developed in Netherlands and is extensively used and tested in the developing nations.[44] In this process the wastewater enters the bottom of the reactor and percolates up through the sludge blanket where organics are converted to methane and carbon dioxide by anaerobic organisms. The gas is collected, and treated effluent is drawn off from the top.[34]

COMMITMENT

The water supply and sanitation schemes in developing countries cannot be successful without the willingness of the government to commit the resources to undertake water supply and sanitation projects. Unfortunately, the governmental priority in this area is quite low. Most of the developing countries set their priorities as follows: (1) agriculture, (2) industries, (3) energy resource development, (4) education, (5) commerce and transportation, (6) family planning, (7) housing and urban development, and (8) water supply, and sanitation, and environmental control.[45] Such priorities have been set on account of necessities. It is necessary that the developing countries give needed priority on water and sanitation.

FINANCIAL RESOURCES

The success of any program in any country depends upon the financial commitments of the respective governments. Developing nations (except for the oil exporting nations) are poor and have very limited funds. A 1975 study indicated that villagers if they have to pay more than 50 cents per month for water supply, they would not participate in the program.[45] Their ability to pay for water and sanitation is very limited. Success of the program will depend upon outside funds, and local labor.

TRAINING AND MANPOWER DEVELOPMENT

It is well recognized that the success of any program requires well trained people. Unfortunately, past experience has shown that many internationally supported training programs were not successful because the emphasis was incorrectly placed. Engineers from developing nations when trained in the western world generally learn the theory and design of most sophisticated unit operations and processes in water and wastewater treatment and environmental engineering. When they return to their countries, they are generally eager to utilize such technologies, although in most cases these may be quite inappropriate. Furthermore, foreign consultant's unfamiliarity with the cultural and operating competence of the people in developing countries have resulted in selection and design of technology that have not functioned.

What is really needed is a program where training at all levels may be effectively provided. The includes training of central government officials and engineers; local government officials, residents and technicians; on-the-job training of operators; and users education. Government officials set the priorities and make funding decisions. Engineers have responsibility to evaluate, select and implement the appropriate technology. Local officials participate and support the program. The technicians and operators have the responsibility for continued operation and maintenance of the system. And finally, users education is essential as they are the ultimate beneficiary.

In order to effectively provide such a complex training opportunity, programs must be developed in the developing nations. Such training programs must be developed as academic programs, short courses, seminars, and internships to address different audiences yet meet the specific needs of each. Brief discussion is given below.

Academic Programs

Academic programs include regular undergraduate and graduate programs at universities, and technology or vocational programs in polytechnics and colleges. Both programs have different objectives.

University Programs Regular academic programs such as master's and bachelor's degrees in civil, sanitary or environmental engineering, biology, public health, and so on. have been modeled along the lines of western universities. These programs often may not be directly applicable to the needs of the developing nations. It is necessary to modify or develop new programs tailored for their national needs. As an example, graduates in civil, sanitary or public health engineering must also know community interactions, socio-economic implications, relationships between service levels and health and well-being, that dictate the success or failure of any water supply and sanitation projects, especially in rural and urban settings of developing nations.

Technology Programs The technology or vocational programs should be directed towards training technicians, health inspectors, and community workers. Specialized training is needed so that these personnel can undertake field work, organize the community, conduct training, and assist the community in the selection and installation of water supply and sanitation facilities, and provide continued operation and maintenance of such facilities. Special training is needed for personnel to motivate people and implement the programs in rural areas where approximately 75% of the population lives.

Short Courses and Seminars

In corporation with the health ministry specialized short courses of one- or two-week duration should be conducted at universities, colleges, and high schools. These short courses should be designed to provide training to full-range of people including decision makers, engineers, operators and community workers.

Decision Makers The government officials who have the responsibility for setting up priorities, and allocation of funds must be aware of the relationship between service levels and health and other benefits, cost of construction and operation, and short-range and long-range implications of the investments in water supply and sanitation projects.

Engineers and Designers The intensive short courses designed for engineers and designers should cover the successes and failures of different projects, improvements needed in future systems, public education, and how to relate health conditions and project objectives in order to develop the most cost-effective system for the specific conditions.

Operators and Community Workers Intensive short courses should be conducted for operators and community workers in basic topics ranging from chemistry, biology, vectors, communicable diseases, to metal working, equipment repairs, operation and maintenance, leak repairs, book keeping, to management of public works projects.

On-the-job Training On-the-job training programs of varying periods with or without short courses should be developed to train field workers and inspectors and local people in construction techniques and operation of water supply and sanitation systems. Such practical training can be a powerful tool for manpower development.

Primary Health Care Workers Health education or sanitation education should be provided at village level. Villagers, literate or illiterate, should be trained as primary health care workers, or *barefoot doctors,* whose training should be emphasized on prevention, and curative care. Six to 12 months intensive training could be provided at local high schools or through health education service.

Users Education Training of villagers in water supply and sanitation must also be provided as part of user or adult education program. This can be achieved by mass campaigns by health education extension services. Health education must also be instituted in primary schools. This will provide the broadest, and most dependable coverage to younger groups. The program must include community sanitation; hygiene habits in home; fly rodent and other vector control; water contamination and use of water supply devices; and proper use and maintenance of various types of privies and latrines.

Training and Educational Material, and Education of Trainers

It is important that training, educational and technical material should be in the form of posters, technical books, manuals, photographs, slides, films and audio-visual aids. Audio-visual equipment has been highly successful in training personnel, and illiterate population in rural areas. A portable power generator may be necessary for rural areas without electrification.

Educational material must be carefully prepared for the specific audience. General health related material covering disease transmission, vector control, water and food protection, personal hygiene, should be prepared for general public with no education. Information on public use of water supply and sanitation devices should also be prepared for general audience. Technical and semitechnical information should cover the design and construction; and operation and maintenance of different types of water supply and sanitation devices presented earlier in this paper.

The other most important of all training program activities is the selection and training of instructors or trainers. Technicians and supervisors should be trained so that they can teach their colleagues, subordinates or interns on the job. Education of trainers should be provided in schools, through short courses, audio-visual aids, and job training.

ROLE OF DEVELOPED NATIONS

Developed nations can play a major role in helping the developing nations to achieve the original goals of the UN

Water Decade. The assistance may come in the form of funding of the international agencies (UN, World Bank, etc.) and international development agencies in their own countries.

The technical assistance and transfer of know-how can also be achieved in the form of advisory services to the governments, utilities, and institutions. The assistance can be provided by short-term and long-term consultants for on-job training, development of educational material, and training of personnel and manpower development.

REFERENCES

1. Kalbermatten, J.M. and F.W. Montanari, Water for the world: The scope of the challenge, *Journal American Water Works Association,* June 1980, p. 308.
2. Muyibi, S.A., Planning water supply and sanitation projects in developing countries, *Journal of Water Resources and Management,* **118,** July/August 1992, pp. 351–5.
3. Rao, S.V.R. and John A. Rasmussen, Technology of small community water supply systems in developing countries, *Journal of Water Resources Planning and Management,* July 1987, p. 485.
4. Witt, Vicent M., Developing and applying international water quality guidelines, *Journal American Water Works Association,* April 1982, p. 178.
5. Anonymous, Human Development Report 1991, UNDP, *Oxford University Press,* New York.
6. Rural water supply, *World Water and Environmental Engineer,* June 1990, p. 46.
7. Theme Interaction, World Water, *Journal American Water Works Association,* June 1988, p. 33.
8. The World Bank, "Development and the Environment", World Development Report 1992, *Oxford University Press,* New York.
9. Churchill, A., *Rural Water Supply and Sanitation: Time for a Change,* World Bank Discussion Paper no. 18. Washington, D.C. 1988.
10. Bris, J. and D. de Ferrantis, *Water for Rural Community: Helping People to Help Themselves,* World Bank, Washington, D.C. 1988.
11. Hofkes, E.H., *Small Community Water Supplies, Technology of Small Water Supply Systems in Developing Countries,* John Wiley & Sons, Great Britain, 1983.
12. Schiller, E.J., The United Nations Water Decade and North American Engineers and Planers, Chapter 1, *Water Supply and Sanitation in Developing Countries,* Ann Arbor science, 1982.
13. The World Bank, "Development and the Environment", World Development Report 1991. *Oxford University Press,* New York.
14. Wolman, Abel, The next five years, *Journal American Water Works Association,* June 1985, p. 12.
15. McJunkin and E.H.A. Hofkes, Hand-Pump Technology for the Development of Groundwater Resources, Chapter 4, *Water Supply and Sanitation in Developing Countries,* Ann Arbor science, 1982.
16. Sattaur, Omar, Developing countries switch to sanappy plastic pumps, *New Scientist,* November 26, 1987, p. 36.
17. Schiller, E.J., Rooftop Rainwater Catchment Systems for Drinking Water Supply, Chapter 7, *Water Supply and Sanitation in Developing Countries,* Ann Arbor Science, 1982.
18. Grover, B., Harvesting precipitation for community water supply, *World Bank,* Washington D.C., 1984.
19. Maddocks, D., An introduction to methods of rain water collection and storage, *Appropriate Technology,* **2,** No. 3, 1975, pp. 24–25.
20. Maddocks, D., Methods of creating low cost membranes for use in the construction of rainwater catchment and storage systems, *Intermediate Technology Publications, Ltd.,* London, 1975.
21. Watt, S.B., Rainwater storage tanks in Thailand, *Appropriate Technology,* **5,** No. 2, 1978, pp. 16–17.
22. Schiller, E.J., Renewable Energy Pumping from Rivers and Streams, Chapter 5, *Water Supply and Sanitation in Developing Countries,* Ann Arbor Science, 1983.
23. Austin, John H. and James K. Jordan, Operation and maintenance of water systems in developing countries, *Journal American Water Works Association,* July 1987, p. 70.
24. Karp, Andrew W. and Stephen B. Cox, Building water and sanitation projects in rural Guatemala, *Journal American Water Works Association,* April 1982, p. 163.
25. Monk, Robert, Terry Hall, and Mohammed Hussain, Real world design: Appropriate technology for developing nations, *Journal American Water Works Association,* June 1984, p. 68.
26. Ehlers, V.M. and E.W. Steel, *Municipal and Rural Sanitation,* McGraw-Hill Book Co., NY, 1965.
27. Ports, D.D., What's in the future for composting toilet, *Compost Science, Journal of Waste Recycling,* **18,** No. 4, July–August 1977, p. 16.
28. Qasim, S.R., Onsite wastewater treatment technology for domestic waste and special applications, *The encyclopedia of Environmental Science and Engineering,* Gordon and Breach, Science Publishers, Inc. New York, Third Edition, 1991, pp. 813–21.
29. Clark and Groff Engineers, Sanitary waste disposal for navy camps in polar regions, *Final Report on Contract NBY-32205,* Salem, Oregon, May 1962.
30. Leich, H.H., UN ignore waterless sanitation devices, *Compost Science, Journal of Waste Recycling,* **18,** No. 3, May–June 1977, p. 7.
31. The Clivus Multrium Composting Toilet, *Compost Science, Journal of Waste Recycling,* **15,** No. 5, Nov.–Dec. 1974.
32. Jounge, L.D., The Toa-Throne- A new compost toilet, *Compost Science, Journal of Water Recycling,* **17,** No. 4, Sept.–Oct. 1976.
33. Leich, H.L. Sanitation for the developing nations, *Compost Science, Journal of Waste Recycling,* **18,** No. 5, Sept.–Oct. 1977.
34. Singh, R.B., Building a biogas plant, *Compost Science, Journal of Waste Recycling,* **13,** No. 2, March–April 1972, p. 12.
35. Methane farming in Kenya, *Compost Science, Journal of Waste Recycling,* **13,** No. 5, Nov.–Dec. 1974, p. 15.
36. Company formed to advice our methane plant, *Compost Science, Journal of Waste Recycling,* **15,** No. 5, Nov.–Dec. 1974, p. 15.
37. Morris, G., Methane for home power use, *Compost Science, Journal of Waste Recycling,* **14,** No. 6, Nov.–Dec. 1973, p. 11.
38. Khandelwal, K.C., Dome-shaped Biogas Plant, *Compost Science, Journal of Waste Recycling,* **19,** No. 2, March–April 1978.
39. Eusebio, J.A. and B.I. Rabiino, Research on biogas in developing countries, *Compost Science, Journal of Waste Recycling,* **19,** No. 2, March–April 1978.
40. Metcalf and eddy, Inc., *Wastewater Engineering: Treatment, Disposal and Reuse,* 3rd Edition, McGraw-Hill Book Co., New York, NY., 1991.
41. Viessman W. and M.J. Hammer, *Water Supply and Pollution Control,* 4th Edition, Harper & Row, Publishers, NY, 1985.
42. Qasim, S.R., *Wastewater Treatment Plants: Planning, Design and Operation,* Technomic Publishing Co., Lancaster, Pa., 1994.
43. Thery, Daniel, The biogas programs in India and China. *Resource Management and Optimization,* **1,** No. 4, Oct. 1981, Harwood Academic Publishers, NY.
44. Anaerobics attack tropical wastewaters, *World Water,* August 1988, p. 31.
45. Dallaire, Gene, UN Launches International Water Decade, US role uncertain, *Civil Engineering, American Society of Civil Engineers,* March 1981, p. 59.

SYED R. QASIM
The University of Texas at Arlington

POLLUTION EFFECTS ON FISH

INTRODUCTION

During the past twenty-five years, man has subjected the earth to an ever increasing variety of chemical substances never before encountered by living organisms. Many of these foreign substances, which originated from industrial, agricultural and municipal sources, are seriously threatening the natural environmental conditions established over geologic time by evolutionary processes. Organisms, including man, are finding it more and more difficult to escape the consequences of exposure to these compounds, many of which are known to interfere with life sustaining biochemical and reproductive mechanisms. Little is known, however, concerning the effects the vast majority of these pollutants will have on plants and animals that inhabit this planet.

The aquatic environment is the ultimate destination of most of these contaminants as a result of rainfall and runoff from the lands. The establishment of water quality requirements for the protection of fish life must, therefore, become an item of first priority. The task is complicated, however, by the fact that different species of fish as well as different developmental states of the same or different species exhibit wide variations in their sensitivity or tolerance to foreign materials. In addition, the basic biochemical and physiological information needed to evaluate the harmful effects foreign substances have on organisms is severely lacking for fish and other aquatic organisms.

It is the purpose of this chapter to briefly discuss the current state of knowledge in regard to the response of fishes to environmental contaminants. Because of the serious lack of information in the area of fish toxicology,[†] it is not possible, nor is it within the scope of this discussion, to present even an outline of the toxic effects of the many foreign substances which fish encounter. Instead, representative compounds have been selected to illustrate the principles that are believed to play a major role in determining how fish respond to environmental contaminants. Unfortunately, these principles are based almost entirely on research performed over the last several decades dealing with the toxicologic responses of mammalian species, including man, to drugs and foreign compounds. Until equivalent research is performed on fishes, we will be forced to rely on these mammalian principles when attempting to evaluate the effects foreign compounds have on fish species.

MECHANISMS OF TOXICITY

Foreign substances can injure organisms through a variety of mechanism but these can generally be grouped into two categories: specific and non-specific injury. Chemicals that produce non-specific injury usually do so at the site of exposure, such as skin, respiratory membranes, oral mucosa, and intestinal mucosa. Frequently, such injury is the result of the caustic or corrosive nature of the chemical with the cellular responses being directly related to the concentration of toxicant within the cell or tissue involved. Agents producing this type of injury are commonly referred to as "primary irritants" because they induce local, minor to severe inflammatory responses and occasionally extensive necrosis (death) of cells. Excellent examples of these chemicals are acids, bases and aldehydes, all of which are strong irritants to mucous and gill membranes.

Other compounds, unlike the non-specific action of the primary irritants, can have a high degree of specificity and act at low dose levels at certain receptor sites to produce pathological change in specific cells with subsequent alteration of the function of organ systems. The degree of injury is dependent on the efficiency of the repair mechanisms for the tissue involved. If a cell is not irreversibly damaged, functional and structural characteristics may return to normal. Cells that are permanently injured at usually replaced by fibrous (scar) tissue. Fibrosis can seriously impair the functional capability of the tissue or organ involved.

Examples of substances that exhibit specific toxicological actions are those that produce mutations by interfering with the genetic machinery of germ cells. Mutation of a dominant gene may express itself immediately and produce fatal abnormalities (teratogenic effects) or early fetal death. Other mutations may not express themselves for several generations and then suddenly appear creating serious abnormalities for that individual and its offspring. Chemicals may be carcinogenic (e.g. diethylnitrosamine and aflatoxin) and initiate the growth of malignant tumors in fish and mammals (Stanton, 1965; Ashley *et al.*, 1964). Many toxicants directly or indirectly affect reproductive mechanism, gonadal growth and development, spawning behavior, and fry survival by specific toxicological actions. This can frequently result in a depletion or extinction of a susceptible species.

[†] Toxicology is the scientific discipline that deals with the quantitation of injurious effects on living systems resulting from chemical and physical agents that bring about alterations in cell or organ structure and function.

FACTORS THAT INFLUENCE TOXICITY

Several biological, physical and chemical factors play a part in determining the ultimate toxicological consequences of a foreign compound on an organism. To produce injurious effects, a toxicant must achieve an adequate concentration at its sites of action. The concentration attained at these sites in the animal obviously depends on the amount of foreign compound present in the animal's environment. Equally important, however, are the extent and rate of the toxicant's absorption, distribution, binding or localization in tissues, inactivation, and excretion. These factors are depicted in Figure 1 (Goodman and Gilman, 1970).

The following sections discuss each of these aspects that play the major role in the response of organisms to foreign compounds.

Absorption of the Toxicant

Generally speaking, a compound with specific toxicological actions must first be absorbed and distributed (biological translocation) in an organism before it can reach its specific site of action and exert its toxic effect. The ease with which foreign substances are absorbed, therefore, is a significant factor in determining the toxicity of foreign compounds. Absorption of substances by fish occurs through the skin, oral mucosa, intestinal mucosa, and gills. Because it is necessary that gill surfaces be exposed to large volumes of water for the maintenance of adequate blood levels of CO_2 and O_2, this organ is an especially vulnerable site for the absorption of foreign materials. Fish are, therefore, exceptionally susceptible to toxicants that readily cross the gill epithelium. In addition, fish acquire many foreign substances from their diet by absorption across the gastrointestinal mucosa. It should be mentioned, however, that studies are severely lacking about this and other mechanisms of translocation of foreign substances in fishes.

The biochemical and physicalchemical properties of a compound determine both its ability to cross biological membranes and its distribution within an organism. In general, the non-ionized, non-polar forms of compounds are significantly soluble in fat, i.e. lipid soluble, and are therefore readily transported across the lipidal components which characterize animal-cell membranes (Whittaker, 1968). The chemical structure of a foreign compound also determines its ability to react with biological molecules as well as its susceptibility to biotransformation (metabolism) by organisms. The ability of an animal to metabolize foreign compounds is important to that organism because the products (metabolites) formed usually are less toxic (but occasionally greater) than the parent compound. A more detailed discussion of biotransformation mechanisms appears later in this chapter.

Distribution of the Toxicant

Once the toxicant is absorbed and has entered an organism's blood or lymphatic system, it is readily transported and distributed to sites of action, centers of metabolic breakdown or detoxication, storage, and excretion.

Most toxicants are transported reversibly bound to blood proteins with only a small portion existing in the free or unbound state. Compounds must usually exist in this unbound state to react with biological molecules (receptors) and interfere with biochemical mechanisms. Therefore, the total amount of plasma protein available for the binding and transport of toxic substances plays an important role in the toxicological consequences of these compounds (Petermann, 1961). It is interesting to note that fish, when compared to mammals, have a distinct disadvantage in this regard since

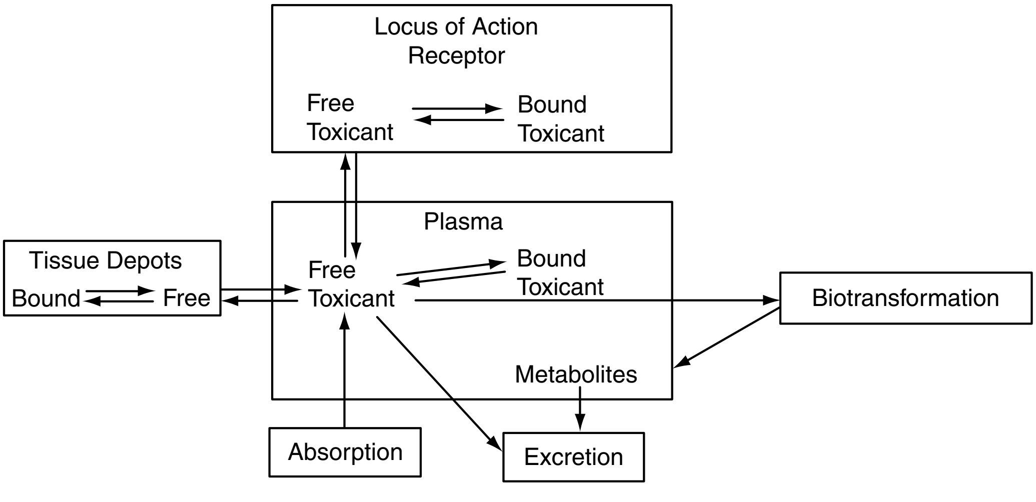


FIGURE 1

they have much less plasma protein. Compounds that have a high tendency to bind to plasma proteins may compete or displace one another from binding sites when they exist together in the blood. This may be the mechanism of dangerous toxic interactions because plasma protein binding sites become so saturated that a greater percentage of unbound toxicant exists in the blood than would normally be present with only one toxicant. When the degree of plasma binding is high and the rate of release is low, plasma proteins can act as a storage depot for the bound substances. Storage depots also frequently result from a particular affinity a toxicant may possess for certain organs or tissues. Examples of this are the chlorinated hydrocarbon pesticides which are stores in body fat and heavy metals such as copper and mercury which are stored in the liver and kidney of fish (Life, 1969).

Many foreign compounds are capable of producing a specific effect, that is, are selectivity toxic, on a specific biological system or systems. These systems are said to be the site of locus of action of the chemical. The site may be confined to one anatomic location within the animal, or may be diffusely located throughout the animal. Two fundamental types of mechanisms are responsible for the selective action of chemicals on cells or cellular mechanisms.

The first type is a result of factors that increase the concentration of the toxicant at specific cell or tissue sites. This is accomplished in the organisms by mechanisms of selective translocation and biotransformation. One good example is the renal tubular (kidney cell) injury produced in fish exposed to copper. Because copper is excreted by the kidney it accumulates at tubular cells. As the concentration rises injurious levels are reached and these cells are damaged or destroyed (Life, 1969).

A second mechanism in the selective toxicity of chemicals on cells involves the presence of specific targets or receptor systems in exposed cells. In this case, the concentration of the toxicant is the same for all cells, but only certain cells are affected. This is due to the specificity of action of the toxicant on receptors that are normally occupied by endogenous hormonal or neurohormonal substances. Organophosphorus compounds such as parathion and malathion are good examples of selectively toxic agents that act in this way. These cholinesterase inhibitors act to inhibit the enzyme responsible for hydrolysis of the neurotransmitter, acetylcholine. In this example cholinesterase is considered to be the receptor. Prevention of the hydrolysis of acetylcholine results in the continuous stimulation of post-synaptic sites throughout the central and peripheral nervous systems and rapid death due to respiratory paralysis is the usual outcome.

Biotransformation and Excretion of the Toxicant

The duration and intensity of injurious action of many foreign compounds are largely determined by the degree and speed at which an organism can eliminate these compounds (Conney, 1967). The kidneys of both marine and fresh water fishes have been shown to share with mammalian kidneys the primary role of ridding an animal of potentially toxic compounds (Forster, 1961, 1967).

The excretion of substances by the kidney is largely determined by lipid solubility characteristics of the compounds as they enter renal tubules. Molecules with a high degree of lipid solubility are readily re-absorbed from the renal tubule through lipoidal membranes back into the circulating blood and consequently are not excreted. It is only through certain specific biochemical transformations of these foreign compounds by the organism itself that lipid solubilities are altered and tubular excretion is successful (Brodie and Erdos, 1962). These reactions or transformations can be classified as oxidations, reductions, hydrolyses and syntheses (conjugation).

Most animals, including fish, transform (metabolize) foreign compounds in two successive phases, the first phase consisting of a variety of oxidations, reductions, and hydrolyses and the second phase of a limited number of syntheses or conjugations (Williams, 1967). Phase I reactions can result in:

1) The inactivation of a toxic compound;
2) The conversion of an initially inactive compound into a toxic compound; and
3) The conversion of a toxic compound into another toxic compound.

The second phase of biotransformation, consisting of synthetic reactions, most often results in the conversion of toxic compounds into inactive excretory products. This concept of the metabolism of foreign compounds can be represented as in Figure 2 (Williams, 1967).

Biochemical reactions of both phases of metabolism are catalyzed by enzymes located in various organ systems, and it is from the study of the qualitative and quantitative variations in these enzymes that an evaluation of detoxifying capacities can be made for an organism (Williams, 1967).

Phase I reactions are carried out by enzymes of normal metabolic routes and by enzymes which occur in the smooth endoplasmic reticulum of liver cells. When these cells are ruptured in the laboratory by homogenization the endoplasmic reticulum undergoes fragmentation. High-speed centrifugation separates these fragments from the remaining cell constituents. These fragments are referred to as microsomes and it is the microsomal enzymes that are involved in the metabolism of many drugs and foreign compounds. Microsomal enzymes do not generally act on lipid-insoluble compounds but rather convert lipid-soluble materials by oxidative and reductive processes to less lipid-soluble metabolites, which are more polar substances and, therefore, readily

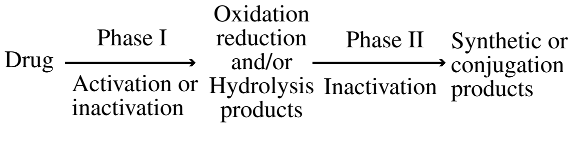


FIGURE 2

excretable by the kidney. Without these biotransformations, the effects of some foreign substances on organisms would last for months (Brodie *et al.,* 1965). Excretion of these lipid-insoluble metabolites can be achieved by active secretion at the tubules of the kidney or by passive transport across the glomerular membrane into the renal tubule (Forster, 1961, 1967). Since the metabolite has been transformed to a less lipid-soluble derivative, it will not diffuse back into the plasma for recirculation subsequent to passing through the glomerulus into the tubule.

The microsomal enzymes were, until recently, thought to be concerned only with the metabolism of compounds which are normally regarded as foreign to the body. Conney (1967) has recently shown that steroid hormones and other normal body constituents are also substrates of the drug-metabolizing enzymes in liver microsomes and he suggests that this enzyme system may play a significant role in their regulation and physiological action. These enzyme systems are thought to operate by a mixed function oxidase mechanism whereby NADPH reduces a component in microsomes which reacts with molecular oxygen to form an "active oxygen" intermediate. The "active oxygen" is then transferred to the drug or toxicant (Gillette, 1963). Key enzymes in the overall reaction are NADP-cytochrome *c* reductase, the flavin enzyme involved in the oxidation of NADP, cytochrome P-450, and NADPH cyctochrome P-450 reductase, which acts to reduce oxidized cytochrome P-450 (Gillette and Sasame, 1966).

Many foreign compounds can alter these key enzymes and enhance or impair the ability of liver microsomal enzymes to metabolize other foreign compounds and steroids (Conney, 1967). Halogenated hydrocarbon insecticides have long been known to be potent stimulators of mammalian drug-metabolizing enzymes (Hart and Fouts, 1963; Hart, Shultice and Fouts, 1963). Buhler (1966) selectively induced drug-metabolizing enzymes in rainbow trout by exposing these animals to DDT or phenylbutazone. Organophosphate insecticides are unlike the halogenated hydrocarbons in that they inhibit, rather than stimulate, the metabolism of drugs and steroids by liver microsomes, when given chronically to rats (Rosenberg and Coon, 1958; Welch *et al.,* 1967). Some heavy metals (Fe^{++}, Cu^{++}, Zn^{++} and Co^{++}) have also been shown to be inhibitory to drug metabolism in mice and rats (Peters and Fouts, 1970). Alterations in microsomal enzyme metabolizing capacities can substantially alter an animal's response to foreign compounds as well as its ability to hydroxylate steroid hormones and other normal body constituents (Conney, 1967). Altered steroid metabolism can directly affect the animal's ability to cope with environmental stresses as well as seriously impair reproductive mechanisms.

Other than Buhler's (1966) work on rainbow trout, little is known about the effects foreign compounds have on the microsomal drug metabolizing capacities of fishes. In addition, the current state of knowledge dealing with metabolism of foreign compounds in fishes is, at best, scanty (Adamson, 1967).

THE TOXICITY OF COMPLEX EFFLUENTS

Most industrial effluents contain mixtures of two or more substances. These complex effluents present special problems in evaluating their toxicities to fish and other organisms. In some cases two agents with similar pharmacologic actions can produce a response that is equal to the summation of the effects of the individual agents or greater than the summation of the independent effects of the two agents. The latter response is called "potentiation" and represents the condition whereby one compound is made more toxic in the presence of another compound which alone may produce minimal or no pharmacologic effect. Potentiation poses a special problem for the aquatic environment as well as the terrestrial environment for it is possible that a certain combination of relatively harmless substances may result in an unpredictable high level of toxicity that would seriously threaten the existence of one or many species. Usually, however, the effect of two agents is the summation of responses to each agent. Occasionally, the effect of a toxic substance is reduced on the addition of another substance, a phenomenon referred to as antagonism. In some cases the antagonistic substance may or may not be toxic when present by itself.

Synergism (potentiation and summation) and antagonism are poorly understood phenomena and greatly confuse the understanding and prediction of the toxic effects of industrial effluents. The majority of toxicological studies conducted in this and other countries on both mammals and fish deal with the effects of single substances on organisms; only a few studies are currently investigating the responses of organisms to complex mixtures of substances. Synergism and antagonism are worthy of further investigation, for little is known about the basic mechanisms governing these processes.

RESISTANCE OF FISH TO TOXICANTS

Resistance of animals to chemicals has been known to exist for some time now and has posed serious obstacles in the control of insects and bacteria. While the mechanisms of resistance remain thus far a mystery, we know that they are genetically based. Most susceptible populations of animals have an occasional individual which exhibits resistance and it is this member that provides the genetic material for selection pressures to act upon.

Resistance has best been demonstrated with various pesticides in several natural populations of fishes (Ferguson *et al.,* 1964, 1966), but these findings have attracted little attention. Many biologists have, in fact, believed this phenomenon to be beneficial to these animals, especially since they are useful to man.

Recently, some disturbing evidence has emerged suggesting that pesticide-resistant vertebrates pose a major hazard in natural ecosystems and that they may be creating serious toxicological problems for man (Ferguson, 1967). The reason for this is some resistant animals carry massive residues of unaltered pesticides in their tissues and aggravate the already serious problem of biological magnification.

Animals whose resistance is due to enhanced abilities to metabolize toxic substances to inactive metabolites do not, however, contribute to biological magnification.

Biological magnification results when a foreign substance enters plants and small animals and is then passed rapidly along food chains to larger animals. As this happens the substance becomes more and more concentrated until it reaches dangerous levels in the large predacious fish, many of which are consumed by birds and mammals including man. Clearly, the more resistance a fish has for the particular toxicant in its tissues the greater the likelihood it will be consumed by animals living on the land. Unfortunately, these animals may not have equivalent levels of resistance and may be unable to adequately deal with these toxicants.

CURRENT METHODS OF EVALUATING TOXICITY IN FISHES

The establishment of water quality standards for concentrations of toxic pollutants that will be safe for fish has recently become a major concern of Federal and State Governments in pollution control (Water Quality Criteria, 1968). Efforts in this regard have centered around determining lethal limits of toxicants by establishing a TLm (tolerance limit, median) of various species exposed to toxicants for periods of time up to 96 hr (Sprague, 1969).

These short term studies have been valuable in defining the upper limits of toxicity but have not considered the subtle deleterious effects of foreign compounds which may not be evident for weeks, months, or longer (Water Quality Criteria, 1963). These responses to toxicity may manifest themselves in appetite changes, metabolic alterations, disorders of the nervous system, reproductive changes, behavioral abnormalities, or alteration of vital functions which are not immediately lethal. For this reason, investigations have only recently been conducted to measure toxicity in terms of survival, growth, and reproductive alterations resulting from long-term exposure to sublethal levels of pollutants (Water Quality Criteria, 1968). The concept of a "maximum acceptable toxicant concentration (MATC)" has originated from these studies and is defined as the highest continuous concentration of a toxicant that does not significantly decrease the laboratory fish production index; an index developed by Mount and Stephen (1967) which takes into account survival, growth, reproduction, spawning behavior, viability of eggs, and growth of fry.

Because the toxicity of most pollutants varies with water characteristics and fish species, Mount and Stephen (1967) proposed the use of an "application factor," (calculated by dividing the MATC of the 96-hr TLm value), to determine safe concentrations of toxic pollutants which, when determined for one species of fish in one type of water, may be applicable to other waters and other species. Studies are currently underway to test the practicality of this approach (Mount, 1968; Mount and Stephen, 1969).

The "application factor" approach may improve upon present methods of estimating safe concentrations of toxicants. At best, however, this approach requires considerable time for collection and evaluation of data and measures only the end result of a multitude of biochemical, physiological, metabolic, pharmacological and pathological responses to toxicants. Little definitive information is gained concerning the mechanism which produces the gross changes upon which the "application factor" is based.

Because of the virtual impossibility of thoroughly assessing the individual, cumulative, synergistic and antagonistic effects of the numerous substances continually being introduced into our environment, it is imperative that we know the basic metabolic, physiologic and toxicologic responses of fish to compounds representative of broad categories of foreign substances. Only then will we be in a position to predict intelligently the biological effects of toxicants and regulate their concentrations to assure protection of this very important biological resource. If fish toxicologists continue to consider only the effect of a substance on the laboratory fish production index without understanding causative mechanisms, they will severely limit the amount of information available for making meaningful decisions so desperately needed in water pollution control programs.

Investigations of the disposition of foreign compounds in fish will shed valuable information on the evolution of enzymes that metabolize drugs, on drug metabolic pathways and excretion, and on factors affecting the biological half-life of foreign compounds in their lower species as well as higher vertebrates including man (Adamson, 1967). The editors refer the reader to numerous studies which have noted the collection of fishes with fin erosion or other deformities. A few concentrated on fresh water streams (Reash and Berra, 1989; Sindermann, 1979) whereas a preponderance focused on marine or estuarine environments with and without pollution (see Cross, 1985; Skinner and Kandrashoff, 1988) for example. Reash and Berra found that the incidence of fish erosion was significantly greater at polluted stream sites compared to unpolluted sites.

REFERENCES

Adamson, R.H. (1967) *Fed. Proc.,* **26**, 1047.

Ashley, L.M., J.E. Halver and G.N. Wogan (1964) *Fed. Proc.,* **23**, 105.

Brodie, B.B., G.J. Cosmides and D.P. Rall (1965) *Science,* **148**, 1547.

Brodie, B.B. and E.G. Erdos (1962) *Proceedings of the First International Pharmacological Meeting,* **6**, Macmillan, New York.

Buhler, D.R. (1966) *Fed. Proc.,* **25**, 343.

Conney, A.H. (1967) *Pharm. Rev.,* **19**, 317.

Cross, J.N. (1985) *Fish Bull.,* **83**, 195.

Ferguson, D.E. (1967) *Trans. Thirty-Second North Am. Wildlife and Natural Resources Conf.*

Ferguson, D.E., D.D. Culley, W.D. Cotton and R.P. Dodds (1964) *Bio Science,* **14**, 43.

Ferguson, D.E., J.L. Ludke and G.G. Murphy (1966) *Trans. Am. Fish. Soc.,* **95**, 335.

Forster, R.P. (1961) Kidney cells, in *The Cell,* Ed. by J. Brachet and A.E. Mirshy, **5**, pp. 89–161. Academic Press, New York.

Forster, R.P. (1967) Renal transport mechanisms, in *Proceedings of an International Symposium on Comparative Pharmacology.* Ed. by E.J. Cafruny, pp. 1008–1019. *Fed. Am. Soc. Expt. Biol.,* Bethesda, Maryland.

Gillette, J.R. (1963) *Prog. Drug. Res.,* **6**, 13.

Gillette, J.R. and H.A. Sasame (1966) *Fed. Proc.,* **25**, 737.

Goodman, L.S. and A. Glman (1970) *The Pharmacological Basis of Therapeutics,* Chapter 1, Macmillan, London.

Hart, L.G. and J.R. Routes (1963) *Proc. Soc. Exp. Biol. Med.,* **114.**

Hart, L.G., Shultice and J.R. Fouts (1963) *Toxicol. App. Pharmacol.*, **5**, 371.

Life, J.S. (1969) Metabolism and pharmacodynamics of copper in the goldfish (*Carassium Auratus*), Ph.D. dissertation, The University of Michigan.

Mount, D.I. (1968) *Water Res.,* **2**, 215.

Mount, D.I. and C.E. Stephan (1967), *Trans. Am. Fish. Soc.,* **96**, 185.

Mount, D.E. and C.E. Stephan (1969), *J. Fish. Res. Bd. Canada* **26**, 2449.

Petermann, M.L. (1961) *Med. Clin. North Am.,* **45**, 537.

Peters, M.A. and J.R. Fouts (1970) *Biochem. Pharmacol.,* **19**, 533.

Reash and Berra (1989) *Air, Water Soil Poll.*, **47**, 47.

Rosenberg, P. and J.M. Coon (1958) *Proc. Soc. Exp. Biol. Med.,* **98.**

Sindermann, C.J. (1979) *Fish Bull.,* **76**, 717.

Skinner, R.H. and W. Kandrashoff (1988) *Water Res. Bull.,* **24**, 96.

Sprague, J.B. (1969) *Water Res.,* **3**, 793.

Stanton, M.F. (1965) *J. Nat. Cancer Inst.,* **34**, 117.

Water Quality Criteria (1963) Federal Water Pollution Control Administration.

Water Quality Criteria (1968) Federal Water Pollution Control Administration.

Welch, R.M., W. Levin and A.H. Conney (1967) *J. Pharmacol. Exp. Therap.,* **155**, 167.

Whittaker, V.P. (1968) *Br. Med. Bull.,* **24**, 101.

Williams, R.T. (1967) *Fed. Proc.,* **26**, 1029.

JOHN E. BARDACH
University of Michigan

POLLUTION FROM MINE DRAINAGE

INTRODUCTION

All forms of mining cause some impact on the aquatic environment, just as any other earth disturbance will impact the local hydrology. Sometimes this impact is very adverse, in which case there is usually a considerable disruption of the natural life cycles in the affected water. Other, and less noted cases may even improve the local waters. Unfortunately, the adverse impacts greatly outnumber the advantageous circumstances so that the result of mining is to severely degrade the aquatic environment.

A generalized characterization of the impact of mining on water is rather difficult, as almost any specific change in the chemical qualities of the affected waters may be found at some specific point. In almost all cases, however, there is an increase in the total dissolved solids in the mine drainage waters. Additionally, the acidity of mine drainage is increased above normal ground water levels for the area, and the level of dissolved metal is increased. In some areas, however, the alkalinity levels are increased by mining. Many forms of mining also increase the suspended solids content of water.

Coal mining has received the greatest amount of attention as the mining which causes water pollution. This is perhaps deserved as the mining of coal has been a major operation for many years and more coal has been mined than any other single mineral. Water pollution from coal mining was known in medieval England and the mines in Wales were known to make the creeks run red. This fact was important to the exploration of the North American continent, as early explorers deduced the presence of large deposits of coal from the natural color and character of some of the streams and creeks. Similar conditions were sometimes noted in relation to other mineral deposits in the US. As the outcrop materials come into contact with the atmospheric conditions, oxidation and solubilization take place and the products are transported into the streams. Hence, the natural production of some mine drainage is a natural phenomena which has existed almost from the beginning of time.

Coal mine drainage may vary from waters pure enough to drink without treatment to waters containing more than 20,000 mg/l acidity with commensurate amounts of iron and other dissolved solids. Drainage from metal mines may vary over almost equally wide ranges of acidity but often contain substantial amounts of dissolved heavy metals. In most respects the acid drainages from metal mines are similar to acid coal mine drainages. This similarity is so great that most of the treatment processes and prevention mechanisms developed and applicable to coal mine drainage can also be applied to metal mine drainage.

ORIGIN OF ACID MINE DRAINAGE

The earth strata associated with and superjacent to coal and many other minerals almost always contain the iron sulfide mineral pyrite (FeS_2). Oxidation of the acidforming pyritic material associated with mining is necessary for the formation of mine acids and as oxygen is a necessary part of the oxidation of these materials, there can be no significant amount of oxidation until these are exposed to air. The process of mining greatly increases the exposure of these materials to atmospheric oxidation. Oxidation of the sulfide mineral begins as soon as it is exposed to the air and continues at a rate characteristic of the geologic and atmospheric conditions. Usually this oxidation causes spalling of the mineral substance with a progressive increase in the amount of surface area available for oxidation. Time then becomes a significant factor in the amount and rate of formation of acid mine drainage.

The exact nature of the pyritic mineral which oxidizes so rapidly and causes the acid drainage from mining has been sought for many years. In appearance, the mineral is often grey like marcasite, and its oxidation rate is even more rapid than that of marcasite. X-ray diffraction studies of the sulfide material associated with coal, however, have confirmed by the crystal structure that the material is pyrite rather than some other crystalline form of iron sulfide.

The intricate mechanism of oxidation of this pyrite and the formation of acid drainage has been the subject of many learned discussions and research efforts extending back over fifty years or so. The reaction will occur at normal room temperature and humidity conditions with the release of SO_2 into the atmosphere. Under more humid conditions, the reaction results in all of the sulfur being converted into sulfate. Typical reactions depicting some of the several paths possible for the reaction of the iron sulfide, air, water and alkali materials are listed.

This list is typical rather than all-inclusive to cover all possible reaction routes.

$$
\begin{aligned}
&\mathrm{FeS_2 + 3O_2 \rightarrow FeSO_4 + SO_2}\\
&\mathrm{2FeS_2 + 7O_2 + 2H_2O \rightarrow 2FeSO_4 + 2H_2SO_4}\\
&\mathrm{4FeSO_4 + 2H_2SO_4 + O_2 \rightarrow 2Fe_2(SO_4)_3 \downarrow + 2H_2O}\\
&\mathrm{4FeSO_4 + O_2 + 2H_2O \rightarrow 4Fe(OH)SO_4 \downarrow}\\
&\mathrm{Fe_2(SO_4)_3 + 2H_2O \rightarrow 2Fe(OH)SO_4 \downarrow + H_2SO_4}\\
&\mathrm{Ca(HCO_2)_2 + FeSO_4 \rightarrow CaSO_4 + Fe(HCO_3)_2}\\
&\mathrm{4Fe(HCO_2)_2 + O_2 + 2H_2O \rightarrow 4Fe(OH)_3 \downarrow + 8CO_2 \uparrow}\\
&\quad \mathrm{Ca(HCO_2)_2 + Fe_2(SO_4)_3 + 4H_2O \rightarrow}\\
&\mathrm{CaSO_4 + 2Fe(OH)_3 \downarrow + 2CO_2 \uparrow + 2H_2SO_4}\\
&\mathrm{Ca(HCO_2)_2 + H_2SO_4 \rightarrow CaSO_4 + 2CO_2 \uparrow + 2H_2O.}
\end{aligned}
$$

There has been a continuing debate among the scientific community over the role which bacteria may play in the formation of acid mine damage. Bacteria of the Ferrobacillus ferroxidans family are almost always found in large pools of acid mine drainage. The bacteria have been shown to have the capability of oxidizing ferrous iron to the ferric state. However, they thrive in a very limited pH range (approximately pH 3.5) and there is no evidence to indicate that they contribute directly to the primary oxidation of the pyretic substance. Ferric ion can oxidize sulfide sulfur and this could possibly provide a mechanism for the bacteria to increase the rate of oxidation of pyrite in some circumstances. Other recent studies have shown that the transfer of oxygen from the atmosphere to the pyrite surface is the rate limiting factor of the reaction, making moot arguments put forth as to whether bacterial or chemical oxidation is the principal mechanism of acid drainage formation.

Bacteria may play a significant role in the oxidation of the ferrous to ferric ion in mine drainage. This fact can be of considerable importance in the treatment of mine drainage as ferrous iron tends to retard the rate of neutralization reactions. It almost appears that this is Nature's first step in the neutralization of acid drainage. The bacterial oxidization of iron allows the alkalinity of the associated earth strata and of diluting waters to be more readily reacted with the acidity of the mine drainage waters. This concept is rather radically at variance with former concepts that considered the sterilization of a mine as a possible method of reducing or eliminating the formation of acid drainage. Bacteria has been used deliberately as a step in the treatment of mine drainage. This process allows the ferrous iron to be oxidized and accelerate the neutralization of the acid salts.

The formation of acid drainages from surface and underground mines is essentially the same process, and the two drainages are indistinguishable on the basis of the chemical qualities of the waters. In general the iron contained in drainage from surface mines and also from coal refuse piles is in the ferric state. The drainage from surface mining may contain very substantial amounts of suspended solids or sediment. Because many of the drainage problems of strip and deep mines are directly interrelated there is almost no rational way of separating the treatment or preventive measures which may be applied to the two types of mining.

EXTENT OF ENVIRONMENTAL IMPACT

The problem of environmental degradation caused by mine drainage is widespread and serious. Some form of mining occurs in each of the fifty states and many states are extensively mined. The aquatic degradation caused by coal mining in the eastern and Appalachian region is best known and has been best documented. For this reason, most statements of the damages caused by mine drainage cite only the degradation in the Appalachian area. The Appalachian Regional Commission reported that some 10,500 miles of streams in that region are affected by mine drainage and acid drainage continually pollutes nearly 5700 stream miles. Since data are not available in many mining areas, particularly in the Rocky Mountain and western areas, firm total statements of the extent of the problem cannot be made. However, enough information is available to indicate that it is very substantial.

The extent and amount of degradation which may be caused by the presence of mine drainage, and the environmental and economic impacts which may be felt, vary, of course, with the type of mine, the strata surrounding it, and other localized conditions.

For example, coal mine drainage, which is usually acidic, kills fish and other forms of aquatic biota by lowering the pH of the water and also may have an adverse economic effect upon the human population. The acidity accelerates the corrosion of bridges, culverts, boats and navigational facilities, making replacement at shorter intervals necessary. High cost water treatment may be required to make the local water supply potable or suitable for industrial use. Water contact sports may not be possible in the area, causing a loss of potential tourist industry revenues.

Although metal mining may cause the same adverse effects, it usually occurs in less densely populated regions. Additionally, metal mine drainage may render the waters toxic to humans as well as aquatic life by the presence of heavy metals.

REMEDIES FOR MINE DRAINAGE POLLUTION

In the past a substantial number of investigators have simply discussed the problem and observed its extent so that more is known about the nature of the problem than about the mechanisms which may be applicable to correcting or mitigating it. Mine drainage may be characterized on the basis of its source and possible remedies considered by this categorization even though the chemical nature and biological impact of the drainage from the several sources is identical. For the purposes of this discussion let us consider three types of mine drainage by the source: drainage from active or operating mines, drainage from non-operating (sometimes called abandoned) mines and drainage which will be generated in the future from mines which have not yet begun operation.

Separating surface and underground mines is not feasible as they frequently occur together and interact to add to the problem.

Presently operating mines have certain characteristics which differentiate mines them from other mines. Primarily because they are now in operation a responsible owner or operator can be located and people, equipment, machinery and power are available at the mine site. This allows consideration of procedures to treat the mine drainage as well as procedures to reduce or minimize the amount of pollutants discharged both during the remainder of the mine's life and after mining has been terminated. Procedures presently available may be employed to minimize the amount of mine drainage pollution which issues from an operating mine, but no procedures are now available which can totally eliminate it.

Non-operating, or abandoned mines, generally do not have any responsible person readily available, or any other resources such as personnel and machinery, which makes abatement techniques more expensive for this type of mine than for one which is operating.

When a mine is still in the planning stage, it is easier to plan for future prevention of pollution and thereby reduce it. For instance, provisions can be pre-planned for a mine to have rapid and complete drainage during the mining operation, thereby reducing the pollutional discharges which otherwise may need to be treated. Additionally, improved mining methods can provide for minimum void spaces after mining; also water level control after mining ceases can be provided only during the mine pre-planning stages.

The methods for alleviating mine drainage may be divided into the two basic categories of treatment of mine drainage, and prevention of the formation of discharge of pollutants.

Treatment removes pollutants by physical/chemical means and generally results in only specific pollutants being removed or reduced. The process must continue for as long as the source produces pollution and usually in the case of mine damage this is tantamount to "treatment in perpetuity." Disposal of wastes and constant usage of power or chemicals made such treatment unduly consumptive of both human and physical resources. Treatment methods for coal mine drainage are summarized in Table 1 and many of the methods are applicable to other types of mine drainage pollution.

Prevention is the total cessation or massive reduction of the formation or discharge of pollutants during and after the operating life of the mine. Prevention of all forms of pollution, including dissolved and suspended materials, is obviously more desirable than simple treatment of pollution with its attendant problems. Provisions for prevention of the formation of pollution cab be planned into a mine while it is still on the drawing board. Specific techniques for prevention and correction of pollution from mines, both operating and non-operating, are shown in Table 2.

TABLE 1
Summary of treatment techniques

Pollutants removed[a]	Neutralization and aeration	Microbial iron oxidation	Reverse osmosis	Ion exchange	Flash distillation[d]	Electro-dialysis	Product recovery	Silt basins
Acidity	XXX	NA	XX	X[c]	XX	X[e]	X[f]	NA
Iron	XX	XX	XXX	X[c]	XXX	?	X[f]	NA
Aluminum	XXX	NA	XXX	X[c]	XXX	X[e]	X[f]	NA
Manganese	X[b]	NA	XXX	X[c]	XXX	X[e]	X[f]	NA
Copper	XX	NA	XXX	X[c]	XXX	X[e]	X[f]	NA
Zinc	X[b]	NA	XXX	X[c]	XXX	X[e]	X[f]	NA
Hardness	—	NA	XXX	X[c]	XXX	X[e]	X[f]	NA
Suspended solids	X	NA	NA	NA	NA	NA	X[f]	X
Dissolved solids	—	NA	XXX	X[c]	XXX	X[e]	X[f]	NA
State of art[g]	CU	PP	PP	R, PP, FSO[c]	FSD	R[e]	R	R, CU
Cost, $/1000[h]	0.05–4.55	UK	0.30–2.57	0.30–2.53	0.33–3.25	0.52–2.52	UK	0.02–1.0
Waste product[j]	S	S	B	B, R	B	B	S,B,R	S

[a] NA—Not applicable, ? Questionable, X-degree to which removed, the greater number of "X"s the higher the removal.
[b] Must be raised to very high pH.
[c] Various ion exchange techniques are under evaluation. Their effectiveness and which pollutant is removed depends on the technique.
[d] 5 mgd plant is under construction in Pennsylvania. Techniques has only a limited potential.
[e] Technique will not operate where iron is present in water.
[f] Various product recovery schemes are under consideration; however, all are still in research stage.
[g] R—Research, PP—Pilot Plant, FSD—Full Scale Demonstration, CU—Common Usage.
[h] Cost depends on the degree of pollution, size of plant, pre- and post-treatment requirements. UK—Unknown.
[j] Each treatment process has a waste product that must be disposed of and creates additional problem. S—insert sludge, B—highly mineralized brine, R—regenerate.

TABLE 2
Cost and effectiveness of various at-source prevention and corrective techniques

Control techniques	Effectiveness %	Cost	Remarks
Surface mine reclamation	25–90	$300–3000/acre	Includes backfilling, regarding, contouring, and water control structures. Prevents acid formation, erosion control, runoff of dissolved soils. Cost and effectiveness controlling factors are nature of surface and overburden, slope of land, and proposed use of land.
Mine sealing (air)	0–50	$1000–5000/seal. Additional cost of $5000–100,000/seal may be required to control entrance of air through subsidence holes, boreholes, outcrop, etc.	Sealing of an underground mine to prevent entrance of air. Prevents acid formation. Cost of effectiveness depend on the ability to locate and seal all air paths to the mine, type and condition of mine operating and type of seal. Most mines cannot be airsealed.
Mine sealing (flooding)	75–99	$1000–20,000/seal. Additional costs of $5000–20,000 may be required to control drainage through bore holes, outcrop, etc.	Sealing of an underground mine to completely and permanently flood the working. Prevents acid formation and sometimes all discharges. Cost and effectiveness depend on the ability to seal all discharges, size of mine, dip of seam, outcrop condition, condition of mine opening, type of seal, and amount of grouting required. Is not applicable to all mines.
Drainage diversion	25–75	$200–20,000/acre	The prevention of water from entering the mine area. Prevention of siltation and flushing of pollutants. Cost and effectiveness depend on ability to divert as much water as possible in properly designed structures.
Impoundment	50–95	$350–1000/acre-ft	Flooding of surface mine pits. Prevents acid production. Cost and effectiveness depend on complete and permanent flooding of the material responsible for acid production.
Refuse pile reclamation (reject material from mining and processing)	25–75	$1000–3000/acre	Stabilizing a refuse pile with soil, chemicals, vegetation, etc. Prevents acid production and siltation. Cost and effectiveness depend upon the availability of the land in which the refuse piles can be filled and also upon the availability of impervious materials such as clay, fly ash, or limestone for compaction over the surface of the filled area.
Reject tailing pond (reject material from mining and processing in slurry form)	25–95	$300–2000/acre	Stabilizing of tailings by flooding, soil covering, chemicals, vegetation, etc. Prevents air pollution and discharge of suspended and dissolved solids. Cost and effectiveness depend upon the size, location.
Revegetation	5–25	$70–700/acre	Establishment of vegetative cover on reclaimed surface mines, reject piles, and pond. Prevents erosion. Cost and effectiveness depend on the type of cover, soil conditioning, and thickness of cover required.
Controlled pumping and drainage	25–75	$0.190.23/1000 gallon	Involves rapid removal of water from a mine before it gets contaminated or discharge of contaminated water at regulated rate so that dilution provides minimum contamination effects. Cost and effectiveness depend upon the characteristics of the material in the mine, contact time between water and exposed materials, rate of pumping, pumping head, and amount of dilution water available.
Inert gas blanket	Under research and development		Filling of an underground mine with an inert gas to prevent acid production.
Sterilization	Under research		Control of the bacteria in a mining environment. Prevents acid production. Technique does not show merit at this time.
Internal sealing	Under research and development		Internal sealing of underground mine to prevent acid production and/or mine drainage discharge.
Longwall mining	Under research and development		Allows complete removal of the coal in an underground mine. Mine roof collapses behind the working face, eliminating "breathing" of the mine.
Daylighting	Under research and development		Strip mining of a previously mined underground seam. Removes pillars left before. Surface must be reclaimed.

It can be seen from Table 2 and Figure 1 that the cost of creating or preventing the formation of mine drainage is a wide ranging variable, dependent on the method(s) selected for use. The estimated costs for the limestone neutralization treatment of mine drainage are shown in Figure 1 as a function of acidity and quantity to be treated.

As can be seen in the figures, the number of methods of alleviation and control of mine drainage which are highly effective are few, and most treatment processes produce undesirable by-products, as well as being expensive over the long run. It would be highly desirable to lessen the sources of mine drainage and to be able to treat it effectively with a gain in desirable by-products.

New techniques for the prevention of pollution from mines are presently in the development and demonstration stage. For example, a new technique known as "daylighting" is being demonstrated. This procedure will use strip mining techniques to remove the residual coal from shallow, non-operating mines and consolidate the overburden to prevent the continued discharge of acid drainage. A variety of abatement techniques related to surface mine reclamation are being demonstrated in the Appalachia region. Also there are two major efforts directed toward the development of non-pollutional mining techniques. These are the mining of coal under oxygen free conditions within the mine to prevent the oxidation of pyrite and the formation of acid drainage and the application of a new mining technique called "longwall stripping." Longwall stripping will apply longwall mining techniques to shallow cover coals, which are now strip mined, to remove the coal without inverting or dismantling the overlying earth strata.

LITERATURE REFERENCES

Most discussions of this type are buttressed by an impressive listing of reference documents citing sources of the many facts contained therein. To the casual reader, these references are useless, except for their creation of an impression of authority. The serious worker will demand to have even more references and documentation. The field of coal mine drainage is somewhat unique in that the literature of the field is regularly collected, abstracted and these abstracts published. This publication, entitled "Mine Drainage Abstracts, a Bibliography" is prepared by the Bituminous Coal Research, Inc. for the Commonwealth of Pennsylvania. Copies may be purchased from B.C.R., Monroeville, Pennsylvania. A listing of the reports of the research in this field may be obtained from the Publications Branch, Office of Research and Monitoring, Environmental Protection Agency, Washington, DC 20460.

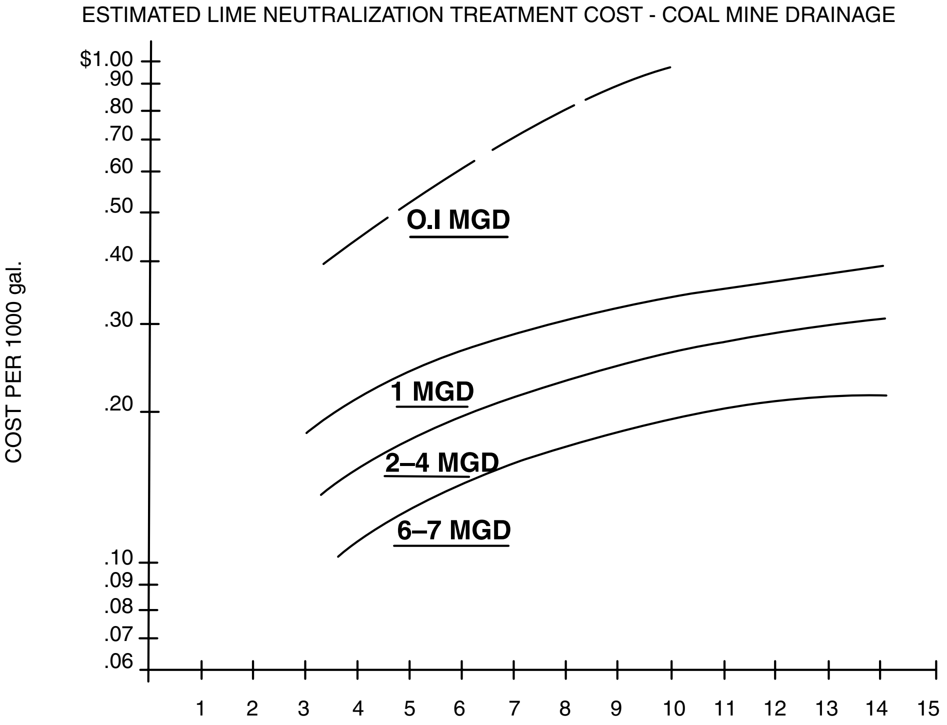


FIGURE 1 Estimated costs for treatment of coal mine drainage waters based upon a composite of published laboratory, pilot plant, and actual plant data. The estimate for the 0.1 MGD plant is preliminary and based upon limited information.

REFERENCES

1. Ramsey, D.L. and D.G. Brannon, Predicted Acid Mine Drainage Impacts to the Buckhannon River, West Virginia, *Water, Air and Soil Pollution,* **39,** 1, 1988.
2. Sobek, A.A., M.A. Bambenck, and D. Meyer, *Soil Sci. Soc. Am. J.,* **46,** 1982.
3. Sullivan, P.J., S.V. Mattigo, and A.A. Sobek, Dissociation of Iron Sulfates from Pyritic Coal Wastes, *Environ. Sci. Tech.,* **20,** 10, 1986.

ERNST P. HALL
Environmental Protection Agency

POLLUTION LAW: see ENVIRONMENTAL LAW

POLLUTION METEOROLOGY: see AIR POLLUTION METEOROLOGY

POLLUTION OF GROUNDWATER: see GROUND WATER RESOURCE

PRIMARY TERRESTRIAL CONSUMERS: see ECOLOGY OF PRIMARY TERRESTRIAL CONSUMERS

PROTECTION OF THE ENVIRONMENT: see ENVIRONMENTAL LAW

PREVENTION OF TOXIC CHEMICAL RELEASE

Purging of process vessels, tanks and piping before startup and after shutdown is imperative for pilot plants and large-scale plants in the process industries or wherever combustible gases and vapors are handled. Gases most often used for purging are nitrogen, carbon dioxide, or gases derived from the combustion of hydrocarbons. The reader is referred to the sections on Vapor and Gaseous Pollutant Fundamentals and to Fossil Fuel Cleaning for unit operations and incineration procedures involved in hazardous gas removal. To show how the purging procedure can be "mapped" a typical purging chart will be constructed and the salient points explained (see Appendix).

Inert gases such as nitrogen have the property of not only depressing or narrowing the explosive range of a combustible gas or vapor, but also of preventing the formation of explosive mixtures with air when these inert gases are mixed in suitable proportions with either air or with the combustible gas or with an explosive mixture of both. By displacing or mixing air contained in a vessel, tank or piping system to be placed into gas service, with a suitable amount of an inert gas such as nitrogen, a combustible gas may subsequently be introduced without the formation of an explosive mixture. Similarly, by displacing or mixing the combustible with a suitable amount of nitrogen, air may later be introduced without causing an explosive mixture to develop. During the purging procedure constant sampling of contents must be pursued using standard accepted methods of chemical analysis.

GAS FLAMMABILITY LIMITS

A flammable mixture of a gas, such as acetylene and air, may be diluted with one of the constituents (acetylene or air) until it no longer is flammable. The limit of flammability due to dilution is the borderline composition: a slight change in the direction will support burning, while in the other direction combustion cannot be supported and maintained.

At the ends of these two extremes there are well defined limits within which self-propagation of flame will take place on ignition. These are known as "upper" and "lower" limits as defined in terms of the percentage by volume of combustible gas present in a mixture of the gas and air. Table 1 below lists these limits for some of the more common gases and vapors for conditions of atmospheric pressure and temperature.

Within these limits, the combustible gas and air mixture liberates sufficient energy to continue to propagate flame from one mixture layer to the other. Mixtures above the upper limit may burn on contact with external air, since these layers are formed in the zone where gases mix. Certain conditions effect a shift in the two limits, either increasing or decreasing the spread between them as we will note later. These conditions include: ignition source, ignition intensity, direction of flame propagation (upward, downward, or across), size and shape of container, vessel, or piping orientation, temperature, pressure and humidity in the containing vessel, oxygen content and turbulence.

EXPLOSIONS

When a chemical reaction is accompanied by the liberation of heat, as the reaction progresses, it is followed by an increase in the amount of heat, which in turn helps to accelerate the reaction. Thus the two advance together, both in highly intimate connection and mutually helpful, until the entire mass has been heated and chemically converted.

When a burning substance, a match for instance, is placed in contact with the extreme outer limit of an explosive mix-

TABLE 1

Per cent gas or vapor in mixture—flammability limits

Gas or vapor	Lower	Upper	Gas or vapor	Lower	Upper
Acetylene	2.5	80.0			
Carbon monoxide	12.5	74.2	Acetone	2.6	12.8
Hydrogen	4.0	74.2	Ethane	3.2	12.5
Ethyl ether	1.9	48.5	Propane	2.4	9.5
Carburetted	6.4	47.7	Butane	1.9	8.4
Hydrogen sulfide	4.3	45.5	Benzene	1.4	8.8
Methyl alcohol	6.7	36.5	Gasoline	1.5	6.2
Coal gas	3.9	29.9	Turpentine	0.8	—
Ethylene	2.8	28.6	Ammonia	16.0	27.0
Pennsylvania	4.9	14.1	Ethyl alcohol	3.3	19.0
Natural gas	—	—	Methane	5.0	15.0

Reference: G. W. Jones, *Chemical review* Vol. 22, 1938. pp. 1–26.

ture of gas and air, ignition takes place at the point of contact, i.e. incites chemical reaction. Combustion proceeds from the outside of the mixture towards the center, and thereby forms a plane of combustion which divides the gaseous mixture into two parts. On the one side are the highly heated products of combustion, and on the other, is the still unconsumed gas mixture.

The velocity at which this plane advances in different for each gaseous mixture, and depends both on the composition of the mixture and on the pressure to which it is subjected. The higher the velocity of propagation the greater the rise in temperature, and this latter, in turn, directly influences gas expansion and the products of combustion, which thereby exert such a high pressure on their environment that any opposing medium, vessels, tanks, piping, walls, etc. is ruptured.

APPLICABLE EXPLOSIBILITY DATA

Not all explosibility data reported in the literature are applicable to purging problems. From the safety standpoint, it is desirable to select the widest explosive limits for the purge operation. In addition, an ample safety factor be applied, especially to the lower explosive limit. For acetylene-air mixtures, at atmospheric pressure and temperature, the published and accepted values for the lower explosive limit (LEL) and upper explosive limit (UEL) are 2.5 and 80.0% acetylene in air respectively. These are the widest limits recording. See Table 1.

In every case involving combustible gas handling or processing, constant awareness of the inherent dangers to life and limb as well as property is imperative. In the case of acetylene, it is a well known fact that it has unstable characteristic at any pressure and whether or not a decomposition would take place depends on the intensity of the initial source of ignition. Higher pressures have the effect of lowering the initial energy necessary to initiate a decomposition.

ACETYLENE—CASE STUDY

Acetylene is inherently an unstable gas at any pressure and whether or not a decomposition would take place depends on the intensity of the initial source of ignition. Higher pressures have the effect of lowering the initial energy necessary to start a decomposition. Higher initial pressures will also result in an increase in the ratio of the maximum pressure developed in the decomposition to the initial pressure. As for acetylene-air mixtures, pretty much the same behavior is manifested. With decreasing ignition energies, the initial pressures must be correspondingly increased so as to bring the total gas mixture volume to decomposition.

The size, shape and orientation of a process vessel as well as its material of construction have a profound effect on the ignition limits of combustible gas mixtures. The same applies to acetylene decompositions. It is a matter of heat balance during the combustion process. Even the relative position of the source of ignition. The widest explosive range is obtained for vertical cylindrical vessels or tanks when the ignition source is located at its base.

The presence of water vapor and high humidity acts as a diluent and an inert gas. This effect is typical for all combustible gas–air mixtures.

As for the effect on explosion limits, the ratio of vessel surface to cross-sectional area from a cooling or heat balance point of view, long slender vessels (high ratio of surface to sectional area) narrow the limits of explosibility. It's all a matter of the rate at which heat is removed from the gas mixture inside the vessel. As for the temperature effect, higher temperatures induce convection currents within the vessel and increase turbulence and widen the limits.

Thus, only tests of actual operating setups can tell the effects on the widening or narrowing of the explosibility range. Then, once established, the LEL and UEL determined experimentally can be used to develop a purging graph hereinafter discussed and developed.

EFFECT OF INERT GAS

The effect of an inert gas on explosibility of combustible gases in air can be shown graphically on Figure 1 for acetylene–air mixtures. Once the conditions for actual operating conditions and configuration are determined experimentally, the graph can be set up as we shall see.

As nitrogen is added to mixtures of the gas and air within the explosive range, a series of new mixtures are formed each of which has a different UEL and LEL than the preceding mixture and the explosive range is narrowed along definite lines of demarcation. As the air and combustible gas contents are reduced by the addition of nitrogen, the line of LELs and the line of UELs converge and meet at a point, i.e., *D*. Here the range has degenerated to zero. No mixture of acetylene, air and nitrogen which contains less oxygen than the lowest point on the line *LDU* (point *D*) is explosive in itself, but all mixtures within the areas bound by *LUD* are within the limits of explosibility and are therefore explosive. Again, in Figure 1, line *ADE* is drawn tangent to *LDU* at *D*. Any mixture represented by point *X* to the right of line *DE* and below the upper explosive limits *DU* is not explosive in itself. However, on dilution of that mixture with air, a new series of mixtures will be formed having compositions falling along line *X A* which passes through the explosive area. Similarly, any mixture represented by a point to the left of line *LD* and *DE* will not form explosive mixtures on further dilution with air.

Note line *LD* is not vertical but swings to the right as falls.

APPLICATION OF GRAPH

At startup when placing gas equipment into service purging from air to inert gas, the object is to reduce the oxygen

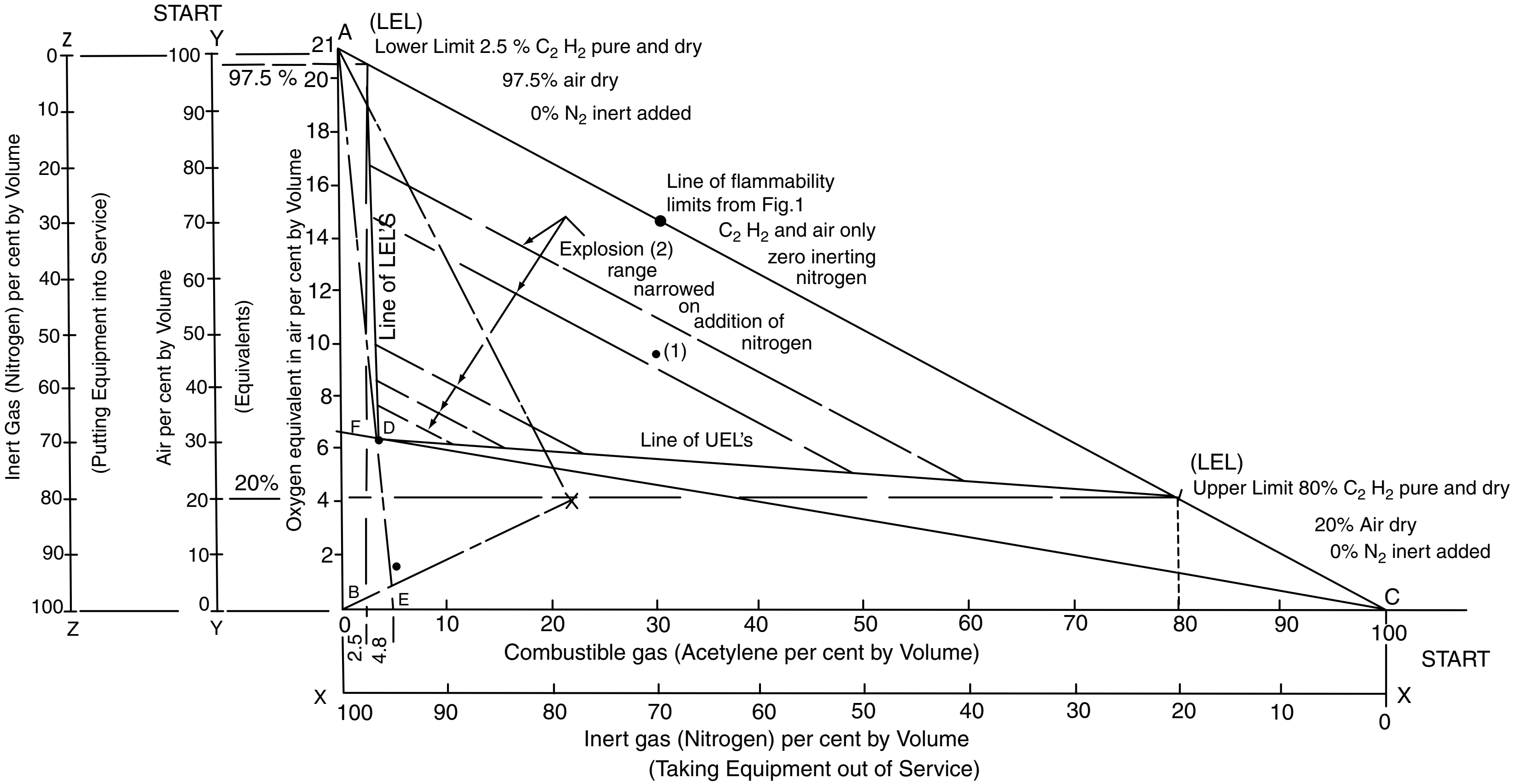


FIGURE 1 Acetylene-Air-Nitrogen purging chart (atmospheric temperature and pressure).

content of the air inside the equipment to a point that acetylene (or other gas) may be subsequently introduced without forming an explosive mixture. From Figure 1 it may be observed that the safe condition will be reached when the oxygen content within the vessel or equipment has been reduced by the introduction of nitrogen to below approximately 6.6% by volume, point *F* where *FC* drawn tangent to *LDU* at *D* intersects the vertical axis *AB*. Any mixture of this combustible gas, nitrogen and air having an oxygen content represented by a point below this line *FC* may be diluted with the combustible gas without forming an explosive mixture (see Appendix).

WHEN SHUTTING DOWN

Likewise, when shutting down or withdrawing equipment containing acetylene from service or when purging from combustible to inert gas, the object is to reduce combustible gas content to such a point that the air may subsequently be introduced without fear of forming an explosive mixture. Figure 1 shows that a safe condition will be reached when the combustible gas content within the equipment or piping has been reduced to below 4.8% by volume, denoted by point *E*.

On occasions the mixtures of acetylene, nitrogen and air are to be diluted with a further quantity of inert gas so as to render them nonexplosive or incapable of forming explosive mixtures on later addition of air. Object here is to add sufficient nitrogen to convert the existing composition of the mixture to some composition corresponding to a point to the left of *LD* and *DE*. If we assume that the mixture is again denoted by point *X*, then on addition of nitrogen, successive mixtures will be formed along *XB*. Point where *XB* crosses *DE* denotes the safe endpoint composition. Then on the further addition of air (intentional or accidental) the line of mixtures will fall to the left of *LDE*.

In practice, it is desirable to decrease the proportions of oxygen and combustible gas below the determined values to provide a factor of safety. To prevent explosion in chemical processing, it is necessary to keep the air–gas or–vapor or mixture of the gases, air and nitrogen, below the *LEL* within the equipment most likely to be affected. An established safe practice is to keep the mixture well blended with air or inert gas so that calculated or experimental values never reach more than 25% of the *LEL* or other low limit as determined from Table 1 in temperatures below 250°F. Placing system into service or out of service should be done when system has been depressurized to atmospheric pressure.

DEPRESSURIZING TO ATMOSPHERE

More often than not, combustible-gas releases to atmosphere do not contain air. When handling acetylene in its pure state a maximum pressure of 15 psig is usually used. Thus on release to atmosphere large quantities of the gas may be emitted through pressure relief valves to flares or high stacks. Then again the gas may be returned to gas holder storage. If released to atmosphere, inert gas purging should be provided. An excellent reference on inert-gas purges of stacks on pressure releasing has been published*; for process equipment see review.[12]

SAMPLING OF MIXTURE CONTENTS

During startup or shutdown of systems, sampling of mixtures within equipment should be conducted downstream of the point of injection of nitrogen to ensure proper mixing of gases before sampling. Taking samples from stratified streams should be avoided. For piping systems sampling should be taken downstream of elbows and open valves. In anticipation of this elbows and turns as well as valves should be designed into the system originally. In actual testing it has been observed that form 6 to 10 volume changes are necessary to purge equipment or piping systems. Sampling should begin after this purge.

Purging to a point below the LEL may be done using an explosimeter[†] that has been properly calibrated for the gas–air–nitrogen mixture. Laboratory testing procedures may also be used where samples are washed[†] with acetone and the volume of acetylene absorbed deducted from the entire sample to give volume of the remainder gases as air or nitrogen.

Spark-resisting tools should be used to prevent sparking and possible ignition of combustible gas. Work shoes with rubber soles and so steel studs should be used. In addition, continuous sampling of the workroom atmosphere should be conducted using explosimeters built for that purpose.

INERT GAS SUPPLY

Where rigid purity of inerting gas is not a requirement, the use of inert gases obtained from the combustion of hydrocarbon fuels may be used. This decision is one to be based on experience and processing needs.

Gas analysis can now be accurately measured and controlled in tenths of a percent range. These inert gas generators are provided with rugged instrumentation and controls for hazardous environments as found in desert and offshore installations. The analysis can even be accurately measured down to the ppm range where instrumentation maintenance is present. The same caliber of technician is needed whether or not the nitrogen source is cryogenic or hydrocarbon.

Such inert gas generators are currently supplied with automatic trimmer control which analyzes the discharge gas and automatically controls the air-to-fuel ratio. For critical installations they are even equipped with inexpensive automatic vent and alarm systems. For more information and their technical manual, write to C.M. Kemp Manufacturing Company, Glen Burnie, Maryland 21061.

[†] Explosimeters are available from Mine Safety Appliances Company, Pittsburgh, PA 15235 and Davis Emergency Equipment Company, Newark, New Jersey.

Skid-mounted units for applications in remote locations, or where onsite cryogenic facilities are too large to be cost-effective, are available. These units provide nitrogen at purities of 97 to 99.5%. They operate by pressure-swing absorption, in which a bed of activated carbon absorbs oxygen from air at high pressure and desorbs it at reduced pressure. The nitrogen produced by this cyclic process contains 5 ppm of water and CO_2. 400 to 20,000 cfh of nitrogen can be supplied, at a pressure of 100 psig. Write Airco Industrial Gases, Murray Hill, New Jersey.

Nitrogen for purging may also be supplied from cascade systems or directly from "bottle" trailers under pressure, normally 2000 psig. When drawn from large cascade systems or reservoirs the nitrogen is piped into a "running tank." The "run" tank is then pressurized and its contents are expanded through pressure regulators as desired for purging. When pressurized trailers are used, the high pressure gas is regulated down to the desired pressure for purging.

ACKNOWLEDGMENTS

This article appeared in its original form in *Chemical Engineering,* December 9, 1980, pp. 65–68. It has now been redone to provide more useful detail in the designation and use of the chart. A special section on checking points and areas on the graph has been included. This author gratefully acknowledges with thanks the permission by *Chemical Engineering* to update and republish.

APPENDIX

STEPS IN SETTING UP PURGING CHART

1) Draw right triangle *ABC* whose size is to accommodate ordinate and abscissa scales.
2) Fill in both ordinate and abscissa scales for the combustible of concern.
3) From Table 1 superimpose the explosive range of the combustible gas of concern on line *AC*, using the abscissa scale values.
4) Using the experimentally determined value equivalent of point, *E*, strike that value on line *BC*.
5) Draw line *AE*.
6) Draw lines *LD* and *UD*.
7) Draw line *FDC*.
8) Establish any point *X* below line *FDC*.
9) Draw lines *AX* and *BX* as shown.
10) Complete the chart by adding in the notes and other embellishments.

Draw chart carefully so that the various points can be determined with accuracy. And remember to use a safety factor in applying the chart figures, i.e. 25% of the chart values (points E and F) when taking equipment out of service and when putting equipment into service, respectively.

PURGING CHART DESIGNATIONS AND USE

Point D = No mixture of acetylene, air, and nitrogen which contains less than 6.4% oxygen is explosive.
LDU = All mixtures within area of triangle LDU are explosive.
Point X = Any mixture such as denoted by point X is not explosive.
Line XA = On dilution of mixture denoted by point X with air, new mixtures along XA are formed.
LDE = Any point to the left of LDE will not form explosive mixtures on dilution with air.
FDC = Any mixture of acetylene, nitrogen and air represented by a point below FDC may be diluted with acetylene without forming an explosive mixture when placing equipment into service.
Point F = fe point reached after air displaced with nitrogen and ready to add acetylene (6.6% oxygen) when placing equipment into service.
Point E = Safe point reached after acetylene displaced with nitrogen and ready to add air (4.8% acetylene). This is the situation when taking equipment out of service.

CHECKING POINT AND AREAS ON GRAPH

Purpose is to become better acquainted with the graph and to check various points.

UEL From UEL on AC read down vertically to 80% acetylene on BC. Then read horizontally left to YY and read 20% air by volume, with Zero inert nitrogen. Then by calculation and since composition of air is 21% oxygen and 79% nitrogen by volume: $O_2 = 4.2\%$; $N_2 = 15.8\%$.

LEL Likewise: 2.5% acetylene and 97.5% air and zero inerting nitrogen. In the air: 20.48% O_2 and 77.02 N_2%.

Point 1 This lies within area bound by LDU and should be explosive. Now from point 1 drop vertically and read 30% acetylene on BC. Move horizontally left to YY and read air per cent = 4.6. Then by calculation, O_2 content in air = 9.66% and nitrogen content = 36.34%. By difference inerting nitrogen is 70–46 = 24% by volume. Since 9.66% O_2 is greater than 6.4%, the mixture is explosive.

Point X Should not be explosive. From X drop vertically and read 21.5% acetylene as before. Move horizontally left and read 19% air on YY. Then by calculation find O_2 in air = 3.99% and N_2 in air = 15.01%, leaving 59.5% inerting nitrogen by volume. Since $O_2 = 3.99\%$ and is less than critical value of 6.4%, mixture is *not* explosive.

Point 2 Should be explosive. Acetylene = 30%, air = 70%. $O_2 = 14.7\%$ and nitrogen = 53.3%. Zero inerting nitrogen. Now since O_2 is greater than 6.4%, mixture is explosive.

X A line and UEL line intersect Acetylene = 19%, air = 29%, O_2 = 6.09%, 52% inserting nitrogen. This is a borderline case and safely assume nitrogen is explosive. Dilute with acetylene of nitrogen to render mixture non-explosive.

XB On dilution of mixture X with nitrogen, new mixture will fall along XB and remain non-explosive. In this process, per cent O_2 will continue to drop below 6.4%. Dilution with acetylene would effect the same results.

When shutting down Add N_2 until acetylene reaches 25% of experimentally determined lower limit. Thus, 0.25 × 4.8 = 1.2%.

When starting up Add N_2 until O_2 reaches 0.25 × 6.4 = 1.6%. When purging with any gas allow from *6* to *10* volume space changes before testing for concentrations. The Design Institute for Emergency Relief Systems (DIERS) Users Group, which is an affiliate of the American Institute of Chemical Engineers, has developed some methodology to design emergency relief systems.[8] The DIERS study was very extensive and complicated. It involved significant developments and applications of complex theories and experiments. Some aspects were reaction kinetics under runaway conditions and multiphase critical flashing flow for viscous and nonviscous systems. A number of DIERS users have attempted to simplify the DIERS technology.[9,10,11]

REFERENCES

1. *Prevention of Acetylene–Air Explosions by Addition of Carbon Dioxide or Nitrogen*. Dr. W. Gliewitzky, Berlin Autogene Metall bearbeitung, 1940, No. 1, 2–5.
2. Bureau of Mines Bulletin 279 (1938). *Limits of Inflammability of Gases and Vapors*. H.F. Coward and G. W. Jones.
3. *Theoretical and Practical Considerations in Purging Practices*. S. S. Tomkins, A.G.A. Proceedings (1934), pp. 799–828.
4. *Precautions in High Pressure Acetylene Work*. B.I.O.S. Final Report No. 1162. Item No. 22, London—H.M. Stationary Office. S.O. Code No. 51–1275–62. Tests conducted during years 1940, 1941 and 1944.
5. *Industrial Explosion Prevention and Protection*, Bodurtha, Frank T., Engineering Department, E.I. du Pont de Nemours & Company, Inc., 1st Ed. 1980, McGraw-Hill Book Company, New York City.
6. *Kemp Gas Generator Technical Manual*, The C.M. Kemp Manufacturing Company, Engineered Gas Systems Division, Glen Burnie, Maryland 21061.
7. *Gascope Combustible Gas Indicator*, Models 60 and 62, Mine Safety Appliances Company, 600 Penn Center Boulevard, Pittsburgh, PA 15235.
8. H.G. Fisher. DIERS Research Program on Emergency Relief Systems. *Chem. Engr. Prog.*, **81**(8), 33–36 (August 1985).
9. H.K. Fauske, G.H. Clare, and M.J. Creed. Laboratory Tool for Characterizing Chemical Systems. *Proceedings of the International Symposium on Runaway Reactions*, Cambridge, MA, March 7–9, 1989. Center for Chemical Process Safety/AIChE, New York, 1989, pp. 372–394.
10. J.A. Noronha, R.J. Seyler, and A.J. Torres, Simplified Chemical and Equipment Screening for Emergency Venting Safety Reviews Based on the DIERS Technology. *Proceedings of the International Symposium on Runaway Reactions*, Cambridge, MA, March 7–9, 1989. Center for Chemical Process Safety/AIChE, New York, 1989, pp. 660–680.
11. D.P. Mason. Highlights of FM Inspection Guidelines on Emergency Relief Systems. *Proceedings of the International Symposium on Runaway Reactions*, Cambridge, MA, March 7–9, 1989. Center for Chemical Process Safety/AIChE, New York, 1989, pp. 722–750.
12. Chatvathi, Kris, and Richard Siwek. Suppression Systems Mitigate Explosions, Chem. Eng. Prog. Vol. 92, March 1996.

JOHN D. CONSTANCE (DECEASED)
Consultant, Cliffside Park, N.J.

PSYCHOLOGICAL ASPECTS OF MAN'S ENVIRONMENT

Man's interaction with his environment is an integral part of the science of psychology. Psychology has been defined as the study of human and animal behavior. Behavior does not take place in a vacuum. It is conditioned by the environment in which it occurs.

The relationship between man and his environment is interactive. Man is not only influenced by his environment; his behavior may also profoundly affect his environment. Where psychology can predict behavior for various environments, it will permit calculation of the impact of behavior on the environment. To the extent that it has developed procedures for controlling such behavior, it can alter the impact of human and animal behavior on the environment by changing that behavior.

Viewed from this perspective, psychology might be expected to have far reaching impact on any area of environmental concern; indeed, for any area of human concern.

It is not surprising then, to find that much work in psychology is relevant to problems of concern to environmental scientists and engineers. Problems in almost any area of environmental study may have roots in psychology. Many substances whose physiological effects are of concern to environmentalists first make their effects on organisms apparent through behavioral indices and tasks studied by psychologists. Many standards problems have psychological aspects. Potable water must be acceptable to the palate, as well as meeting biological and chemical standards based on health considerations. It has been shown that existing ventilation standards for inhabited buildings originated in part from considerations of odor control. The impact of a field depends upon its achievements, as well as its domain of concern, however. One of the aims of this paper will be to consider the extent to which psychology can accomplish its aims in areas of environmental interest.

This paper will concern itself first with psychological studies relevant to man's physical environment. It will then go on to consider his psychological and social environment. These categories are to some extent artificial. The relevant environment for an organism consists of those aspects of the physical environment to which it is sensitive. This environment is not necessarily the same as the physical environment as measured by physicists or engineers. A colorblind organism will not respond to variations in the wavelength of light in the same way as one with color vision, though the physical stimulus is the same in both cases. Psychophysics is the branch of psychology concerned with the relationship between man's sensory processes and his physical environment and over one hundred years of psychophysical research has shown that the relationship is both subtle and complex. Nevertheless, such division conforms to some natural boundaries within the fields of psychology, and of environmental studies, and so it will be adopted here.

THE PHYSICAL ENVIRONMENT

The earliest studies in psychology were concerned with the relationship between measurements of the physical environment and perceptual processes. Since one cannot determine the nature of a percept directly, what was actually measured were certain observable responses, which were taken to be indices of these percepts. Although the concept of behaviorism was not originally enunciated in connection with psychophysics, but came later in connection with studies of learning, it is well to note that problems in all areas of psychology can be formulated in behavioral terms. In the author's opinion, this should always be done. It avoids innumerable arguments over concepts and processes if one recognizes that only the behavior is measurable, and that other concepts, such as motives, feelings, images, etc., are, from this point of view, theories developed to facilitate description of the relationships between the environment and behavior, however real these constructs may seem in other contexts.

Psychological Methods for Environmental Study

Psychological procedures relevant to environmental questions can be conveniently divided into the following categories: studies of sensory thresholds and simple perceptual and psychomotor tasks; behavior in simple learning situations; more complex-learning and decision tasks; methods for studying opinion; and evaluation of such complex processes as emotion and personality changes in response to environmental factors.

Psychophysical Procedures The study of sensory thresholds is one of the oldest areas of psychology, going back to the work of Weber and Fechner towards the end of the nineteenth century. Classically, two types of thresholds have been determined; absolute thresholds and difference thresholds. The absolute limen, or threshold (AL) is the minimum amount of stimulus energy required to elicit a response; the difference limen (DL) or threshold, is the minimum energy change required to perceive a change in the stimulus. Because the amount of energy related to that required for a perceived change is often related to the absolute value of the stimulus, the different limen is often given as a fraction of the absolute value.

Since the range of intensity for most of the senses is very large, it is customary to use a logarithmic scale in expression of the stimulus values. In audition, the decibel is a common unit. Intensity (I) is given in decibels (dB) by

$$I = 10 \log P_1/P_r \mathrm{dB},$$

where P_1 is the stimulus power, and P_r the reference level. Measurement in decibel units is less common for the other senses, but some type of logarithmic scale is frequently used.

The measurement of thresholds presents certain technical problems, due to variability in the response, and the possibility of response bias. Two major methods were devised to deal with the problem: the method of limits and the method of average error. In the method of limits, series of stimuli of varying intensity are presented and the subject is asked to make a response, according to whether the subject does or does not perceive the stimulus, or stimulus change. Since threshold values obtained typically depend on whether the approach to the threshold is up (from unperceived to perceived) or down (from perceived to unperceived), several series of trials in both directions may be averaged. The method of average error is an adjustment method. The subject is asked to adjust a variable to match a standard, and the variability of the response is used as a measure of the limen. It is most suitable to difference limen measurements.

Recently, the concept of a sensory threshold has been questioned, and the most popular view at present is to view the internal process as continuous, with a criterion applied to produce a dichotomous response. The view derives from the signal detection theory of sensory processes. The extent to which this model of the threshold differs from older ones is arguable; the arbitrary nature of the threshold concept has always been recognized. However, it has led to a more systematic treatment of the problems of noise and bias in threshold measurements. Procedurally it stresses the measurement of thresholds by yes/no judgments in which the probability of a false alarm is explicitly estimated, and by forced choice techniques. Receiver operating characteristic (ROC) curves are used to analyze the data. The resulting measure of sensitivity is usually referred to as *d*. From the point of view of reliability, either the classic methods or new signal detection procedures will usually lead to satisfactory results.

The sensory dimensions have been divided into two types: intensive dimensions and qualitative dimensions. Examples of intensive dimensions are the intensity of light and sound. Examples of qualitative dimensions are color and pitch, which often have recurrences in them, such as the scale relation in pitch, which complicate analysis. Sensory intensity normally increases logarithmically with increases in the physical stimulus. Qualitative dimensions may have more complex relationships with their physical correlates. The measurement of sensory thresholds may be extended to animals by means of conditioning procedures to be described in the next section.

Several thresholds other than the absolute and differential limens are important for environmental studies. Often, concern will lie in thresholds for discomfort, and for tolerance. It may also be desired to obtain information on the degree of discomfort produced by different environmental conditions. Direct judgments of discomfort or tolerance tend to be highly variable; individual tolerances, or at least statements about these tolerances, differ from individual to individual. One technique which is sometimes used is to obtain comparison judgments about the relative discomfort caused by different stimuli. One method of measuring pain thresholds is based on this principle; the discomfort caused by various stimuli is compared with the pain caused by a known exposure to infrared light on the forehead.

A fixed standard need not be used; cross comparison of the set of stimuli of interest is possible. Techniques have been developed for taking such paired judgments, and using them to scale the stimuli along the dimension of judgment. Variations of this technique can be applied to ranked data: ranking is less time consuming than a full set of paired comparison judgments, but it requires more assumptions to generate a scale. Multidimensional scaling techniques are usefully applied to preference or similarity judgments, however obtained. They are especially useful where the relevant dimensions of eth stimulus set are not clear, since they can be analyzed to obtain dimensions of judgment. It is also possible to analyze such judgments for clusterings among the subjects.

Classical and Operant Conditioning Classical conditioning is a type of learning pioneered by I. Pavlov, and still extensively studied by psychologists. In its simplest form, the procedure consists of presenting some stimulus which consistently elicits a response, known as the unconditioned stimulus (UCS) in combination with some other stimulus which does not normally evoke the response, known as the conditioned stimulus (CS). After a number of trials, the CS will be found to elicit the response in the absence of the UCS. Such a response is known as a conditioned response (CR).

Various complexities can be introduced into this paradigm and it has been the subject of extensive study and theoretical discussion. Its primary importance for the environmental field lies in the fact that the technique can be used to determine sensitivity to various stimuli. This can occur in two ways.

1) The effect of various physiologically active stimuli on the course of acquisition and extinction of a known CR can be studied. Changes in the course of learning and extinction are evidence of the effect of the stimulus.
2) The procedure can be used to determine threshold sensitivity to a stimulus. If a CR can be established to a given concentration or intensity of a stimulus, it is evidence that the organism is sensitive to the stimulus, though the reverse is not necessarily

> true. Thus, classical conditioning techniques can be used to measure sensory thresholds in animals.

A second major type of conditioning is known as operant conditioning. In operant conditioning, there is nothing equivalent to the UCS; a response which occurs is followed by some subsequent event, such as the occurrence of food, which will modify the frequency of the response. The desired response may have a low probability of occurrence and procedures such as shaping and the creation of special environments may be necessary. Shaping consists of reinforcing successively closer approximations to the desired response. A common example of a special environment is the Skinner box, a small enclosure whose most prominent feature is the response mechanism, usually a bar or button. The rat or pigeon in a Skinner box, pushing a bar or pecking or at a lever to obtain food or avoid electric shock are example of operant conditioning situations.

Operant conditioning procedures can be used for the same purposes as classical conditioning methods, but they have some advantages. Classical conditioning is relatively difficult to establish, is easily disturbed by any change in the environment and the CR ceases to occur (undergoes extinction) rapidly once the UCS is removed. Operant conditioning, on the contrary, is relatively easy to establish. One reinforcement often makes a marked difference in response rates. It is less sensitive to extraneous environmental interference and it can be maintained for long periods of time with a very low reinforcement rate. Records of cumulative responses to various reinforcement schedules show definite regularities. Changes in these patterns provide a sensitive measure of the effect of various chemicals on behavior and have been extensively used for this purpose in drug research. Threshold measurements in operant conditioning require a slightly more complex situation than in classical conditioning if the substance to be studied does not act as a reinforcer of behavior. However, by using it as a cue to indicate when some positive reinforcement is to be given, the thresholds of such substances can be studied by operant techniques. Both operant and classical conditioning techniques are applicable to humans as well as other organisms.

Other Simple Learning Situations Other types of learning situations which have been studied sufficiently to provide a basis for the study of behavioral effects of environmental factors are maze learning, discrimination learning, and a variety of simple verbal learning tasks. The former tasks may be used with either humans or animals. The verbal tasks are limited to humans. The complexity of the stimulus situation is generally greater in these tasks. Specifying the relevant dimensions of even simple mazes for learning situations is complicated, and an enormous and continuing effort has been devoted to the study of the features of verbal stimuli which affect simple learning tasks such as paired associate learning (where people are presented with pairs of words and then asked to recall one of the words when shown the other) and serial learning (where subjects are asked to learn lists of words). They are therefore less easy to quantify and compare across experiments than are conditioning responses and thus introduce additional problems for the study of environmental effects. However, with sufficient care, reliable measurements can be obtained.

These tasks come closer than simpler tasks to being of a complexity relevant to behavior in real life situations. It is quite possible that some stimulus will have no measurable effect on a simple process, but will effect more complex behaviors; the reverse is sometimes also the case; environmental conditions which will affect simple and monotonous tasks may not affect more intrinsically interesting tasks. For this reason, learning tasks other than simple conditioning are often studied when the effects of various environments are under investigation.

More Complex Tasks The number of human performances which have been studied by psychologists is very large; no satisfactory classification has been worked out for procedures going beyond the simplest. Various bases for categorization have been proposed; for example, sensory-motor to verbal, depending on the type of response; or in terms of complexity, from simple conditioning to complex problem solving. Probably the best ultimate solution will consist of a task analysis for more complex behaviors in terms of their dependence on more simple processes, but this has not been achieved in any general way at present. The particular type of task employed and the dimensions along which it is to be analyzed can be best considered at present in terms of the particular experimental situation. If an investigator is concerned with the effects of certain pollutants on complex judgments required by some task, he will do well to construct a task as much like the criterion task as possible and test the effects by means of a suitable design. If his concern is with performance in a vigilance situation, then a vigilance type task should be selected. Generalizations from effects on simple behaviors to more complex behaviors are hazardous at present.

Standardized Tests of Behavior Standardized tests have been developed for a wide variety of human abilities. IQ tests are the most famous, but many others exist. Such tests have as their primary purpose either selection in connection with personnel work, or evaluation for counseling for guidance, but they may be used in the same way as other tests for the evaluation of the effect of environments on performance.

Tests of this sort may be divided into tests of general ability, tests of specific ability, and tests of personality. Both of the former have been used in connection with environmental work. Occasionally, the latter have been also, though their suitability for the purpose of evaluating temporary shifts in emotional state is often dubious. Most standardized tests are designed to measure some relatively permanent characteristic. However, some of these characteristics, such as various types of manual dexterity, are sensitive to various environmental forces in a fairly straightforward way. The effects of depressant drugs on personality tests are considerably less clear. There is also a serious problem of validity with regard to some tests, especially personality tests. The objective tests, such as the Guilford-Zimmerman or the Minnesota Multiphasic Personality Index (MMPI), have been extensively validated against certain criteria, such as psychiatric diagnosis. However, the criteria themselves are open to argument in many cases. Also the criteria used for validation are seldom those of interest if a test is to measure temporary changes in emotional state due to changed environment. The tests often appear likely to be insensitive to the types of mood changes which may be of interest to environmentalists. Projective personality tests appear, on the surface, to be more sensitive for this purpose, for they permit much freer response than objective tests, and they have been used, for example, in studies of aggression. However, they are exceedingly difficult to validate, even against the type of criteria used for objective personality tests, and their predictive value for other types of behavior is almost entirely unknown.

Comments on Psychological Methods In evaluating these procedures for use in the environmental field, there are some general considerations to keep in mind. First, the psychophysical procedures are generally reliable, if carefully done, and generalization to other individuals from a small sample is good. This work is often done with two or three observers, and the results are repeatable. The extreme sensitivity of the sensory systems means that very precise control over the stimulus is needed, which is often expensive. Certain stimuli, especially odors, are difficult to handle; special equipment, such as odorometers, is necessary. Timing down into the millisecond range is typically needed in visual and auditory experiments.

Other types of behaviors tend to be more variable. Only in simple conditioning situations are such small numbers of experimental subjects likely to be found, and even here, for most purposes, more than three or four subjects are necessary. The requirements for stimulus control are somewhat less rigid, in most cases. In verbal learning studies, timing to .01 sec is usually adequate; less precision is often tolerable. However, the substantial variability of organisms presented with even relatively simple learning tasks, such as mazes or simple discrimination learning, makes a larger sample desirable. Individual experimental conditions may be tested on 10–12 subjects, or more; it is not uncommon to find verbal learning experiments which use several hundred subjects. Generalization to other populations is sometimes more of a problem and the equivalence of stimuli in different experiments is more difficult to establish. However, within these limitations, reliable results can be obtained in simple conditioning situations and most other simple learning tasks, including simple verbal learning tasks.

Both classical and operant conditioning procedures are very sensitive to the action of certain agents. Effects of levels of ionizing radiation have been reported as low as 5.0 r in conditioning experiments. Behavioral effects of several drugs of environmental importance, such as alcohol and amphetamines, can be found for relatively low dosages. While the information on the sensitivity of conditioning procedures to agents of interest such as SO_2 and NO_2 is sparse, there is no reason for believing that behavioral measures will not be equally applicable to many of these substances. Behavioral studies done in Russia emphasize the use of classical conditioning techniques in studying various environmental factors. In such research behavioral effects for low levels are frequent, though it is difficult to evaluate due to the obscurity of much of this literature by American standards of reporting.

Problems arise when it is desired to interpret behavioral measures in a more refined fashion than simply as indicators of some physiological activity of the substances. Temporary threshold shifts from noise exposure have been extensively investigated, but their implications for permanent damage is still being debated (see Noise). Similar problems exist for other measures. The problems which have faced psychopharmacologists in interpreting effects of drugs on animal behavior in terms of drug effects on humans have their counterpart in the environmental field; granted that a certain level of ionizing radiation can serve as an aversive stimulus in operant conditioning experiments on saccharin preference, for example, how are the results to be interpreted in terms of public health standards? How long lasting must such effect be in order to be indicators of possible physical damage? If no physical damage is to be expected, there is still the question of the costs of possible functional impairment. Does impairment of operant conditioning imply that the agent will also impair more complex functions in humans, with a possible increase in errors or accidents as the result? At present, answers are not available to these questions.

More complex tasks raise more serious problems of reliability. It is not uncommon in educational research for several experiments to yield contradictory results, for reasons which are often obscure. The complexity of the materials and other variations in subjects and procedures which are difficult to control or characterize are probably responsible. Thus, a single study done on the environmental effect of a substance on complex learning should be viewed with caution. Only when a series of experiments done under varying conditions converge on a result should the result be accepted.

Certain special problems which often arise are the effects of novelty (the so-called "Hawthorne" effect) and habituation effects. Individuals may perform well under a certain test procedure, which, when put into practice proves no better than the alternative. The problem is that subjects on whom new procedures are tried may make a special effort which is not sustained under routine conditions. A similar effect arises in regard to habituation. Individuals may respond badly to a new environment, but adjust to it over time without final decrement in performance. Since adaptation for some tasks can take months, adaptation effects are often inadequately tested.

Most research involving the possibility of exposure to physically hazardous conditions is done on animals. However, a good deal of research which involves physical discomfort or emotional stress is done on humans. Also, in some research, personal information of a private nature may be obtained. Concern over possible abuses of subjects has arisen in recent years, and several professional societies have or are developing codes of ethics to deal with these problems. There are also legal aspects to the use of human and animal subjects of which any potential research worker in these fields should make himself aware.

THE PHYSICAL ENVIRONMENT AND BEHAVIOR

Sensory Processes

Judged purely in terms of their adequacy as detectors, sensory systems are very efficient for many types of stimuli. The auditory system, for example, is capable of detecting sound near the level of Brownian motion under suitable test conditions. The number of photons required to obtain a visual response has been estimated as low as 5–14 photons under optimal conditions. The olfactory system is also exceedingly sensitive. As has been noted, the intensity of a sensation goes up approximately as a logarithmic function of increases in intensity of the physical stimulus (the Weber–Fechner law). It applies to such sensory domains as vision, audition, etc. In specific applications, of course, this information is inadequate for most purposes and more precise information is needed. The apparent loudness of a noise will depend on its frequency as well as its intensity, for example.

When qualitative factors are under consideration, it is more difficult to make any statements. The annoying properties of a noise are a complex function of specific combinations of frequency and intensity and it is still not well understood. See the article on noise in this volume for a more complete discussion of this problem.

There are, however, a few general statements about sensory systems which may be of use to environmentalists. First, of course, the primary channels of information to human beings are the visual and auditory systems. Senses such as touch, taste, smell, and so on are of far less importance than in the case of many animals. Of the major channels, vision has the larger channel capacity, in the information theoretic use of that term. However, this is not equivalent to an expression of their social importance. The auditory channel is the primary means of face-to-face communication between human beings, and studies indicate that loss of hearing may produce more profound personality changes than loss of vision. Early loss of hearing can produce a disastrous effect on language learning. For many purposes in adults, they may be equivalent for information transmission. When one is dealing with tasks which are purely informational, the visual channel may be more efficient, since in tasks like reading, humans can process more words than they can listen to in a given period of time.

Knowledge of visual and auditory parameters is important in the proper design of a number of environments. Consider the hazards associated with sources of glaring light along highways at night, for example. The visual system adapts relatively rapidly (usually within a minute) to large increases in light. The process of dark adaptation is much slower, however; it may take as long as half an hour to fully dark adapt the visual system. Rapid changes in the external light levels at night can therefore constitute a hazard, since they may keep the eyes of drivers in a state of adaptation to relatively high illumination levels even though the average level is relatively low.

Effects of Various Pollutants on Behavior

Ionizing radiation Perhaps the pollutant whose behavioral effects have been studied most extensively is ionizing radiation. Studies on both lethal and sublethal doses of both particulate and non-particulate radiation have been done. The conclusion has been that there is relatively little effect of either on the learning behaviors for most organisms. Insofar as tests are available, this seems to be true of man also. However, the activity levels of young rats have been reported to be affected when the mothers had been irradiated with one dose of 200 r. during pregnancy.

This does not mean, however, that organisms are not sensitive to radiation. Radiation has been used to condition avoidance to a saccharine solution (normally preferred by rats). It was found to produce an 80% decrease in response rate under some conditions. The amount of the decrement has been shown to be dose dependent. Russian research, using classical conditioning techniques, has reported a variety of behavioral effects, occasionally for doses as low as 5 r.

Air Pollutants The effects of carbon monoxide (CO) on behavior have received a good deal of study. Among other effects, 90 min of exposure to 50 ppm CO has been shown to impair ability to discriminate among relative lengths of short time intervals. The effect of CO on the electroencephalogram (EEG) patter in rats has been investigated. The results suggest a possible depressive effect of CO on central nervous system function. Significant reductions in tests of manual dexterity have been observed with blood carboxyhaemoglobin (COHb) levels in excess of 25%; such levels may be reached in garages and other enclosed areas containing motor vehicles. CO is known to elevate the visual threshold.

Since CO and some other pollutants have as one of their effects the reduction of the oxygen level in the blood supply, studies on the effects of hypoxia (oxygen deficiency) are relevant to the behavioral effects of gaseous pollutants. The visual threshold is very sensitive to hypoxia; significant impairment can occur at altitudes as low as 5000 ft, and at 15,000 ft twice as much light is required for minimum visibility as is required at sea level. Auditory and speech perception thresholds may also be affected slightly, especially for long exposures. Tasks requiring complex coordination, such as handwriting, deteriorate above 10,000 ft, and higher altitudes can affect simpler tasks and reaction times. Various cognitive tasks, such as card sorting, are affected, usually more than simple psychomotor tasks.

The physiological effects of a number of other pollutants such as NO_2 and SO_2 have been studied and information about their olfactory effects is also available. The olfactory system is very sensitive, and may give adequate warning for some pollutants. Sulfur dioxide, though extremely toxic is difficult to inhale at lethal concentrations without warning because of its intensely irritating properties. Habituating effects do occur with repeated exposure, however, and severe poisonings have occasionally been reported among workers who have adapted to high levels of this pollutants. It is important to note that the olfactory sense adapts very rapidly to most odors. This consideration is important when using odor tests, for chlorine in water, for example.

A number of other substances are not readily detectable by smell, however. Carbon monoxide is odorless, and nitrogen dioxide can cause fatal poisoning at concentrations which are not physically annoying.

Other Chemical Pollutants The behavioral effects of a number of other chemical substances which may be of environmental significance have been studied. These include alcohols, various psychoactive drugs, and various industrial chemicals. Ethyl alcohol has been found to increase reaction time by about 10% when the level of alcohol in the blood reaches 0.35%. Blood alcohol levels of 100 mg/100 ml have been reported to affect monitoring and tracking tasks, and lower levels have been reported to affect various complex tasks.

It has been possible to study the effects of a number of such chemicals in conditioning experiments. Several drugs have been shown to increase eating or drinking behavior in animals; these are usually drugs which act to reduce anxiety. Meprobamate and similar drugs usually increase eating; the barbiturates chiefly affect drinking. A number of such drugs have been shown to produce decrements in escape or avoidance conditioning, an effect usually attributed to their fear or anxiety reducing properties. Morphine addiction can be induced in animals, and the drug can serve as a positive reinforcer for behavior, permitting the study of various aspects of addiction, such as withdrawal, in animals, using conditioning concepts and procedures.

Depressant drugs, such as the barbiturates, have generally been found to impair memory in humans. Retrograde amnesia (failure to recall events just before administration of a drug or some traumatic incident, such as a blow to the head) has been reported for ether and for CO_2, administered post-trial. However, conflicting evidence exists. Post-trial administration of ether can also facilitate retention, and 30% nitrous oxide has also been reported to facilitate retention in humans when given after learning, though other studies have found impairment with nitrous oxide.

Facilitation effects on memory have been reported for a number of other drugs. Strychnine has been shown to improve maze learning for food rewards, discrimination learning, and other tasks. Various other substances have also been studied. However, the results are conflicting at present.

The various effects of chemicals on retention have been interpreted in various ways. It has been hypothesized that a certain period of time is necessary following learning for consolidation of the memory trace. Drugs or other events, such as a blow on the head, or the administration of electroconvulsive shock, may interfere with this process and prevent the retention of learned responses over time. Other possible explanations lie in facilitation or interference with nerve transmission or with protein synthesis. The inhibiting effects of puromycin and other antibiotics on some learned responses has been attributed to their interference with protein synthesis, though other explanations are possible. Explanations in terms of the drug effects on the animals' perception of the cues in the learning situation have also been suggested.

Environmental Stressors

Unusual conditions of temperature, noise, lighting, vibration, etc. may be encountered in some environments. Extreme values of such stimuli are often regarded as environmental stressors, and their physiological effects have been extensively studied. Concern has also been felt about their behavioral effects, and a number of studies have been done on this question.

Noise Excessive noise levels are known to be capable of damaging the auditory system. Noise can also be annoying, and its annoying properties are a complex function of its spectral composition (see Noise). It can also seriously interfere with tasks which depend on auditory communication. However, studies of effects on non-verbal tasks have reported mixed results, often negative. While interference through distraction may occur with sudden unpredictable sounds, in general, noise seems to have a minimal effect on most non-verbal tasks. It should be kept in mind, however, that adapting to noisy environments may impose additional stress on the organism. Noise can induce audiogenic seizures in susceptible strains of rats, and there have been reports of sound-induced seizures in epileptics.

Temperature A number of studies have shown decremental effects of elevated temperatures on a variety of tasks. Telegraph operators who had been acclimatized to a hot environment (Singapore) were studied at several temperatures above that of their usual environment. Errors were found that were related logarithmically to the temperature increase in the receipt of coded messages made up of assorted letter and number strings. Similar results have been reported by artificially acclimatized subjects in England. Subjects' judgments of comfort are not necessarily reliable guides to the temperature which gave best task performance. In general, the relationship between behavior decrement and heat stress is more straightforward for tasks involving physical effort than for other tasks, but effects on both have been reported.

Cold has also been studied, chiefly in relation to its effects on manual dexterity, which it decreases. Adaptation effects are also reported to cold.

Studies have been reported which show decrements in learning and retention with elevated temperatures, and improvements with cold environments. The latter have mostly been reported with invertebrates, and it is unclear whether the effect is directly temperature related, or due to restriction of activity with cold.

Vibration Excessive vibration has been shown to be physically damaging, especially to the circulatory system, in addition to causing symptoms such as seasickness in susceptible individuals. In terms of effects on other behaviors, the chief effect has been shown on visual acuity; most other studies have reported other behaviors to be unaffected. However, vibration has been reported to be unpleasant, if prolonged, even though the levels are below those believed to cause physiological damage, and secondary effects might be expected on monotonous tasks. The vibrations studied have varied from 0–60 Hz, these being the range of practical interest for most environmental purposes. Interactions with frequency are known to exist. Frequencies in the range of 10–24 Hz most severely impair visual acuity, while manual tracking is most seriously affected at frequencies of about 5 Hz. This frequency also produces the greatest subjective discomfort.

Inadequate Lighting Lighting may be inadequate either because it is too low, or because it causes glare. Excessive light intensities, such as those generated by lasers, can cause physical damage to the eye. While the eyes can adapt to a wide range of light conditions, the thresholds of the photopic receptors in the fovea of the eye are higher than in the periphery (scotopic vision). The density of receptors in the fovea is also higher, and the functioning of these receptors is important for tasks requiring good visual acuity. Therefore, if the light levels fall below the photopic threshold, such tasks will suffer. Visual acuity may be tested by means of Landolt rings (circles) with small breaks whose orientation the subject must detect) or by means of grids of varying fineness and orientation. Effects of light composition have also been shown on some tasks, even well above the photopic threshold. For most visual tasks, light levels well above the photopic threshold should be maintained.

Glare causes discomfort, and may interfere with various tasks. Two kinds of glare are distinguished; direct glare from a light source in the visual field; and specular glare due to light reflected from surfaces within the visual field. Both types of glare can be controlled by suitable adjustment of the environment. Direct lighting offers maximum light at the working surface when the source is directed downward, but it is likely to produce shadows, glare and undesirable brightness contrasts. Indirect lighting avoids these problems, and while it may lessen visual efficiency in reading over very prolonged periods, it has been shown superior to direct lighting in tests of 3 hr in length.

It should also be realized that color vision depends on the receptors in the fovea. Therefore, tasks requiring the use of colors must be performed with adequate lighting. The color of surfaces depends on the spectrum of the light which falls on them, so illuminants must be chosen to provide the proper colors for tasks involving color coding.

Special Environments

Changes in Sensory Input Early reports on the effect of isolated environments from polar expeditions and the like are in agreement with later experimental findings on the undesirable effects of restricted sensory inputs. Studies on extreme sensory restriction as carried out at McGill University, where subjects wore translucent goggles and stayed in low variability environments, have shown that such effects as loss of ability to concentrate and hallucinations may result. Polar expeditions have reported compensatory behavior in the form of the development of unusually structured groups, and special interests and hobbies on the part of individual members. Similar patterns sometimes develop in such isolation situations as solitary confinement.

Studies on early sensory restriction in animals have suggested that deficits of varying degrees of permanence may develop in subsequent behavior as a result. Extreme deprivation of sensory input and social contact is known to result in a variety of deficits in human infants, even though the physical needs of the infants have been met. The monotonous effects of long periods at tasks such as monitoring sonar screens for infrequent signals, and other vigilance tasks have shown deleterious effects in terms of missed targets.

The effects of excessive sensory input are less clear. Any situation which overloads the information processing capacities of an individual will be likely to result in decremental performance. However, apart from such short-term effects, it is not clear that there are long-term effects from an overly complex environment.

The reticular activating system in the brain stem, which has been shown to have an alerting effect, has been suggested as a possible locus for some of the effects of sensory isolation.

Other Special Environments Some other special environments that have received a good deal of study are submarines and simulated space capsules. The outstanding feature of the submarine environment is its limited space. Most studies have focused on physical problems associated with this environment, but a few have been concerned with the psychological effects. In general, adaptation to crowded conditions is seems to be good, though there are some losses of personnel attributed the confined conditions. One study reported a loss of 7% of 187 men over 6 months due to this factor. Complaints about lack of privacy have been reported also.

The problem of diurnal cycles arises in connection with such environments. The usual night–day shifts are lacking, of course, in extended undersea cruises, and time confusion has been reported: for example, inability to remember if it is 8.090 a.m. or p.m. It has been found desirable to introduce features, such as news reports and evening entertainments, into the daily routine which will avoid the sensation of being cut off from the rest of the world and which serve to mark diurnal cycles. It also appears that, although such cycles can be influenced by the environment to some extent, there are limits to the process. Alluisi reports evidence of diurnal cycling in performance under a variety of work–rest schedules, including some as extreme as 4 hr of work followed by four of rest.

Simulated space capsule environments have many of the features associated with sensory deprivation conditions, plus the added factor of fatigue. Prolonged exposure to such environments has been shown to produce decrements in a number of tasks, from vigilance tasks to complex decision making functions. Diurnal rhythms in efficiency have also been reported. However, it has also been found that trained pilots, who were familiar with the concept of diurnal rhythms, did not show such effects. Both motivational and adaptation processes may be at work here. It has been shown repeatedly that measures of "can" and "will" are often very different for special environments.

Diurnal cycles may be important in less exotic environments. Efficiency in a number of tasks has been shown to vary with the diurnal cycle. These variations appears to be related to the body temperature, which reaches its lowest point after several hours of sleep, and does not reach its maximum until the evening of the next day. These variations create problems for jobs with several shifts, since operator efficiency is constantly varying. While individual cycles show adaptation to changes in times spent awake and asleep, this adaptation takes time, so that several days may be required to adapt fully to a change in shift.

THE DESIGN OF ENVIRONMENTS

This section will be concerned with the design of physical environments for maximum human comfort. A later section will be concerned with psychology's role in the design of environments in a broader sense, as in urban planning.

A great many human engineering studies have been done on special environmental problems. Anthropometric measurements have been gathered to facilitate the design of objects used by humans, and a great deal of information is available on the integration of this information with special problems, such as designing handles for use with gloves. Information about the size and movement of various parts of the body are important to proper design of a wide variety of furnishings and equipment, from chairs to work areas.

The proper display of information is another important aspect of proper environmental design. Studies have shown that certain types of dials are read more easily than others. Airplane roll indicators which used a fixed artificial horizon with the position of the plane varying have been shown to be more easily interpreted than indicators which keep the plane steady and move the horizon, though the latter design had been commonly used. A good deal of information about desirable features of scales has been obtained, and some useful principles obtained. For example, operators should not have to transform information before using it. Jet aircraft tachometers, for example, may be calibrated in percentage rpm rather than actual rpm so that the pilot need not remember the maximum rpm for different engines in order to interpret the information.

Many other factors go into the proper display of information. For example, if color coding is used, it must be remembered that, while the visual system is able to discriminate a very large number of colors, in the sense of being able to say they are different if both are presented simultaneously, the number that can be correctly distinguished when only one is presented and memory must be used is far more limited. The exact number will depend on the colors, viewing conditions, and other factors such as training, but is unlikely to reliably exceed ten in many situations and may be much lower. Knowledge of these and many other aspects of visual and auditory perception are necessary for the safe and efficient design of many facets of our environment.

Another important problem is the proper integration of human operators into complex man–machine systems. Here again, knowledge of man's information processing capacities is important. Such questions as the speed with which an operator can respond, the probability of his detecting warning signals under various conditions of display and frequency, his tracking abilities with various displays and types of targets, are merely samples of the types of questions which arise and which have been extensively studied.

It should not be assumed, however, that such considerations are important only in military or industrial situations. Proper design considerations are important for the safety and comfort of the consumer as well. One study showed, for example, that different burner-control linkages on stoves resulted in different error rates even after a number of trails. Agreement among auto manufacturers on the arrangement of controls in cars with automatic transmissions is an example of the importance of establishing conventions for the design of widely used products. The list of consumer products with features whose design will affect either comfort or safety is very long indeed.

The information processing capacities of humans are important considerations in the design of any environment. The information processing capacities of any organism, including man, are limited. In such areas of environmental concern as urban planning, traffic control, etc., such considerations can be very important. Man is capable of remembering enormous amounts of information, and performing feats of processing which defy the most complex machines, but his capacity for short term information processing is distinctly limited. A large number of studies have been done on the problem of human attention and the effect of information over-loads on various performances. The results are quite consistent in showing very great limitations to man's short term information processing capacities. Exceeding these capacities will result in more or less serious performance decrements. For example, DAF (delayed auditory feedback) studies have shown that mismatches in input information can greatly degrade performance. Individuals have only a very limited capacity for responding to different information input simultaneously to the two ears.

Consideration of man's information processing capacities should be made in planning facilities for human beings. In particular, if new environments are being proposed, a very careful study should be made of their information processing demands. However, concern should not be limited to unusual environments. Badly planned highway signs, which require a driver to attend simultaneously to too much information and make too many decisions too rapidly are an obvious, and unfortunately all too common, example of failure to consider man's information processing capacities in designing the environment. The adequacy of instructions and cautions in connection with the use of equipment or potentially hazardous substances may also require consideration of certain of man's information processing capacities. Court cases have hinged on the adequacy of cautions regarding the use of equipment to inform potential users of hazards.

There are psychological aspects to the design of larger aspects of the environment. A classic study done on a student housing project at MIT showed that social groupings were significantly influenced by the physical arrangement of the apartments. Other aspects of modern building design may be important. For example, most large modern buildings are air-conditioned today. Air conditioning may affect the ionic balance of the air and some concern has been felt that it may produce symptoms such as depression, though the evidence to date is inconclusive.

MAN'S SOCIAL ENVIRONMENT

Decisions made regarding the physical environment have implications for social aspects of behavior, as we have seen in the previous section. In addition, some areas of environmental concern, such as urban planning, must consider man's psychological and social environment more or less explicitly. Population growth, for example, is greatly affected by social attitudes, as the dramatic change in the birth rate in Japan after World War II shows.

The social sciences, including psychology, have roles as information sources and in providing techniques for implementing decisions about the social environment, especially where these involve inducing changes in human behavior patterns. A vast amount of research has gone into the study of social factors which influence human behavior, and much of this information may be relevant in certain areas of environmental planning. Psychological techniques can be used to gain information about people's needs and preferences. It is not easy to design questionnaires which will provide reliable answers and many problems arise in obtaining representative samples of the target population. Psychologists and sociologists have studied these problems extensively. A great deal is known today about proper sampling techniques for such studies. General answers to designing questionnaires are more difficult to come by, but extensive experience has provided some useful practical advice. It appears likely that many difficulties which have arisen in the past in connection with such matters as public housing in urban areas, have been due in part to a failure to accurately assess the feelings and concerns of those affected by such projects. More extensive use of survey techniques might eliminate some of these problems.

Certain reservations about research on man's social environment should be mentioned, however. Large scale experimentation is obviously very difficult to do, and so reliance must often be placed on field studies. Direct measurement of behavior may be difficult, so recourse may be made to rating scales and other indirect measures whose relation to actual behavior is often uncertain. Smaller scale experiments often introduce varying degrees of unreality into a situation which may make generalization to the real world situation difficult. While all of these comments apply in some degree to all psychological research, they are especially apt to apply in areas relevant to man's social environment.

Any discussion of man's social environment must also consider areas in which psychological techniques can be employed to alter behavior. Consider, for example, the impact of problems such as crime, drug addiction, and so forth. on urban planning. Informational aspects of these problems certainly exist. For example, what is the evidence for a relationship between juvenile delinquency and housing conditions? Such questions have been extensively studied, and while the answers are still being debated, a great deal of information is available on observed correlations.

However, in addition to such informational aspects, the question of the role of psychologists in dealing with such questions must be considered. Plans must be made for dealing with such problems, and this means that resources must be allocated to them. Obviously, the proper allocations must depend in part on available options and their effectiveness. It is therefore reasonable to ask what techniques psychologists have available for modifying behavior and to inquire into their effectiveness.

A number of techniques, generally classified as psychotherapies, have been developed to deal with neurotic behaviors. It is difficult to characterize these procedures in any simple way, since they have diverse theoretical roots and practical implementation, but in general they depend on verbal interactions between therapist and patient to effect change.

A number of variations on the classic one-to-one patient–therapist relationship have developed in recent years. Various forms of group therapy have developed. In addition, sensitivity training procedures have been developed. These are usually aimed at changing the behavior of those not considered to be suffering from any personality disturbance, and they are likely to involve varying degrees of role playing and direct confrontation.

The effectiveness of psychotherapies has long been a matter of debate and the question has not yet been resolved. Difficulties in evaluating them in regard to neurotic behaviors stem from the high spontaneous remission rates, the difficulties of diagnosis, and the problems of defining what is meant by a cure and determining when one has occurred. The record with regard to disorders such as psychopathic personality, drug addiction, etc., is even less optimistic. Very little evidence of effectiveness is available, and many reports are discouraging.

The situation is similar for sensitivity training procedures and group therapy. While enthusiastic reports can be found, so can negative reports. Especially disturbing are reports of undesirable reactions to some kinds of sensitivity training, which can be rather aversive.

To those familiar with other health areas, these problems of evaluation may seem surprising. Clinical research has a long history in medicine, after all, and while problems such as placebo effects exist, methods have been developed to overcome them. However, in addition to the usual problems associated with any clinical research, much of the research on psychotherapies suffer from severe underlying problems associated with the entire concept of mental illness. Szasz has gone so far as to deny the utility of the concept, and has developed some interesting analogies between the concept of mental illness and the concept of witchcraft during the Middle Ages. While his views are hardly universally accepted, it is clear that the problems go beyond the merely procedural.

Behavior therapies, derived from conditioning techniques fare somewhat better. Classical conditioning techniques have had some success with alcoholics, though problems exist. Operant condition techniques have been applied extensively in the treatment of retardates, and have been used as the basis of token economies in mental institutions and homes for delinquents. It is usually somewhat easier to evaluate the behavior therapies, in part because the specification of the procedures in terms of reinforcement contingencies provides information about details of the process often lacking for other methods, and also because they focus on changing behavior. Such changes in behavior can usually be measured directly, whereas changes in personality or emotional state must be measured indirectly, and the appropriate tests are often uncertain. Behavior therapies are not invariably successful, of course, and a current weakness of the procedures is that it is difficult to establish systematic criteria for use. Another problem which has been reported are unpredicted effects on behaviors which were not under explicit control by reinforcement. These changes are not necessarily negative, but since they were not under intentional stimulus control, they may create problems. Finally, these procedures are relatively new, and the history of psychology is replete with procedures whose early promise has not been fulfilled.

Although the major use of explicit control of behavior by reinforcement contingencies has been in dealing with abnormal behaviors, the principles are not limited to such applications. An interesting example of an application in a different context is the use of reinforcement techniques to reduce littering behavior in theatres and campgrounds.

In addition to the techniques described above, psychologists and sociologists have been concerned with the effectiveness of communication in changing behavior. Much research has gone into such questions as the effect of the status of a speaker on his persuasiveness, and the role of various social pressures in changing opinion. In a sense, much educational research can be looked upon as the study of a special kind of behavior control where the concern is with inducing behaviors which result in learning in an educational setting. Rothkopf has done some interesting experiments on methods of inducing students to adopt appropriate behaviors to facilitate learning, which he refers to as mathemagenic behaviors.

In evaluating the usefulness of such techniques in the design of the social environment, it is important to take cost-effectiveness criteria into consideration. Judged by such criteria, the more traditional psychotherapies do not come off well. They require highly trained manpower, in a relatively large supply, and effectiveness is questionable, especially in situations where the cooperation of the target population is in doubt. Modifications such as group therapy may dilute the manpower requirements somewhat, but the other problems remain. The behavior therapies show up somewhat better in such an analysis. At least, it is possible to monitor behavior change to determine whether they are effective, and once a program has been established, it can often be carried out by personnel with little formal training.

In evaluating psychology as an information source for social planning, the need for sound information from the behavioral sciences cannot be overstressed, and every effort should be made to obtain more of such information. At the same time, the many obvious limitations of our present data suggest that they be used with due caution.

In particular, with the increasing use of psychological consultants in various aspects of environmental planning, the bases for various expert opinions should be scrutinized with great care. The judgment of experts in the social sciences, when it is not based on carefully controlled research, can be very misleading. The nature of the subject matter and the conditions of observation are such that it is very easy to find apparent support for a wide variety of hypotheses, many of which may be proved incorrect with more careful observation. It is always desirable to ascertain the factual basis for any recommendation, and to obtain the views of as many professionals from different backgrounds as possible.

FUTURE DEVELOPMENTS

The future cannot, of course, be predicted, but certain trends seem likely to continue. First, techniques for controlling behavior derived from operant conditioning procedures appear likely to come into increasing use. Environmental design will be likely to take into account more explicitly such facts in its design procedures. Human engineering procedures will also have an increasing impact on aspects of design for the environment. Certain fundamental questions about human behavior in complex situations will become better understood, and will provide the informational basis for superior design of learning environments and other special purpose environments. However, the 1984 or Brave New World fears seem unjustified at present. The more extensive attempts to control human behavior are more notable for their weaknesses than their power. Even predictive power is inversely associated with the behavioral complexity of the response. Skinnerian techniques seem to offer the best road at present to more powerful approaches. However even here caution should be noted. Most of the basic research on current contributions has been done on animals in very restricted environments. It has also been done on animals in more or less severe states of need; pigeon experiments using a food reinforcement typically maintain the birds at some percentage of *ad lib* body weight, such as 85%. Such animals are very well motivated, and the powerful effects of positive reinforcements found in these studies should be interpreted in this light. It is possible that most humans remain in a similar state of need with regard to the kinds of social reinforcers most likely to be manipulated in planning of larger environments, but the point has yet to be established. The limited environment is another factor which may lead to extensive modification of conditioning principles in larger situations, where a much wider choice of response alternative exist. However, the positive value for effective use of environmental resources which can result from a better understanding of human needs and response to various environments is very great, and it is in this area that the greatest impact of psychology should be anticipated.

REFERENCES

Alloway, Thomas M. (1969) Effects of low temperature upon acquisition and retention in the grain beetle (*Tenebrio molitor*), *J. Comparative and Physiological Psychology,* **69**, 1.

Alluisi, E.A. and W.D. Chiles (1967) Sustained performance, work–rest scheduling, and diurnal rhythms in man, *Acta Psychologica,* **27**, 436.

American Public Health Association (1990) *Standard Methods for the Examination of Water and Waste Water,* 17th Ed.

Atkinson, J.E., T.D. Crocker and H.L. Needleman, The importance of specification uncertainty and intolerance to measurement error in a study of the impact of dentine lead on children's IQ's, *Int. J. Env. Studies,* **29**, 2/3, 1987.

Baker, Rodney R. (1968) The effects of psychotropic drugs on psychological testing. *Psychological Bull.*, **69**, 377.

Bare, Carole E. and R.R. Mitchell (1972) Experimental evaluation of sensitivity training, *J. Appl. Behavioral Science,* **8**, 263.

Bellinger, D.C. *et al.*, Extra sensory-motor development and prenatal exposure to lead, *Neurobehavioral Toxicology and Teratology,* **6**, 1984.

Bennett, Corwin A. (1971) Toward empirical, practicable, comprehensive task taxonomy, *Human Factors,* **13**, 229.

Bierman, Ralph, R.R. Carkhuff and M. Santilli (1972) Efficacy of empathic communication training groups for inner city preschool teachers and family workers, *J. Appl. Behavioral Science,* **8**, 188.

Biersner, Robert J. (1972) Selective performance effects of nitrous oxide, *Human Factors,* **14**, 187.

Blake, Robert R. and Jane Syrgley Mouton (1971) A behavioral design for the development of society, *J. Appl. Behavioral Science,* **7**, 146.

Breisacher, Peter (1970) Neuropsychological effects of air pollution, *Am. Behavioral Scientist,* **14**, 837.

Brozek, J. (1964) Recent developments in Soviet psychology, in *Annual Review of Psychology,* Ed. by P.R. Farnsworth, O. McNemar and Q. McNemer, **15**, pp. 493–594.

Burgess, Robert L., R.N. Clark and J.C. Hendee (1971) An experimental analysis of anti-litter procedures, *J. Appl. Behavior Anal.,* **4**, 71.

Burns, Neal M. and D. Kimura (1963) Isolation and sensory deprivation, in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler, The Free Press of Glencoe, Division of the MacMillan Co., New York, pp. 167–192.

Buros, Oscar K. (Ed.) (1965) *The Sixth Mental Measurements Yearbook,* Gryphon Press, Highland Park, NJ.

Byrd, J.F. and A.H. Phelps (1968) Odor and its measurement, in *Air Pollution,* Ed. by A.C. Stern, (2nd Ed.), **I**, Academic Press, New York, pp. 305–327.

Cahoon, R.L. (1972) Simple decision making at high altitude. *Ergonomcis,* **15**, 157.

Calhoun, J.B. (1962) Population density and social pathology. *Scientific American,* **206**, 139.

Cannell, Charles F. (1968) Interviewing, in *The Handbook of Social Psychology,* Ed. by G. Lindzey and E. Aronson (2nd Ed.). **II.** Addison-Wesley, Reading, Mass., pp. 526–595.

Cartwright, Dorwin and Alvin Zander (Eds.) (1960) *Group Dynamics; Research and Theory,* (2nd Ed.). Row, Peterson and Co., Illinois.

Chapanis, A. (1959) *Research Techniques in Human Engineering,* The Johns Hopkins Univ. Press, Baltimore, Md.

Chiles, W.D. and A.E. Jennings (1970) Effects of alcohol on complex performance, *Human Factors,* **12**, 605.

Chiles, W.D., P.F. Iampietro and E.A. Higgins (1972) Combined effects on altitude and high temperature on complex performance, *Human Factors,* **14**, 161.

Clark, Roger N., R.L. Burgess and J.C. Hendee (1972) The development of anti-litter behaviour in a forest campground, *J. Appl. Behavior Analysis,* **445**, 1.

Cole, Sherwood O. (1967) Experimental effects of amphetamine: A review, *Psychological Bull.*, **68**, 81.

Colquhoun, W.P., M.J.F. Blake and R.S. Edwards (1969) Experimental studies of shift-work III: stabilized 12-hour shift systems, *Ergonomics,* **12**, 865.

Colquhoun, W.P. (1970) Circadian rhythms, mental and efficiency and shift work, *Ergonomics,* **13**, 558.

Cullen, Joseph W. (1969) Modification of salt-seeking behavior in the adrenalectomized rat via gamma-ray irradiation, *J. Comparative and Physiological Psychology,* **68**, 524.

Deese, J. and S.H. Hulse (1967) *The Psychology of Learning,* McGraw-Hill, New York.

Deutsch, J.A. (1969) The physiological basis of memory, in *Annual Review of Psychology,* Ed. by P.H. Mussen and M.R. Rosensweig, **20**, pp. 85–104.

Dill, D.B., E.F. Adolph and C.G. Wilber (Eds.) (1964) *Handbook of Physiology, Section 4: Adaptation to the Environment.* American Physiological Society, Washington, DC.

Faith, W.L. (1983) Effects of atmospheric pollutants, *The Encyclopedia of Environmental Science and Engineering,* 2nd Ed., Gordon and Breach, New York.

Fiske, D.W., H.F. Hunt, L. Luborsky, M. Orne, M.B. Parloff, M.F. Reiser and A.H. Tuma (1970) Planning of research on effectiveness on psychotherapy, *Am. Psychologist,* **25**, 727.

Fleishman, E.A. and C.J. Bartlett (1969) Human abilities, in *Annual Review of Psychology.* Ed. by P.H. Mussen and M.R. Rosenzweig, **20**, pp. 349–380.

Fowler, Frank D. (1972) A consumer ombudsman: The human factors engineer, *Human Factors,* **14**, 7.

Freeman, H.E. and J.M. Giovannoni (1969) Social psychology of mental health, in *The Handbook of Social Psychology,* Ed. by G. Lindzey and E. Aronson, (2nd Ed.) V. Addison-Wesley, Reading, Mass., pp. 660–719.

Freeman, H., ed., Mental health and the environment, Churchill Livingston, London, 1984.

Furchtgott, Ernest, J. Reid Jones, R. Stephen Tacker and John Deagle (1970) Aversive conditioning in prenatally X-radiated rats. *Physiology and Behavior,* **5**, 571.

Goldsmith, J.R. (1968) Effects of air pollution on human health, in *Air Pollution,* Ed. by A.C. Stern, (2nd Ed.), **I**, Academic Press, New York, pp. 305–327.

Goodstein, Leonard D. and Ronald K. Boyer (1972) Crisis intervention in a municipal agency: A conceptual case history, *J. Appl. Behavioral Science,* **8**, 318.

Goranson, R.E. (1970) Media violence and aggressive behavior: A review of experimental research, in *Advances in Experimental Social Psychology,* Ed. by L. Berkowitz, **5**, Academic Press, New York.

Gordon, Robert A. (1967) Issues in the ecological study of delinquency, *Am. Sociological Rev.*, **32**, 927.

Graham, C.H. (Ed.) (1965) *Vision and Visual Perception,* John Wiley, New York.

Grether, Walter F. (1971) Vibration and human performance, *Human Factors,* **13**, 203.

Hall, R. Vance, Saul Axelrod, Lucille Tyler, Ellen Grief, Fowler C. Jones and Roberta Robertson (1972) Modification of behavior problems in home with a parent as observer and experimenter, *J. Appl. Behavior Analysis,* **5**, 53.

Hendler, Edwin (1963) Temperature effects on operator performance. in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler, The Free Press of Glencoe, Division of The MacMillan Co., New York, pp. 321–352.

Herbert, E.W. and D.M. Baer (1972) Training parents as behavior modifiers: Self-recording of contingent attention, *J. Appl. Behavior Analysis,* **5**, 139.

Hopkins, E.S., W.M. Bingley and G.W. Schucker (1970) *The Practice of Sanitation,* (4th Ed.), Williams and Wilkins, Baltimore, Md.

Jones, J.R., E.N. Zeigler and J.R. Pfafflin (1983) Effects of chemical pollutants on human ecology, *The Encyclopedia of Environmental Science and Engineering,* 2nd Ed, Gordon and Breach, New York.

Kling, J.W. and L.A. Riggs (Eds.) (1971) Woodworth and Schlosberg's *Experimental Psychology,* (3rd Ed.). Holt, Rineheart and Winston, New York.

Kish, Leslie (1965) *Survey Sampling,* John Wiley and Sons, New York.

Klauss, A.K., R.H. Tull, L.M. Roots and J.R. Pfafflin (1970) History of the changing concepts in ventilation requirements, *ASHRAEJ,* 51.

Kobrick, J.L. (1972) Effects of hypoxia on voluntary response time to peripheral stimuli during central target monitoring, *Ergonomics,* **15**, 147.

Korbel, Susan F., Kenneth D. Johnson and Paul S. Rowland (1971) After-effects of low intensity UHF radiation, *Psychonomic Science,* **23**, 50.

Kristt, Donald A. and Jeri A. Sechzer (1969) Behavioral thermoregulation at high environmental temperatures: The intra-session pattern of cooling and resting, *Psychonomic Science,* **16**, 151.

Kuman, R., I.P. Stolerman and Hannah Steinberg (1970) Psychopharmacology, *Ann. Rev. Psychology,* **21**, 595.

Leibowitz, Herschel W., Charles N. Abernethy, E.R. Buskirk, Oded Bar-Or, and Robert T. Hennessy (1972) The effect of heat stress on reaction time to centrally and peripherally presented stimuli, *Human Factors,* **14**, 155.

Lieberman, Morton A., Irvin D. Yalom and Matthew B. Miles (1972) The impact of encounter groups on participants: Some preliminary findings, *J. Appl. Behavioral Science,* **8**, 29.

Lion, Judith S., E. Richardson and R.C. Brown (1968) A study of the performance of industrial inspectors under two kinds of lighting, *Ergonomics,* **11**, 23.

Loeber, Rolf (1971) Engineering and behavioral engineer, *J. Appl. Behavior Analysis,* **4**, 321.

Loftus, J.P. Jr. and Lois Hammer (1963) Weightlessness, in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler. The Free Press of Glencoe, Division of The MacMillan Co., New York, pp. 353–377.

Lorenz, K. (1966) *On Aggression,* Harcourt, Brace and World, New York.

Loborsky, L.,M. Chandler, A.H. Auerbach, J. Cohen and H.M. Bachrach (1971) Factors influencing the outcome of psychotherapy: A review of quantitative research, *Psychological Bull.*, **75**, 145.

Magill, P.L., F.R. Holden, C. Ackley and F.G. Sawyer (Eds.) (1956) *Air Pollution Handbook,* McGraw-Hill, New York.

Maling, G. (1983) Noise. *The Encyclopedia of Environmental Science and Engineering,* 2nd Ed. Gordon and Breach, New York.

Martin, J.C. (1968) Sex differences in the activity wheel as a function of fetal age at irradiation, *Psychonomic Science,* **13**, 249.

Mason, W.A. and A.J. Riopelle (1964) Comparative psychology, in *Annual Review of Psychology,* Ed. by P.R. Farnsworth, O. McNemar and Q. McNemar, **15**, pp. 143–1801.

Matarazzo, J.D. (1965) Psychotherapeutic processes, in *Annual Review of Psychology,* Ed. by P.R. Farnsworth, O. McNemar and Q. McNemar, **16**, pp. 181–204.

Maxcy, K.F. (Ed.) (1956) Rosenau's *Preventive Medicine, and Public Health,* (8th Ed.), Appleton-Century-Crofts, New York.

Miller, James M. and Thomas H. Rockwell (1972) The use of human subjects in human factors research, *Human Factors,* **14**, 35.

Moore, W.E. (1969) Social structure and behavior, in *The Handbook of Social Psychology,* Ed. by G. Lindzey and E. Aronson, (2nd Ed.), **IV.** Addition-Wesley, Reading, Mass., pp. 283–319.

Morgan, Clifford T., Jesse S. Cook, A. Chapanis and Max W. Lund (Eds.) (1963) *Human Engineering Guide to Equipment Design,* McGraw-Hill, New York.

Nadel, Aaron B. (1963) Vibration, in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler, The Free Press of Glencoe, Division of The MacMillan Co., New York, pp. 379–394.

Peters, George A. (1972) The role of the human factors engineer in product liability litigation, *Human Factors,* **14**, 13.

Phillips, E.L., E.A. Phillips, D.L. Fixsen and M.M. Wolf (1971) Achievement place: Modification of the behaviors of pre-delinquent boys within a token economy, *J. Appl. Behavior Analysis,* **4**, 45.

Roe, A. (1970) Community resource centres, *Am. Psychologist,* **25**, 1033.

Rothkopf, E.Z. (1971) Experiments on mathemagenic behavior and the technology of written instruction, in *Verbal Learning Research and the Technology of Written Instruction,* Ed. by E.Z. Rothkopf and P. E. Johnson. Teachers College Press, New York.

Ruff, George E. (1963) Psychological and psychophysiological indices of stress, in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler, The Free Press of Glencoe, Division of The MacMillan Co., London, pp. 87–125.

Schoenfeldt, L.F. (1970) Data archives as resources for research, instruction, and policy planning, *Am. Psychologist,* **25**, 609.

Scott, J.P. (1969) The social psychology of infrahuman animals, in *The Handbook of Social Psychology,* Ed. by G. Lindzey and E. Aronson, (2nd Ed.), Vol. IV, Addison-Wesley, Reading, Mass., pp. 611–642.

Scott, Wm. A. (1968) Attitude measurement, in *The Handbook of Social Psychology,* Ed. by G. Lindzey and E. Aronson, (2nd Ed.), **II**, Addison-Wesley, Reading, Mass., pp. 204–273.

Shontz, William D., Gerald A. Trumm and Leon G. Williams (1971) Color coding for information location. *Human Factors,* **13**, 237.

Simons, David G., D.E. Flinn and B. Hartman (1963) Psychophysiology of high-altitude experience, in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler, The Free Press of Glencoe, Division of The MacMillan Co., New York, pp. 127–164.

Sommer, R. (1969) *Personal Space: the Behavioral Basis of Design,* Prentice-Hall, Englewood Cliffs, NJ.

Stokinger, H.E. and D.L. Coffin (1968) Biologic effects of air pollutants, in *Air Pollution,* Ed. by A.C. Stern, (2nd Ed.), **1.** Academic Press, New York, pp. 445–546.

Sweets, John A. (1964) *Signal Detection and Recognition by Human Observers: Contemporary Readings,* John Wiley, New York.

Szasz, Thomas (1970) *Manufacture of Madness: A Comparative Study of the Inquisition and the Mental Health Movement,* Harper and Row, New York.

Tobias, Paul R. (1963) The effect of radiation on integrated behavior, in *Unusual Environments and Human Behavior,* Ed. By Neal M. Burns, R.M. Charnbers and E Hendler, The Free Press of Glencoe, Division of The MacMillan Co., New York, pp. 395–417.

Trumbull, Richard (1966) Diurnal cycles and work-rest scheduling in unusual environments, *Human Factors,* **8**, 385.

Twardosz, S. and T. Sajwaj (1972) Multiple effects of a procedure to increase sitting in a hyperactive, retarded boy. *J. Appl. Behavior Analysis,* **5**, 73.

Van Cleave, C.D. (1963) *Irradiation and the Nervous System,* Rowman and Littlefield, New York.

Warner, Harold D. and Norman W. Heimstra (1972) Effects of noise intensity on visual target-detection performance, *Human Factors,* **14**, 181.

Welch, B.L. and A.S. Welch (Eds.) (1970) *Physiological Effects of Noise,* Plenum Press, New York–London.

Weybrew, Benjamin B. (1963) Psychological problems of prolonged marine submergence, in *Unusual Environments and Human Behavior,* Ed. by Neal M. Burns, R.M. Chambers and E. Hendler, The Free Press of Glencoe, Division of The MacMillan Co., New York pp. 87–125.

Wohlwill, Joachim (1970) The emerging discipline of environmental psychology, *Am. Psychologist,* **25**, 303.

Woodworth, R.S. and H. Schlosberg (1954) *Experimental Psychology,* (Rev. Ed.) Henry Holt and Co., New York.

Woollel, J. (1972) The measurement of driver performance, *Ergonomics,* **15**, 81.

Young, R.D. (1967) Developmental psychopharmacology: A beginning. *Psychological Bulletin,* **67**, 73.

FURTHER READING

Bachmann, M.O., (1995) Influences on sick building syndrome symptoms, *Social Science and Medicine,* Vol. 40, no. 2, pp. 245–251. Jan.

Baldassare, M., (1992) The threat of environmental problems as predictors of environmental practices, *Environment and Behavior,* Vol. 24, no. 5, pp. 602–616, Sept.

Dunn, J.R., (1994) Psychological aspects of PCB contamination and remediation: The use of Smithville, Ontario, *Social Science and Medicine,* Vol. 39, no. 8, pp. 1093–1104, Oct.

Dwyer, W.O., (1993) Critical review of behavioral interventions to preserve the environment: Research since 1980, *Environment and Behavior,* Vol. 25, no. 3, pp. 275–321, May.

Dykman, J.B. (1992) Is ignoring ergonomic warnings courting disaster?, *Modern Office Technology,* Vol. 37, no. 8, p. 10, Aug.

Fernberg, P.M., (1994) Are workplace ailments all they're claimed? *Modern Office Technology,* Vol. 39, no. 3, pp. 27–36, March.

Fisher, A., (1996) Toward a more radical ecopsychology, *Alternatives,* Vol. 22, no. 3, pp. 20–26, July.

Fischoff, B., (1990) Psychology and public policy, *American Psychologist,* Vol. 45, no. 5, pp. 647–653, May.

Gatchel, R.A., (1991) Psychophysiological effects of toxic chemical contamination exposure: A community field study, *Journal of Applied Social Psychology,* Vol. 21, no. 4, pp. 1961–1976, Dec. 16.

Hedge, A., (1992) Ecological ergonomics: The study of human work environments, *Impact of Science on Society,* Vol. 42, no. 1, pp. 53–64.

Kitchen, R., (1995) Strange bedfellows: Geography and psychology, *Geographical,* Vol. 67, no. 9, p. 10, Sept.

Larsen, K.S., (1995) Environmental waste: Recycling attitudes and correlates, *Journal of Social Psychology,* Vol. 135, no. 1, pp. 83–88, Feb.

Meyer, J.P., (1995) Survey of heat stress in industry, *Ergonomics,* Vol. 38, no. 1, pp. 36–46, Jan.

Millbrath, L.W., (1995) Psychological, cultural and informational barriers to sustainability, *Journal of Social Issues,* Vol. 51, no. 4, 101–120, Winter.

Oskamp, S., (1995) Applying social psychology to avoid ecological disaster, *Journal of Social Issues,* Vol. 51, no. 4, pp. 217–239, Winter.

Parson, K.C., (1995) Ergonomics of the physical environment: International ergonomics standards concerning speech communications, danger signals, lighting, vibration and surface temperatures, *Applied Ergonomics,* Vol. 26, no. 4, pp. 281–292, Aug.

Rich, R.C., (1995) Citizen participation and empowerment: The case of local environmental hazards, *American Journal of Community Psychology,* Vol. 23, no. 5, Oct.

Richards, J.M., (1991) Units of analysis and psychometrics of environmental assessment scales, *Environment and Behavior,* Vol. 23, no. 4, pp. 423–437, July.

Staples, S.L., (1996) Human response to environmental noise: Psychological research and public policy, *American Psychologist,* Vol. 51, no. 2, pp. 143–150, Feb.

Stern, P.C., (1992) Psychological dimensions of global environmental change, *Annual Review of Psychology,* Vol. 43, pp. 269–302.

Stokols, D., (1995) The paradox of environmental psychology, *American Psychologist,* Vol. 50, no. 10, pp. 821–837, Oct.

Sundstrom, E., (1996) Environmental psychology 1989–1994, *Annual Review of Psychology,* Vol. 47, pp. 485–512.

Vaughn, E., (1996) Individual and cultural differences in adaptation to environmental risks, *American Psychologist,* Vol. 48, no. 6, pp. 673–680, June.

SHEILA M. PFAFFLIN
AT&T

PUBLIC HEALTH: see COMMUNITY HEALTH; ENVIRONMENTAL HEALTH

R

RADIATION ECOLOGY

Radiation Ecology or Radioecology is a term that came into common usage in 1956 to denote that area of the broad field of ecology concerned with the assessment of radioactivity in the environment. More specifically, radiation ecology has come to be recognized as that area of ecology concerned with radioactive substances, radiation and the environment. The development and subsequent expansion of nuclear energy for military and peaceful purposes has been accompanied by environmental problems, some of which are typical of other facets of industrialization and some unique to atomic energy. The unique problems primarily concern the fate and ecological effects of radionuclides released into the environment.

The major environmental problems introduced by the Atomic Age may be grouped into several areas of scientific and public concern. Underlying each of these is the worry about the effects of ionizing radiation—on man, his domesticated plants and animals, and on the environment and its living components, Fallout from weapons testing, reactor radioactive waste effluents, radioactive waste disposal, nuclear war, and use of nuclear explosives for major engineering and related technological projects of large scale comprise the activities which have concerned society and which, because of potential impact on the environment and man, have stimulated the development of radiation ecology. Understanding the manner in which our ecological systems (ecosystems) distribute, assimilate, and affect the environmental behavior of radioactive substances, and the effects of radiations emitted from those substances, are the concern of the radioecologist.

RADIONUCLIDES OF ECOLOGICAL IMPORTANCE

Radionuclides which are of interest to the ecologist are listed in Table 1. These radioactive elements represent the major naturally-occurring and man-made sources of radiation in the environment. Principal sources of exposure from background (natural) radiation are represented by the uranium, thorium and actinium decay series. Internal exposure to man results primarily from ^{40}K, ^{14}C, ^{226}Ra, and ^{228}Ra and their daughter products that are deposited in the body. Radionuclides such as ^{222}Rn and ^{220}Rn and their daughter products represents sources of internal radiation exposure to man from inhalation.

Radionuclides produced by the fissioning of uranium (fission products) are of the greatest current concern. These man-made isotopes are not essential to organisms, but they constitute the major sources of radiation in the environment whether it be from fallout or waste disposal from reactor operations. All of these radionuclides may enter ecosystems where they become part of the flux of systems that are being circulated within and between systems.

Some of the fission products which are chemically similar to biologically essential elements are of special interest. They vary greatly in their physical half-life and in the extent to which they participate in metabolic processes of living organisms. The most important radionuclides affecting plants and animals on land are strontium-90, cesium-137, and iodine-131.

Strontium-90 remains in the environment for a long time. Its half-life is 28 years. Strontium is chemically similar to calcium, and it can enter living organisms as a replacement for calcium. In weapons fallout strontium-90 is usually deposited in the root systems of plants. the amount of ^{90}Sr that a plant absorbs from soils depends on several factors, particularly on the quantity of calcium in the soil, the relative quantities of calcium and strontium at the depth where the roots are located, and the ability of the plant to discriminate between the two elements. The plant is the base for ^{90}Sr to enter the human food chain. This chain is a short and simple one consisting of plants, cows, and man, with cow's milk being the chief source of entry into man. There is considerable discrimination against the transfer of strontium at each step in this food chain, but the small amount that is transported to man tends to concentrate in bone tissue. It remains there, undergoing radioactive decay and emitting its radiation. Its danger is related to the fact that bone contains blood forming (erythropoetic) tissue. In sufficient quantities the radiation can cause leukaemia and bone cancer.

TABLE 1
Radionuclides of ecological importance

Category	Major radionuclides	Ecological importance
Naturally occurring radionuclides	Uranium Thorium Actinium series elements Potassium-40 Carbon-14	Major contributors to background radiation (long half-lives)
Fission products	Strontium-89, 90, 91 Yttrium-90, 91 Zirconium-95 Niobium-95 Ruthenium-103, 106 Rhodium-106 Iodine-131 Cesium-137 Barium-137, 140 Lanthanum-140 Cerium-141, 144 Praeseodymium-143, 144 Neodymium-147 Promethium-147	Enter ecological systems through fallout or waste disposal (half-lives ranging from a few hours to 30 yrs)
Radioisotopes of elements essential to organisms	Hydrogen-3 Cobalt-60 Carbon-14 Sodium-22, 24 Phosphorus-32 Sulfur-35 Potassium-42 Calcium-45 Manganese-54 Iron-59 Copper-64 Zinc-65	Used as tracers in both radionuclide cycling and radiation effects studies on organisms and ecological systems

Strontium-90 accumulation normally is greater in children than in adults, because growing children are building bone at a greater rate and use a larger amount of calcium. A study of species in the deer family (cervidae) graphically demonstrated the effects of ^{90}Sr fallout. The levels of ^{90}Sr in the antlers of deer rose continually from 1947 through 1955, then remained constant for 2 yrs, and rose again in 1958. The concentration of ^{90}Sr was more than 8 × as high in 1958 as it was in 1947. With the cessation of major weapons tests in the 1960's, the levels began to drop off.

Cesium-137 is another major fission product that is found in fallout and untreated radioactive waste effluents. Cesium behaves chemically very much like potassium and follows the same metabolic route in plants and animals as potassium does. It enters plants directly through the leaves after being deposited by rain, and so it appears in plant tissues more quickly than does ^{90}Sr.

From there on the route of ^{137}Cs is much the same as ^{90}Sr; it appears in the milk and muscles of cattle that eat contaminated plants, and it enters the human body in food. Once in the human body, it becomes part of muscle tissue and so has an almost uniform distribution throughout the body. It stays there for only about 4 months. Since its half-life, like that of ^{90}Sr, is about 28 yrs, little of the ^{137}Cs undergoes radioactive decay while in the body.

The transfer of ^{137}Cs and ^{90}Sr from plants to animals also has been observed in species that are not important as food sources for man. The coconut crab, a land animal that lives on a diet of fruit and nuts on islands in the Pacific, was found to have accumulated radioactive materials as a result of the Pacific test explosions. Strontium-90 was found in the skeleton, and ^{137}Cs was found in the soft body parts—direct results of eating contaminated vegetation.

The third radionuclide or fission product of importance in fallout is radio-iodine (^{131}I). The chemistry of radioactive ^{131}I is exactly like that of natural ^{127}I, which is not radioactive. Therefore, its concentration in the body depends only on the concentration in the source material. Iodine becomes concentrated in the thyroid glands of vertebrate animals, where it can cause cancer of the thyroid and damage to other tissues. Like ^{90}Sr and ^{137}Cs, it enters plants as a result of radioactive fallout and then enters humans either by way of the plants themselves, or by way of contaminated milk from cows that eat the plants. The radioactivity of thyroid glands removed from certain animals can serve as a sensitive indicator of ^{131}I in the environment, because the concentration of ^{131}I in the thyroid can be as much as 10,000 × higher than the concentration in nature. The black-tailed jack-rabbit is a useful animal for such measurements. It has a large thyroid gland that is easy to remove and weigh. The level of radioactivity in each gram of its thyroid tissues varies directly with the fallout on vegetation.

Finally there are man-made radionuclides (activation products) which are important because they are the isotopes of elements which may be essential to plants and animals. Some of these also may enter the environment as activation products resulting from reactor operations or nuclear explosives. Examples of activation radionuclides are cobalt-60 and zinc-65. In aquatic or marine environments these radionuclides have been found to accumulate in food organisms especially shellfish and mollusks. Generally ^{60}Co can be anticipated to be accumulated by organisms or to be retained in organically enriched materials such as forest floor humus and organic sediment. Zinc-65 is of particular concern in marine environments where it is likely to be accumulated in clams and oysters. However, being an activation rather than a fission product its presence depends more on appropriate stable elements present which in turn are exposed to fast neutrons than on fissionable material.

RADIOSENSITIVITY OF ECOLOGICAL SYSTEMS

Although there is much current concern about the possible effect of low level, chronic radiation on ecological systems, there is relatively little comprehensive scientific data on

TABLE 2
Comparative radiosensitivity of groups of organisms

Group	Lethal dose range[a] (rads)
Bacteria	100,000–1,000,000
Insects	5,000–100,000
Fish	1,000–300
Mammals	300–1,200
Herbaceous plants	5,000–70,000
Coniferous trees	800–3,000
Deciduous trees	4,000–10,000

[a] Estimated acute whole body gamma radiation doses required to kill 50% or more of the adult organisms.

this problem. Much more is known about the radiosensitivity of organisms exposed to radiation doses which are much higher than we expect to contend with in the normal environment. In general, higher animals are far more sensitive to radiation than are lower animals, and the very young and the aged are more sensitive than mature, healthy animals. For example Table 2 gives estimated acute doses of gamma of X-radiation necessary to kill 50% or more of the adult members of several groups of organisms. These data should be considered only as an indication of relative radiosensitivity as they represent generalized ranges. Half the humans exposed to a single dose of 500 R will die. For other mammals the lethal dose ranges from less than 300–1200 R. Frogs and newts can survive higher radiation levels, depending on their body temperatures at the time of exposure. Insects can survive doses of up to 100,000 R in a few instances; most have lethal doses in the 10,000–20,000 R range. These kinds of data do not reflect the more complex responses of organisms subjected to ionizing radiation under natural conditions. For example, most organisms go through several stages of development from egg to adult. These stages may take place in different parts of the ecosystem. Likewise, the radiosensitivity of these organisms may differ in different stages of the life cycle. In radioresistance groups such as insects, 10% or less of the lethal dose to adults may be effective at juvenile or egg stages. For example, in the bagworm a dose of 450 R is sufficient to kill 50% of 1-day old eggs, whereas a dose of approximately 10,000 R was required to produce the same effect in the larvae.

Effects other than lethality also may be produced by radiation, especially in ecosystems where all organisms are linked through various interactive processes. Aside from genetic or reproductive effects, changes in number, growth, disease resistance, life span or response to physical environmental factors are of interest to the ecologist. Radiation-induced changes can affect the role of organisms or populations within the ecosystem. Predator–prey relationships, food chain transfers, and other ecological processes which depend on the continuing interaction between different organisms may be altered by the impact of ionizing radiation.

The effect of ionizing radiation on plants has been studied both outdoors and in greenhouses. One indoor test field consists of 10 acres of land with a ^{60}Co source located at the center. It is installed in a vertical tube, which can be raised to different heights for irradiation and then lowered by remote control into a lead case when not in use. Various species of plants grow in the soil in concentric circles around the source. Each species is arranged in a wedge-shaped area so that the plants are located at various distances from the source and receive various intensities of radiation. Plants are exposed to radiation for 20 hr a day.

Radiation effects on plants are complex and depend on a number of factors, including the plant species, the maturity of a plant, its physical condition, the parts of the plant exposed to radiation, the kind and amount of radiation, and the rate at which the radiation dose is applied. Woody plants generally are more sensitive to radiation than are herbaceous plants (Table 2). Gymnosperms are more sensitive to radiation than angiosperms. A pine tree shows severe growth inhibition at a level of about 10 R/day, while the same degree of inhibition in a gladiolus plant requires about 5000 R/day. Some ecologists have speculated that radiation from a nuclear attack would destroy all pine trees and other gymnosperms in irradiated areas, leaving other plants relatively unharmed.

It is possible to predict some radiation effects in plants. The meristematic or growth regions in plants are the most radiosensitive tissues. It is the absorption of radiation energy in these regions that alters plant growth and development. Ecologists and botanists have shown that the response of plants to ionizing radiation is directly proportional to the interphase chromosome volumes in meristematic tissues. That is, plant species with large chromosomes are more sensitive; those with small chromosomes are more resistant to radiation. In general, this is an extremely useful concept, and it has been applied to predict and assess probable radiation effects on vegetation (natural and agricultural) from military uses of nuclear devices.

Seeds are far less sensitive to radiation than are growing plants. A stand of pine trees exposed to a total of 12,000 R of gamma radiation was 90% destroyed, yet 95% of the seeds taken from cones on the same trees were viable. The high resistance of seeds to radiation damage is probably associated with their low water and oxygen content. The sensitivity of dry seeds varied widely among species, however. Lily seeds show practically no ability to sprout after receiving a dose of 2000 R. Yet the seeds of other plants seem to be stimulated to sprout more vigorously than normal under the same amount of radiation or more. Such differences favor the growth of certain species over others in areas where radiation is a factor in the environment.

EFFECTS OF RADIATION ON ECOSYSTEMS

The effects that large scale ionizing radiation such as from a nuclear attack would have on plants and animals living together in an ecosystem have concerned radioecologists ever since the first use of atomic bombs. Several studies have been conducted involving small ecosystems in an attempt to determine what would happen on a large scale. In one study, 10,000 acres of land surrounding a nuclear reactor where exposed to radiation ranging from lethal levels to levels no higher than the natural background radiation. The ecosystem on this land consisted mainly of an oak-hickory-pine climax forest. The forest was exposed to a mixture of gamma radiation and neutrons, with an intensity similar to that expected from fallout after a nuclear attack. The radiation reached about 37,00 trees, plus many more herbaceous species and many more shrubs. Ecologists examined thousands of plants in order to differentiate between the effects of ionizing radiation and the effects of frost, disease, insect damage, drought, and other natural factors. In still another study, a community of spring and summer annuals was exposed to gamma radiation for nearly 4 months during one growing season and then was observed over the next 3 yrs. In still another study, gamma radiation was applied daily throughout one winter and spring to a forest and to an open field with a well-established cover of annual plants.

On the basis of these and other studies, radioecologists have formulated a scenario depicting how ionizing radiation would affect the plants and animals of our forests and fields if a nuclear attack occurred during the summer growing season. People emerging from shelters several weeks after the attack would find little change in their surroundings, except in areas of extremely high radiation. All plants and animals, both large and small, would have been killed in these high radiation areas, and as the plants died they would subject the surrounding areas to further danger from fire. However, most fields and woodlands would appear unchanged by radiation when viewed from a distance. Closer inspection would reveal more clearly the extent of the damage. The ground would be littered with the bodies of birds and animals killed by the radiation. Inspection of lakes, streams, ponds and marshes would show that the lower animals had fared better. Fish, frogs, toads and salamanders would be alive and healthy. The sound of insects would be heard as before. Among the plants the damage would be least serious to those that appear early in a natural succession pattern. Mosses and lichens would be undamaged, annuals would be somewhat affected, shrubs more so, and trees most of all. The damage to pine trees would be most apparent. Pines nearest the radiation zone would have turned a brilliant red brown within a few days after the attack. Other plants in the forest and fields would undergo little change during the remainder of the summer.

In the autumn the oaks, hickories and other hardwood trees would lose their leaves earlier than usual—perhaps as much as 7 weeks earlier in areas nearest the high-radiation zone. The following spring these areas would remain in their state of winter dormancy 7 or 8 weeks longer than usual. Examination of the hardwoods (oaks, hickories, etc.) at this time would reveal severe damage to the buds, resulting in the development of fewer leaves and of abnormal leaves. Near the high-radiation zones, the trees might be leafless. The distribution of annuals in the open fields and on the forest floor would also be changed from previous years. Certain species would grow in greater numbers, partly as a result of the stimulation of their seeds by radiation and partly as a result of the radiation in seed germination among other competing species. The delay in development of leaves on the trees would give these annuals an extra long growing season. In the abundance of sunshine, weeds would grow on the forest to heights of 8 ft or more. The absence of a leafy canopy would also cause changes in the forest soil. With greater wind flow through the bare trees and higher temperatures from direct sunlight, the soil would become drier and harder during sunny weather. In rain storms the harder impact of rain drops would wash away topsoil in areas not covered by weeds or shrubs. Throughout the first summer following the attack, birds and animals from outside the irradiated areas would move in to replace those that were killed.

RADIOACTIVE TRACERS

With the threat of nuclear war receding, and nuclear reactors being equipped with ever more elaborate safe-guards to reduce radioactive releases to the environment, the thrust of radioecology is changing. Activation products in controlled quantities are now being used as radioactive tracers to follow the pathways of chemical elements in the bodies of organisms and in the complex interactions of ecosystems. The radioactive materials have the advantage of being easily detected and quantitatively measured in biological materials without elaborate chemical separation of the elements otherwise necessary.

For example, the radioisotope ^{137}Cs was added to the upward flow of water in trunks of yellow poplar trees in Tennessee about 18 years ago. In the ensuing years radioecologists have followed the movement of this relatively inert tracer into leaves of the trees, into leaf-eating insects, into the insect eating birds, into the forest litter as the dead leaves fell, into soil insects, and so forth. Periodic sampling has confirmed the recycling of natural materials in this forest ecosystem. The radioecologists, in concert with systems analysts, are currently developing computer simulation models to mimic the ecological cycles revealed by this cesium tagging experiment. Comparable information on the exchange of materials from one component of the ecosystem to another could never have been obtained without the knowledgeable use of radioactive materials by these trained radioecologists.

Other activation products such as calcium-45 or phosphorus-32 find use in studies of metabolic processes in organisms, populations or communities. Such studies lead to an understanding of regulatory processes and structural characteristics of living systems. Other examples of experimental use of radioactive tracers may range ecologically, from studying the uptake of ^{45}Ca tagged fertilizer by corn, to following the pathways of ^{32}P in a stream, including its distribution in the non-living as well as the living components.

Radiation ecology is now an area of ecological research and teaching that encompasses far more of the impacts of man on his environment than the atom bomb and nuclear reactors. Understanding of nearly all pollutant chemicals in the environment is being enhanced by use of techniques and principles of ecological cycling developed by radioecologists.

CURRENT DEVELOPMENTS IN RADIATION ECOLOGY

The last several years have witnessed a major decrease of interest in, and hence support of, research in radiation ecology. In the United States the research programs and projects initiated primarily under the Atomic Energy Commission (AEC) have been mostly dismantled. The rationale behind these policy shifts is difficult to comprehend; however, it seems to have been associated with a perception that most of the scientific challenges associated with the ecological aspects of radiation are either sufficiently understood or can contribute little to those practical issues related to radiation protection that are still of concern. Despite this rationale, there has been little change in the long-standing public fear of ionizing radiation and its potential consequences. In addition, the recent major accident at the Russian nuclear power station at Chernobyl (1986), in which 50–100 Mci was released into the environment, not only raised or exacerbated fears in those public sectors already concerned about radiation problems associated with nuclear power but also served to galvanize resistance in large groups (e.g., the Soviet public and other East European populations) that hitherto had either accepted nuclear power or manifested little if any public resistance.

The Chernobyl accident underscored both the ineffectuality of political boundaries against environmental contamination and the role of food chains, both natural and agricultural, in exposing humans and other organisms to potentially harmful levels of radionuclides. Likewise, Chernobyl focused interest on the direct consequences of radiation on ecosystems in the zones of high contamination (within a radius of 18 km of the reactor site). The release of large quantities of ^{134}Cs and ^{137}Cs resulted in the contamination of lakes, streams, and forests in the path of the plume. The need to understand the rates of transfer and patterns of bioaccumulation of these radionuclides in different ecological pathways became manifest in many European countries located thousands of kilometers away from the reactor. In Sweden, for example, high concentrations of ^{137}Cs were found in reindeer and moose (1,000 to 10,000 Bq/kg) and in several species of freshwater fish. The relatively rapid buildup of radionuclides in these organisms was the result of processes which can affect both the rate and extent of bioaccumulation in food chains. Thus the Chernobyl accident has emphasized an increased need for additional research in radiation ecology.

Food chains are the ecological pathways by which many substances are moved in terrestrial and freshwater environments. In the case of radionuclides, these pathways are important in the assessment of radiation exposure to critical population subgroups and human populations. Until recently the uptake and transfer coefficients used in regulatory models were mainly generic default values intended for use in lieu of site-specific information. The Chernobyl accident demonstrated the importance of and need for geographic-specific data on individual radionuclide behavior in terrestrial and fresh-water pathways.

Unlike the United States, most other countries are involved in extensive radioecological research. This research is aimed at obtaining data for predicting exposure resulting from transport of radionuclides in agricultural food chains. The processes of interest in terrestrial environments are those involving atmospheric deposition onto soils and vegetation; resuspension and leaching from these surfaces; uptake from soils by the edible portions of vegetation; and transfer into meat, milk, and other animal products utilized by humans. In the aquatic environment the key processes involve the bioaccumulation of radionuclides from sediments, water, and algae into the edible components of aquatic biota.

The assessment of the environmental and health impacts resulting from radiation exposure is dependent on the use of mathematical models, which, like all other models, are prone to uncertainty. The best method for evaluating uncertainties in the predictions of dose-assessment models is to test predictions against data obtained under real-world conditions. The large extent of contamination following the Chernobyl accident has provided exactly this type of opportunity. Currently an international cooperative effort known as BIOMOVS (BIOspheric MOdel Validation Study) is under way to test models designed for the calculation of environmental transfer and bioaccumulation of radionuclides and other trace substances. More than 20 assessment models are now being tested against data collected from numerous sites throughout the Northern Hemisphere. Upon completion of the initial model testing effort of the BIOMOVS project, additional long-term testing is being planned and organized by the International Atomic Energy Agency (IAEA).

Another issue of concern that has not received research attention recently in the United States in the direct effect of ionizing radiation on populations and communities of organisms. This issue invariably arises whenever there is a nuclear-related incident. In the case of the Chernobyl accident, radiation exposures in the immediate vicinity of the reactor resulted in 28 human fatalities, with a larger number or persons (209) suffering varying degrees of radiation sickness. Pine forests within several kilometers of the reactor site received sufficient contamination to result in an accumulated dose of more than 1000 rads. According to Soviet reports, pronounced morphological damage to pine foliage was visible within 5 months after the accident in the zones where the doses ranged from 300 to 1000 rads. Lethal effects in the 1000 rad zone were also manifest by this time, and by winter (7 months postaccident) 400 ha of forest was destroyed. An ecological preserve has been established in one of the natural areas subjected to high levels of radionuclide contamination. The Soviet government has announced its intention to carry out long-term radioecological observations and studies in this preserve to assess the long-term impacts, if any, on the resident flora and fauna.

To the Soviets' credit, they have recognized both the need and the opportunity to obtain data on the long-term effects of ionizing radiation on plant and animal populations as manifested through genetic mechanisms. To this end, they have established experimental facilities at the accident site to carry out this research. Radioecologists have long recognized the need for hard data on the long-term consequences of exposure to chronic radiation to populations of organisms. Little is known about the interaction of ionizing radiation and environmental stress on populations that are subject to competitive pressures, predation, and other factors that affect survival. We need to be concerned with the effects that a buildup of radionuclides in the environment would have on the eventual fate of the organisms inhabiting such an environment.

Thus, despite the current lack of attention given to research issues in radiation ecology in the United States, much can be learned by collaborating with scientists in Europe and Asia who are now engaged in investigating the fates and effects of radioactive substances deposited from the Chernobyl accident.

REFERENCES

1. Lansdell, Norman, *The Atom and the Energy Revolution,* Philosophical Library, New York, 1958.
2. Curtis, Richard and Elizabeth Hogan, *Perils of the Peaceful Atom,* Doubleday and Co., New York, 1970.
3. Bryerton, Gene, *Nuclear Dilemma,* Ballantine, New York, 1970.
4. Ravelle, Roger *et al.,* The ocean. *Scientific American,* September 1969.
5. Russell, R. Scott *et al., Radioactivity and Human Diet.* Pergamon Press, London, 1966.
6. Auerbach, Stanley I. A Perspective on Radioecological Research, *J. Soc. Radiol. Prot.* **4**(3): 100–105, 1984.
7. Izrael, Yu, A. *et al.,* Ecological Consequences of Radioactive Contamination of the Environment in the Chernobyl Emergency Zone. Moscow, 1987.
8. Peterson, R. C., Jr., *et al.,* Assessment of the Impact of the Chernobyl Reactor Accident on the Biota of Swedish Streams and Lakes, *Ambio* **15**(6): 327–334, 1986.
9. BIOMOVS Progress Report No. 6 Swedish National Institute for Radiation Protection, Stockholm, 1988.
10. International Atomic Energy Agency. Coordinated Research Project on the Validation of Terrestrial, Aquatic, and Urban Radionuclide Transfer Models and Acquisition of Data for that Purpose. IAEA, Vienna.

STANLEY I. AUERBACH
Oak Ridge National Laboratory

RADIOACTIVE WASTE MANAGEMENT: see MANAGEMENT OF RADIOACTIVE WASTE

RADON

Following the discovery of very high levels of indoor radon associated with homes in northeastern Pennsylvania in December of 1984, the New Jersey Department of Environmental Protection was designated as the lead state agency to address the potential for radon problems in New Jersey. The Department was charged with characterizing the problem as it might exist in New Jersey and developing a coordinated state response to the problem.

Due to the potentially serious public health implications of exposure to high levels of radon, the characterization of New Jersey's radon problem proceeded simultaneously with the development of a program to educate the public regarding radon and respond to residents' needs for reliable radon testing and radon remediation services.

The current data indicates that New Jersey's indoor radon levels are among the highest known and that as many as 1. 6 million homes in New Jersey may be at risk for radon problems. Approximately one-third of homes studied thus far test at or above the 4 picocuries per liter guidance level recommended by the United States Environmental Protection Agency as a feasible goal for remediation of high concentrations of indoor radon. New Jersey's response to this problem is characterized by an integrated multi-agency approach and the growth of firms providing reliable, regulated radon testing and remediation services in the private sector.

BACKGROUND

Radon-222 is the single gaseous isotope among the fourteen principal decay products composing the 4n + 2 natural decay series beginning with uranium-238 and terminating in lead-206. The principal decay products of this series, their half lives and principal associated emissions are given in Table 1.[1]

The atmospheric concentration of radon varies with location, but is typically on the order of 0.1 picocuries per liter (0.1 pCi/1) over land masses.[2] The average radon concentration indoors in the United States in not well known, but has been estimated[3] to be in the range of 1 to 2 pCi/1.

In December of 1984, previously unheard of levels of radon were found in the home of an engineer, Stanley Watras who worked at the Limerick Nuclear Generating Station and lived in the Boyertown, Pennsylvania area.[4] The engineer had been tripping portal monitor radiation alarms indicating radioactive material on his clothing when leaving work. Investigations conducted by the utility and the Pennsylvania Department of Environmental Resources demonstrated that the radioactive material on the engineer's clothing consisted of the short lived decay products of radon-222 (Po-218 through Po-214 in Table 1) which had originated in his home. The concentrations of radon in the home ranged as high as 2,600 pCi/1. Subsequent investigations of the area around the discovery home demonstrated that elevated levels of indoor radon were common to the area and that the probable source of the radon was the underlying gramitec rock formation known as the Reading Prong.[5]

In May of 1985, an article[6] by Phillip Shabecoff appeared in the New York Times reporting the Watras incident and noting that the Reading Prong ran in a northeasterly direction out of Pennsylvania, through northern New Jersey and into New York. The article also identified an increased risk of lung cancer as a possible outcome of exposure to elevated levels of radon and its decay products. Immediately, state agencies in New jersey, particularly the Department of Environmental Protection, were deluged by telephone and written requests for information on radon and assistance in obtaining radon testing services.

Numerous studies have demonstrated the link between exposure to radon and its decay products in mining situations and an increased risk of lung cancer.[7] Research efforts directed at establishing a link between residential exposure to radon and its decay products have been far more limited in scope than the mining studies and the data derived from them are, at best, equivocal.[8] Studies underway now in Sweden, New Jersey, Pennsylvania and New England may have sufficient power to elucidate the residential exposure risks.

Based primarily on the results of the mining studies he United States Environmental Protection Agency (EPA) has estimated that from 5,000 to 20,000 of the 130,000 lung cancer deaths annually in the United States are attributable to radon.[9] EPA has further estimated that 1 to 5 persons in 100 exposed in their homes to radon at 4 pCi/1 over a 70 year lifetime, with 75% occupancy, will develop lung cancer.[10] As this risk is assumed to be linear, the risk at 200 pCi/1 under similar conditions of exposure is estimated to be from 44 to 77 in 100. These risks are two or more orders of magnitude greater than other environmental health risk levels which traditionally trigger actions to reduce them.

It is important to note that radon itself is not believed to be the major contributor to the possible development of lung cancer.[11] As radon is an inert gas there is not preference for deposition with the body. When radon decays in ambient

TABLE 1

The Uranium Decay Series

Nuclide	Half Life	Alpha Energy (MeV)	Beta Energy (MeV)	Gamma Energy (MeV)
U-238	4.5×10^9 y	4.1–4.2	—	—
Th-234	24 d	—	0.06–0.2	0.09 (6%)
Pa-234 m	1.2 m	—	2.3	—
U-234	2.5×10^5 y	4.7–4.8	—	—
Th-230	8.0×10^4 y	4.6–4.7	—	—
Ra-226	1600 y	4.6–4.8	—	—
Rn-222	3.8 d	5.5		—
Po-218	3.05 m	6.0	—	—
Pb-214	26.8m	—	0.7–1.0	0.08 (17%) 0.24 (7%) 0.30 (9%) 0.35 (37%)
Bi-214	19.7 m	—	0.4–3.3	0.61 (45%) 0.77 (5%) 1.12 (15%) 1.24 (6%) 1.38 (5%) 1.76 (15%) 2.20 (5%)
Po-214	1.5×10^{-4} s	7.7	—	—
Pb-210	22 y	—	0.1	—
Bi-210	5.0 d	—	1.2	—
Po-210	138 d	5.3	—	—
Pb-210	Stable	—	—	—

air, however, the decay products formed are metals which initially associated with aggregates of water molecules in the air. These aggregates in turn adhere to airborne dust particles. When inhaled, the dust particles, if of appropriate size, will preferentially deposit in the tracheobronchial region of the lung. Radioactive decay of the short lived alpha emitting progeny of radon will result in intense local irradiation of the tissue underlying the deposited dust. It has been calculated that the vast majority of the total radiation dose to the lung tissue is due to the alpha decay of Po-218 and Po-214.[12] It is this localized irradiation which is believed to trigger the development of lung cancer.

In order to have a radon problem in a home, three conditions are necessary. First, the immediate parent of radon-222, radium-226, must be present in the underlying soil or rock to yield a radon source term. Second, the home must have penetrations in the basement or the ongrade slab which provide entry points for the gas. Third, the pressure inside the home must be less than the surroundings so as to provide a pressure differential to facilitate the entry of radon into the home. The first condition is geologic in origin and is met in many locations in the United States. The second condition is a typical consequence of standard construction techniques, particularly in homes where a peripheral channel in the basement slab and an associated sump pump are used to control wet basement problems. Other points of entry for radon in a home include unsealed openings around piping, at or below grade, cracks in the walls and floor, and porous concrete. The third condition is met by a combination of factors which include the normal stack effect associated with the house, combusting fuels for heating, and the use of appliances such as dryers which are vented to the outside. These conditions are met in homes in much of the northern portion of New Jersey, and to a lesser extent, in specific locations in southern New Jersey.

THE NEW JERSEY RADON PROGRAM

Shortly after the Watras incident, Pennsylvania Department of Environmental Resources personnel contacted their counterparts in the New Jersey Department of Environmental Protection (DEP) to advise them of the problem. In the beginning months of 1985, DEP staff working with the New Jersey Department of Health began to assess New Jersey's potential for a radon problem and develop a response to it.

A review of available data demonstrated that New Jersey had significant potential for a random problem and that it was likely that the problem would not be confined solely to New Jersey's section of the Reading Prong. The National Uranium Resource Evaluation data[13] indicated that significantly elevated levels of uranium, and hence radon, could exist both north and south of the Reading Prong. Outside the Reading Prong elevated levels of uranium were associated with limestone, shales and clays. Additionally, a literature survey[14] previously conducted by the New Jersey Geological Survey had found reports of more than 100 locations of radioactive minerals in New Jersey, many well outside the Reading Prong.

Early State planning activities focused on the need to educate the public with regard to radon and develop legislation providing specific programmatic mandates and associated funding.

In January of 1986, Governor Kean signed a bill providing $3.2 million to begin New Jersey's radon program. A second bill, signed into law in August of 1986, provided an additional $1 million, mandated the development of certification for testing and remediation firms and also guaranteed the confidentiality of random test results. The major components of the radon program developed using this funding include:

Radon Information Program

By December 1, 1987 the DEP's toll free information line had received and responded to more than 52,000 telephone inquiries. As a result of these inquiries more than 20,000 informational packages have been mailed out to New Jersey residents. As the knowledge of radon has grown among New Jersey residents so has the sophistication and complexity of the telephone inquires. Currently, a great deal of staff time is spent providing telephone consultation to residents who wish to discuss the interpretation of their radon test results or who wish an in depth discussion of the relative effectiveness of alternative remediation methods. The Department of Health also provides telephone consultation on specific health issues related to radon. The Department of Environmental Protection and the Department of Health provide speakers for groups interested in learning about radon. By December 1, 1987, more than 250 presentations have been made to interested groups. Also, a slide presentation on radon was developed for use with general audiences throughout the state. This slide presentation is currently being expanded and re-packaged in a video tape format for distribution to local libraries.

Statewide Study of Radon

In 1986 the DEP contracted with the environmental engineering firm of Camp, Dresser and McKee, Inc. (CDM) to perform a $1.3 million statewide study of radon. The specific tasks within this study are to:

a. Map areas of potential radon problems using all available geographic, geologic, radiometric and demographic data.
b. Develop a model to predict indoor radon exposure statewide.
c. Test the model developed in Task b using a sampling program including approximately 6,000 structures statewide.
d. Perform a risk assessment for New Jersey residents of the dangers posed by exposure to radon gas.
e. Develop a computerized database for data collected in the study.
f. Propose and prioritize further research suggested by the statewide study.
g. Provide a summary report containing the most salient findings of the study for presentation to the legislature and the public.

Epidemiologic Study

The New Jersey Department of Health is conducting a case-control study of female lung cancer victims and controls in New Jersey. The primary aim of this study is to determine whether or not the lung cancer risk estimates derived from studies of uranium and other miners are applicable to individuals exposed to radon in their homes. The DEP has supported this effort by performing radon tests in more than 800 homes of cases and controls.

Confirmatory Monitoring

As the radon program was being planned the spring of 1985, it was recognized that as many as 1.6 million New Jersey homes might be at risk for elevated levels of indoor radon. It was clearly outside the resources of state government to offer free testing to each potentially affected household. Preliminary cost analyses of no-charge testing for 1.6 million homes indicated a cost ranging from $40 million to $200 million. A conscious decision was therefore made to allow, and even assist in, the development of a radon testing industry in the private sector. It was also recognized however, that the unregulated growth of such a new industry carried with it a significant potential for error and for consumer fraud. Therefore, to protect New Jersey citizens, it was decided to offer confirmatory testing to those residents who had commercial radon tests performed in their homes and obtained radon concentrations equal to or greater than the 4 pCi/1 guidance level at which the EPA recommends consideration of remedial action.

Certification of Radon Testing and Mitigation Firms

The DEP is in the process of promulgating regulations for the mandatory certification of radon testing and mitigation firms. In the interim, a voluntary certification program has been implemented. This program currently enrolls 58 radon testing firms and 24 radon mitigation firms. Respectively, these total represent about one-half the number of firms applying for enrollment in the two programs. The firms are required to document specific expertise, refrain from scare-tactic advertising, participate in the EPA Radon Proficiency

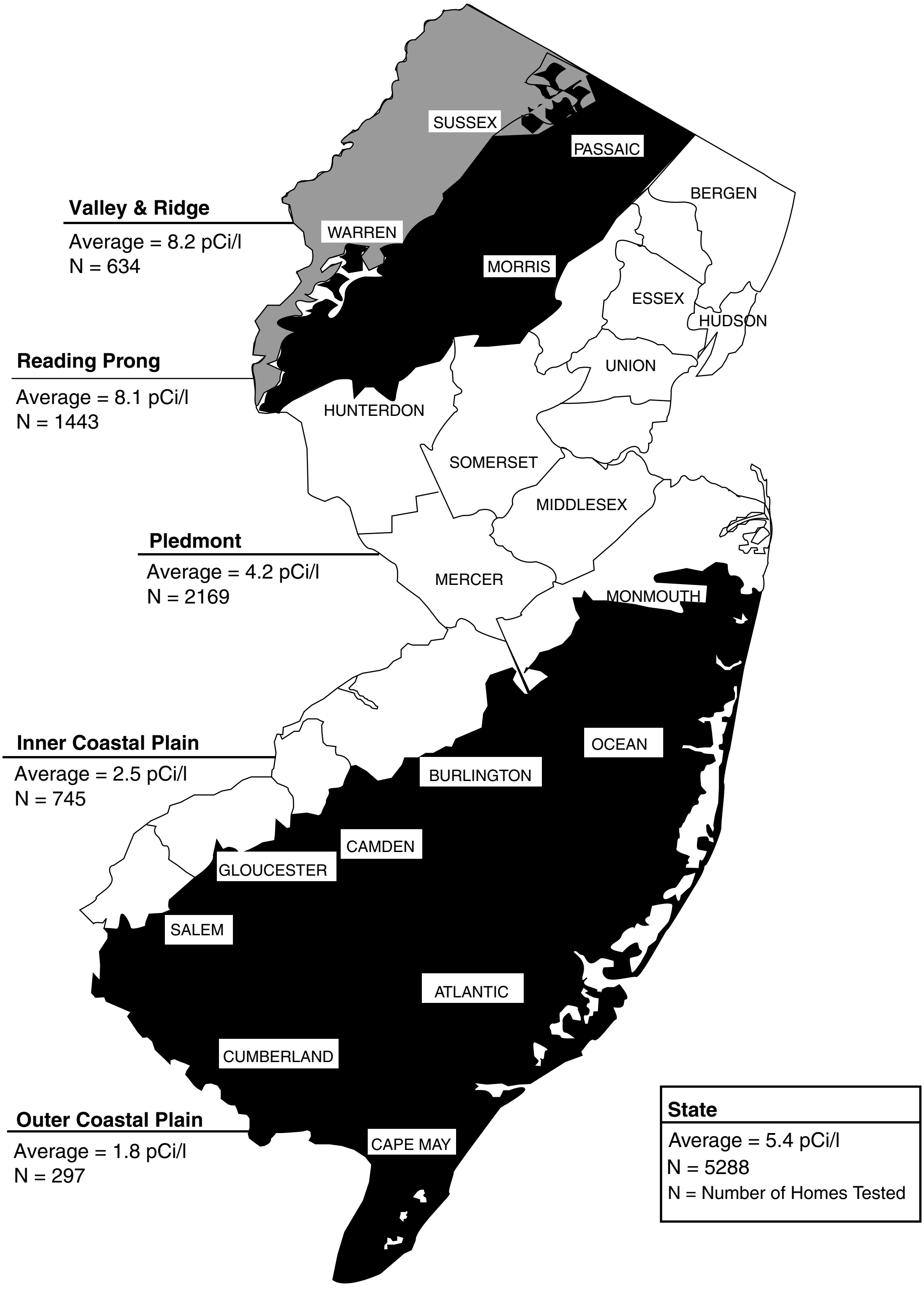


FIGURE 1 New Jersey Department of Environmental Protection, preliminary report on statewide radon study.

TABLE 2
Initial results statewide scientific study of radon, home radon testing results by county

County	Average Radon concentration (pCi/1)	Median radon concentration (pCi/1)	Number of samples	Range (pCi/1)
Atlantic	0.75	0.60	23	0.4–3
Bergen	1.85	1.14	234	0.3–43
Burlington	1.80	1.20	245	0.3–28
Camden	2.27	1.55	66	0.3–9
Cape May	1.43	0.50	18	0.4–8
Cumberland	1.79	1.10	29	03–12
Essex	1.23	0.90	116	0.4–5
Gloucester	3.12	1.10	27	0.4–34
Hudson	2.47	1.00	29	0.4–34
Hunterdon	6.88	3.40	732	0.1–112
Mercer	4.46	1.80	371	0.1–83
Middlesex	2.16	1.10	327	0.2–28
Monmouth	2.68	1.30	305	0.3–38
Morris	5.13	2.30	625	0.3–121
Ocean	0.95	0.70	35	0.3–2
Passaic	3.67	1.70	223	0.3–35
Salem	2.49	1.40	78	0.4–36
Somerset	5.20	2.30	469	0.3–91
Sussex	6.47	3.90	620	0.2–127
Union	2.32	1.30	183	0.3–22
Warren	11.83	6.20	603	0.4–246
Statewide	5.4	2.2	5288	0.1–246

Measurement Program[15] and meet other quality assurance objectives.

Follow-Up Testing

As residents tested for radon and remediated their radon problems, they requested access to reliable, unbiased radon testing which would enable them to determine whether or not the remediation methods employed were effective. In response to these requests, DEP began a program of testing remediated homes at no charge to the owner. As of December 1, 1987, a total of 774 homeowners had availed themselves of this service. In general, these follow up tests document the effectiveness of the remedial actions employed. When they do not, DEP staff advise both the homeowner and the contractor and then assist in working out a mutually acceptable solution.

Radon Cluster Identification

In March of 1986, through the Confirmatory Monitoring program, the DEP became aware of several homes in Clinton, New Jersey which had unusually high concentrations of indoor radon. The DEP and Department of Health worked closely with municipal and county officials to identify other homes in the area which also had very high radon levels. In one neighborhood in Clinton, 105 homes were found to be above 4 pCi/1. Forty of the 105 homes were found to be above 200 pCi/1 and five of the forty homes were found to be above 1000 pCi/1.

Through the cooperation of government at all levels, what could have been a disaster for the community became a solvable problem. The DEP, the Department of Health and local officials provided information on radon and access to testing for those judged to be at significant risk. The municipal government participated fully in all the decision making with regard to the State activities in Clinton and offered access to random testing at reduced cost to all residents. The EPA was invited to extend their radon remediation research program begun in Boyertown, Pennsylvania into New Jersey. EPA remediated ten of the highest radon level homes at no cost to the occupants and provided detailed specifications for the remediation of twenty more homes. The successful remedial techniques developed in Clinton have been adapted to hundreds of other homes in northern New Jersey.

Due to the experience of finding a clustering of high radon level homes in Clinton, the DEP and Department of Health

TABLE 3
Summary of initial results for radon testing in non-residential buildings

Province	Type of building	Number of buildings	Average radon concentration (PCI/L)	Range (PCI/L)
Valley & Ridge	School	8	1.5	1.4–2.9
	Hospital	1	0.7	0.7
	Municipal	10	2.4	0.7–5.8
Highlands	School	12	2.3	0.3–6.9
	Hospital	5	1.6	0.4–4.8
	Municipal	10	4.4	0.3–36.2
Piedmont	School	18	1.7	0.3–4.3
	Hospital	20	1.4	0.3–5.7
	Municipal	21	11.4*	0.4–192
Inner Coastal Plain	School	14	1.6	0.3–4.7
	Hospital	7	0.6	0.4–0.9
	Municipal	10	1.9	0.3–6.1
Outer Coastal Plain	School	11	1.4	0.3–5.6
	Hospital	8	0.7	0.7–1.1
	Municipal	10	0.6	0.3–1.1
Totals	School	63	1.7	0.3–6.9
	Hospital	41	1.1	0.3–5.7
	Municipal	61	5.5 **	0.3–192
	All	165	2.9	0.3–192

* Becomes 2.4 if the single 192 pCi/1 value is excluded.

** Becomes 2.3 if the single 192 pCi/1 value is excluded.

developed a program of responding to reports of high radon level homes (i.e., homes which are confirmed to have radon concentrations at or above 200 pCi/1). A report of such a home triggers the Radon Cluster Identification Protocol under which the local governing body is advised of the confirmation of such a home in their community and their cooperation in sponsoring a public information meeting solicited. Following the public meeting, DEP staff test for radon in homes in the area around the "discovery" home. If this survey yields one or more additional homes with radon concentrations at or above 200 pCi/1, the survey is repeated in the area around these homes. This process is repeated until all homes in the area with radon levels at or above 200 pCi/1 have been identified. The specific concern with homes having this level of exposure to radon is that the lung cancer risk associated with a lifetime of such exposure at 75% occupancy is estimated by EPA to be in the range of 44 to 77 in 100. EPA has recommended that individuals finding such levels in their homes remediate them within several weeks or temporarily relocate until such remediation can be accomplished.[16]

Radon Remediation Research

In addition to supporting EPA's remediation research in Clinton, DEP is also cooperating with the EPA on its House Evaluation Program under which fifteen New Jersey homes, predominantly in the Piedmont region, are undergoing in-depth evaluation of their radon problems at no charge to the homeowner. This program is expected to yield important information on how specific house characteristics and weather patterns influence the radon levels in homes. Additionally, the DEP has contributed $185,000 in funding to a radon remediation research program being undertaken cooperatively by EPA, DEP, Lawrence Berkeley Laboratory of the University of California, Oak Ridge Associated Universities and Princeton University. The primary aim of this research is to fine tune existing remediation methods so as to make them easier to implement and more cost effective.

Radon Database Development

Radon testing firms participating in the voluntary certification program submit radon testing data they have collected summarized by municipality name and zip code. This data is entered into a computerized database by DEP and used as an aid to decision making.

Radon in Non-Residential Structures

The DEP issued interim guidance for the testing of non-residential structures such as schools, hospitals, government buildings and businesses in 1986. The Department of Health issued health related guidance for such structures in November of 1987.

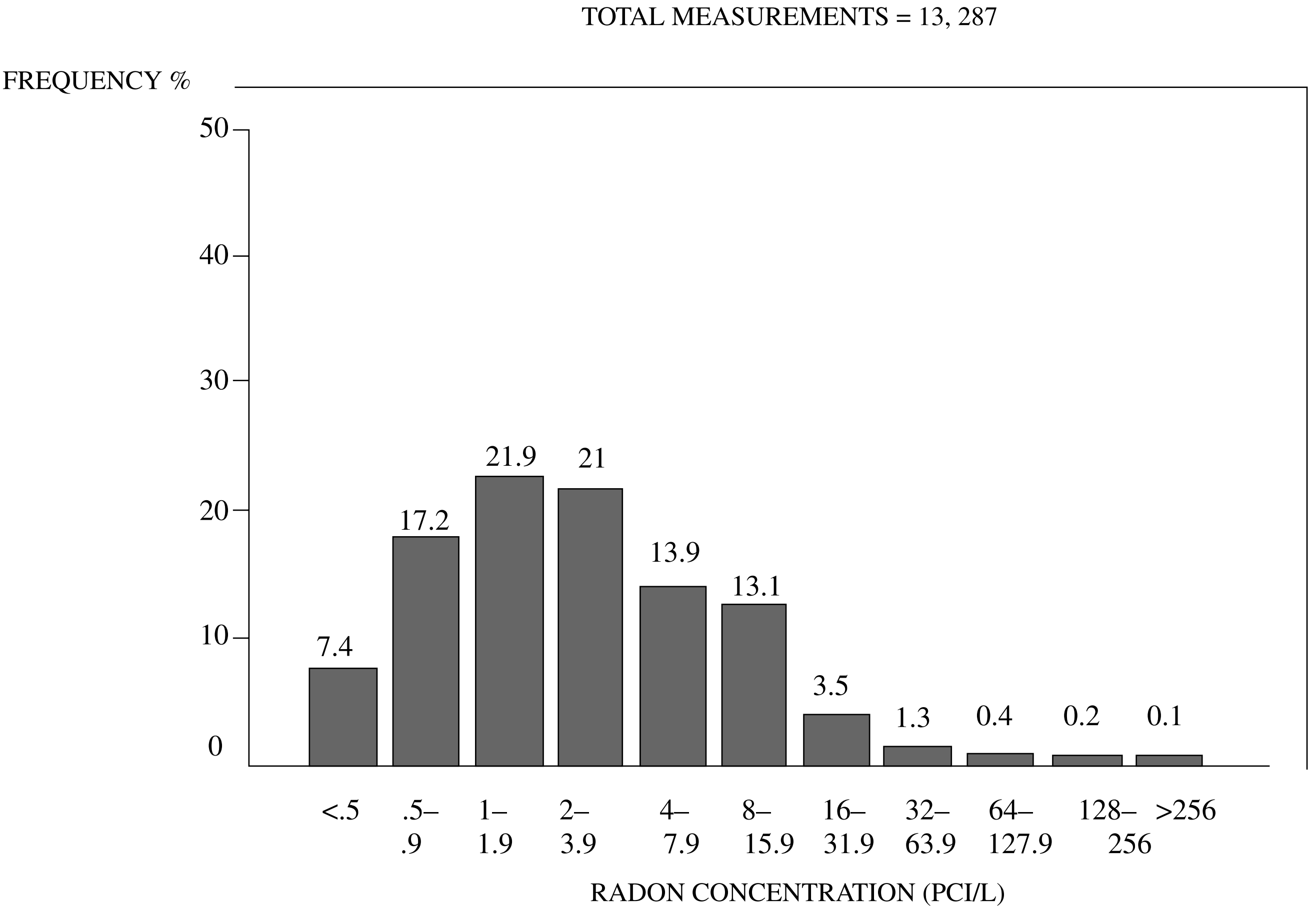


FIGURE 2 Commercial firm radon test data distribution, basement measurements.

Radon Registry

In order to maintain an open channel of communication with individuals who have already been exposed to elevated levels of radon, the Department of Health developed a Radon Registry for New Jersey residents. Participants in the survey are advised of the most current information regarding radon and lung cancer via a newsletter. Over 200 New Jersey residents are participating in the registry.

Low Interest Loans

In comparison to the equity value of most New Jersey homes, the cost of remediating a significant radon problem (i.e., initial test results in excess of 20 pCi/1) is relatively small, currently averaging about $1,200. There are, however, some New Jersey resident would find a single expenditure of this amount quite burdensome. In response to this potential problem, the New Jersey Housing and Mortgage Finance Agency developed a low interest loan program for affected residents. There are no income requirements associated with the program and the rate is at or below the best available loan rates, currently 7.875%. For residents with severe radon problems (e.g., radon levels at or above 200 pCi/1) an expedited loan application processing procedure is available within three working days.

Additional Research

External to the funding provided in the radon legislation, the DEP's office of Science and Research and the New Jersey Geological Survey are conducting research to determine radon in water levels in selected northern New Jersey aquifers.

ASSESSMENT

The various state radon efforts provide a significant body of data from which an assessment of New Jersey's radon problem can be drawn.

Preliminary radon concentration data from homes tested as part of the statewide study are listed by country in Table 2. The same data is organized by physiographic province in Figure 1. These data are almost exclusively for single family homes.

Data on the testing on non-residential structures obtained as part of the statewide study is summarized in Table 3.

All the radon test data presented here were obtained using charcoal canisters[17] exposed for a period of four days under the EPA recommended protocols.[18] All testing was conducted during the period from November 15, 1986 through March 15, 1987 when closed house conditions

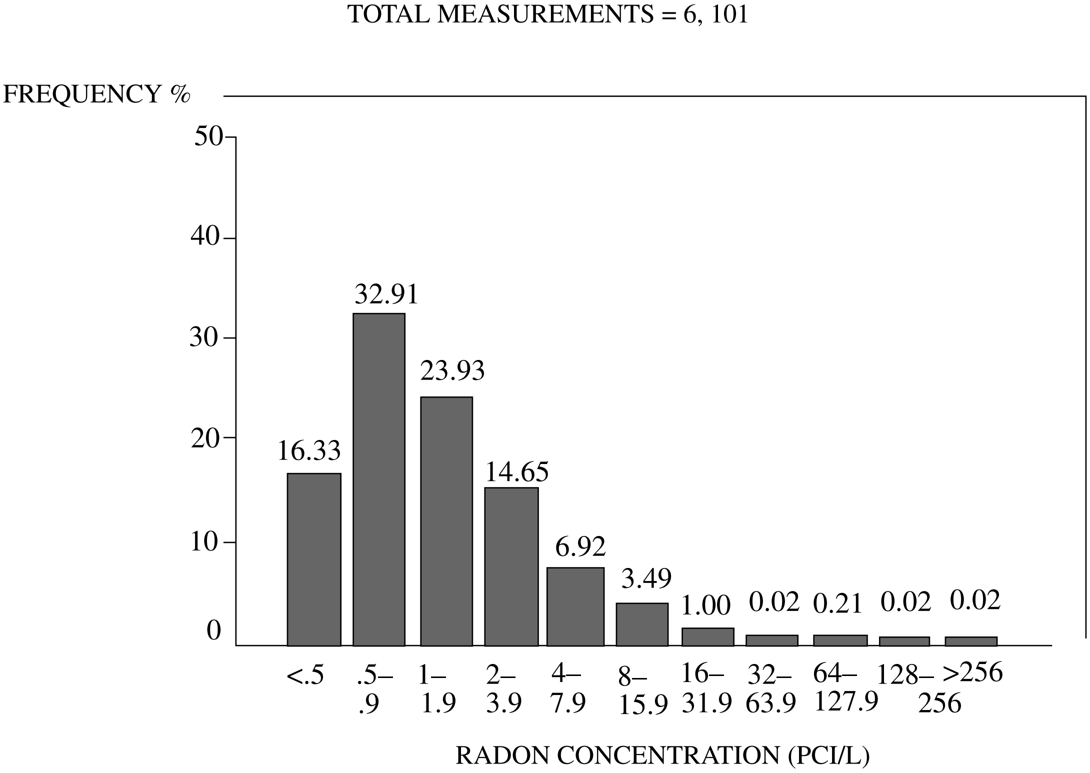


FIGURE 3 Commercial firm radon test distribution, ground level measurements.

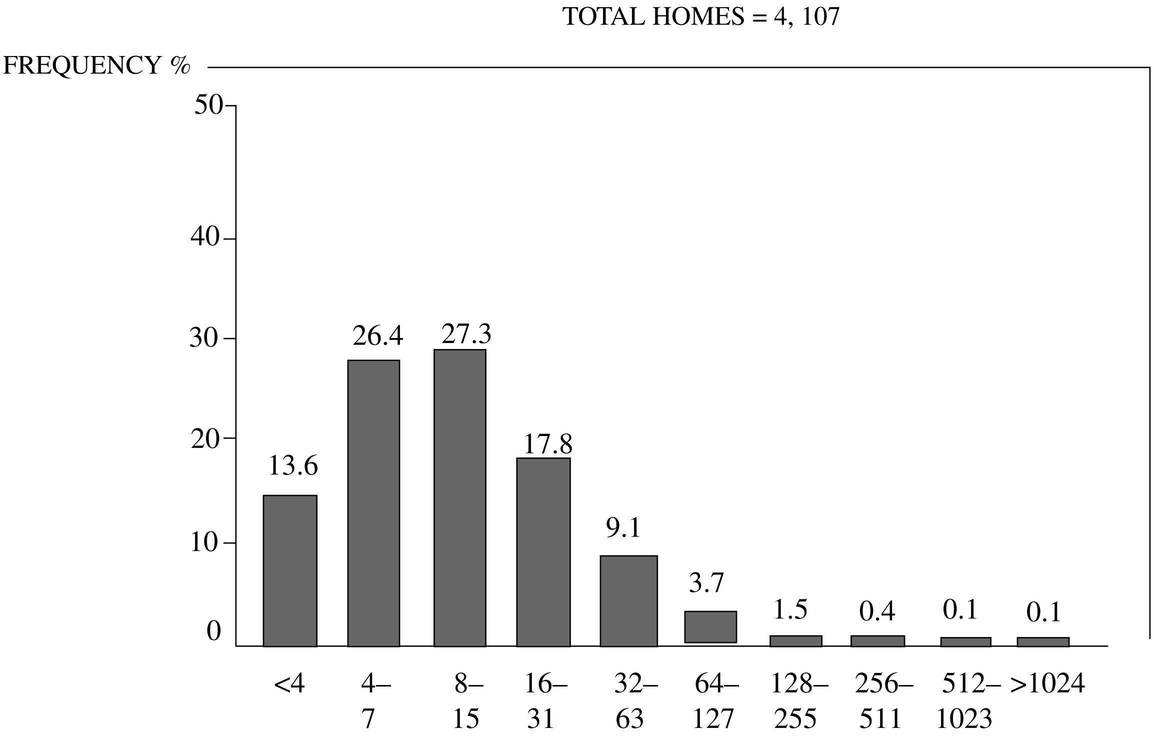


FIGURE 4 Confirmatory monitoring data distribution.

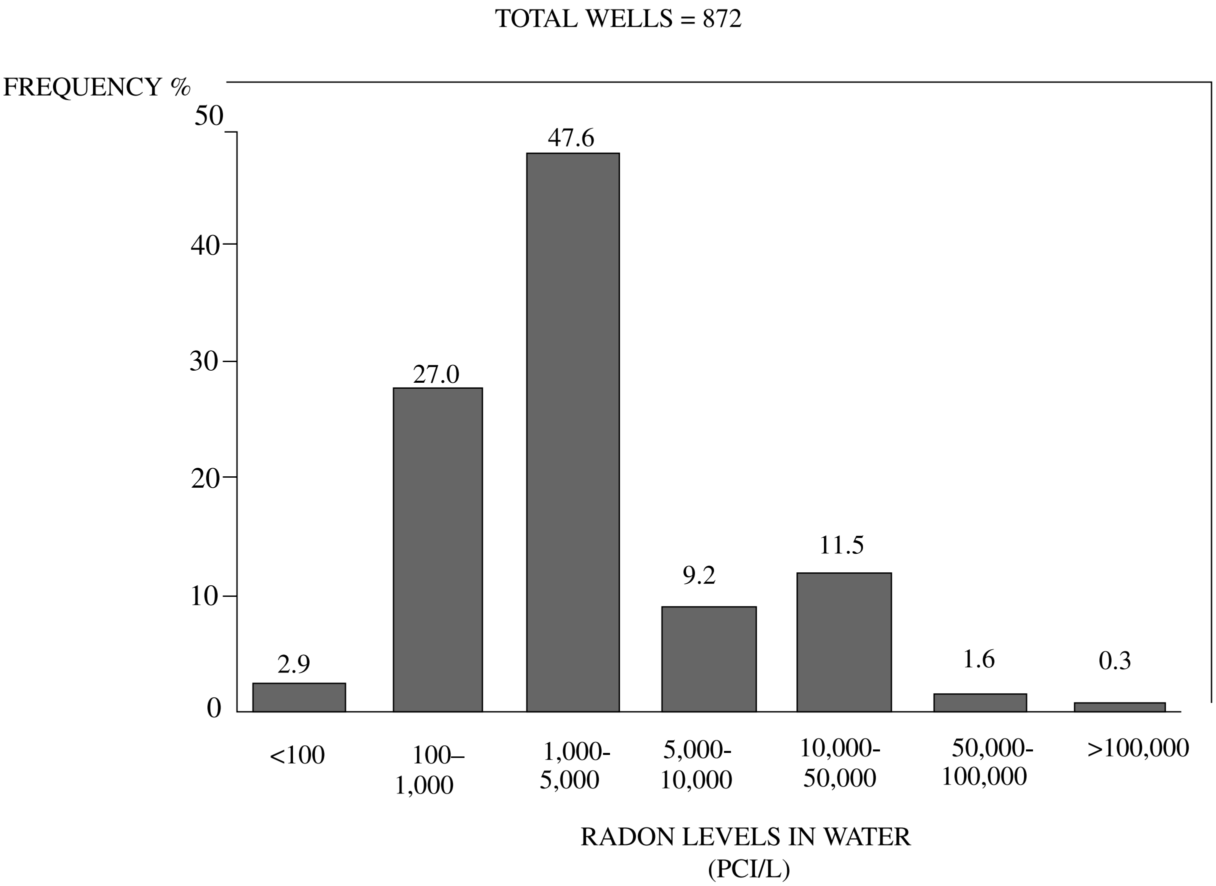


FIGURE 5 Radon in domestic well water obtained in the confirmatory monitoring program.

TABLE 4
Radon cluster identification program data distribution

Location	<4 pCi/1	4–19pCi/1	20–199 pCi/1	200 pCi/1	N
Clinton	4	20	54	41	119
Montgomery	5	6	3	1	15
Ewing 1	8	22	18	1	49
Ewing 2	10	8	10	3	31
Princeton	12	22	12	1	47
Bethlehem	16	26	16	7	65
Bernardsville	6	20	17	4	47
Mansfield	5	10	25	3	43
Hampton	0	5	7	2	14
TOTALS	66	138	162	63	429
% OF TOTALS	15.3	32.3	37.7	14.7	100

generally prevail in New Jersey homes. In accordance with the EPA protocols, the canisters were placed in the level of the home closed to the underlying soil. Unless individuals live at the same level as the tests were made, this testing method tends to overestimate the average concentrations actually breathed by residents by a factor of about two. Most participants in the study lived one floor in the home above the testing level. An exception to the two fold reduction in radon concentration with a one floor increase in distance from the underlying soil occurs in homes with forced air heat. Openings in duct work in the basement make possible the circulation of basement air throughout the home.

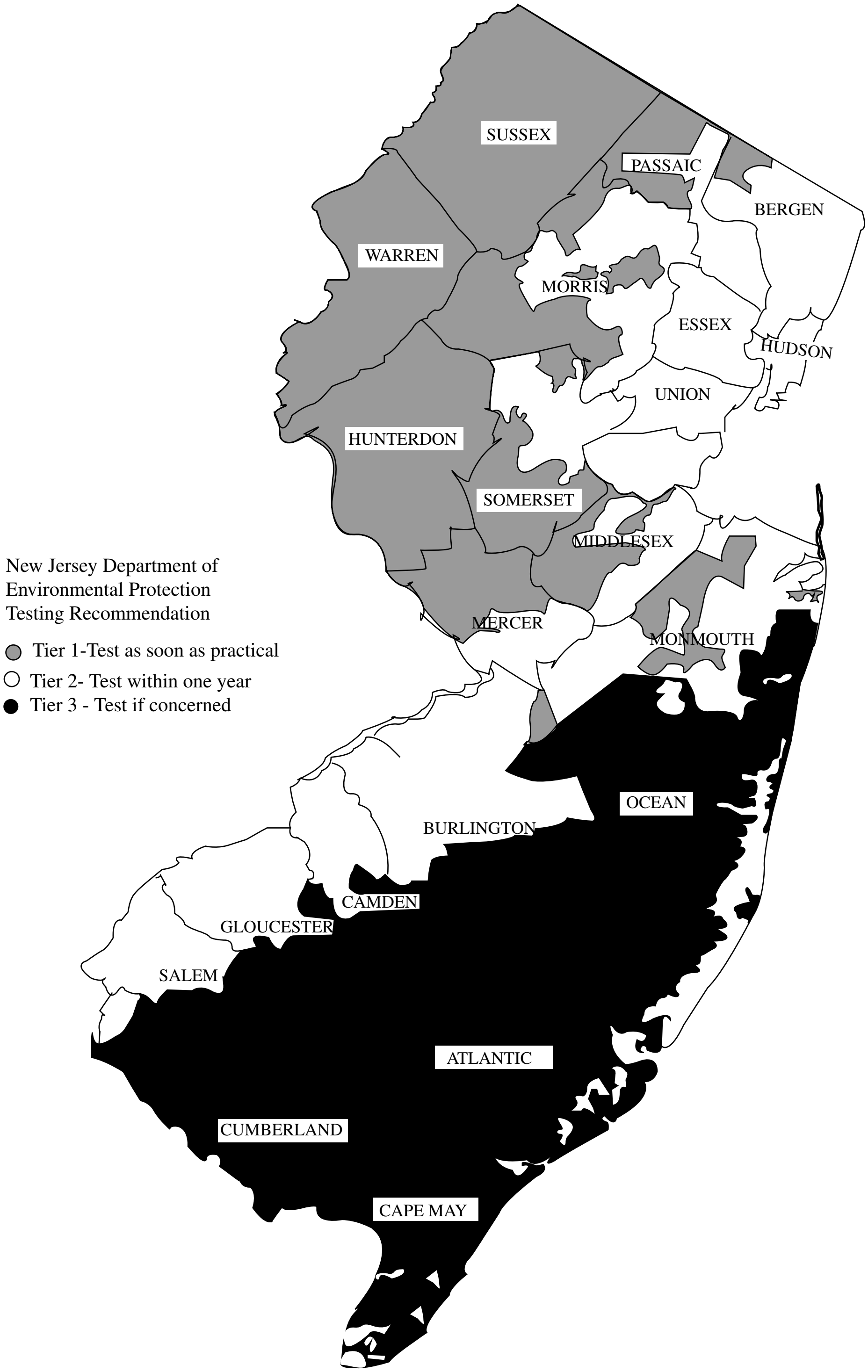


FIGURE 6 New Jersey Department for Environmental Protection, preliminary recommendations for radon testing.

Under these conditions, radon levels on upper floors have been observed to equal or, in a few cases, even surpass basement level radon concentrations.

It should be noted that the statewide study was designed to locate areas of potential radon problems and hence the area sampling density was varied depending on anticipated results. Approximately 20% of the total number of samples were set aside for use later in the sampling period to better define potential problem areas identified in the initial sampling. Hence, the sampling is somewhat biased in terms of problem prone areas, and the values somewhat overestimate the true population based averages.

A total of 35.0% of the homes in the study were found to have radon levels at or above the EPA's 4pCi/1 guidance level. This is in very good agreement with the 32.5% of the homes tested by commercial firms participating in the voluntary certification program. This data, which is approximately log-normally distributed, is shown in Figure 2. Data obtained by commercial firms using primarily charcoal canisters in ground levels of homes are shown in Figure 3. Approximately 11.7% of these results are equal to or exceed 4 pCi/1.

Data collected as part of the confirmatory monitoring program are shown in Figures 4 and 5. Figure 4 shows that, in 86.4% of the homes tested, the DEP measurements corroborated the commercial testing result to the extent that both indicated a radon concentration at 4 pCi/1 or greater. In the 13.6% of the homes where this corroboration did not occur, the difference between the two results is believed to be attributable to natural variations in the radon level, variation in the testing method, and the fact that many residents take action to reduce radon levels immediately upon receiving the initial result.

Residents participating in the confirmatory monitoring program who obtain their water from private wells also have their well water tested for them as part of the program. This data is shown in Figure 5.

The EPA has estimated that a radon in water concentration of 10,000 pCi/1 produces an increase in the radon concentration in the ambient air of approximately 1 pCi/1 in the home where that water is used. Approximately 100 to 1,800 of the 5,000 to 20,000 lung cancer deaths attributed by EPA to radon annually in the United States are believed to arise from exposure to radon entering the home via the domestic ground water supply. Inhalation is the only source of risk as drinking water containing radon is not considered a health threat.[19]

Through the Radon Cluster Identification Program, a number of high level radon homes have been identified. The data derived from the program is presented in Table 4. As indicated, approximately 85% of the homes tested as part of this program have radon levels at or above 4 pCi/1 Approximately 15% of the homes tested had a radon concentration of 200 pCi/1 or greater. Thus, identification of a high level home and subsequent sampling of surrounding homes is a means of identifying radon problem prone areas.

Based on the available data, the DEP made recommendations to residents regarding radon testing in September of 1987. These recommendations are summarized in Figure 6. The three tiers given in the figure reflect the probability for finding an indoor radon problem in a given area of the state based on available testing and geologic evidence.

CONCLUSIONS

Data obtained from the statewide study of radon, through the voluntary certification program for radon testing firms, and from the radon cluster identification program clearly demonstrate that New Jersey has a significant radon problem. Residents do, however, have ready access to testing and remediation services from State certified vendors in the private sector. Further, State programs exist to provide information, confirm test results, identify problem prone areas and support research specific to New Jersey needs.

REFERENCES

1. Adapted from National Council on Radiation Protection and Measurements, *Exposures from the Uranium Series with Emphasis on Radon and Its Daughters,* NCRP Report No. 77, Bethesda, MD 1984, pp. 2 and 9–11.
2. *Ibid.* p. 25.
3. Nero A.V., M.B. Schwekr, W.W. Nazaroff and K.L Revzan, "Distribution of Airborne Rn-222 Concentrations in U.A. Homes" *Science* Vol. 234, November 1986, pp. 992–996.
4. Margaret Reilley, untitled feature article, *Newsletter, Conference of Radiation Control Program Directors,* vol. 19, no. 2, July 1985.
5. *Ioc. cit.*
6. Phillip Shabecoff, "Radiation Gas in Soil Raises Concern in Three State Areas", New York Times, 19 May 1986, p. 1.
7. Thomas D.C., K.G. McNeill and C. Dougherty "Estimates of Lifetime Lung Cancer Risks Resulting from Rn Progeny Exposure" *Health Physics* vol. 45 no. 5, November 1985, pp. 825–846.
8. Judith Klotz "Estimating Lung Cancer Risks of Indoor Radon: Applications for Prevention" *Indoor Radon Proceedings of an APCA International Specialty Conference,* Air Pollution Control Associates, February 1986, pp. 37–38.
9. United States Environmental Protection Agency, *A Citizen's Guide to Radon,* EPA document OPA-86–004, August 1984, p. 1.
10. *Ibid.*, p. 9.
11. National Council on Radiation Protection and Measurements, *Evaluation of Occupational and Environmental Exposure to Radon and Radon Daughters in the United States,* NCRP Report No. 78, Bethesda, MD, 1984, p. 19.
12. *Ibid,* pp. 42–56.
13. Popper G.H.P. and T.S. Martin, "National Uranium Resource Evaluation, Newark Quadrangle, Pennsylvania and New Jersey", Berrdix Field Engineering Corporation Report PGTIF-1243 (83) for the U.S. Department of Energy (1983).
14. Christy Bell, *Radioactive Mineral Occurrences in New Jersey,* New Jersey Geological Survey Open File Report No. 83–5, Trenton, NJ, 1985.
15. United States Environmental Protection Agency, *Implementation Strategy for the Radon/Radon Progeny Measurement Proficiency Evaluation and Quality Assurance Program,* EPA document EPA 520/1–86–03, February 1986.
16. United States Environmental Protection Agency, *A Citizen's Guide to Radon,* EPA document OPA-86–004, August 1984, p. 11.
17. Andreas C. George, "Passive, Integrated Measurement of Indoor Radon Using Activated Carbon", *Health Physics* vol. 46 no. 4, pp. 867–872, 1984.
18. United States Environmental Protection Agency, *Interim Indoor Radon and Radon Decay Protocols,* EPA document EPA 520/1–86–04, April 1986.
19. United States Environmental Protection Agency, *Removal of Radon From Household Water,* EPA document EPA-87–011, September 1987, pp. 1–2.

RICHARD T. DEWLING
Dewling Associates

DONALD A. DEIESO
Research Cottrell Co.

GERALD P. NICHOLLS
New Jersey Department of Environmental Protection

RECYCLING WASTE MATERIALS

Reclamation of waste materials for the purpose of reuse is a commonly accepted definition of recycling. Although water is the substance reused in greatest volume, it will not be covered here. (See Biological Treatment of Wastewater; Municipal Wastewater; Physical and Chemical Treatment of Wastewater; Water Reuse; Water Treatment). Impurities in the air are considered only after removal. Baghouses are an example.

Recycling, or reuse, is based on economics. A used material will be discarded if reclamation costs exceed the cost of new (virgin) material. Disposal costs, however, have increased greatly in the last decade and it is no longer as easy to "throw away" many things. Acceptable disposal sites increasingly are in short supply and transportation costs have become a much more significant part of the expenses of disposal. A typical municipal solid waste can be expected to be somewhat close to the following, by weight:

Paper	35%
Yard wastes	16%
Food wastes	15%
Metals	10%
Glass	10%
Plastics, rubber and leather	7%
Textiles	2%
Miscellaneous	5%

These percentages will vary with the season.

More acute public awareness of environmental pollution has added to the pressure for recycling. Resources are not limitless and many non-renewable materials will be seriously depleted in the not-too-distant future. Raw materials for plastics are petroleum based. Oil has not had a very stable price since the early 1970's.

Landfills have served as repositories for municipal and industrial solid wastes since before the turn of the century. Waste was placed in a dump without regard to separation or segregation of materials. Odors and rodents were common problems. Little planning went into site selection of solid waste disposal areas. As urban areas grew, many dumps at the edge of town were bypassed and were then within the city limits. In many cases, these sites were covered and homes and buildings constructed. These structures all too frequently settled unevenly on the poorly compacted material. At present there is interest in use of decomposition gases from old landfills. Removal of these gases, primarily methane, gives a saleable product and also aids in compaction. Results have been mixed, at best.

Kitchen wastes, organic in nature, appear in unsegregated garbage and trash. Some authorities view these kitchen wastes as suitable for incineration. Older incinerators give rather poor performance, due to high water content in the garbage lowering process temperatures. Problems were common. Improved technology promises higher temperatures and cleaner stack effluents. Ideally, such installations will burn wastes, produce steam and generate electrical energy. The stream and electricity can then be sold and generate sufficient income to pay for the whole operation. It has been proposed that suitably prepared sewage sludge might be added to the combustible mixture. However, the NIMBY (Not in my backyard!) Syndrome complicates the situation. Every group wants to dispose of waste from an upstream vantage point. In the 1990's everyone is downstream from someone else. It is obvious that properly designed and implemented recycling programs can aid in alleviating both technical problems and public perception difficulties.

Ease of separation is fundamental in conversion of waste to a profitable product. Industrial wastes are more easily separated than are municipal solid wastes. Industrial wastes can be collected unmixed and in bulk quantities. Municipal solid waste requires separation techniques, increasing costs to a degree depending on the number of steps.

Reducing the volume of the waste stream reduces ultimate disposal costs by making recyclable materials more readily available and lessening the volume necessary to contain the wastes. About 2000 communities in the US are practicing curbside pick up of recyclables or plan to do so in the near future. The most important consideration in planning such a program is selection of the materials to be included. Decisions must be made as to which materials will be collected, both at startup and later in the program. If all materials are included at the planning stage, the equipment will be adequate, even if lesser numbers are initially processed. Trucks will be of sufficient size and will have enough compartments. Processing equipment will be in place and able to handle all the materials. Recyclable materials are usually segregated by the householder or industrial generator. It is necessary to make economic projections for the waste. Glass requires separation as to color.

In the 1960's plastic (polymeric) containers began to replace glass containers in many uses. Milk and soft drinks had been sold in glass bottles and deposits helped ensure that these containers were returned to the place of purchase. Imposition of deposits effectively kept much of the glass from the waste stream. The plastics have, however, added significantly to the volume of waste and have made waste management more difficult.

A controversial subject is “Beverage container deposit legislation” (BCDL), commonly called bottle bills. It has been found that such bills are effective in controlling litter.

The US Environmental Protection Agency estimates that the weight percentage of plastics in municipal solid waste is about 7–9% but this material occupies 25–30% of the total waste volume.

Many plastic containers today are stamped with symbols as an aid to recycling. These are shown in Figure 1 and are described in Table 1. Only PET and HDPE are commonly collected for recycling.

In 1989 plastics were a part of 800 state and local waste management related bills. This was indicative of the “greening” of consumer buying patterns and more stringent waste disposal regulations. Many major industrial users of plastics are now looking closely at further uses of plastics as a part of “design for disassembly” for ease of separation and reuse. Public attitudes seem to have changed. Once recycled materials in products were regarded as shoddy. Now recyclability is seen as a good selling point. Prices for recycled plastics are now close to those for virgin materials. It is expected that recycled plastics prices will stabilize at 80–90% of those for virgin materials. Major plastics manufacturers have become involved in plastics recycling efforts. In Europe, petrochemical producers have joined to form recycling research units. It is felt that industry must become actively involved in recycling. If it does not, legislation curbing applications and seriously affecting producers may be introduced. In Japan, which incinerates much of its municipal solid waste, industrial plastics producers have also become deeply involved in plastics recycling research. Much work is being done on conversion of plastics to fuel oil.

A major problem in plastics recycling is the low weight-to-volume ratio. It is unprofitable to transport uncrushed containers over long distances. As mentioned earlier, separation at the end user level is most important. Also, different types of plastics are mixed in the waste stream and their similar characteristics make them difficult to separate. Material recovery facilities (MRF), sometimes called municipal recycling facilities, have been established to address the latter problem.

Plastics for reuse are ground or chopped, washed and dried prior to use in extruding machines. Ground plastics may be remelted. This allow additional purification and assists in obtaining a more uniform color. Plastic pellets as the end product can be molded or extruded the same as virgin material. Plastic chips are separated by density differential in liquid flotation processes. Water is commonly used but hydrocarbon liquids have been tried. PET and HDPE are the plastics most commonly collected for recycling from municipal solid waste and inclusion of polyvinyl (PVC) bottles is not desirable. PVC melts and decomposes at reprocessing temperatures. Methods for automatic detection and separation are under development. One company, requiring consistent container color, has sandwiched recycled polyethylene between layers of virgin material. It is generally recognized that supply and regulatory problems must be solved in order to make plastics recycling a viable industry. At present the overall recovery rate for plastics is about 1–2%.

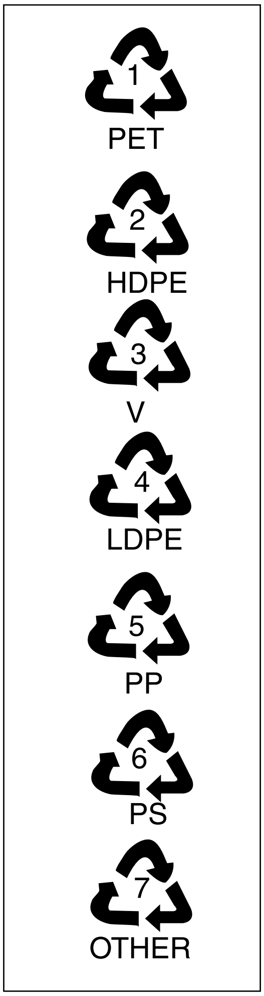


FIGURE 1

The majority of the paper recycling market is newsprint and corrugated paper. However, more valuable grades of paper are found among the 70 commercial grades of wastepaper. Many of these are now incinerated or landfilled.

By the end of the 1990’s about one quarter of the pulp supply in the US came from recycled sources. It is expected that capacity for “secondary” fiber (including recycled and byproduct material) will grow at a rate twice as great as that for virgin (primary) fiber. Chemicals used in making paper products are being reformulated or used differently in order to assist recycling efforts.

The European Economic Community (EEC) averages a recycle rate of about 35% for paper, about the same as that of

TABLE 1

1-PET	Polyethylene Terephthalate. Beverage bottles, frozen food boil-in-the-bag pouches, microwave food trays. About 7% of the plastic waste stream.
2-HDPE	High Density Polyethylene. Milk jugs, trash bags, detergent bottles, bleach bottles, aspirin bottles. About 31% of the waste stream.
3-V	Vinyl. Cooking oil bottles, meat packaging. About 5% of the stream.
4-LDPE	Low Density Polyethylene. Grocery store produce bags, bread bags, food wraps, squeeze bottles. About 33% of the total stream.
5-PP	Polypropylene. Yoghurt containers, shampoo bottles, straws, syrup bottles, margarine tubs. About 9% of the total stream.
6-PS	Polystyrene (Styrofoam^{TM}). Hot beverage cups, fast food clamshell containers, egg cartons, meat trays. About 11% of the total stream.
7-Other	All other plastic resins or mixes of 1–6 in the same product. About 4% of the stream.

Source: Council for Solid Waste Solutions.

the U.S. Japan recycles about 50% of its paper and expects to reach 55% by 1995.

The US EPA estimates that as much as 40% of the municipal solid waste stream is composed of paper and related fiber products. Fluctuations in supply and demand are serious obstacles to paper recycling. This is particularly true in areas which are near pulp producing regions. Once newsprint brought $20 per ton. Now many municipalities must pay for disposal and frequently it is hauled to landfills.

Proposed legislation will mandate that recycled paper make up a portion of the product mix. If enacted, demand will more closely correspond to supply. The U.S., at the beginning of the 90's, is the world's largest exporter of wastepaper.

There is a growing demand for better grades of recycled paper. Washing techniques for ink removal from paper are quite water use intensive and remove only ink particles larger than 30 micrometers. Flotation, on the other hand, removes about 85% of particles smaller than 30 micrometers and uses only half the water. Surfactants added to the bath aid in producing a cleaner pulp. Flotation produces a cleaner end product and reduces the amount of resulting waste sludge for disposal. There is a move towards surfactant producers to formulate products for a recycler's specific needs.

Magazines, in the past, were difficult to recycle. Coated paper in magazines had much more calcium carbonate or clay filler and less wood pulp than did newsprint or other paper grades. It has been found that filler makes it easier to separate ink from fiber in a flotation process. Some recyclers find it advantageous to add "old mixed grade" (OMG) to the new material entering a pulping system. It is estimated that, on the basis of 30% OMG and 70% newsprint, the recoverable supply of OMG could be exceeded by the end of the 90's. One new process shows promise of elimination of the deinking step by application of 400 psi steam at 400°F to newsprint.

Metals have been reclaimed on a regular basis from industrial sources. Now a significant source of scrap metal is municipal solid waste. The recovery rate for aluminum is about 60%. Typically, recycled metal is remelted, along with virgin ore, in furnaces similar to those used in metal production. The focus in metal recycling is on separation techniques and not on new smelting technology.

The Resource Conservation and Recovery Act (RCRA) will be reauthorized in 1991 and the final form of this Act will determine which wastes will be acceptable for reprocessing. One difficult question to be addressed is, "what is true recycling and what is disposal disguised as recycling?" There is no doubt that recycling reduces both the industrial and municipal solid waste streams. However, many Superfund sites were formerly recycling facilities. Of these, some were "sham" recyclers and some were legitimate. The knowledge of future effects of the wastes handled was not adequate when these facilities were put into operation. Some operators did not care.

The reader is directed to *Recycling, Fuel and Resource Recovery; Economic and Environmental Factors* for specific information on various industries.

The future of recycling is bright. Public acceptance of the need for recycling is probably the most important single development of the past two decades. This acceptance has given rise to legislative recognition of the importance of the subject.

REFERENCES

1. Coakley, B.C., Recycling waste materials. The Encyclopedia of Environmental Science and Engineering, Vol. 3, Second Edition. Gordon and Breach Science Publishers, New York, 1983.
2. Decision-makers guide to solid waste management. US EPA, EPA/530-SW-89-072 Washington, D.C. 1989.
3. Documented data from deposit States. Environmental Action. 1986.
4. Lees, B., Can we afford to waste municipal waste? National Society for Clean Waste. Brighton. 1990.
5. *Metals recyclers warily eye sources.* Chemical Engineering. pp. 29–35. Dec. 1990.
6. Paper recyclings new look. Chemical Engineering. pp. 45–48F. March 1991.
7. *Plastics recycling gains momentum.* Chemical Engineering. pp. 37–43. Now. 1990.
8. Recycling, Fuel and Resource Recovery; Economic and Environmental Factors. Encyclopedia Reprint Series. Wiley Interscience. New York, 1984.
9. Smith, S.J. and L.C. Shera, Material selection; Avoiding the pitfalls.
10. Solid Wastes and Power. pp. 36–42. Feb. 1991.
11. U.S. industrial outlook. Department of Commerce. Washington D.C. 1990. Report. *Task force on solid waste management.* National Governors, Association. 1990.

MARK A. TOMPECK
Hatch Mott MacDonald

REMOTE SENSING

Remote sensing is the act of acquiring information about an object from a distance. Environmental applications of remote sensing typically involve the collection of photographic or electronic images of the Earth's surface or atmosphere from airborne or spaceborne platforms. Visual interpretation or computer processing can then be used to analyze these images.

The history of remote sensing dates back to the 19th century, when the first aerial photographs were taken from balloons and kites. The invention of the airplane provided a better platform and aerial photography advanced rapidly after World War I. With the dawn of the Space Age in 1957, the field of photographic remote sensing expanded to include pictures taken from satellites and other space platforms, as demonstrated by pictures taken with a variety of camera types during the *Apollo* missions.

At the same time, non-photographic remote sensing systems such as electronic multispectral scanners (MSS) were developed for use on airborne platforms as well as on meteorological satellites, Earth resources satellites, and other spaceborne platforms. The 1990s have seen a further expansion of the field of remote sensing, with the appearance of new imaging radar satellites, imaging spectrometers, and high-resolution MSS systems, and with the improvement of methods for computer processing of remotely-sensed data.

TYPES OF REMOTE SENSING SYSTEMS

Several types of remote sensing systems can be differentiated based on the principles employed for measuring electromagnetic radiation. The most common types fall into four broad categories: photographic systems, videographic systems, multispectral scanners, and imaging radar systems. Within these categories, particular instruments are designed to operate in specific portions of the electromagnetic spectrum. The Earth's atmosphere scatters and absorbs many wavelengths of electromagnetic radiation, limiting the portion of the spectrum that can be used for remote sensing.

Photographic Systems

Many types of cameras have been used to acquire photographs of the Earth's surface from airplanes and from space. Common formats include 35-mm, 70-mm, and 9 × 9-inch film sizes, although specialized cameras that employ other film sizes are also used. Film types include black and white panchromatic, black and white infrared, color, and color infrared, covering the visible and near-infrared portions of the electromagnetic spectrum from approximately 0.4 to 0.9 μm. (Photography in the ultraviolet range, from 0.3 to 0.4 μm, is also possible but is rarely done due to atmospheric absorption and the need for quartz lenses.) Once the film has been processed, photographs can be electronically scanned at a variety of resolutions for use in a digital environment. Photographic systems provide relatively high-resolution images, with the nominal scale of a vertical aerial photograph being dependent on the focal length of the camera and the flying height of the sensor.[1]

Videographic Systems

Video cameras can be used to record images in analog form on videotape. Video systems have been designed to operate in the visible, near-infrared, and mid-infrared portions of the electromagnetic spectrum. The advantages of video systems include low cost, near-real-time image availability, and the ability to collect and store many image frames in sequence. The primary disadvantage of video is its low spatial resolution, with approximately 240 lines per image for standard video cameras.[2,3]

Multispectral Scanners

MSS systems use electronic detectors to measure electromagnetic radiation in selected bands of the spectrum from approximately 0.3 to 14 μm, including the visible and near-, mid-, and thermal-infrared regions. These individual bands may be fairly wide (greater than 0.2 μm in width) or quite narrow (less than 0.01 μm in width). The designs used for MSS systems fall into two categories. Across-track scanners employ a rotating or oscillating mirror to scan back and forth across the line of flight. Along-track ("push-broom") scanners use a linear array of charge-coupled devices (CCDs) to scan in parallel along the direction of flight. Distinct subcategories of MSS systems include thermal scanners, which measure emitted radiation in the thermal infrared portion of the spectrum, and imaging spectrometers, or "hyperspectral scanners," which generally collect data in over 100 continuous, narrow spectral bands, producing a complete reflectance spectrum for every pixel in the image.[1,4]

Airborne and satellite MSS systems have become widely used in many environmental science and resource management applications. Examples of different types of MSS system include the following:

- The U.S. Landsat satellite series includes two MSS systems. The original Landsat MSS (on board systems launched between 1972 and 1978) includes four spectral bands in the visible and near-infrared portions of the spectrum, with a spatial resolution of 80 m. The Landsat Thematic Mapper (TM) instrument (since 1982) includes six bands in the visible, near-, and mid-infrared regions, with a spatial resolution of 30 m (and a thermal infrared band with a resolution of 120 m). Both instruments operate in an across-track configuration with a swath width of 185 km, and a current orbital repeat cycle of 16 days.
- The HRV instrument on the French SPOT satellite series can collect data in either a single wide "panchromatic" band, with 10m resolution, or in three narrower bands in the visible (green and red) and near-infrared, with 20 m resolution. The orbital repeat cycle is 26 days, but the sensor's ability to be rotated (via ground command) up to 27°— left or right allows more frequent imaging of a given location on the Earth's surface. The HRV is an along-track scanner, with a swath width of 60 to 80 km depending on the viewing angle. Two identical HRVs are included on each SPOT satellite.

Imaging Radar Systems

Whereas the previous types of remote sensing systems operate in the visible and infrared portions of the electromagnetic spectrum, imaging radar systems operate in the microwave portion of the spectrum, with wavelengths from approximately 1 cm to 1 m. At these wavelengths, radar is unaffected by clouds or haze (shorter wavelength systems are used for meteorological remote sensing). In addition, radar systems are active sensors, transmitting their own radiation rather than passively measuring reflected solar or emitted radiation; thus, they can be operated at any time of day or night. Imaging radar systems are sensitive to the geometric structure and dielectric properties of objects, with the primary determinant of an object's dielectric properties being its liquid water content. Current satellite radar systems include the European ERS-series and the Canadian Radarsat, which each have a single 5-cm wavelength band, and the Japanese JERS-1 system with a 23-cm band. Several airborne radar systems have been developed, such as the NASA/JPL AIRSAR, which operates at multiple wavelengths.[5,6,7]

Photographic cameras, video cameras, and multispectral scanners can be operated in a vertical configuration to minimize the geometric distortion of the image, or at an oblique angle to provide a side view of the landscape. Imaging radar systems are not operated vertically, but in a side-looking configuration with a broad range of possible look angles.

ENVIRONMENTAL APPLICATIONS OF REMOTE SENSING

Remote sensing has been used for a wide variety of applications in the environmental sciences. Among the earliest uses of remote sensing was geologic mapping, including the discrimination of rock and mineral types, lineament mapping, and identifying landforms and geologic structures. Today, many types of remotely-sensed data are used for geological applications at a variety of spatial scales, ranging from high-resolution aerial photography, to thermal-scanner images, to lower-resolution Landsat images covering large areas.

Agricultural applications of remote sensing are also common. Aerial photography and other remotely-sensed data are widely used as a base for soil mapping, while multispectral and thermal images are used for soil moisture mapping. Imaging radar systems, with their sensitivity to moisture-related dielectric surface properties, can also be used to measure soil moisture. Multispectral visible and infrared data are used for crop classification and assessment, including monitoring the health and productivity of crops, with the goal of predicting yields and identifying areas of crop damage.

In forestry, aerial photographs are used to delineate timber stands and to estimate tree heights, stocking densities, crown diameters, and other variables relating to timber volume. Color infrared photography and multispectral imagery can be used to map forest types and to identify areas of stress due to pest infestations, air pollution, and other causes. Aerial and satellite imagery can be used to map the effects of wildfires, windthrow, and other phenomena in forested regions. Wildlife habitat can be assessed using remote sensing at a variety of scales. High-resolution aerial photography can also be used to assist in wildlife censuses in non-forested areas such as rangeland.

Many aquatic and hydrological applications make use of remote sensing. Water pollution can be monitored using aerial photography or MSS systems, and imaging radar can be used to detect oil slicks. Thermal imagery is used to study currents and circulation patterns in lakes and oceans. Both optical and radar data are used to monitor flooding, including flooding beneath a forest canopy in the case of radar. Wetlands delineation and characterization can both be assisted by remote sensing. Radar systems are used to measure ocean waves, and both radar and optical images have been used to detect sea and lake ice.

Remote sensing is often used to assist in site selection and infrastructure location, urban and regional planning, and civil engineering applications. Aerial photographs are often acquired with a significant overlap between adjacent photos, allowing heights to be measured using the stereoscopic effect. This process is extensively used for topographic mapping and for creation of geometrically-correct orthorectified photographs to serve as base maps for other applications. Radar interferometry is also being used on an experimental basis for topographic mapping.

IMAGE INTERPRETATION AND ANALYSIS

Many environmental applications of remote sensing rely solely on visual image interpretation. In many cases, visual analysis is improved by stereo viewing of overlapping pairs of images. Increasingly, however, some degree of digital image processing is used to enhance and analyze remote sensing data. Simple image enhancement techniques include data stretches, arithmetic operations such as ratioing and differencing, statistical transformations such as principal components analysis, and image convolution, filtering, and edge detection. More complex image processing techniques include automated land use/land cover classification of images using spectral signatures representing different land cover types.[8]

Most remote sensing applications require the collection of some form of reference data or "ground truth," which is then related to features or patterns in the imagery. For example, pixels in a remotely-sensed hyperspectral image might be compared to a series of mineral spectra acquired from ground samples. Ground measurements of soil moisture, crop productivity, or forest leaf-area index (LAI) could be related to observed reflectance in a satellite image using linear regression. Often, ground truth locations are established using the Global Positioning System (GPS) to facilitate the relation to a georeferenced image.

One significant advantage of digital remotely-sensed imagery, whether collected electronically or as scanned photographs, is the ability to use digital data in a geographic information system (GIS). Once a digital image has been georeferenced, it can be combined with a variety of other types of spatial data. This combination of image and non-image data can be used for a wide range of purposes from simple map updates to complex spatial analysis.[9,10,11,12]

Remote sensing is a rapidly changing field, with more than twenty new satellite systems scheduled for launching in the next decade. Major sources of new data will be high-resolution (approximately 1 m) commercial systems and the various sensors comprising the Earth Observing System (EOS).

REFERENCES

1. Lillesand, T.M. and R.W. Kiefer, 1994. *Remote sensing and image interpretation*, John Wiley and Sons, Inc., New York.
2. Mausel, P.W., J.H. Everitt, D.E. Escobar and D.J. King. 1992. Airborne videography: current status and future perspectives. *Photogrammetric Engineering and Remote Sensing*, vol. 58, no. 8, pp. 1189–1195.
3. Meisner, D.E. 1986. Fundamentals of airborne video remote sensing. *Remote Sensing of Environment*, vol. 19, no. 1, pp. 63–79.
4. Vane, G. and A. Goetz, 1993. Terrestrial imaging spectrometry: current status, future trends. *Remote Sensing of Environment*, vol. 44, no. 2/3, pp. 117–126.
5. Waring, R.H., J. Way, E.R. Hunt, Jr., L. Morrissey, K. Jon Ranson, J.F. Weishampel, R. Oren and S.E. Franklin, 1995. Imaging radar for ecosystem studies. *Bio-Science*, vol. 45, no. 10, pp, 715–723.
6. Way, J. and E.A. Smith, 1991. Synthetic aperture radar systems and their progression to the EOS SAR, *IEEE Transactions on Geosciences and Remote Sensing*, vol.29, no. 6, pp. 962–985.
7. Elachi, C., 1988. *Spaceborne Radar Remote Sensing*, IEEE Press, New York.
8. Jensen, J.R., 1986. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Englewood Cliffs, NJ.
9. Pequet, D.J. and D.F. Marble, Eds. 1990. *Introductory Readings in Geographic Information Systems*, Taylor and Francis, New York.
10. Star, J. and J. Estes. 1990. *Geographic Information Systems*, Prentice-Hall, Englewood Cliffs, NJ.
11. K. Hsu, X. Gao, S. Sooroshian and H.V. Gupta, 1997. Rainfall estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteorl.*, **36,** 1176–1190.
12. Sooroshian, S., S.K. Shu, X. Gao, H. Gupta, B. Imam and D. Braithwaite, 2000. Evaluation of the PERSIANN system satellite-based estimates of tropical rainfall, *Bull. Am. Hydrometeorol. Soc.*, **81,** 2035–2046.

JONATHAN CHIPMAN
University of Wisconsin

S

SEDIMENT TRANSPORT AND EROSION

INTRODUCTION

The literature on the subject of erosion and sediment transport is vast and is treated in the publication of such disciplines as civil engineering, soil science, agriculture, geography and geology. This article provides a brief introduction to the subjects of soil erosion, transport of detritus by streams and the response of a stream channel to changes in its sediment characteristics.

The agriculturalist is concerned with the loss of fertile land through erosion. Sheet, gully and other erosion mechanisms result in the annual movement of about five billion tons of sediment in the United States.[1] By this process, plant nutrients and humus are washed away and conveyed to the streams, reservoirs, and lakes.

The sediment characteristics of a stream also affect its aquatic life. Changes in the character of the sediment load will normally tend to change the balance of aquatic life. Fine sediment, derived from sheet erosion, causes turbidity in the waterways. This turbidity may interfere with photosynthesis and with the feeding habits of certain fish, thus favoring the less susceptible (often less desirable) varieties of fish. The resulting mud deposits may have similar selective results on spawning. The plant nutrients (phosphates and nitrates) that accompany erosion from farmlands may contribute to the eutrophocation of the receiving waters.

Turbidity also makes waters less desirable for municipal and industrial use. Mud deposits may ruin sand beaches for recreational use.

From the engineer's point of view an understanding of sediment transport processes is essential for proper design of most hydraulic works. For example, the construction of dam on a stream is almost always accompanied by, a reservoir siltation or aggradation problem and a degradation problem. A reasonable prediction of the rate of reservoir siltation is necessary in order to establish the probable useful life and thus the economics of a proposed reservoir. The degradation or erosion of the downstream channel and the consequent lowering of the river level may, unless properly accounted for, endanger the dam and other downstream structures (due to under-cutting). After construction of the Hoover Dam the bed of the Colorado River downstream from the dam started to degrade. In 12 years the bed level dropped about 14 feet (Brown[1]).

In addition the downstream channel may change its regime (i.e. its dominant stable geometry). For example, a wide braided channel or delta area may become a much narrower and deeper meandering channel thus affecting the prior uses of the stream. This appears to have happened as a result of the Bennett Dam on the Peace River in British Columbia.[2]

CLASSIFICATION OF STREAMBORNE SEDIMENTS

Terminology

The materials transported by a stream may be grouped under the following type of load:

1) dissolved load,
2) bed load,
3) suspended load.

The dissolved load, although a significant portion of total stream load, is generally not considered in sediment transport processes. According to Leopold, Wolman and Miller,[3] the dissolved load in US streams increases with increasing annual runoff, reaching a maximum of about 125 tons/sq.mile/year for runoffs of 10 inches/year or more.

Bed load consists of granular particles, derived from the stream bed, which are transported by rolling, skipping or sliding near the stream bed. Einstein[4,5] defines the bed load as the sediment discharge within the bed layer which he assumes to have an extent of two sediment grain diameters from the bed.

Suspended sediment load is that part of the sediment load which is transported within the main body flow, i.e. above the bed layer in Einstein's terminology. Turbulent diffusion is the primary mechanisms of maintaining the sediment particles in suspension. The suspended load may be subdivided into:

1) Wash load which consists of fine sediments mainly derived from overland erosion and not found in

significant quantities in alluvial beds; often wash load is arbitrarily taken to be sediments finer than 0.062 mm, i.e. silts and clays.

2) Suspended bed material load which is the portion of the suspended load derived primarily from the channel bed; generally the bed material is assumed to be the sediment coarser than 0.062 mm, i.e. sands and gravels.

Table 1 indicates the terminology used by the American Geophysical Union in describing various sizes of sediment.

Properties of Sediments

An excellent review of the properties of sediments is presented by Brown.[1] He discussed the determination and significance of the following:

a) properties of the individual particle,
b) particle size distribution and
c) bulk properties of sediments.

Properties of the Particle Neglecting interaction effects, the behavior of an individual particle in a stream depends on its size, specific weight, shape, and the hydraulics of the stream.

Two commonly used methods of determining particle size are: (1) mechanical sieve analysis and (2) the fall velocity method. The sieve analysis method differentiates particle size on the basis of whether or not the particle will pass through a certain standard square opening in a sieve or mesh. This method is applicable for sands or coarser particles. Except in the case of spheres, “sieve size” will only be an approximation to the true equivalent diameter of the particle since the results depend to some extent on the particle shape.

The fall velocity method of determining the effective sediment size is gaining popularity in sediment transport research. On the basis of the particle’s terminal velocity, in a specified fluid (water) at a specified temperature, the particle is assigned a fall or sedimentation diameter equal to the diameter of the quartz sphere which has the same terminal velocity in the same fluid at the same temperature.[1] This particle size integrates the effects of grain size, specific weight and shape into a single meaningful parameter for sediment transport studies. Researchers at Colorado State University have developed a Visual Accumulation Tube to aid in the determination of the fall diameter distributions for sediment samples.

Particle Size Distribution On the basis of a sieve analysis of fall diameter analysis, of a sediment sample, a cumulative frequency curve for the particle size can be drawn. Figure 1 shows typical particle size frequency curves for a sample taken from a sandy stream bed and for a sample of suspended load over the same bed.[6] The frequency curves are usually plotted on logarithmic-probability paper.

TABLE 1
Sediment grade scale

Group	Particle size range, mm
Boulders	4096–256
Cobbles	256–64
Gravel	64–2
Sand	2–0.062
Silt	0.62–0.004
Clay	0.004–0.00024

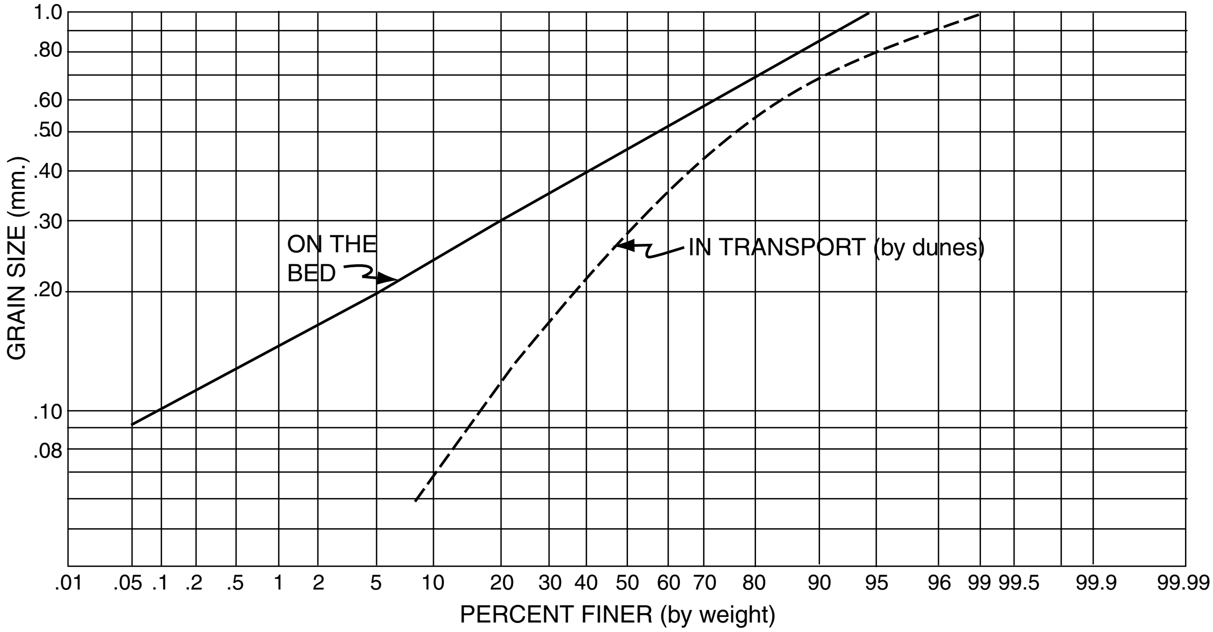


FIGURE 1 Typical particle size frequencies curves for stream sediments (after Bishop *et al.*).

Some important descriptors of the frequency distribution are: (1) the median size or d_{50}, that is, the size for which 50% by weight of the sample has smaller particles; (2) the scatter of particle size as indicated, for example, by the standard deviation or perhaps the geometric deviation; (3) the characteristic grain roughness which has been associated with the d_{65};[5,7,8] (4) the d_{35} has also been used as characteristic sediment size.[9]

Bulk Properties The determination of bulk, in place specific weights of sediments is discussed under *Reservoir Sedimentation.*

EROSION

Most of the sediment in streams is produced by the following processes:[1]

1) Sheet erosion,
2) Gully erosion,
3) Stream channel erosion,
4) Mass movements of soil (e.g. landslides and soil creep),
5) Erosion to construction works,
6) Solids wastes from municipal, industrial, agricultural and mining activities.

Morisawa,[10] using a system diagram, similar to Figure 2, summarizes the inter-relation of climatic and geologic factors that influence soil erosion and runoff. Figure 2 also shows man's influence on the system.

Langbein and Schumm[3,17] proposed the correlation shown in Figure 3 between annual sediment yield and effective annual precipitation for the United States. The effective precipitation is the adjusted precipitation which would have produced the observed runoff for an annual mean temperature of 50°F.

A recent paper by Saxton *et al.*[11] relates total runoff, surface runoff and land use practices to the sediment yield from loessial watersheds in Iowa. This paper compares erosion and surface runoff from contoured-corn watersheds and from pastured-grass and level-terraced areas. In a 6-year study the contoured-corn areas yielded, annually, about 19,000 tons/sq. mile of sheet erosion plus 3000 tons/sq. mile of gully erosion while the level terraced and grassed watersheds yielded about 600 tons/sq. mile. Similarly the surface runoff from the contoured-corn areas was approximately 5 inches compared with 1.5 inches for the level-terraced and grassed areas. The experimental watersheds were of the order of 100 square miles.

Other land use factors are discussed in a paper by Dawdy[12] who presents sediment yields for the state of Maryland. The annual sediment yield from heavily wooded areas is about 15 tons/sq. mile compared with 200 to 900 for crop land. The annual sediment yields from urban development areas (usually only a few acres) varied from about 1000 to 140,000 tons/sq. mile.

Curtis[13] obtained annual sediment yields of 390 and 290 for two watersheds (264 and 651 square miles) in the Miami Conservancy District, Ohio.

A number of empirical formulae have been developed[1,14,15] to permit estimation of rates of overland erosion. The US Department of Agriculture developed the universal soil-loss equation (for upland areas),

$$E = RKLSCP, \tag{1}$$

where E = soil loss/unit area; R = rainfall runoff factor; K = soil erodibility factor; L = slope length factor; S = slope steepness factor; C_1 = crop management factor; and P_1 = erosion control practice factor. Details for estimating the above factors are given by Meyer.[15]

MEASUREMENT OF SEDIMENT DISCHARGE

Samplers have been developed to measure both suspended and bed load in streams. However bed-load samplers are not

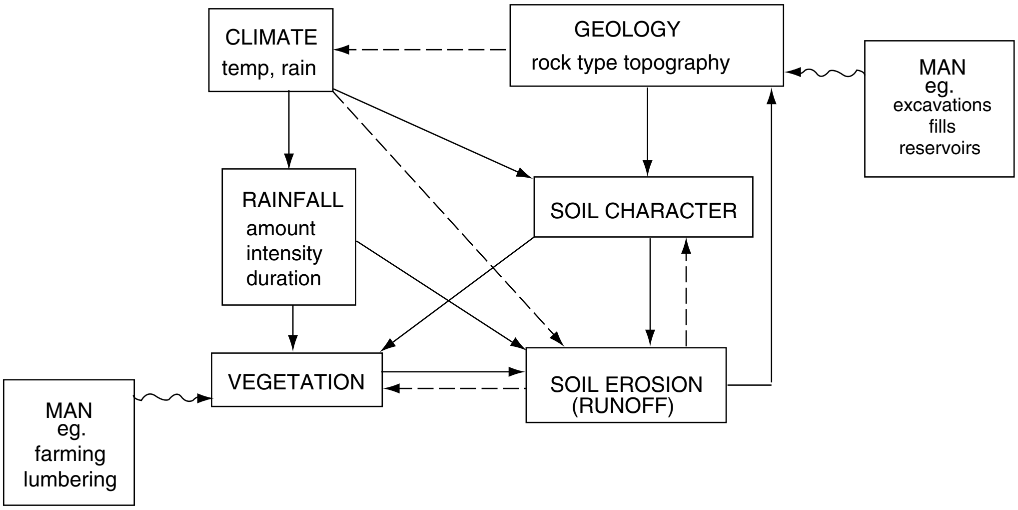


FIGURE 2 The relationship of climate and geology to soil erosion (adapted from Morisawa).

widely used because of their doubtful accuracy. Generally, only suspended load samples are collected in samplers of the type shown schematically in Figure 4. This sampler is designed so that the intake velocity is nearly the same as the local stream velocity. The extent of the suspended sampled zone is limited by the size of the sampler. Methods or extrapolating these measurements and estimating bed load are discussed in the next section.

For more details of sediment measurement techniques and equipment, the reader is referred to Nordin and Richardson,[15]

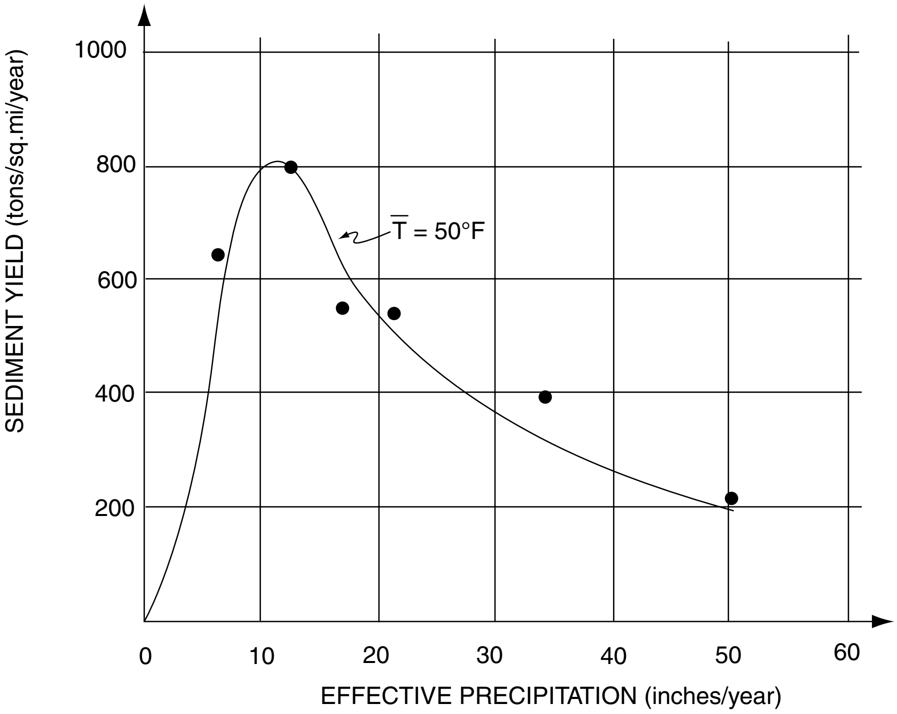


FIGURE 3 Sediment yield in the United States (after Langbein and Schumm).

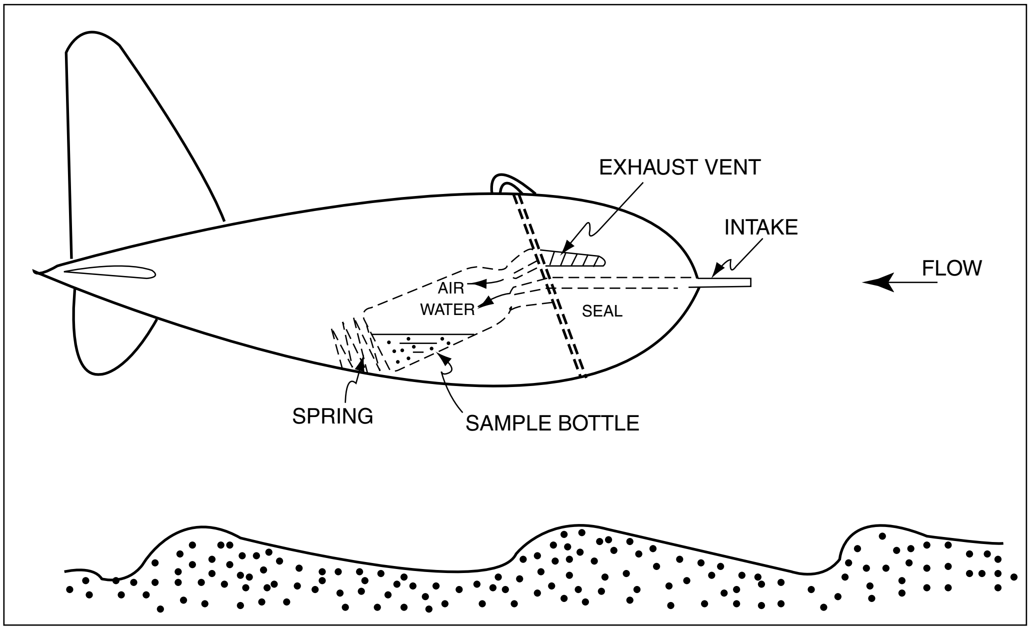


FIGURE 4 Sketch of a suspended load sampler.

Shen,[15] Karaki,[15] Graf,[16] Brown,[1] Simons, and Senturk and the ASCE Sedimentation Engineering Manual.

In some instances[15] turbulence flumes (a concrete lined reach with baffles to create severe turbulence) have been constructed across a stream channel in order to suspend the bed load and thus to sample it by suspended load techniques.

THE MECHANICS OF SEDIMENT TRANSPORT IN A STREAM

General

The nature of sediment transport in a stream depends on the shear intensity of the flow and the type of bed material. The diagram in Figure 5 shows the sequence of bed forms (waves) associated with increasing levels of shear on a fine granular bed material.[3] This figure also shows, schematically, the typical changes in the Darcy friction factor and the sediment concentration with increasing flow velocity. The primary mode of transport of particles, in the case of ripples,[17] is discrete steps along the bed; however with increasing shear more of the bed material becomes suspended until the particle motion is nearly continuous for anti-dunes.

Dunes and ripples are triangular in shape with relative flat upstream slopes and sleep downstream slopes. The water surface waves are out of phase with the dune formation while ripple formations appear to be independent of the free surface.

Dune wave lengths, λ_d, are related to the depth of flow and in general,

$$\lambda_d > 3 \text{ feet} \tag{2}$$

whereas ripple wave lengths λ_r are shorter,

$$2'' \lesssim \lambda_r \lesssim 18'' \tag{3}$$

Dune heights, H_d, are related to the depth of flow, with the limiting height approaching the average flow depth. The ratio of dune length to height is given by[17]

$$8 \lesssim \frac{\lambda}{H_d} \lesssim 15. \tag{4}$$

The maximum ripple height is about 0.1 feet.

Both ripples and dunes progress downstream. The transition from dunes to anti-dunes occurs at a Froude number close to 1.0.

Anti-dunes, as indicated by Figure 5, are in phase with the surface wave. The may be stationary or move upstream. The maximum height of an anti-dune is approximately equal to the flow depth at the trough of the surface wave.

Simons,[17,18] on the basis of experimental data, developed the relationship shown in Figure 6 between stream power and bedform for varying fall diameters. Simons and Richardson[19] also studied the variation of Chézy's C with bed form. Their results are summarized in Table 2.

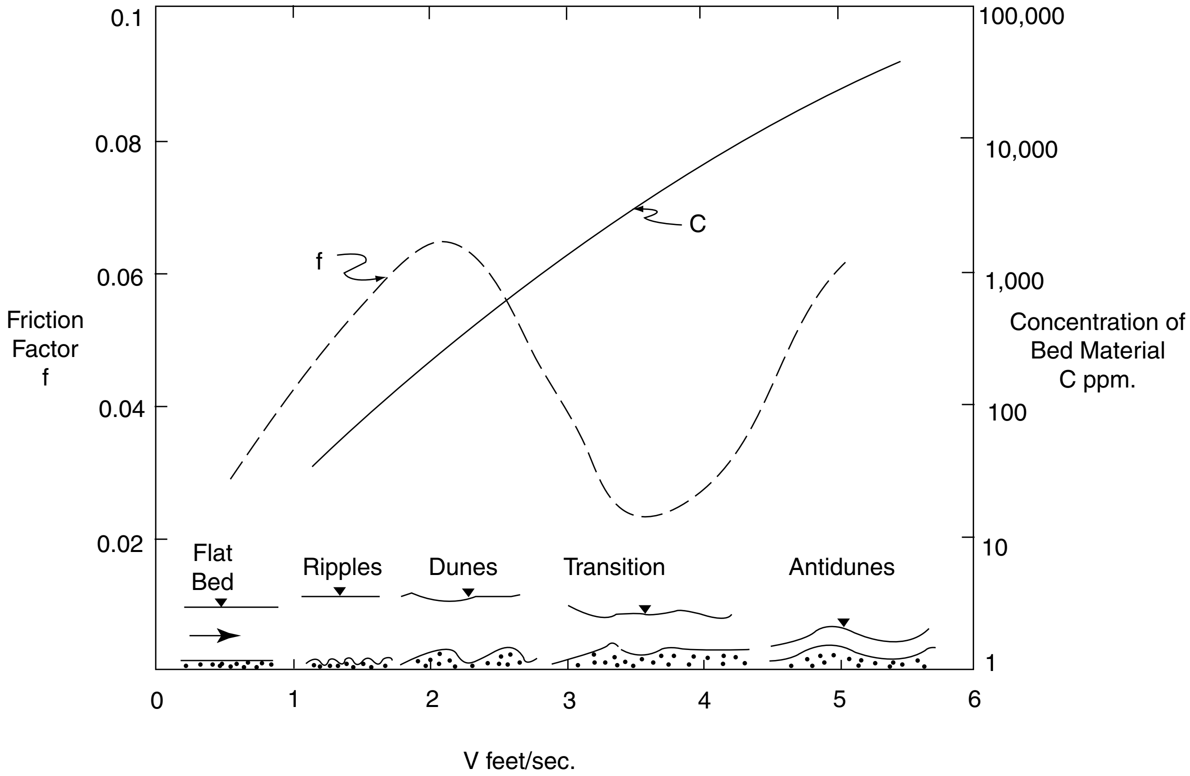


FIGURE 5 The behavior of a mobile stream bed (adapted from Leopold, Wolman, and Miller).

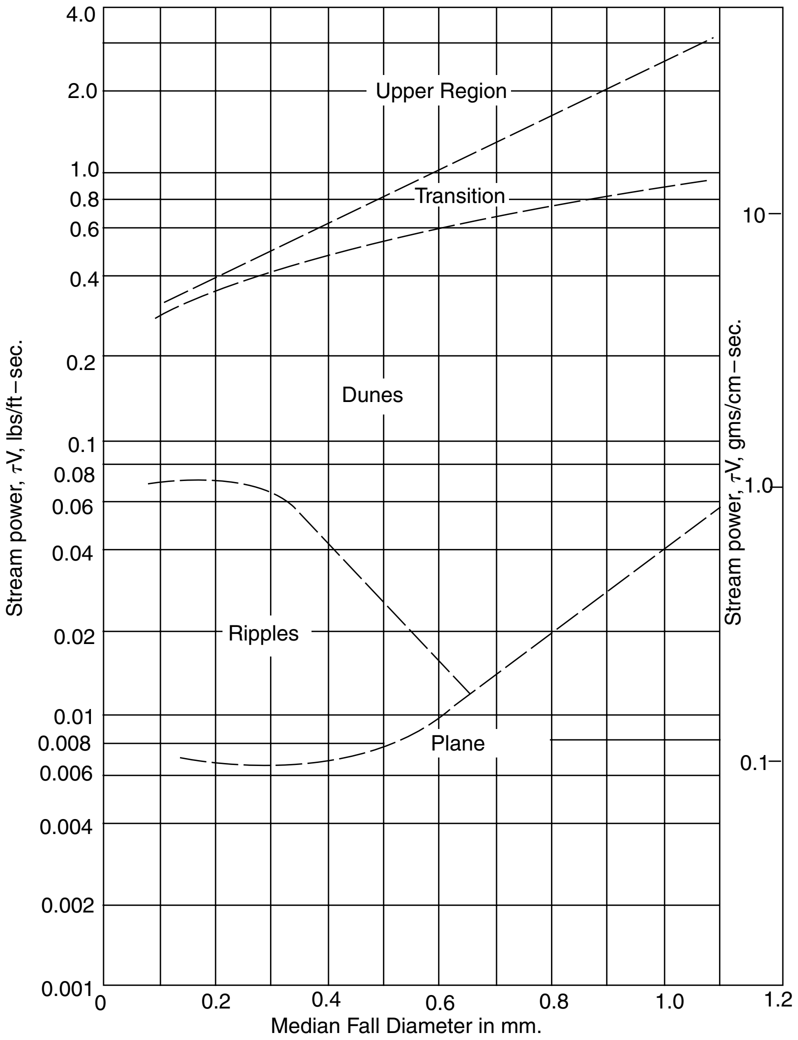


FIGURE 6 Relation of stream power and median fall diameter to bed form (after Simons).

TABLE 2
Chézy *C* in sand channels

Regime	Bed Form	*Cl* √g (where *C* is Chézy *C*)
Lower regime		
	ripples $d_{50} < 0.6$ mm	7.8 to 12.4
	dunes	7.0 to 13.2
	transition	7.0 to 20
Upper regime	plane bed	16.3 to 20
	anti-dune {standing wave	15.1 to 20
	{breaking wavechutes and	10.8 to 16.3
	pools“slug” flow	9.4 to 10.7
		—

Initial Motion

White, in 1940,[20,21] using an analytical approach, showed that, for sufficiently turbulent flow over a granular bed, the critical shear or shear to initiate grain movement is

$$\tau_c = \kappa_c \gamma_f (S_s – 1)d, \tag{5}$$

in which κ_c ; 0.06; γ_f = fluid specific weight; S_s = specific gravity of sediment grain; d = grain diameter.

Shields,[21] using an experimental approach, obtained the more general equation

$$\tau_c = \gamma_f (S_s – 1)\, d f(R_*), \tag{6}$$

in which $R_* = U_* d/\nu$; U_* = friction velocity; ν = kinematic viscosity; and $f(R_*)$ is defined in Figure 7.

Permissible or allowable tractive stresses for use in channel designs with granular or cohesive boundaries are given by Chow.[7]

Bed Load Formulae

When the bed shear, τ_o, due to the flowing stream exceeds the critical shear, τ_c, a part of the bed material starts to move in a layer of the stream near the bed, i.e. the bed layer. Experimental studies[3] indicate that this sediment discharge, known as the bed load, q_B, is a function of the excess of τ_o above τ_c or

$$q_B \propto fcn\,(\tau_o—\tau_c). \tag{7}$$

Figure 8 illustrates a typical experimental relationship between q_B and $(\tau_o—\tau_c)$.

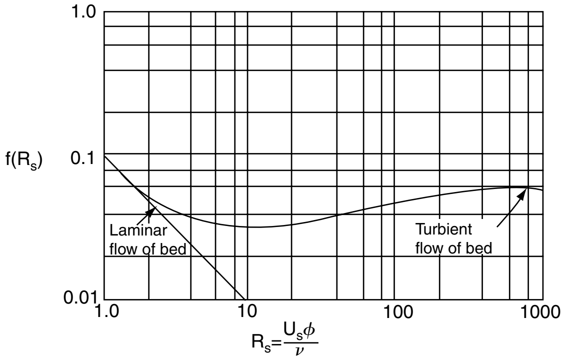


FIGURE 7 Shields' critical shear function (adapted from Henderson).

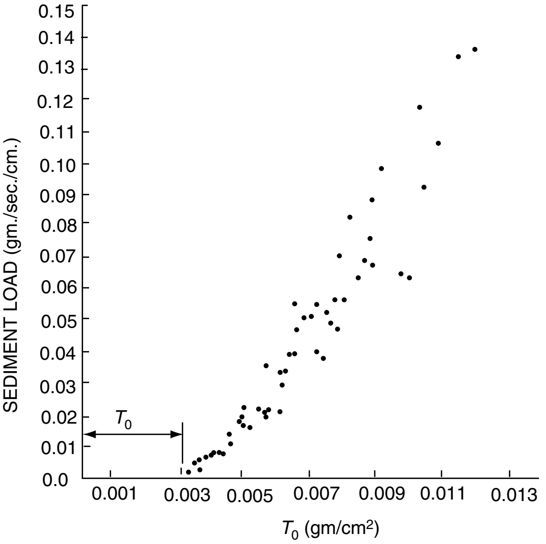


FIGURE 8 Typical relation of shear and sediment load (adapted from Leopold, Wolman and Miller).

DuBoys in 1879[22] treated the bed material, involved in the bed load, as it if consisted of sliding layers which respond to and distribute the applied stress τ_o. He proposed the relation

$$q_B = C_s\tau_o\,(\tau_o–\tau_c) \tag{8}$$

Both C_s and τ_c depend on particle size as indicated in Table 3.

Chang,[1] Schoklitsch,[1,16] MacDougall,[1] and Shields[1] have presented bed load formulae similar to Eq. (8).

The theoretical bed load model developed by Einstein[4,5,8,24] has formed the basis for a number of researches in sediment transport.[6,15,24,36] Einstein utilized: (1) the statistical nature of turbulent flow; (2) the fact that in steady uniform flow there is an equilibrium between the processes of erosion and deposition, that is, (probability of erosion) = (the probability of deposition); (3) the fact that grains near the bed are more in quick "steps" interrupted by "rest" periods; (4) a separate hydraulic radius, R', associated with grain roughness and another hydraulic radius, R'', associated with the bed form.

Einstein obtained the erosion probability function by assuming that the lift force, on a grain, consists of an average component [related to $(U_*)^2$] and a normally distributed random component. Einstein thus obtained the "bed load" equation

$$\frac{A_*\Phi_*}{1+A_*\Phi_*} = \frac{1}{\sqrt{\pi}}\int_{-B_*\psi_*-(1/\eta_o)}^{B_*\psi_*-(1/\eta_o)} \mathrm{d}t \tag{9}$$

in which

$$\Phi_* = \frac{i_B q_B}{i_b\gamma_s\sqrt{gd^3(S_s-1)}} \tag{10}$$

is Einstein's bed transport function; $A_* = 43.5$; $B_* = 0.143$; $\eta_o = 1/2$;

$$\psi_* = \xi Y\left\{\frac{\log\,10.6}{\log\dfrac{10.6X}{\Delta}}\right\}^2 (S_s-1)\frac{d}{S-R'} \tag{11}$$

in the Einstein flow intensity function; i_B = fraction of q_B in the size range associated with d; d = geometric mean of particle

TABLE 3
Typical values of C_s and τ_c (after Straub[22,23])

d mm	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4
$\frac{C_s\text{ft}^6}{16^2-\text{sec}}$	0.81	0.48	0.29	0.17	0.10	0.06
$\tau_c\,\frac{\text{lb}}{\text{ft}^2}$	0.016	0.017	0.022	0.032	0.051	0.09

size range being considered; S_e = energy slope; i_b = fraction of bed sediment in specified range; ξ = hiding factor; Y = lift correction factor; $\Delta = d_{65}/X$; X = correction factor for hydraulically smooth flow; $\delta' = 11.6\, \nu/U_*$;

$$U'_* = \sqrt{gR'S_e} \tag{12}$$

$X = 0.770$ if $\Delta/\delta > 1.80$; $X = 1.39\,\delta$ if $\Delta/\delta < 1.80$.

Schen[15] gives an up-to-date review of the modern stochastic approaches to the bed material transport problem.

The Suspended Load

Equations of Motion of the Fluid The flow in natural streams is almost always turbulent and may be assumed to be incompressible; consequently the applicable equations of motion for the fluid are the Reynolds[25] equations

$$\rho\left[\frac{\partial \overline{U}_i}{\partial t} + \overline{U}_j \frac{\partial \overline{U}_i}{\partial x_j}\right] = \frac{\partial}{\partial x_j}(\sigma_{ji}) + \overline{F}_i, \tag{13}$$

in which $\bar{U}_i$ = ensemble mean point velocity in the direction i; σ_{ij} = stress tensor $=\{-\overline{P}\delta_{ji} + \mu\overline{D}_{ji} - \rho u_i u_j\}$ = average pressure; δ_{ji} = Kronecker delta; μ = dynamic viscosity; $\bar{D}_{ji}$ = deformation tensor; $-\rho u_i u_\xi$ = turbulence or Reynolds Stresses; ρ = fluid density ; u_i = random component of velocity in the i direction; $\overline{F}$ = body force in the i direction. The first term in the stress tensor represents the normal stresses due to the average pressure at a point; the second term represents the viscous shear forces; the last term or Reynolds stress has both normal and tangential components. A common method of simplifying equations involves the introduction of an eddy viscosity, ε_m such that

$$\rho\varepsilon_m \overline{D}_{ji} = -\rho\overline{u_i u}_{j(j\neq i)}. \tag{14}$$

The requirement that $(i \neq j)$ in Eq. (14) eliminates the normal stresses due to turbulence; in order to account for these normal stresses an average turbulence pressure $\overline{P}_i$ is added to $\overline{P}$ thus yielding the simplified stress tensor

$$\sigma_{ji} = -(\overline{P} + \overline{P}_t)\delta_{ji} + (\mu + \rho\varepsilon_m)_i D_{ji}. \tag{15}$$

The fluid continuity equation is

$$\frac{\partial \overline{U}_i}{\partial x_i} = 0. \tag{16}$$

Equations (13), (15), and (16) may be solved in a few cases by methods developed to solve the Navier-Stokes equations.

Transport of a Scalar Quantity in Turbulent Flow In an incompressible turbulent fluid the conservation of a scalar quantity requires that the rate of change of a scalar (say $\overline{c}$ plus the rate of generation of $\overline{c}$ at the point or

$$\frac{D\overline{c}}{Dt} = \frac{\partial}{\partial x_i}\left(h\frac{\partial \overline{c}}{\partial x_i} - \overline{u_i c'}\right) + \overline{F}_c \tag{17}$$

in which $c = \overline{c} + c'$; $\overline{c}$ = ensemble average of c; c' = random component of c; D/Dt = substantial derivative; $\overline{F}_c$ is the generation term; h = molecular diffusion coefficient. It is usual to introduce, into Eq. (17), an "eddy" transport coefficient ε_j, such that

$$-\overline{u_i c'} = \varepsilon_c \frac{\partial \overline{c}}{\partial x_i}. \tag{18}$$

Since in most practical problems $\varepsilon_c >> h$, then Eq. (17) can be reduced to

$$\frac{D\overline{c}}{dt} = \frac{\partial}{\partial x_i}\left(\varepsilon_c \frac{\partial \overline{c}}{\partial x_i}\right) + \overline{F}_c. \tag{19}$$

Equations (3) and (19) are valid for low sediment concentrations. A review paper by Vasiliev[26] discusses the governing equation which account for various levels of sediment concentrations. For example a first order correction to the Reynolds equations is

$$\rho\frac{D\overline{U}_i}{Dt} = \frac{\partial}{\partial x_i}(\sigma_{ji}) + (1 + r\overline{c})F_t. \tag{20}$$

The volume continuity equation is the same as Eq. (16) while the mass continuity equation becomes

$$\frac{D\overline{c}}{Dt} = -\frac{\partial}{\partial x_i}(\overline{\overline{cu_i}}) + \left\{v_s \frac{\partial \overline{c}}{\partial x_3}\right\} \tag{21}$$

in which $r = (S_s—1)$; $\overline{c}$ = average ensemble concentration at a point (mass/mass); ν_s = settling viscosity; x_3 = vertical coordinate (opposite to the direction of ν_s).

The Vertical Concentration Profile There is no general solution for Eqs. (3), (16), and (19) or (18), (20), and (21); however a few special cases, of practical interest, have been solved.

Using the simplifications which result for steady, uniform flow in two dimensions (as shown in Figures 9 and 10), it is possible to obtain a solution for the vertical velocity, and concentration profiles. The following assumptions are typical of those required to solve Eqs. (13), (14), and (21):

a) $\overline{c} >> 1$;

b) $\varepsilon_c = \beta\varepsilon_m$ where β ; 1;

c) $\overline{F}_x$; $g\rho s_0$;

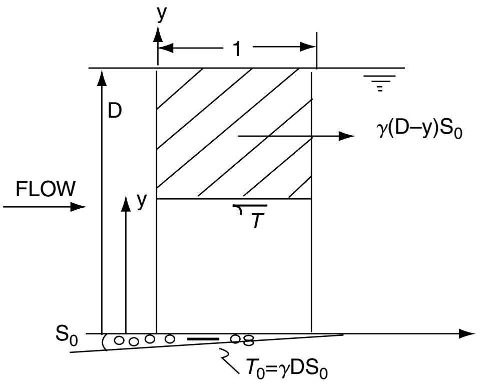


FIGURE 9 Defining sketch for uniform flow.

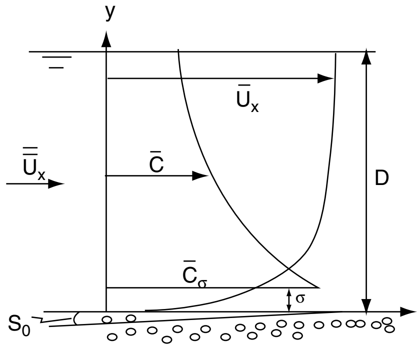


FIGURE 10 Defining sketch for velocity and concentration profiles.

d) $\dfrac{\partial(\bar{P}+\bar{P}_t)}{\partial x}=0;$

e) $\tau o = \rho g S o D$ and $\tau = \rho g S o (D-y)$;

Chang *et al.*, used the above assumptions to obtain: (a) the vertical velocity profile,

$$\bar{U}_x=\frac{2U_*}{\kappa}\left\{\sqrt{1-\varepsilon}-1\mathrm{n}\left(\frac{\varepsilon^{1/2}}{1-\sqrt{1-\varepsilon}}\right)+\frac{1}{3}\right\}+\bar{\bar{U}}_x \tag{22}$$

and (b) the vertical concentration profile

$$\bar{c}(y)=\bar{c}_a\left\{\left(\frac{y}{a}\right)^{1/2}\frac{\left(\sqrt{D}-\sqrt{D-a}\right)}{\left(\sqrt{D}-\sqrt{D-y}\right)}\right\}^{2z}, \tag{23}$$

in which U_* = friction velocity $=\sqrt{gDs_0}$; $\kappa = fcn\ (U_*d/v) \div 0.4$; $\varepsilon = y/D$; $\acute{U}$ = average velocity in vertical; a = reference height; $Z = v_s/(\beta U_*\kappa)$. Using the Keulegan velocity distributions

$$\bar{U}_x=5.75\log_{10}\left(\frac{9.05yU_*}{v}\right)\ \text{(for smooth boundaries)} \tag{24a}$$

$$\bar{U}_x=5.75\log_{10}\left(\frac{30.2y}{d_{65}}\right)\ \text{(for smooth boundaries)} \tag{24b}$$

Einstein and others[5,1] have obtained a slightly different equation for $c(y)$, i.e.

$$\bar{c}(y)=\bar{c}_a\left[\frac{a(D-y)}{y(D-a)}\right]^z. \tag{25}$$

The Suspended Sediment Load The suspended sediment discharge q_s (weight/unit time/unit width) above a reference level $y = a$ is given by

$$q_s=\gamma\int_a^D\bar{U}_x\bar{c}dy, \tag{26}$$

where $\bar{U}$ and $\bar{c}$ are given by Eqs. (22) and (23) or (24) and (25).

Longitudinal Dispersion Another problem which has received some attention is that of longitudinal diffusion and dispersion in natural streams and estuaries. Several researchers[16,33,34,35] have sought analytical and numerical solutions for the longitudinal variation in the mean concentration in the vertical, $\bar{c}$.

Considering two dimensional longitudinal dispersion, Eq. (17) can be approximated by[16]

$$\frac{\partial\hat{c}}{\partial t}+\bar{\bar{U}}_x\frac{\partial\hat{c}}{\partial t}=E_L\frac{\partial^2\hat{c}}{\partial x^2} \tag{27}$$

in which E_L ; coefficient of longitudinal diffusion. A typical[16] value for E_L is

$$E_L=5.9\bar{U}_*D.$$

The solution of Eq. (27) for an initial step change, M_o, in concentration, is

$$\hat{c}(x,t)=\frac{M_o}{\sqrt{4\pi E_L t}}e^{-(x-\bar{U}_x t)^2/4E_L t}. \tag{29}$$

Other treatments of the dispersion problem may be found in the works of Holley[28] Householder *et al.*,[29] Chiu *et al.*,[30] Conover *et al.*,[31] Sooky,[32] Fischer,[33] Harleman *et al.*,[34] and Sayre.[35]

The Total Sediment Load

Einstein[5] developed a unified total bed material formulae by converting his computed bed load, q_B to a reference concentration at $y = a = 2d$. Inserting into Eqs. (25) and (26) he obtained an estimate of q_{sB} the suspended bed material load. Hence the total bed material load per unit width, q_{TB} is

$$\begin{aligned} q_{TB} &= q_B + q_{sB} \\ &= \sum_{\substack{\text{all size}\\ \text{ranges}}} i_B q_B\,(1 + P_e I_1 + I_2) \end{aligned} \tag{30}$$

in which $I_1 = K\int_A^1 (1/y-1)^z\,\mathrm{d}y$; $I_2 = K\int_A^1 (1/y-1)\ln y\,\mathrm{d}y$; $K = 0.216\,A_e^{Z'-1}/(1-A_e)^{Z'}$; $A_\mathrm{e} = 2\mathrm{d}/D$; $P_\mathrm{e} = 2.3\log 30.2\chi d/d_{65}$; $Z' = v_\mathrm{s}/(\beta U_*\kappa)$; ($X$ *see Bed Load Formulae*).

The total sediment load per unit width in a stream q_T is

$$q_\mathrm{T} = q_\mathrm{TB} + q_\mathrm{w}, \tag{31}$$

where q_w = wash load (fines) which must be obtained independently, e.g. by direct measurement. The Einstein method requires a knowledge of: grain size distribution in the bed; the grain density; the energy slope, S_e; and the water temperature, in order to compute both bed material load and water discharge for a given depth and width of flow.

Colby and Hembree[9,36] modified Einstein's method in order to compute total sediment load (q_T). Their procedure utilizes: the sampled suspended load Q_s; measured discharge; measured depths and sampler depths, the extent of the sampled zone; and all the data used by the Einstein procedure except S_e. Their main modifications are:

1) The finer portion of the total suspended load, Q_s, is based on extrapolation of the actual sampled load Q'_s (using Eqs. (25) and (26)).
2) The coarser part of the total load (including the bed load) is computed from a simplified Einstein procedure (using a modification of Eq. (30)).
3) Einstein's grain shear velocity U'_* is replaced by an equivalent shear velocity U_m based on the Keulegan equations and the measured discharge.
4) Einstein's flow intensity function ψ_*, is replaced by the larger of

$$\Psi_\mathrm{m} = 1.65 g d_{35}/(U_\mathrm{m})^2 \text{ or } \Psi_\mathrm{m} = 0.66 g d/(U_\mathrm{m})^2 \tag{32}$$

5) The modified term Φ_m is used to enter Einstein's Eq. (9) to obtain a bed transport function ψ_*; the modified bed transport function is

$$\Phi_* = \Phi_*/2. \tag{33}$$

The value of Φ_m is used to compute the bed load associated with a size range *d*, i.e.

$$i_\mathrm{B} q_\mathrm{B}\ ;\ 1200 d^{3/2} i_\mathrm{B} \Phi_m \,\mathrm{lb/sec/ft}. \tag{34}$$

6) Using the computed bed load for a certain size range, $i_\mathrm{B}Q_\mathrm{B}$, the measured suspended load in the same size range, $I_s Q'_s$, and Einstein's Eq. (30) one can obtain a value for Z' in Eq. (25) which should be better than a Z' based on an estimated v_s.

Bishop, Simons, and Richardson[6] simplified the Einstein procedure for determining total bed material load. They introduced a single sediment transport function ΦT which includes both suspended bed material and bed load. Their flow intensity term is

$$v_T = (S_s - 1)\frac{d_{35}}{R' S_e}. \tag{35}$$

The experimental relationship shown in Figure 11 were established for actual river sediments of various median sizes. Using ΦT from Figure 11 the total bed material load per unit width, is

$$q_{TB} = \Phi_{T\rho s}(gd)^{3/2}(S_s - 1)^{1/2}. \tag{36}$$

The wash load must be added to q_TB to obtain the total sediment load.

Colby[37] analysed extension laboratory and field data to establish the empirical relationship, shown in Figure 12, for the determination of sand discharge. Figure 12 is valid for a water temperature of 60°F and a flow to moderate wash load ($\hat{c} < 10{,}000$ ppm). Colby provides adjustment coefficients for water temperature and wash load. For example, at a flow depth of 10 feet a ±20°F change in temperature would result in about +25% change in the sand discharge and an increase in the concentration of fines from 0 to 100,000 ppm could cause up to 10 fold increase in the sand load.

The reader is referred to Graf,[16] Shen,[15] and Chang *et al.*,[27] for other contributions to the determination of total bed material load.

The Annual Sediment Transport

In general it is not practical to continuously sample the sediment in a stream; instead, representative samples are taken for various flow conditions and a sediment load versus water discharge or sediment rating curve is established. A typical sediment rating curve is shown in Figure 13. A number of factors contribute to the scatter of data points in Figure 13. The sediment load is out of phase with the discharge hydrograph as illustrated by Figure 14. The sediment load depends on the season of the year.

Using the available stream flows and the sediment rating curve an average annual sediment transport can be estimated.

Often the bed load is not included in the sediment rating curve; if this is the case, the bed load may be computed as discussed under *Bed Load Formulae* and added to the annual sediment transport.

THE RESPONSE OF A CHANNEL TO CHANGES IN ITS SEDIMENT CHARACTERISTICS

Lane's Model

Lane[39] proposed the relationship

(sediment load) × (sediment size) ∞ (stream slope) × (stream flow)

or

$$(Q_\mathrm{s} \times d) \infty (S \times Q) \tag{37}$$

to describe qualitatively to behavior of a stream carrying sediment.

Lane used the following terms in referring to streams: (1) "grade ≡ equilibrium or regime slope; (2) "aggrading" ≡ rising of the stream bed due to deposition; (3) "degrading" ≡ losing of the stream bed due to scouring

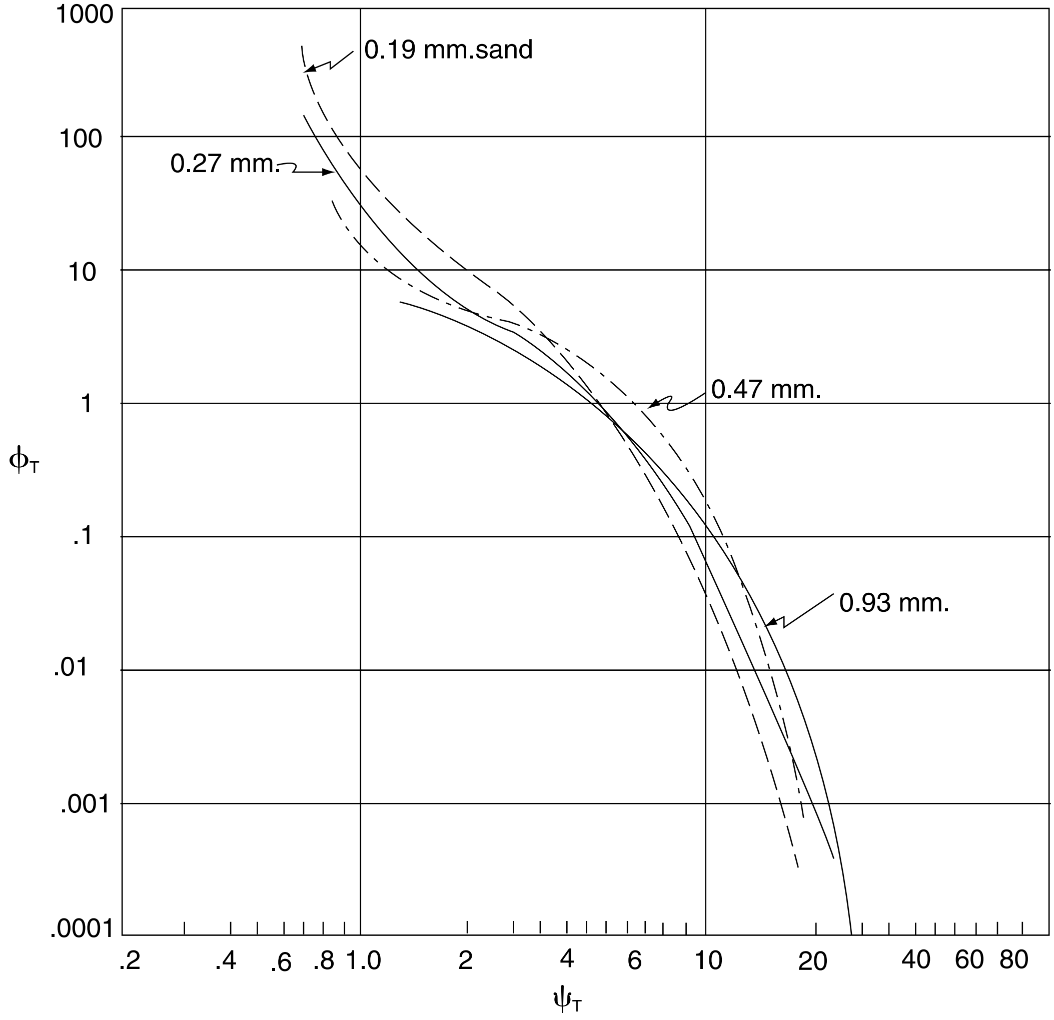


FIGURE 11 Bed material transport function (after Bishop *et al.*).

(4) "base level" ≡ the local level to which a stream tends to cut its bed.

Lane contended that there is a natural tendency for a balance between the products in Eq. (38). For example, if one of the factors, Q_s, is decreased then, in order to balance the equation, the stream slope might also tend to decrease, i.e. degradation. On the other hand an increase in Q_s could lead to an increase in S, i.e. aggradation.

The Regime Approach

"The dimensions (width, depth, and slope) of a channel to carry a given discharge, with a given silt load, are fixed by nature, i.e. uniquely determined." A channel whose dimensions are so established is said to be in regime.

In the geological sense[3,10] a river system is never really in equilibrium. According to W.M. Davis, who postulated a geomorphological cycle,[3,10] the agents of uplift and gravity (represented mainly by steam erosion) are always opposing each other. However, from the engineering point of view, a stream can be considered to be in "equilibrium" over a period of a few decades if its average behavior or average dimensions remain unchanged. There are always fluctuations, of the channel geometry, about this average; thus the steam is sometimes said to be in "dynamic equilibrium."

Of course, a stream may be aggrading or degrading (on the average in Engineering time) and thus it is not in equilibrium. The regime theory could be used to predict the ultimate dimensions of a stream that is not in regime.

Kennedy[40] and Lindley[41] collected data from canals in India (Pakistan) and proposed an equation for the non-filtering, non-scouring velocity, v,

$$v = C_1 y^n, \tag{38}$$

where $C_1 = 0.84$; $n = 0.64$; and y = depth of flow.

Kennedy was followed by Lacey, Inglis, and Blench who developed equations for channel slope and width.

Lacey[42,21] introduced the equations

$$v = 1.17\sqrt{fR} \tag{39}$$

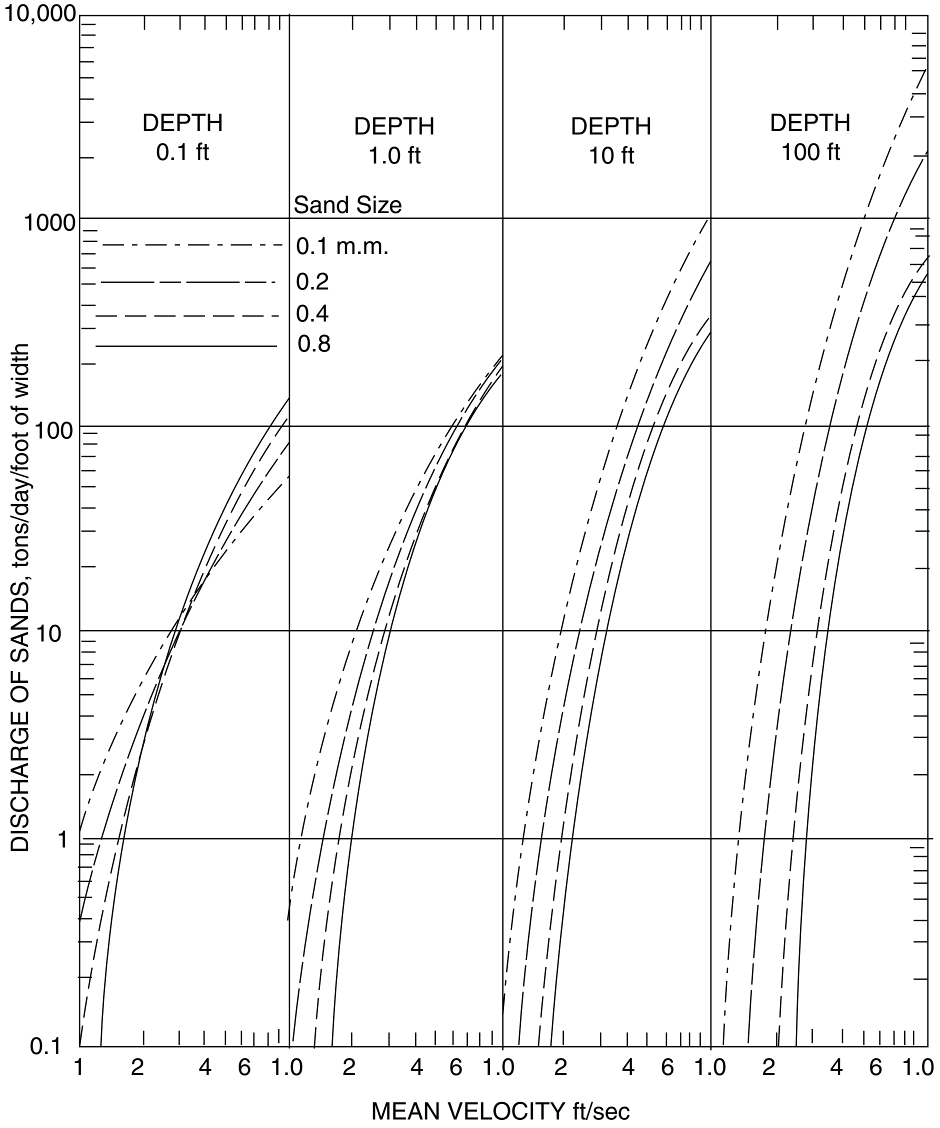


FIGURE 12 Sand discharge (after Colby).

$$f = 8\sqrt{d} \tag{40}$$

$$v = 16R^{2/3}S^{1/3} \tag{41}$$

$$P = 2.67\sqrt{Q} \tag{42}$$

in which R = hydraulic radius (ft.); d = medium grain diameter (inches); P = wetted perimeter (ft.) Q = dominant discharge = υPR(cfs.).

The dominant discharge in the case of canal flow is the design flow. In the case of a natural stream it is the channel forming flow which is often taken to be the bankfull discharge[3] which has an approximate return period of, between 1.25 and 2.33, years.

Blench[43] advanced the work of Lacey and introduced bed and side factors to better describe the depth and width of a channel in regime. His equations are:

$$f_b = v^2/y = \text{bed factor} \tag{43}$$

$$f_s = v^3/b = \text{side factor} \tag{44}$$

and

$$S = \frac{(f_b)^{5/6}(f_s)^{1/12}1^{1/4}}{3.63gQ^{1/6}(1+c/2330)}, \tag{45}$$

in which c is the concentration in ppm; b = mean width (ft—sec. units). The recommended f_b is

$$f_b = 9.6\sqrt{d(\text{in})}(1+0.012c) \tag{46}$$

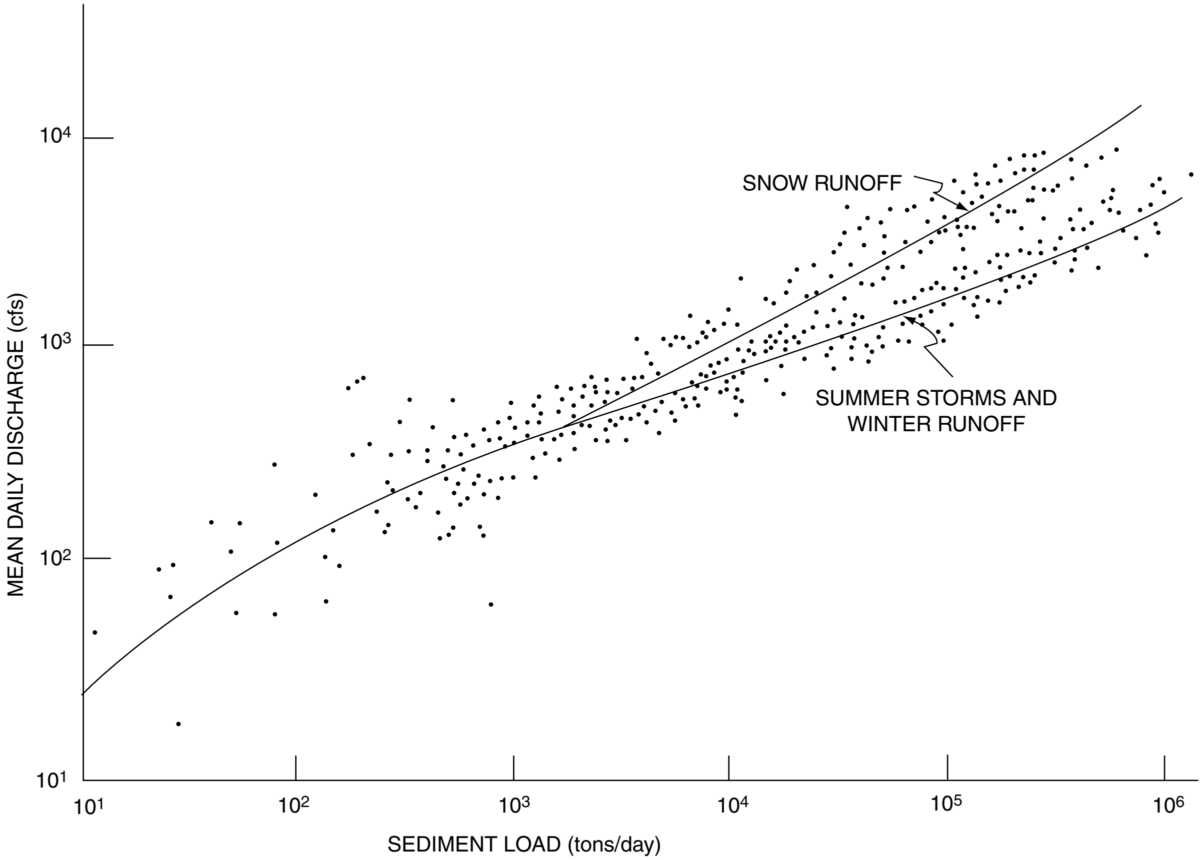


FIGURE 13 Typical sediment rating curve (adapted from Ref. 38).

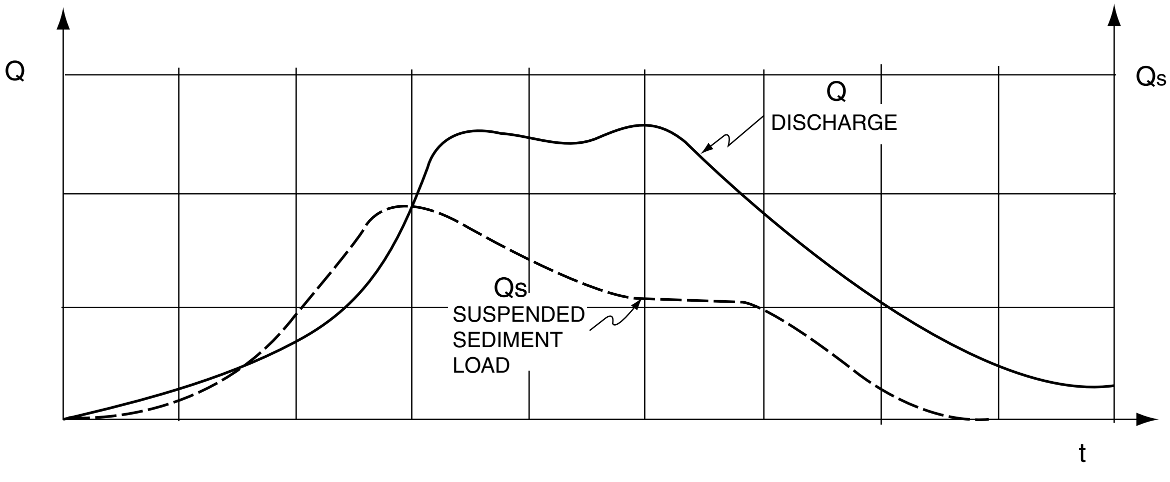


FIGURE 14 Variation of sediment load with time (adapted from Graf).

and f_s depends on the cohesiveness of the material in the channel banks. For example, f_s is approximately 0.1, 0.2, and 0.3 for low, medium, and high cohesiveness respectively.

The data of Lacey and Blench were mainly from canals, with moderate to low sediment loads ($c < 500$ ppm) and with sand beds and slightly cohesive banks.[17]

Simons and Albertson[44,21] have extended the regime equations to make them applicable to the five channel classifications shown in Table 4.

They proposed the following general equations:

$$P = K_1 Q^{1/2} \tag{47}$$

TABLE 4
Canal classification[21,44]

1) Sand bed and banks.
2) Sand bed and cohesive banks.
3) Cohesive bed and banks.
4) Coarse noncohesive material.
5) Same as for 2, but with heavy sediment loads, 2000–8000 ppm.

$$b = 0.9\,P \tag{48}$$

$$b = 0.92B–2.0 \tag{49}$$

$$R = K_2Q_{0.36} \tag{50}$$

$$y = 1.21\,R \text{ for } R < 7 \text{ ft.} \tag{51}$$

$$y = 2 + 0.93\,R \text{ for } R > 7 \text{ ft.} \tag{52}$$

$$v = K_3(R^2S)_m \tag{53}$$

$$\frac{C^2}{g} = \frac{v^2}{gyS} = K_4\left(\frac{vb}{v}\right)^{0.37} \tag{54}$$

in which C = Chézy coefficient; B = surface width; the values of the K's and m are defined in Table 5.

Schumm[17,45] studied the Great Plains rivers in the US and similar rivers in Australia; he correlated stream geometry with discharge (mean annual flood Q_{ma}) and the percent silt-clay, M, in the channel boundary. Some of his correlations are:

$$\text{Full Bank width} = W = \frac{2.3Q_{ma}^{0.58}}{(M)^{0.37}} \tag{55}$$

$$\text{Width: depth ratio} = \frac{W}{D} = \frac{21Q_{ma}^{0.18}}{(M)^{0.74}} \tag{56}$$

$$\text{Sinuosity} = 0.94M^{0.25}, \tag{57}$$

The high correlation of stream geometry and M led Schumm to classify channels (see Table 6) using M as an index to the ratio of coarse load to total load.

Schumm[45,17] associated meandering channels with high values of M and low bed load and braided channels (a relatively straight, steep main channel consisting of a maze of sub-channels sometimes separated by bars or islands) with low values of M and high bed load.

RESERVOIR SEDIMENTATION

A common objective of many sediment transport studies is the prediction of reservoir "siltation" rates. Reservoir siltation depends on: (1) the average annual sediment load entering the reservoir; (2) the grain size distribution of the sediment load; (3) the reservoir trap efficiency; (4) the bulk dry specific weight of the deposited sediments in the reservoir.

TABLE 5
Constants in Eqs. (47) to (54) (ft-sec. units)

	Channel type				
Coefficient	1	2	3	4	5
K_1	3.5	2.6	2.2	1.75	1.7
K_2	0.52	0.44	0.37	0.23	0.34
K_3	13.9	16.0	—	17.9	16.0
K_4	0.33	0.54	0.87	—	—
m	0.33	0.33	—	0.29	0.29

TABLE 6
Schumm's classification

Stream type	Bed load	Mixed load	Suspended load
M	Braided		Meandering
	> 5%	5–20%	> 20%
Coarse load Total total	> 11%	3–11%	< 3%

The determination of annual sediment yields was discussed in Sections *Erosion* and *The Mechanics of Sediment Transport in a Stream.* It is important to predict the possible effects of land development and/or sediment control measures on future sediment yields.[46]

An estimate of the percentage sand, silt, and clay for the incoming sediment can be obtained on the basis of grain size analyses of the existing load.

Brune in 1953[47,8] presented the trap efficiency curve shown in Figure 15, which applies to reservoirs which normally ponded. Detention type reservoirs would have lower trap efficiencies.

Since annual sediment yield is usually determined in terms of weight it is necessary to know the bulk dry specific weight, and the trapped sediments, in order to estimate the volumetric decrease in reservoir storage. To accomplish this Lane and Koelzer[48,8] developed an equation to describe the change in bulk dry specific weight, $\gamma*$, of reservoir deposits with time, i.e.

$$\begin{aligned}\gamma_* = (\gamma^*_{\text{sand}(1)} + K_{\text{sand}} \log_{10} T)X_{\text{sand}} \\ + (\gamma^*_{\text{silt}(1)} + K_{\text{silt}} \log_{10} T)X_{\text{silt}} \\ + (\gamma^*_{\text{clay}(1)} + K_{\text{clay}} \log_{10} T)X_{\text{clay}}\end{aligned} \tag{58}$$

in which T = time in years; X_{sand} = fraction of sand in deposit; $\gamma^*_{\text{sand}(1)}$ etc. are the respective bulk dry specific weights at $T \leqslant 1$ year (see Table 7); K_{sand} etc. are constants (see Table 7).

After a period of N years a reservoir will contain deposits varying in age from less than one year to N years with the consequence that the true bulk dry specific weight, γ^*, for the entire deposit should be found by averaging Eq. (58)

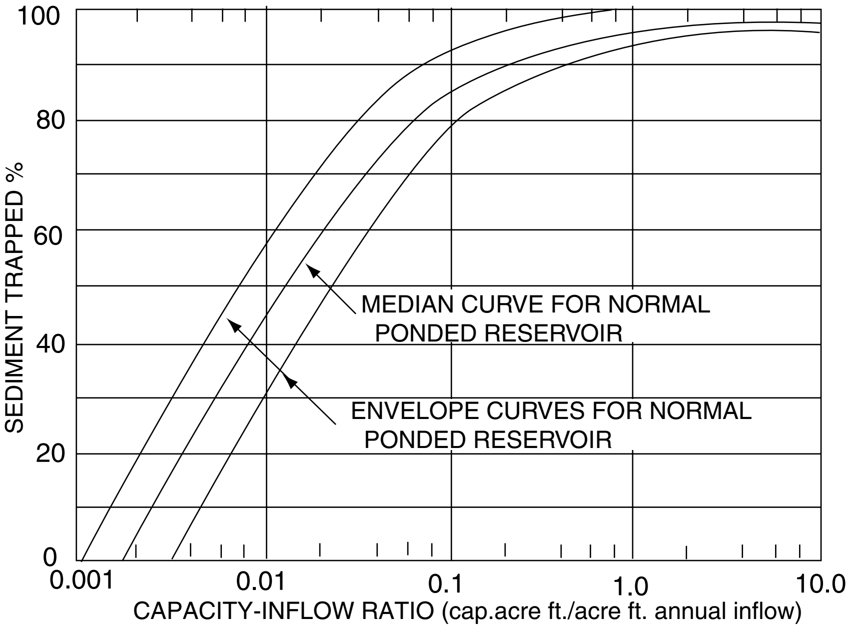


FIGURE 15 Trap efficiency curve (after Brune).

TABLE 7
Constants for Eq. (59) (after Lane and Koelzer)

Reservoir operation	Sand		Silt		Clay	
	γ^*	K	γ^*	K	γ^*	K
Normally submerged	93	0	65	5.7	30	16.0
Moderate reservoir drawdown	93	0	74	2.7	46	10.7
Considerable reservoir drawdown	93	0	79	1.0	60	6.0
Reservoir normally empty	93	0	82	0.0	78	0.0

over the N year period. Some typical[8] averaged values of $\bar{\gamma}^*$ for various time periods are:

$N = 10$ years, $\bar{\gamma}_{*10} = \gamma_{*T=1} + 0.675K$

$N = 50$ years, $\bar{\gamma}_{*50} = \gamma_{*T=1} + 1.298K$

$N = 100$ years, $\bar{\gamma}_{*100} = \gamma_{*T=1} + 1.588K$.

The decrease in storage (in acre-ft) during an N year period is given by:

$$\Delta v = \left(\frac{\overline{TE}}{100}\right)\frac{(SY)(DA)\times N}{\bar{\gamma}_{*N}}\times\left(\frac{2000}{43{,}560}\right) \tag{59}$$

in which

$\overline{TE}$ = average trap efficiency in percent;
$\overline{SY}$ = average annual sediment yields in tons/sq. mile;
(DA) = drainage area in sq. miles;
$\bar{\gamma}_{*N}$ = average bulk specific weight for the N-year computation period in lbs/ft^3.

A typical distribution of reservoir sediments is shown in Figure 16. The coarse sediments (sands and gravels) form a delta at the upstream section of the pond. Finer sediments are deposited downstream from the delta. Very fine sediment (clay particles) may pass through the reservoir or be deposited in the downstream portion of the pond.

THE DEGRADATION PROBLEM

The response of a stream to changes in its sediment load was described qualitatively in Section 6. The problem of computing the probable rate and extent of degradation downstream from a proposed reservoir has recently received the attention of a number of researchers.[15,49,50]

Numerical estimates of the degrading channel profiles may be obtained by solving, simultaneously, the following equations (see Gessler, Ref. 15):

1) the equation of sediment continuity—

$$\frac{\partial z}{\partial t} + \frac{1}{b(1-m)}\frac{\partial (q_{TB} b)}{\partial x} = 0 \tag{60}$$

2) the bed material transport equation which may have the form—

$$q_{TB} = C'\,(\tau_o - \tau_c)P \tag{61}$$

3) the bed shear equation—

$$\tau_o = \gamma R S_e \tag{62}$$

4) a friction equation, e.g. the Chézy equation—

$$v = C(RS_e)^{1/2}, \tag{63}$$

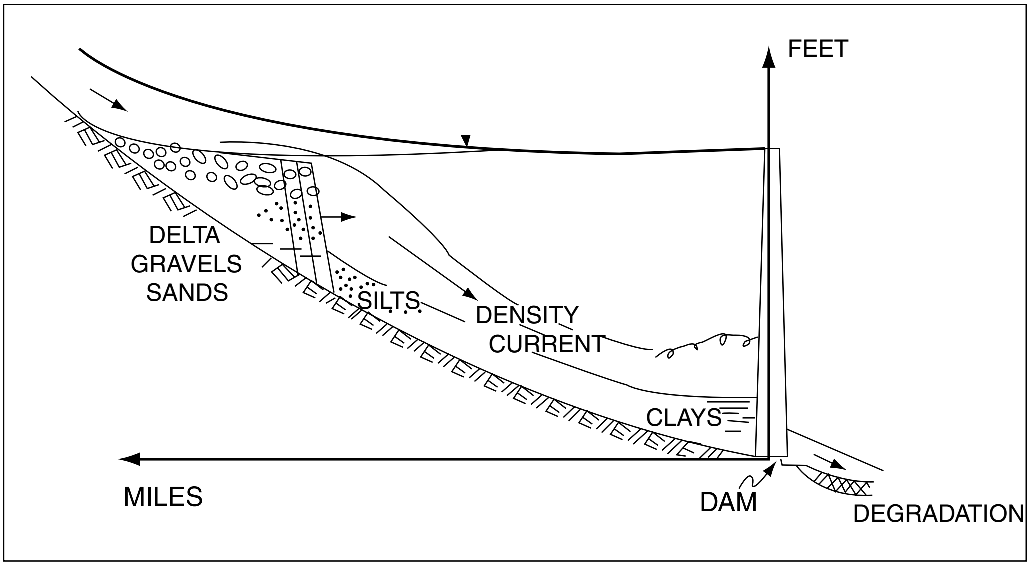


FIGURE 16 Distribution of deposits in a reservoir.

in which z = bed elevation; qTB = volumetric bed material load; m = bed porosity; C' and p are constants for a given bed material and bed form.

As the bed is soured, an armouring process occurs due to the fact that the finer bed material is more readily removed than the coarser particles in the bed (See Figure 1). This may be accounted for by adjusting τ_c in Eq. (62).

SEDIMENT CONTROL

The adverse effects of excessive sediment loads in reservoirs, navigation channels, harbors, and aquatic life may be alleviated by a sediment control program which may include measures such as prevention of over-hand erosion or containment of eroded soil near its source. On the other hand the removal of an established sediment load from a stream may also lead to undesirable consequences such as channel degradation or changes in the established aquatic life.[2]

The ASCE Sedimentation Tank Committee[46] has classified sediment control measures under: (1) land treatment, (2) structural.

Land treatment measures are used to reduce wash load (fines) resulting mainly from sheet erosion. Structural measures are most effective for reducing sediment load derived from stream-channel erosion, gully erosion, and sediment associated with mining and construction work.

The main *land treatment* measures are summarized below:[46]

1) *Vegetative treatment* includes changes in the existing land use towards more: use of cover crops and crop rotation, maintenance of effective vegetative cover in critical areas, leaving of straw and stubble in the field, use of long-term hay stands, mulching, pasture planting, and re-forestation.
2) *Protecting existing vegetative cover* involves protection of existing forest sands from excessive fire losses and the protection of all vegetated areas from over grazing.
3) *Mechanical field practices* are used in conjunction with (1) and (2) and include contour farming, contour furrowing of range land, contour strip-cropping, use of gradient and level terraces, use of diversions to divert runoff away from critical areas, use of grassed waterways and ditch and canal linings, and the use of grade stabilization structures in areas subject to possible gully erosion.

The ASCE Task Committee[46] outlined three commonly used *structure* measures:

1) *Reservoirs,* either detention or multi-purpose, decrease flood stages and consequently may decrease downstream damages due to deposition of flood borne sediments; however it should be remembered that reservoirs themselves create sediment problems which may be just as serious as the problem being solved.
2) *Stream channel improvement and stabilization*—these measures may include straightening, cleaning, deepening, and widening of existing channels in order to decrease local flood stages and related sediment damages; again, complications, such as increased channel erosion, can be expected. Other channel improvement methods are: lining the channel on bends and other erodible areas, and providing spur dykes to deflect the flow away from erodible banks.
3) *Debris and sedimentation basins* are usually relatively small reservoirs designed to trap debris and sediment near its source. This method is particularly useful in controlling sediment yields from construction sites or mining operations.

LIST OF SYMBOLS

a = $2d$ = reference distance from the stream bed;
A = flow area;
A_e = a/D;
b = average channel width;
B = top width of channel;
c = point concentration;
$\bar{c}$ = average concentration at a point;
c' = random component of concentration at a point;
$\hat{c}$ = average concentration in the vertical;
C = Chézy coefficient;
C_s = Du Boys' coefficient;
C = constant;
D = flow depth;
$\bar{D}_{ij}$ = deformation tensor;
E_L = coefficient of longitudinal dispersion;
E = erosion rate;
f = friction factor;
$fcn()$ = function of ();
$f()$ = function of ();
f = silt factor (Lacey);
f_b = bed factor (Blench);
f_s = side factor (Blench);
F_c = generation term:
F_t = body force;
g = acceleration due to gravity;
h = coefficient of molecular diffusion;
H = height;
i = index in tensor notation;
i_B = portion of material in the bed within a specified size range;
i_B = portion of the bed load in specified size range;
i_s = portion of the suspended load in specified size range;
j = index in tensor notation;
k = constant;
K = constant;
m = porosity;
m = exponent (Simons and Albertson);
M = percent silt-clay;
M_o = initial concentration;
n = exponent;
N = time period;
p = exponent;
P = wetted perimeter;
$\bar{P}$ = average pressure;
P_t = turbulence pressure;
q_B = bed load/unit width;
q_s = suspended load/unit width;
qTB = total bed material load/unit width;
qT = total sediment load/unit width;
qw = wash load/unit width;
Qs = total suspended load;
QT = total sediment load;
Q = water discharge;
R = hydraulic radius;
R' = hydraulic radius associated with grain roughness;
R'' = hydraulic radius associated with bed form;
r = (S_s-1);
S_s = specific weight of sediment grains;
S_o = bed slope;
S_e = energy slope;
SY = average annual sediment yield;
t = time;
T = time period;
u_i = random component of the velocity U;
U_i = instantaneous velocity in direction i;
$\bar{U}_i$ = average velocity in direction i;
U_* = friction velocity;
$\bar{U}_x$ = average point velocity in the x-direction;
U_x = average velocity in the vertical;
v = stream velocity;
v_S = terminal settling velocity of sediment particles;
w = wash load;
W = channel width;
x = coordinate;
y = coordinate;
z = coordinate;
Z = $\frac{v_s}{\beta U_* \kappa}$;
β = $\varepsilon_m/\varepsilon_c$;
γ = specific weight of water;
γs = specific weight of sediment grains;
γ^* = bulk dry specific weight of a deposit;
δ = laminar sub-layer thickness;
δ_{ij} = Kronecker delta;
ε_m = kinematic eddy viscosity;
ε_c = kinematic eddy transport coefficient;
κ = von Karman constant; 0.4;
λ = wave length;
μ = dynamic viscosity;
v = kinematic viscosity;
ρ = density;
σ_{ji} = stress tensor;
τ_o = bed shear stress;
τ_c = critical shear stress;
Φ = sediment transport function;
ψ = flow intensity function.

REFERENCES

1. Brown, C.B., Sediment Transportation, Chapter XII, *Engineering Hydraulics,* H. Rouse, Editor, J. Wiley and Sons, New York, 1950.
2. Turner, D.J., Dams and ecology, *Civil Engineering,* ASCE, Sept. 1971, pp. 76–80.
3. Leopold, L.B., M.G. Wolman, and J.P. Miller, *Fluvial Processes in Geomorphology,* W.H. Freeman and Co., San Francisco, 1964.
4. Einstein, H.A., Formulas for the transportation of bed load, *Trans. ASCE,* **107,** 1942.
5. Einstein, H.A., The bed-load function for sediment transportation in open channel flows, US Dept. of Agriculture Technical Bulletin No. 1026, 1950.
6. Bishop, A.A., D.B. Simons, and E.V. Richardson, Total bedmaterial transport, *Proc. ASCE,* **91,** HY2, March 1965, pp. 175–191.
7. Chow, and Ven Te, *Open-Channel Hydraulics,* McGraw-Hill Co., New York, 1959.
8. Chow, and Ven Te, *Handbook of Applied Hydrology,* McGraw-Hill Co., New York, 1964.

9. US Bureau of Reclamation, Step method for computing total sediment load by the modified Einstein procedure, Sedimentation Section, Hydrology Branch, July 1955.
10. Morisawa, M., *Streams, Their Dynamics and Morphology,* McGraw-Hill Book Co., New York, 1968.
11. Saxton, K.E., R.G. Spomer, and L.A. Kramer, Hydrology and erosion of loessial watersheds, *Proc. ASCE,* **97,** HY11, Nov. 1971, pp. 1835–1851.
12. Dawdy, D.R., Knowledge of sedimentation in urban environments, *Proc. ASCE,* **93,** HY6, Nov. 1967, pp. 235–245.
13. Curtis, L.W., Sedimentation in retarding basins, *Proc. ASCE,* **94,** HY3, May 1968, pp. 739–744.
14. Wigham, J.M., Sediment transportation, Section XI, *Handbook on the Principles of Hydrology,* D.M. Gray, Editor, National Re-Search Council of Canada, 1970.
15. Shen, H.W., *River Mechanics,* **I** and **II,** Editor–Publisher H.W. Shen, P.O. Box 606, Fort Collins, Colorado, 1971.
16. Graf, W.H., *Hydraulics of Sediment Transport,* McGraw-Hill Book Co., New York, 1971.
17. Lecture notes; Summer institute in fluvial geomorphology, Colorado state University, Lectures S.A. Schumm and D.B. Simons, 1967.
18. Simons, D.B. and E.V. Richardson, Forms of bed roughness in alluvial channels, *Trans. ASCE,* **128,** Part 1, 1963, pp. 284–323.
19. Simons, D.B. and E.V. Richardson, Resistance to flow in sand channels, *Proc. XII Congress of the International Association for Hydraulic Research,* **1,** Sept. 1967, pp. 141–150.
20. White, C.M., The equilibrium of grains on the bed of a stream, *Proc. Roy. Soc. A,* **174,** 1940, pp. 322–334.
21. Henderson, F.M., *Open Channel Flow,* The MacMillan Co., New York, (Chapter 10), 1966.
22. Raudkivi, A.J., *Loose Boundary Hydraulics,* Pergamon Press, New York, 1967.
23. Straub, L.G., Approaches to the study of the mechanics of bed movement, *Proc. of Hydraulics Conference,* Univ. of Iowa Studies in Engineering, Bulletin 20, 1940.
24. Shen, H.W., (Editor), Sedimentation', Symposium to honor Professor H.A. Einstein, Berkley., California, June 1971.
25. Hinze, J.O., *Turbulence,* McGraw-Hill book Co., New York, 1959.
26. Vasiliev, O.F., Problems of two-phase flow theory, General lecture XIII Congress of the International Association for Hydraulic Research, **5**–3, Kyoto, 1969.
27. Chang, F.M., D.B. Simons, and E.V. Richardson, Total bed-material discharge in alluvial channels, *Proc. XII Congress of the International Association for Hydraulic Research,* **1,** Fort Collins, Sept. 1967, pp. 132–140.
28. Holley, E.R., Unified view of diffusion and dispersion, *Proc. ASCE,* **95,** HY2, March 1969, pp. 621–631.
29. Householder, M.K. and V.W. Goldschmidt, Turbulent diffusion and Schmidt number of particles, *Proc. ASCE,* **95,** EM6, Dec. 1969, pp. 1345–1369.
30. Chiu, C.L. and K.C. Chen, Stochastic hydrodynamics of sediment transport, *Proc. ASCE,* **95,** EM5, Oct. 1969, pp. 1215–1230.
31. Conover, W.J. and N.C. Matalas, Statistical model of turbulence in sediment-laden streams, *Proc. ASCE,* **95,** EM5, Oct 1969, pp. 1063–1081.
32. Sooky, A.A., Longitudinal dispersion in open channels, *Proc. ASCE,* **95,** HY4, July 1969, pp. 1327–1346.
33. Fischer, H.B., Dispersion predictions in natural streams, *Proc. ASCE,* **94,** SA5, Oct. 1968, pp. 927–943.
34. Harleman, D.R.F. and C.H. Lee, Numerical studies of unsteady dispersion in estuaries, *Proc. ASCE,* **94,** SA5, Oct. 1968, pp. 897–911.
35. Sayre, W.W., Dispersion of silt particles in open channel flow, *Proc. ASCE,* **95,** HY3, May 1969, pp. 1009–1038.
36. Colby, B.R. and C.H. Hembree, Computation of total sediment discharge, US Geol. Survey, Water Supply Paper 1357, 1955.
37. Colby, B.R., Discharge of sands and mean-velocity relationships in sand-bed streams. US Geol. Survey, Prof. Paper 462. A, 1964.
38. US Bureau of Reclamation, Analysis of flow duration, sedimentrating curve method of computing sediment yield, Sedimentation Section, Hydrology Branch, Denver, Colorado, April 1951.
39. Lane, E.W., The importance of fluvial morphology in hydraulic engineering, *Proc. ASCE,* **81,** 1955.
40. Kennedy, R.G., The prevention of silting in irrigation canals, *Minutes of Proc. Inst. of Civil Engineers (UK),* **119,** 1894–95.
41. Lindley, E.S., Regime channels, *Proc. Punjab. Eng. Congress,* India, **7,** 1919.
42. Lacey, G., A general theory of flow in alluvium, *Proc. Inst. of Civil Engineers (UK),* **27,** Nov. 1946.
43. Blench, T., *Mobile Bed Fluviology,* University of Alberta Press, Edmonton, Alberta, Canada, 1966.
44. Simons, D.B. and M.L. Albertson, Uniform water conveyance channels of alluvial material, *Proc. ASCE,* **86,** HY5, May 1960, pp. 33–71.
45. Schumm, S.A., River metamorphosis, *Proc. ASCE,* **95,** HY1, Jan. 1969, pp. 255–273.
46. Vanoni, V.A., Chairman of the Task Committee for Preparation of Manual on Sedimentation, Sediment control methods—Introduction and watershed area, (Chap. V), *Proc. ASCE,* **95,** HY2, March 1969, pp. 649–675.
47. Brune, G.M., Trap efficiency of reservoirs, *Trans. Am. Geophysical Union,* **34,** No. 3, 1953, pp. 407–418.
48. Lane, E.W. and V.A. Koelzer, Report No. 9, Density of sediments deposited in reservoirs, US Government and Iowa Inst. of Hydraulic Research, A study of methods used in measurement and analysis of sediment loads in streams, 1943.
49. Komura, S. and D.B. Simons, River-bed degradation below dams, *Proc. ASCE,* **93,** HY4, July 1967, pp. 1–14.
50. Komura, S., Prediction of river-bed degradation below dams, *Proc. XIV Congress of International Association for Hydraulic Research,* **3,** C31, Sept. 1971.
51. Simons, D.B. and F. Senturk, Sediment transport technology *Water Resources Publications,* For Collins, Colorado, 1977.
52. Vanoni, V. (Editor), Sedimentation engineering, Prepared by ASCE Task Committee for preparation of the Manual on sedimentation, ASCE, New York, NY, 1975.
53. Quick, M.C., Sediment transport by waves and currents, *Can. J. Civ. Eng.,* Vol. 10, no.1, 1983.
54. Kamphius, J.W., On the erosion of consolidated clay material by a fluid containing sand, *Can. J. Civ. Eng.,* Vol. 10, no. 2, 1983.

J.A. McCORQUODALE
University of New Orleans

SEWAGE: see MUNICIPAL WASTEWATER

SOLID WASTE DISPOSAL: see MANAGEMENT OF SOLID WASTE

SOURCE OF ENERGY: see ENERGY SOURCES—ALTERNATIVES

SOURCES OF POLLUTANTS: see AIR POLLUTION SOURCES

SMALL FLOW WASTEWATER TREATMENT TECHNOLOGY FOR DOMESTIC AND SPECIAL APPLICATIONS

Recent concern regarding water pollution and general worldwide public awareness of the problem associated with pollution has resulted in increased pressures on all waste dischargers to provide effective treatment and disposal of their waste streams. Small subdivisions, motels, resorts, mobile homes, watercraft, railroads, and the like, have not escaped these pressures, although in many respects their total waste contribution to the environment may be small. However, as a result of the public awareness of pollution and increasing regulatory pressures, there has been a rapid commercial growth in recent years in plants designed and built specifically for such applications. In this article the state of the art of small flow wastewater treatment systems developed for such special applications is presented.

INDIVIDUAL HOMES

The 1990 census figures indicate that there are more than 25 million onsite residential wastewater treatment systems in the United States, often because wastewater collection sewers are not available.[1] These systems include a variety of components and configurations. Among the most common systems are the anaerobic and aerobic biological treatment. In recent years constructed wetlands have also been investigated for home application. However, no treatment systems as yet has been entirely satisfactory either to homeowners or to health officials.[2,3]

Anaerobic System

Septic Tank Septic tank is the most commonly used individual waste disposal system. The US Environmental Protection Agency estimates that there are approximately 18 million housing units in the US that use on-site wastewater and disposal systems. This is about 25% of all housing units. Additionally, about one-half million new systems are being installed each year.[2,3]

A septic tank consists of a tank in which wastes are accumulated and digested under anaerobic conditions. The effluent from septic tanks is malodorous, and the bacterial count is often quite high. Subsurface absorption field is necessary to absorb the effluent from the septic tank. Capacity, hydraulic design and soil conditions are most important factors influencing the septic tank performance. A detailed discussion on septic tank design, performance, and economics is available in several publications.[1–7]

A conventional septic tank removes about 40–50 percent TSS. In recent years many improvements have been made in the design of septic tank, gravel filter, and soil absorption trenches that enhance their performance significantly. An important variation in conventional septic tank design is currently being manufactured. The *Ruck* system requires wash waters be separated from sanitary and kitchen wastes. The sanitary and kitchen wastes are thus held in an upper compartment for a longer period of anaerobic digestion, providing more concentrated treatment to the sanitary wastes.[3,8] The wash water is treated in a lower compartment for shorter periods where effluent from the upper chamber is mixed. The system is sold as a package unit in a fiberglass housing. A soil absorption system is also included in this package which is designed to make the system independent of the natural soil characteristics. Andreadakis reviewed the performance of an on-site treatment and disposal system using a septic tank, gravel filter, and soil absorption trenches.[9] BOD_5 and TSS removal efficiencies averaged over 92 and 93 percent respectively, and up to 70 percent nitrogen was removed due to nitrification followed by denitrification.[9,10] Many researchers believe that the reduction in hydraulic loading by water conserving devices will improve the performance of on-site treatment and disposal systems.[11]

Intermittent Sand Filters The main purpose of the intermittent sand filters is to reduce the BOD_5 and TSS prior to soil infiltration. Currently, many intermittent sand filters are used throughout the United States to treat wastewater from individual homes. The process is highly efficient, and requires a minimum of operation and maintenance. Intermittent sand filters are beds of granular material underlain by graded gravel and collecting tiles.[2] Uniform distribution is normally obtained by dosing or flooding the entire surface of the bed. Recirculation has also been used. The filters may be buried or may have some free access.

Disposal Methods Under favorable conditions the effluent from treatment devices are safely disposed of by (a) subsurface absorption, (b) evaporation, and (c) discharge into surface waters. The subsurface soil absorption is usually the best method of wastewater disposal from homes because of its simplicity, stability and low cost. Partially treated wastewater is discharged below ground surface where it is absorbed and treated by the soil as it percolates into the ground. Nearly one-third of the homes in the United States dispose of their wastewater in this way.[9]

Evaporation systems utilize techniques to evaporate the effluent without infiltration. The system utilize evaporation and evapotranspiration beds. Direct discharge of onsite treatment system effluent is also a disposal option if an appropriate receiving water is available and if the regulatory agencies permit such discharges. Various onsite disposal methods are listed below. Detailed discussion may be found in Ref. 2. Typical design of septic tank and intermittent filter are shown in Figure 1. Typical trench, seepage pit and mound systems are shown in Figure 2.

- Subsurface soil absorption systems
  - trench and bed
  - seepage pit
  - mound
  - fill
  - artificially drained systems
  - electro-osmosis*
- Evaporation systems
  - evapotranspiration and evapotranspiration—percolation
  - evaporation and evapopercolation ponds
- Surface water disposal
  - outfall in stream
  - outfall in lake

Aerobic System

Several types of household treatment systems are available which utilize aerobic stabilization of organic wastes. Most systems are designed for continuous flow.

The raw wastewater enters an anaerobic tank where solids are settled and partially digested. The liquid enters an aerobic compartment. Air is supplied either by mechanical aerators or by diffusers. The bacterial action thus produced is similar to that in an activated sludge plant. The solids in the aerated liquid are settled into a separate tank. This tank most commonly has a sloping bottom to return the settled sludge into the aeration tank by gravity. System components of such units are schematically shown in Figure 3.

Some of the manufacturers' variations in the aerobic system include (1) absence of anaerobic digestion tank, (2) different methods of aeration, (3) packed bed media, (4) trickling filter, (5) rotating biological contactor, and (6) use of tube settlers for increasing the sedimentation rate.[2,12] Many of these variations are shown in Figure 4.

The effluent from an aerobic system is generally better than that from a septic tank. Manufacturers claim BOD and suspended solids removal of about 90%. The effluent from aerobic systems has lower clogging effect on soil absorption system. If the system operates properly, the effluent is suitable for surface drainage. The disadvantages of the aerobic system are higher operating costs, susceptibility to shock loading, and variation in effluent quality. The design criteria, operational characteristics and cost data on aerobic systems are extensively available in the literature.[2,3,13,14]

Constructed Wetlands

Constructed and natural wetland rely solely on natural process to treat wastewater and are most often used for secondary treatment. In a single-residence system, for example, a septic tank generally provides partial treatment. The effluent flows to the wetlands where it is distributed into the system. Organic matter is stabilized by microorganisms attached to the plant roots. Aquatic plants deliver oxygen, provide shade, metabolic nutrients and surface area for microbiological growth.

Constructed wetlands are designed either as a "discharge" system or as an "non-discharge" system.[1] Two basic design approaches exit for constructed wetlands. These are developed by (1) Tennessee Valley Authority (TVA) for systems less than 75,700 L/d, and (2) EPA for larger municipal wastewater treatment plants. The design equations are based on hydraulic loading, organic loading, and Darcy's equation. The design procedure for constructed wetland can be found in Refs. 1, 15–17. The suggested residential wetland design details are: length = 12.7 m, width = 4.3m, depth = 0.3 m, detention time = 3d, and hydraulic loading criteria = 1.3 $m^2/L \cdot d$.

The aquatic plants are generally chosen from indigenous species of *Typhaceae* (cattail family), *Cyperaceae* (sedge family), *Gramineae* (grass family), *Scirpus validus* (softstem bulrush), or *Phragmites australis* (giant reed). Care should be taken to avoid plants that "choke out" each other, or those eaten by animals. The choice of vegetation is dependent upon wastewater characteristics, solar radiation, temperature, aesthetics, wildlife desired, indigenous species, and the depth of constructed wetlands. Vegetation harvesting may be necessary when it becomes too dense, cause obstruction to the natural flow and create anaerobic conditions.[1]

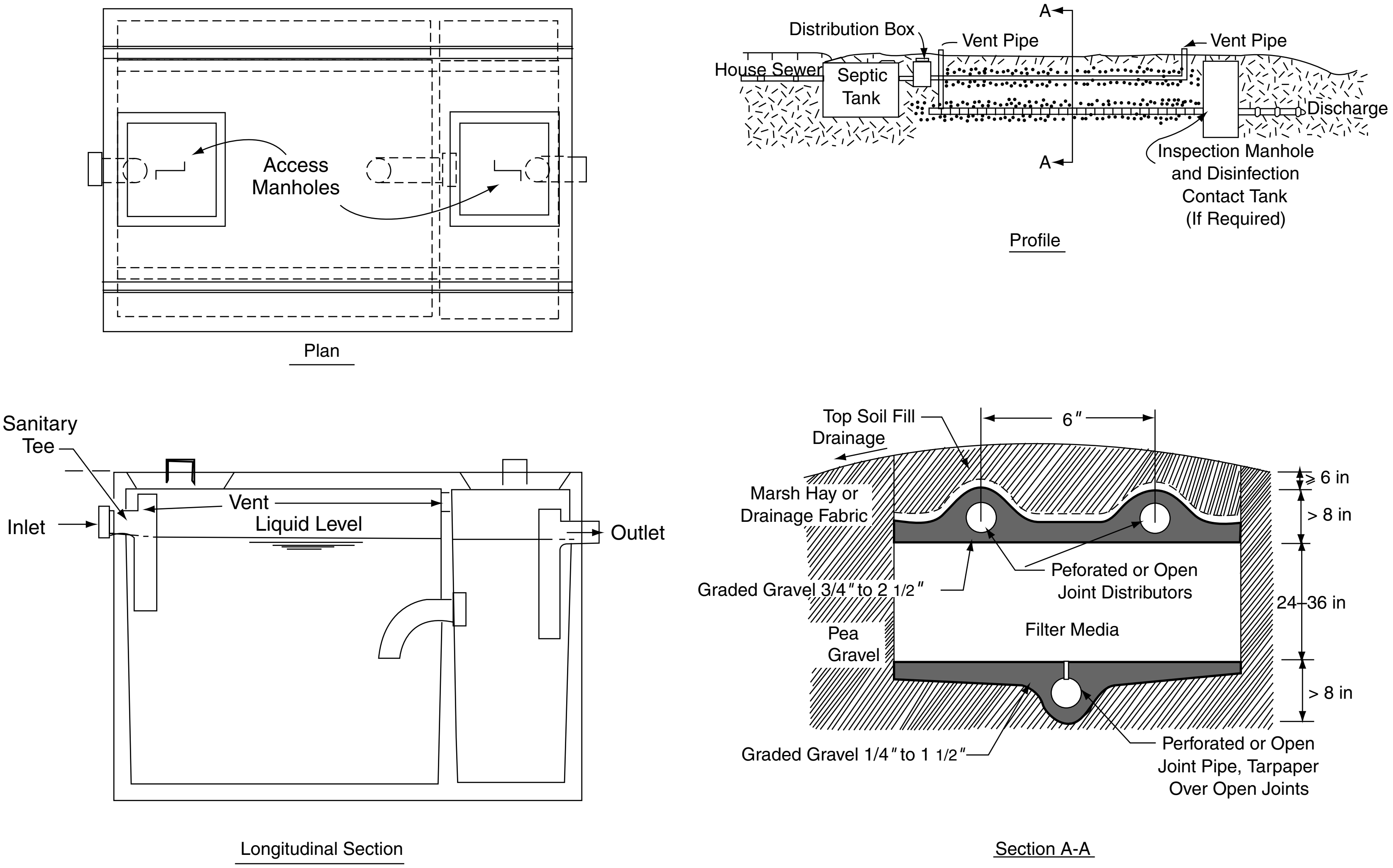
SMALL ESTABLISHMENTS

Small establishments generally include motels, restaurants, stores theatres, clubs, camps, rest areas, institutions, apartment houses, small factories, subdivisions, small communities, etc. Disposal of wastewaters from these establishments in suburban areas certainly poses serious problems. Construction of public sewer systems may not be economically feasible to convey the wastewaters to an existing treatment plant. Since the flows are relatively large, often septic tanks and subsurface absorption fields may not provide safe disposal, causing a very serious public health hazard or pollution of ground or surface waters.

Package treatment plants were first introduced about thirty-five years ago to treat wastes from small establishments. During this time, there has been a rapid commercial growth in package plant industry. The package plants are prefabricated in the factory and, in most cases, completely assembled prior to delivery to the plant site. Most of the plants are made of steel but many concrete plants are also

---

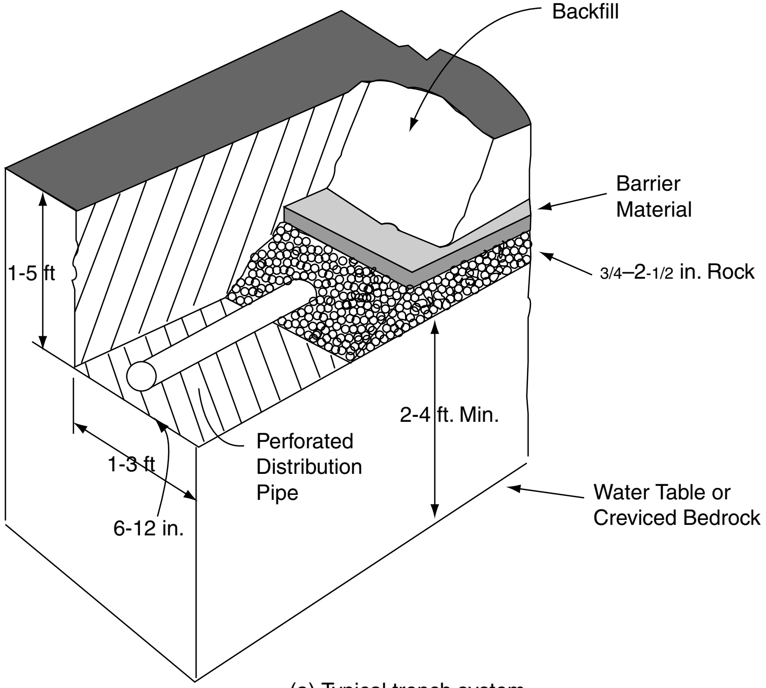
[*] Electro-osmosis is a technique used to drain and stabilize slowly permeable soils during excavation. A direct current is passed through the soil which draws the free water in the soil pores to the cathode.



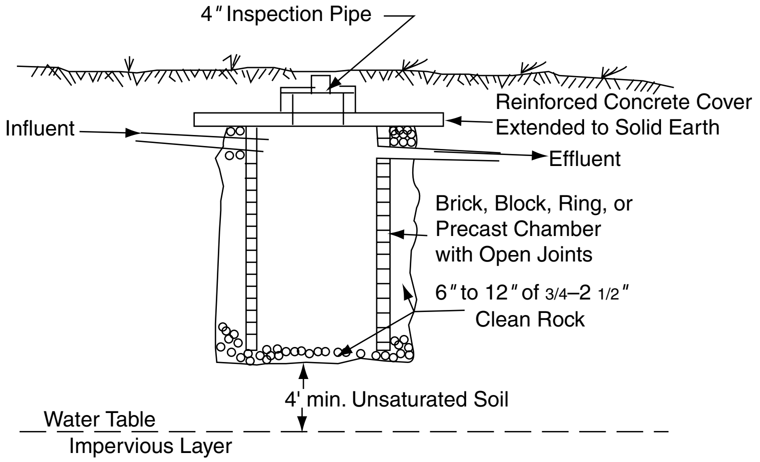
(a) Typical design of two-compartment septic tank

(b) Typical design of buried intermittent filter installation

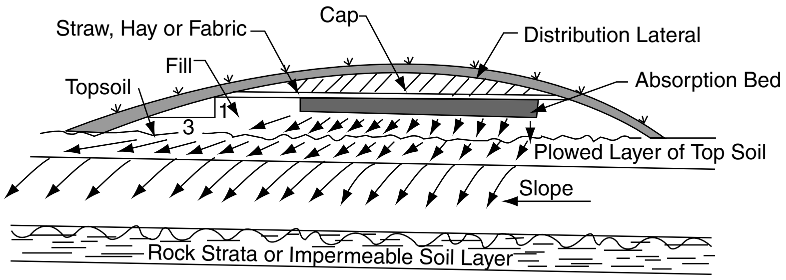
FIGURE 1 Typical design of septic tank and intermittent filter. (Adapted from Ref. 2.)



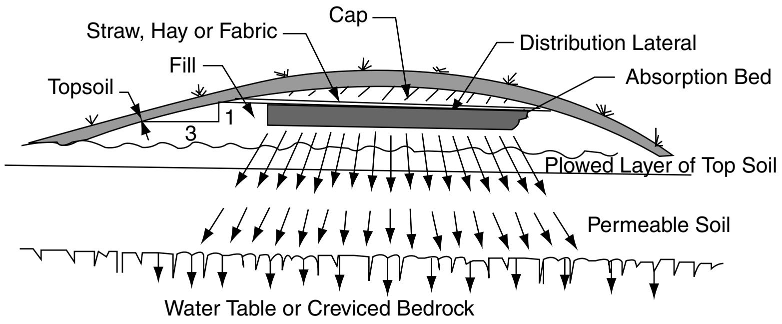
(a) Typical trench system



(b) Typical seepage pit system

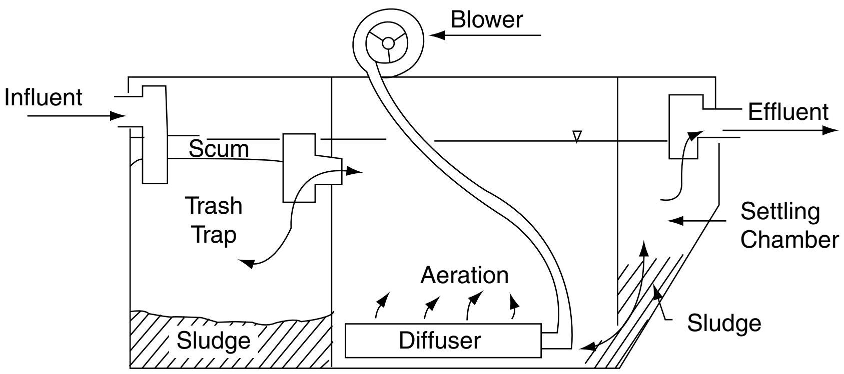


(c) Typical mound for a slowly permeable soil on sloping site

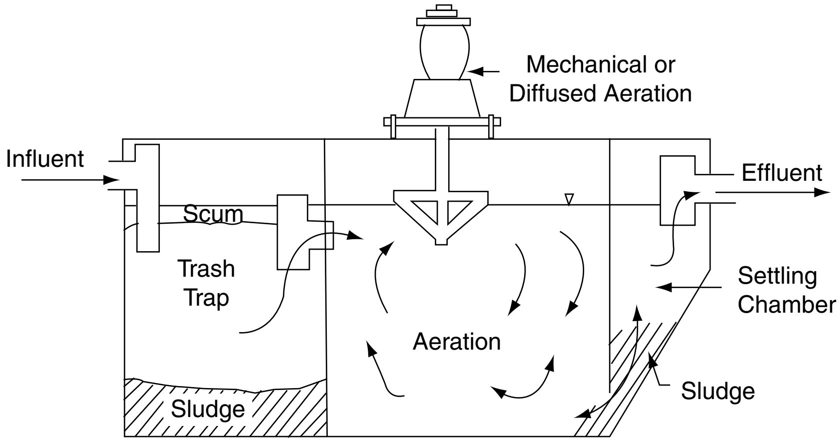


(d) Typical mound system for a permeable soil with high groundwater or shellow creviced bedrock

FIGURE 2 Subsurface absorption system. (Adapted from Ref. 2.)



(a) Diffused aeration



(b) Mechanical aeration

FIGURE 3 System components of aerobic suspended growth biological treatment process. (Adapted from Ref. 2.)

available. During earlier years of package plant usage, the plant size was usually limited to the maximum size that could be shipped by truck. This was about 190–228 m^3/d (50,000 to 60,000 gallons per day (gpd)) maximum plant size. Currently the manufacturers fabricate the plants in the field. Thus package plants with capacities over 3800 m^3/d (one million gallons a day (mgd)) can be obtained.

The earlier package plants were designed as extended aeration plants. Currently available package plants utilize many treatment processes. Some of these processes are listed below.

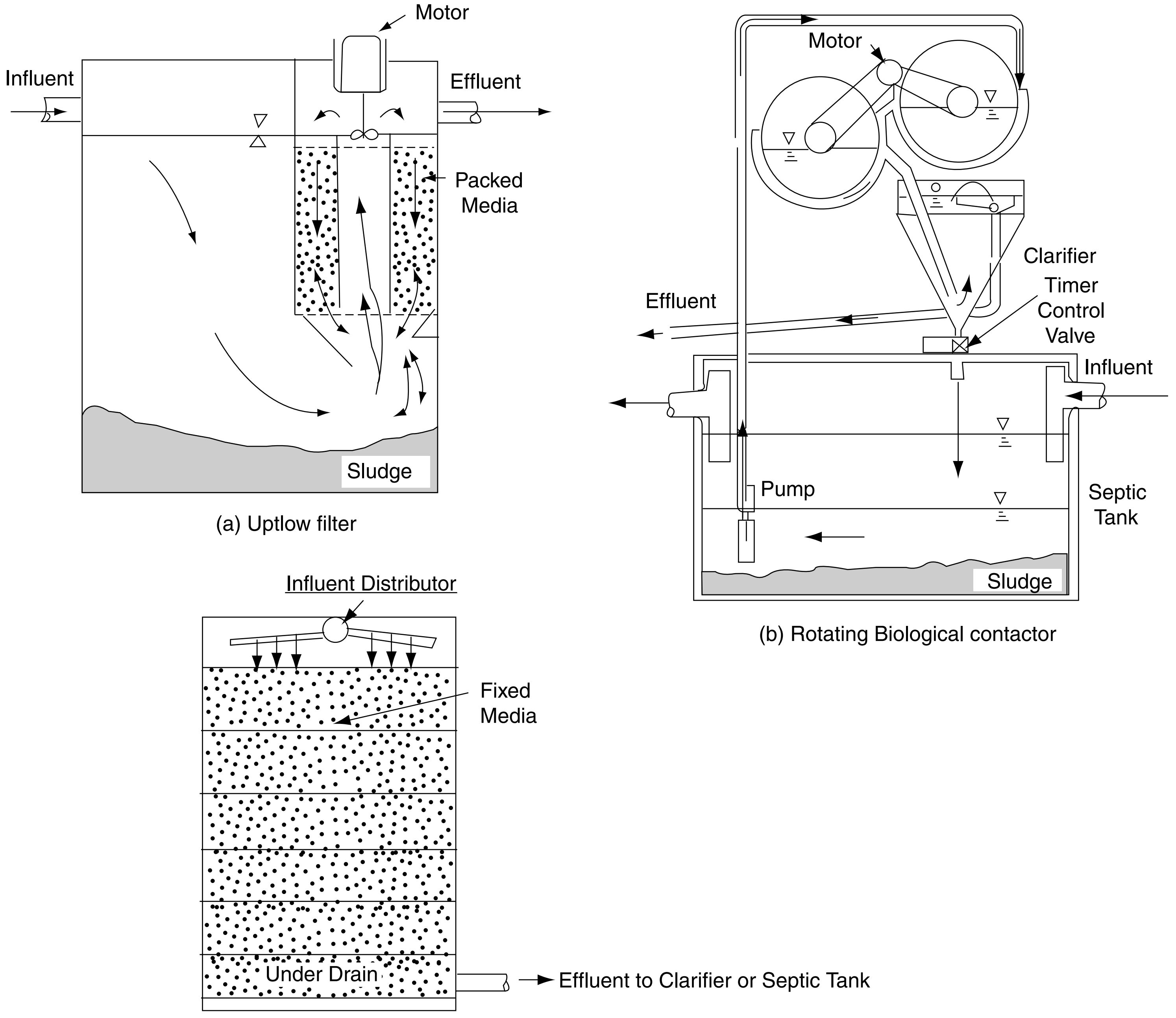
Biological

Extended aeration
Contact stabilization
Completely mixed
Step aeration
Trickling filter
Rotating biological contractor
Sequencing batch reactor (SBR)

Chemical

Chemical precipitation
Electrochemical flotation
Ultrafiltration

Detailed discussions on package plants including manufacturers, process alternatives, size, weight, design criteria, and cost is extensively available in the literature.[18–22] These plants have been successfully applied in the treatment of wastewaters in the suburban areas. The range of these plants has expanded from small establishments to large municipal and industrial applications. Manufacturers broadly group the models or design series into flow capacity, BOD loading, dimensions, air supply, motor horse power, etc., which has simplified the job of unit selection.



(a) Uptlow filter

(b) Rotating Biological contactor

(c) Tricking filter

FIGURE 4 System components of aerobic attached growth biological treatment process. (Adapted from Ref. 2.)

In recent years natural systems are also used for wastewater treatment from small establishments and sub-divisions. Among these are land treatment, and natural and constructed wetland. Information on natural system may be found in Refs. 1, 4, 15–17, 23 and 24.

POLAR REGIONS

Waste handling in polar regions is a complex problem. Conventional wastewater collection and treatment systems have severe limitations due to cold weather. Sewers constructed through snow, ice or permafrost[2] must be insulated to transport the wastes without freezing. Waste treatment by chemical or biological methods may not be possible due to retarded reaction rates.

The predominant sanitary facility used in small Alaskan villages is manual collection of human fecal waste and their disposal to ground, snow or ice. In many cases, pit privies, vaults, bored holes, straddle trenches, box and can, and crude chemical toilets are also used. Problems with these systems are: inconvenient to use especially in cold climates, unaesthetic features such as odor and unsightly conditions, and health hazards.[25,26]

Perhaps the most comprehensive analysis of waste problems in arctic areas was conducted by the Federal Housing Administration about 50 years ago.[27,28] Subsequently, a number of research programs were conducted by the US Public Health Service, National Research Council, US Navy, US Air Force, and US Army in developing suitable systems for use in Greenland, Alaska and most recently in the Antarctic.[29] Prime requirements for a system suitable for installation in these areas were: (1) minimum water use as year-round water supply in these areas may be lacking, (2) non-electric operation, (3) freeze-free, (4) minimum final disposal problems, (5) odor-free, and (6) minimum of maintenance. Systems investigated in these research programs included: (1) incinerating toilets, (2) chemical toilets, and (3) chemical and biological toilets with recirculation.

Incinerating Toilets

Incinerating toilets have been designed to destroy human body wastes. The thermal energy required may be obtained from electricity, fuel, oil, or liquified petroleum gas (LPG). Several designs of incinerating toilets are now commercially available. Although these toilets provide complete prevention of pollution from human wastes, they are relatively inefficient in terms of fuel consumption and often fail to provide complete burning which may result in a noxious odors or excessive smoke.[28] Two incinerating toilets *Incinolet* and *Stornburn* are manufactured for application in Alaska.[26]

Chemical and Composting Toilets

Chemical toilets require addition of chemicals into the waste storage tank. These chemicals liquify the fecal wastes, produce bactericidal effect and suppress fecal odors. Most commonly used chemicals are:[3] (1) halogens and their compounds, (2) coal-tar distillate (phenols and cresols), (3) heavy metals and their salts (zinc, copper and silver salts), (4) quaternary ammonium compounds, and (5) alkaline substances such as sodium and potassium hydroxide and lime. Various chemicals are sold in the market under different trade names.

One mechanical-flush chemical toilet was developed by Naval Engineering Laboratory, Port Huneme, California (US Patent No. 3,460,165) for use at remote Antarctic stations. This toilet was extensively tested by the military personnel and found satisfactory for polar applications.

A variety of composting toilets are on the market for application in cold region. These toilets use no water, and eventually produce compost that can be used on home garden. Three composting toilets named in the literature are *AlasCan,* the *Phoenix,* and the *Sun-Mar.*[26,29,30]

Chemical and Biological Toilets with Recirculation

A special application of waste treatment unit for polar use is a unit where treated effluent could be recirculated as a flush fluid. Such concepts with both chemical and biological treatment systems were investigated. Walters developed a recirculating chlorinator toilet which used standard toilet fittings.[31] The heavily chlorinated effluent from the storage tank was reused for flushing the toilet bowl. The system was tested in Alaskan single homes and results were found esthetically acceptable.

Another study evaluated three extended aeration plants in which the effluent was recirculated for toilet flushing.[32] The findings of this investigation showed that the flushing fluid turned brown during the first week and remained that color for the entire test period of several months. However, no odors were detected.

Both the chemical and biological treatment system with effluent recirculation have great promise in developing toilets for use in polar regions. These toilets have unique features such as: no water requirement, non-electrical (with hand pump), non-freezing, no odor, conventional toilet design, and a minimum of maintenance.

A number of innovative wastewater collection, treatment, reuse and disposal system have also been applied for community and individual home applications in polar region. Among these are pressurized and vacuum collection systems, low flush, ultra-low flush, micro-flush toilets, and mineral oil flush systems.[26,33–38]

WATERCRAFT

Use of the waterways for pleasure boating has increased enormously in this country. Because of a substantial increase in the number of recreational vessels, the public has become aware of the potential seriousness of the waterborne pollution resulting from this source. Many recreational watercraft and commercial and government vessels discharge wastes into the water.[34] Recreational watercraft are highly mobile, and may reach beaches, commercial fisheries, shell-fish growing areas and may seriously contaminate the waters, thereby rendering them dangerous for public water supplies and water contact sports. Contamination may also seriously affect the commercial fisheries and the shellfish industry.

The discharge in territorial waters of the United States is regulated by the Clean Water Act.[40] This Act also specifies allowable types of marine sanitation devices (MSD), and mandates, that the US Coast Guard test the various types of MSD and certify them for use aboard water craft. These tests are found in 33CFR 159, page 492, 500 and 501.[41,42] Type I MSD utilize macerator and disinfection. Both commercial disinfectants and onsite electrochemical devices are included. Type II MSD include biological treatment or fiber filtration (microscreening). Type III MSD consists of storage tank.

Sewage equipment for use aboard watercraft is already available in the form of (1) maceration-disinfection devices, (2) holding tanks and recirculating toilets, (3) incinerator devices, and (4) chemical and biological treatment facilities. A brief discussion of these systems is given below. For details on manufacturers, cost and unit variations, readers are referred to several sources in the literature.[43–49]

Maceration-Disinfection Devices

Maceration-disinfection devices utilize a mechanical macerator to grind the human fecal wastes, mix a disinfecting chemical (usually hypochlorite) and retain the disinfected sewage mixture for a brief period before discharging it into the water. A number of companies manufacture such units for installation aboard virtually every type of watercraft. Such units are small, lightweight, and relatively easy to install. However, their performance in terms of BOD and suspended solids reduction, and degree of disinfection achieved may be questionable.

Holding Tanks and Recirculating Toilet

A holding tank is a closed container for retaining sewage onboard a watercraft until it can be properly emptied, usually into an onshore sewage receiving facility. Holding tanks include chemical toilets, recirculating flush toilets, classic holding tanks, and any variation which simply retains the sewage for later disposal at an appropriate site.

One potentially useful variation of the holding tank is the recirculating flush toilet. This device requires a small amount of precharge of chemically treated water in the retention tank integrated into the toilet design. Waste deposited in this toilet accumulates in the retention tank. For subsequent flushing, an internal separation mechanism recovers a fraction of this precharge/waste mixture for continued reuse as flushing fluid. The tank retains sewage from 80 to 100 toilet usages before it must be emptied. It uses minimum space and requires no water for flushing.

Although holding tanks completely prevent the discharge of sewage from watercraft, they require extensive shore support facilities for emptying and cleaning. Low flush and vacuum flush toilets are desirable because they minimize storage and treatment requirement.[34]

Incinerator Devices

As described earlier, several types of incinerating toilets have been developed to reduce human waste to a small amount of ash. The most common problems encountered with incinerator devices with watercraft include the difficulty of supplying electrical power as most small craft do not have the generating equipment required. If gas or oil burner type incinerators are used, space is also required. Such burners also provide fire hazards if improperly designed. Regardless of the type of fuel used, however, burning of human wastes may result in emission of odor from the venting stack.

Chemical and Biological Treatment Plants

Wastewater treatment plants similar to those frequently used for land-base sewage treatment have been adopted for vessel use. The mot successful of the biological treatment systems are the extended aeration activated sludge process.[48] Attempts have been made also to adapt thermally heated aeration systems for increased biological activity. Trickling filter-type biological treatment systems with forced air aeration have also been adopted to shipboard applications.[18] These systems, which provide treatment to all waste streams generated aboard the watercraft, are relatively large and heavy, and easily upset due to change in salinity of flush water.

Among the chemical systems, an electrochemical floatation plant for shipboard waste treatment was designed and built.[49,50] This system utilized chlorine gas for disinfection, and partial oxidation and flotation of organic matter. Another chemical process oriented system utilized a comprehensive approach to the management of wastewater on board ship and is concerned not only with an improved treatment but also with an innovation and improved collection system. The system employs two main elements: (1) a recirculating chemical toilet, and (2) an evaporation system for solid/liquid separation.[31,51–53] Condensed liquid is discharged overboard after chlorination. Concentrated sludge is stored for subsequent disposal to shore facility or into unrestricted waters.

COMMERCIAL AIRCRAFT

Until about 1940, waste management problems aboard aircraft were considered minor; human wastes were simply discharged overboard. In January of 1943, the US Public Health Service published the "Sanitation Manual for Land and Air Conveyance Operating in Interstate Traffic," which formulated policy regarding "Discharge of Wastes from Conveyance En Route."[54] Following this publication, the International Sanitary Convention for Aerial Navigation issued a publication in February, 1945, forbidding aircraft to throw or let fall matter capable of producing an outbreak of infectious diseases.[55] Federal laws and regulations now prohibit air-planes indiscriminately discharging untreated human wastes.[56]

Early waste management practices for aircraft included carry-out pail methods of waste collection within aircraft and hand-carrying them to the ground servicing facilities.

Subsequently, this system was replaced by built-in retention tanks. These retention tanks utilized wash and galley water for toilet flushing purposes. The system posed several operational problems, including the fact that the flushing water taxed the capacity of the retention tanks on longer flights, and was later replaced by recirculating chemical toilets.

Recirculating chemical toilets currently being used aboard commercial aircraft have highly improved features to fulfill the requirements of the modern planes. Reinforced fiberglass retention tanks for lightness, teflon coated toilet bowls for cleanliness, timer assemblies to control the flush cycle, improved reversible-motor-operated pumps, and a number of filter and filter-cleaning devices have all been developed to provide trouble-free operation and an aesthetic facility.

The recirculating chemical toilets provide efficient operation but, depending on the capacity of the water tank, the system requires frequent ground servicing. Furthermore, the amount of space and weight available for waste storage in aircraft is quite limited. Therefore, a system to concentrate the wastes during the flight is considered highly desirable. To achieve this, several waste-volume-reduction techniques have been investigated for use aboard commercial aircraft. These include:

1) Evaporation of the liquid to yield dry or highly concentrated solids to reduce the waste-storage space in the aircraft and eliminate the frequent ground servicing need.
2) Incineration devices which utilize electrical and fuel energy for waste incineration. Several improved incineration systems for aircraft application have been built and evaluated.[57–59]
3) Evapo-combustion system to burn the macerated waste into a combustion chamber of the jet engine. Vacuum toilets have also been successively installed on large commercial aircraft. These toilets reduce the waste accumulation.[34, 60–64.]

RAILROAD TRAINS

Historically, wastes from railroad trains have been discharged into the environment without benefit of any treatment. This primitive practice poses a threat to the public health. Although passenger traffic on trains in the US has declined in recent years, large numbers of persons, including railroad employees still use toilet facilities on trains. According to Food and Drug Administration (FDA), about 23 million pounds of human excrement or 16 million gallons of wastewater are discharged annually from locomotives and cabooses, and about 9.5 million gallons of "untreated human wastes" were discharged in 1968 from intercity and commuter passenger train cars.[56] A history of some of waste disposal practices of Amtrak was presented in a hearing before Congress in 1988.[65]

Federal laws and regulations now prohibit buses (42 CFR 72.156) from discharging untreated human waste.[56,57] As a result, the passenger buses are equipped with suitable types of chemical or recirculating toilets.

Currently, several types of waste treatment and disposal systems are being marketed which are designed and built for railroad trains. These include incinerating toilets, retention tanks, and recirculating toilets.

Incinerating toilets built for railroad cars operate on natural gas, propane, diesel fuel, or electricity. These toilets operate without the use of water or chemicals and require no holding tanks or plumbing fixtures.

Recirculating toilets of various types are also available for railroad use. One system built for locomotives, cabooses and crew cars uses a vacuum system. In this system air, rather than water, is used to carry waste from the toilets to a centrally located tank.[56,61] This system enables locating the holding tank elsewhere in the railroad car and two or more toilets can be connected to this tank.

PICKUP CAMPERS, TRAVEL-TRAILERS, TENT CAMPING

Various types of portable recirculating toilets are currently manufactured. These units have suitcase-style handles molded into their cases for easy carrying; can be used in tents or in camper.[66] A small family can get a few day's use before the facility must be emptied and recharged.

A unique system for reducing the volume of wastes from recirculating toilets was developed. In this system, the fecal wastes were liquefied by adding chemicals. The liquid mixture is pumped to a sanitizer which is a short, stainless steel tank connected to the exhaust pipe of the vehicle. The sanitizer operates at about 500°F. At this operating temperature, the waste is concentrated and the microorganisms are destroyed.[66] The operating temperature of the sanitizer is reached at a vehicle speed of about 35 mph.

ENVIRONMENTAL CONTROL AND LIFE SUPPORT SYSTEMS (ECLSS) FOR SPACE STATION

NASA has sponsored programs to develop efficient, compact equipment to handle the various aspect of environmental life support for spacecraft and for the planned space station. The tasks include CO_2 removal, O_2 regeneration, temperature and humidity control, the purification of water recovered from the dehumidifier condensate, hygiene uses, and in the future, from urine. Also the removal of trace contaminants from the air, the maintenance of the air composition and pressure, and the storage of solid wastes pending their return to earth are included. A wide variety of techniques have been evaluated depending upon their prospects for meeting the desired performance specifications.[67–73] Table 1 provides a list of ECLSS technologies used or evaluated.

The space systems have grown in complexity and comprehensiveness as both the duration of the missions and the size of the crew have increased. With the possibility of long duration space missions to other planets, and also the establishment of bases on the moon, NASA is in the early stages of testing technologies for solid waste treatment and recycling, and the intensive agricultural technique necessary to grow and process food in very confined spaces under low or zero gravity.

Some of the technologies developed for the space program have possible application on earth. Many of the applications will only be relevant to submarines and hyperbaric chambers. In some cases there may be more widespread uses, such as the recovery and reuse of dehumidification and hygiene water in arid areas or in very cold climates, removing CO_2 and regenerating O_2, and controlling temperature and humidity in deep mines or while drilling long tunnels. Certain technologies may be applicable to the treatment of industrial emissions and/or effluents. It is likely that the basic technological knowledge will be applied to terrestrial problems, rather than the actual hardware developed for the space program.[74–77] Possible terrestrial application of space craft environmental systems are presented in Table 2.

TABLE 1
ECLSS technologies used or evaluated

ECLSS subsystem category	Used/Evaluated	Technology
Atmosphere revitalisation	Used	LiOH
	Used	Molecular sieve
	Used	Sabatier reactor
	Used	Static feed water electrolysis
	Evaluated	Solid amine fixed bed
	Evaluated	Liquid sorbent closed loop
	Evaluated	Bosch system
	Evaluated	Algal bioreactor
	Evaluated	Growing green plants
Trace contaminant removal	Used	Activated charcoal
	Used	Catalytic oxidiser
	Used	Particulate filters
Water recovery and management	Used	Vapor compression distillation
	Used	Chlorine
	Used	Sodium hypochlorite injection
	Used	Iodine injection
	Used	Heat sterilisation
	Used	Fuel cell byproduct water
	Evaluated	Unibed filter
	Evaluated	TIMES membrane filter
	Evaluated	Reverse osmosis
	Evaluated	Electrodialysis
	Evaluated	Electrooxidation
	Evaluated	Supercritical water oxidation
	Evaluated	Electrodeionisation
	Evaluated	Air evaporation
	Evaluated	Vapor phase catalytic ammonia removal
	Evaluated	Immobilised cell or enzyme bioreactors
	Evaluated	Plant transpiration and water recovery
Temperature and humidity control	Used	Condensing heat exchangers
	Used	Water cooled suits
Atmosphere control and supply	Used	Compressed gas storage
	Used	Cryogenic gas storage
Waste management	Used	Urine stored in bags
	Used	Feces stored in bags
	Used	Urine vented
	Used	Feces stored in bags and vacuum dried
	Used	Urine stored in tank and vented
	Used	Feces stored in bags and compacted

TABLE 2
Possible terrestrial applications of spacecraft environmental systems

ECLS systems	Possible earthbound uses
Water reclamation	Submarines, arid area operations, supply of filtered and sterile water for medical, experimental uses. Cleaning industrial effluent.
Oxygen generation	Submarines, medical use in oxygen enriched atmosphere, hyperbaric chambers, under water habitats and deep undergroundwork locations.
CO_2 removal	Submarines, rescue and scuba equipment, under water habitats, hyperbaric chambers. Cleaning industrial emmissions. Revitalising atmosphere in deep underground work locations.
Trace contaminant removal	Submarines, under water habitats, hyperbaric chambers.
Temperature, humidity control	Passive systems developed for the control of the space station temperature gain and loss from and to space itself may have some terrestrial applications. Active systems used in controlling temperature in locations where there are localised sources of heat may be useful for cooling dense electronics, in satellites, possibly in aircraft, and perhaps in advanced super computers.

A joint NASA and NSF life support system utilizing some of the water recovery, waste treatment, plant growth, and energy efficiency technologies is to be installed at the US research station at the South Pole. This will provide a real world opportunity to use the planned technologies on a realistic scale.[78]

REFERENCES

1. Sauter, G. and K. Leonard, Natural Home Remedy, *Water Environment & Technology, Water Environment Federation,* **7,** No. 8, pp. 48–52 (August 1995).
2. US Environmental Protection Agency, *Design manual, Onsite wastewater treatment and disposal systems,* Office of water Program Operations, Municipal Environmental Research Laboratory, Cincinnati, Ohio, EPA 625/1-800-012, October 1980.
3. US Environmental Protection Agency, *Handbook of Septage Treatment and Disposal,* USEPA 625/6-84-009, Center for Environmental Research Information, Cincinnati, OH, October 1984.
4. Qasim, S.R., *Wastewater Treatment Plants: Planning, Design and Operation,* Technomic Publishing Co., Lancaster, PA (1994).
5. *Manual of septic-tank practice,* USPHA, Publication No. 526, US Department of the Interior, Washington D.C. (1967).
6. Coulter, J.B., The septic tank system in suburbia, *Pub. Health Rept.,* **73,** 488 (1958).
7. Winneberger, J.H., *et al.*, *Biological aspects of failure of septic tank percolation systems,* Final report, Sanitary Engineering Research Laboratory, University of California, Berkeley (August 1969).
8. Laak, Rein., Wastewater Engineering Design for Unsewered Areas, Second edition. *Technomic Publishing Co.*, Lancaster. PA. (1986).
9. Andreadakis, A.D., Organic matter and nitrogen removal by an on-site sewage treatment and disposal system, *Water Research* (G.B.), **21,** 559 (1987).
10. Scalf, M.R. *et al., Environmental effects of septic tank systems,* US Environmental Protection Agency, Ada, Oklahoma, EPA-600/3-77-096, 1977.
11. Sharp, W.E. *et al.,* Restoration of Failing On-site Wastewater Disposal Systems Using Water Conservation, *Journal Water Pollution Control Federation,* **56,** No. 7, pp. 858, (July 1985).
12. Reid and Leroy C. Jr., Design of wastewater disposal systems for individual dwellings, *Journal Water Pollution Control Federation,* **43,** No. 10, pp. 2004–2010 (October 1971).
13. Report on *Individual household aerobic sewage treatment systems,* National Research Council–National Academy of Science, Publication No. 586 (February 1958).
14. Thomas, H.A. Jr. *et al.*, Technology and economics of household sewage disposal systems, *Journal Water Pollution Control Federation,* **32,** No. 2, p. 113 (February 1960).
15. U.S. Environmental Protection Agency, *Design Manual on Constructed Wetland and Aquatic Plant System for Municipal Wastewater Treatment Plants,* Office of Research and Development, Center for Environmental Research Information, Cincinnati, Ohio, (September 1988).
16. U.S. Environmental Protection Agency, *Subsurface Flow Constructed Wetland for Wastewater Treatment–A Technology Assessment,* Office of Water, EPA 832-R-93-00, (July 1993).
17. U.S. Environmental Protection Agency, Subsurface Flow Constructed Wetlands, Proceedings of the *Subsurface flow Constructed Wetlands Conference,* August 16 and 17, 1993, University of Texas at EL Paso, (1993).
18. Qasim, S.R., N.L. Drobny, and B.W. Valentine, Process alternatives available in packalaged wastewater treatment plants, *Water and Sewage Works,* **177,** Reference Number, pp. R-99 (November 1970).
19. Mahoney, J.A., *Summary of commercial wastewater treatment plants,* report No. AD 860 067, Air Force Systems Command, Kirtland Air Force Base, New Mexico, (September 1969).
20. Seymour, G.G., Operation and performance of package treatment plants, *Journal Water Pollution Control Federation,* **44,** No. 2, pp. 274 (February 1972).
21. Defne, R.R., G.E. Elbert, and M.D. Rockwell, Contact stabilization in small package plants, *Journal Water Pollution Control Federation,* **44,** No. 2, p. 255 (February 1972).
22. Sancier, J.W., The application or package wastewater treatment plants to suburban areas, *Water and Wastes Engineering,* p. 33. (December 1969).
23. Water Pollution Control Federation, Natural Systems for Wastewater Treatment, Manual of Practice F9–16, *Water Pollution Control Federation,* Washington D.C. (1990).
24. Dinger, R., *Natural Systems for Water Pollution Control,* Van Nostrand Reinhold Co. New York, (1982).
25. Mellor, M., *Utility on Permanent Snowfields,* Monograph III-A2d, U.S. Army Corps of Engineers, Cold Regions Research Laboratory, Hanover, NH (1969).
26. An Alaskan Challenge: *Nature Village Sanitation,* US Department of Housing and Human Services, Y3.T 22/2:2AL 1S, (1994).
27. Alter, A.J., *Arctic sanitary engineering,* Federal Housing Administration, Washington, DC (1950).
28. Clark and Groff Engineers, *Sanitary waste disposal for navy camps in polar regions,* final report to Naval Civil Engineering Laboratory on Contract NBy-32205 (1962).
29. A New Device for Cold Region Sanitation Disposal—the Alas-Can, *The Northern Engineer,* **22,** No. 1, (Spring 1990).
30. Product Literature, Lehman Hardware and Appliances, Inc., Kidron, Ohio (1995).
31. Walters and Charles F., Promise disposal system, *Water and Sewage Works* (August 1962).
32. Walters, Charles F. and J.A. Anderegg, *Water conservation through reuse of flushing liquid in an aerobic sewage treatment process,* paper presented at 11th Alaska Science Conference (August 20, 1960).
33. Kouric, R., Toilets: Low Flush/No Flush, *Garbage: The Practical Journal for the Environment,* **2,** No. 1, pp. 16–24, (January 1990).
34. Water-wise Toilets, *Technology Review,* **97,** (1994).

35. Pollen, M.R. and D.W. Smith, Water Conservation in Borrow Alaska, Proceedings of the Third Symposium on Utilities Delivery in Cold Regions, University of Alberta, Edmonton, Alberta, Canada (1982).
36. Viraraghavan, T. and G. Mathavan, Effect of Low Temperature on Physicochemical Processes in Water Quality Control, *Journal of Cold Regions Engineering,* American Society of Civil Engineers, **2,** No. 3, (1988).
37. U.S. Environmental Protection Agency, *Water Related Utilities for Small Communities in Alaska,* EPA-600/3-76-104, Environmental Research Laboratory, Office of Research and Development. Corrallis, Oregon, (1976).
38. Bowen, P.T. and J.C. Lan, Wastewater Treatment of the South Pole, *Water Environment and Technology,* **5,** No. 3, (March 1993).
39. Federal Water Pollution Control Administration, Wastes from watercraft, US Senate, 90th Congress 1st Session, Document No. 48, *US Government Printing Office,* Washington, DC (August 7, 1967).
40. *The Clean Water Act Amendments of 1981,* PL 97–117, Title III, Sec. 311.(b)(1).
41. 33 CFR, Part 159, *Marine Sanitation Devices,* (1994).
42. Wachinski, A.M. and M. Knofczynski, Managing Shipboard Wastes, *Water Environment and Technology, Water Environment Federation,* **4,** pp. 58–62 (August 1992).
43. Hertzbery, R.H., Waste disposal from watercraft, *Journal Water Pollution Control Federation,* **40,** No. 12, p. 2055 (December 1968).
44. Oertle, Lee, Plan a head, monomatic self-contained recirculating head, *Popular Mechanics,* p. 140 (December 1968).
45. Manning, A.P., Headspace and horse sense on chlorinator, *Motor Boating,* pp. 42–47 (March 1968).
46. Dugan, P.R. and F.M. Pfister, *Evaluation of marine toilet chlorinator units,* New York State Department of Health, Project No. GL-WP-1 (July 1962).
47. Yacht Safety Bureau, Inc., *Requirements of marine sewage incinerating devices* (February 1968).
48. Stewart, M.J. and H.F. Ludwig, Shipboard extended-aeration wastewater treatment plant design, *Naval Engineering Journal,* **109,** pp. 985 (December 1963).
49. Singerman, H.H. and E.T. Kinney, The Navy's technological progress in water pollution abatement, *Naval Engineers Journal,* pp. 795 (October 1968).
50. Bockris, J.O. and Z. Nagy, Electrochemistry for Ecologist, *Plenum Press,* New York (1974).
51. Qasim, S.R. and N.L. Drobny, *Development of a waste treatment system for naval vessels phase I-laboratory studies,* research report, Battelle Memorial Institute, Columbus Laboratories (October 1970).
52. Droby, N.L., S.R. Qasim, and A. Cornish, *Waste management systems for naval vessels,* proceedings of the 17th Annual Meeting and Equipment Exposition, Institute of Environmental Sciences, Los Angeles, California (April 1971).
53. Joungs and Jim, *et al.*, Accessory Guide, *Trailer Boats,* pp. 70–78, (April 1995).
54. Sanitation manual for land and air conveyance operation in interstate traffic, Public Health Service report, **58,** No. 5, pp. 157–193 (January 29, 1943).
55. *Handbook on sanitation of airlines,* US Department of Health, Education and Welfare, Public Health Service Publication No. 308, Vol. 31, Washington, DC (1953).
56. *Environmental pollution: Discharge of raw human wastes from railroad trains,* 32nd report by the Committee on Government Operations, Washington, DC (October 8, 1970).
57. U.S. Congress, *Discharge of Waste on Air and Highway Conveyances,* Public Law, 21 CFR, 40 CFR 56524, February 1975, and as amended 48CFR 11432, March 18, 1983.
58. Fredric, E.W. and B. Mangus, *Waterless toilet,* US Patent 3, 110,037 (C14-131) (November 1963).
59. Krukeberg, C.W. and R.J. Jauch, Assignors to Tokheium corporation, Indiana, *Incinerating Toilet,* US Patent 2,995,097 (C1 110-9) (August 1961).
60. Duncan, Leon, L., Assignor to William J. King, Palos Park, III., *Total disposal unit and method,* US Patent, 3, 319,588 (C1 110-9) (May 1967).
61. Watkins, A.M., Sanivac-Revolutionary vacuum toilets, *Popular Science,* p. 86 (July 1970).
62. Kourik, R., Toilets: The Low-flush/No Flush Story, *Garbage,* p. 16, (January/February 1990).
63. Research Products/Blankenship, Incinolet: *That Electric Toilet,* Research Products/Blankenship (August 1994).
64. *Alternative Toilets,* National Small Flows Clearing House, US Environmental Protection Agency, West Virginia University, Morgantown, W.Va. (March 1994).
65. Committee on Government Operations, Government Activities and Transportation Subcommittee, *Dumping of Human Waste from Amtrak Trains,* Hearing Before a Subcommittee of the Committee on Government Operation, House of Representatives, 100th Conferee, Second Session, Washington, D.C., September 27, 1988.
66. For the travelling family: Take along comfort station, *Popular Science,* p. 99 (August 1970).
67. Hightower, T.M., *Recycling and Source Reduction for Long Duration space Habitation,* Proceedings of the 22nd International Conference on Environmental Systems, Seattle, WA, July 13–16, 1992. *Society of Automotive Engineers,* Warrendale, PA (1992).
68. Wieland, P.O. *Designing for Human Presence in Space: An Introduction to Environmental Control and Life Support Systems.* George C. Marshall Space Flight Center, NASA RP-1324 (1994).
69. Ramananathan, R., J.E. Straub, and J.R. Shultz. *Water Quality Program Elements for Space Station Freedom,* Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991. *Society of Automotive Engineers,* Warrendale, PA, pp. 1–21 (1991).
70. McElroy, J.F., T.M. Molter, and R.J. Roy, *SPE Water Electrolyzers for Closed Environment Life Support,* Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991. *Society of Automotive Engineers,* Warrendale, PA, pp. 261–270. (1991).
71. Howard, S.G. and J.H. Miernik, *An Analysis of Urine Pretreatment Methods for Use on Space Station Freedom,* Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991. *Society of Automotive Engineers,* Warrendale, PA, pp. 157–166 (1991).
72. Herrman, C.C. and T. Wydeven, *Physical/Chemical Closed-loop Water Recycling for Long Duration Missions,* Proceedings of the 21st Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990. *Society of Automotive Engineers,* Warrendale, PA, pp. 233–245(1990).
73. Miernik, J.H., B.H. Shah, and C.F. McGriff. *Waste Water Processing Technology for space Station Freedom: Comparative Test Data Analysis,* Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991. *Society of Automotive Engineers,* Warrendale, PA, pp. 229–239 (1991).
74. Rossier, R.N., *Nuclear Powered Submarines and the Space Station: A Comparison of ECLSS Requirement,* Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986. *Society of Automotive Engineers,* Warrendale, PA, pp. 321–329 (1986).
75. Straight, C.L. *et al.*, *The CELSS Antarctic Analog Project: A Validation of ECLSS Methodologies at the South Pole Station,* In Proceedings of the 23rd Intersociety Conference on Environmental Systems, Colorado Springs, CO, July 12–15, 1993. *Society of Automotive Engineers,* Warrendale, PA, pp. 1–11 (1991).
76. Sribnik, F., R.C. Augusti, and E.G. Glastris, *Life Support System for a Physically Isolated Underground Habitat.* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990. *Society of Automotive Engineers,* Warrendale, PA, pp. 1–11 (1990).
77. Dempster, W.F., Biosphere II, Engineering of Manned Closed Ecological Systems. *Journal of Aerospace Engineering,* **4,** No. 1, Jan 1991, pp. 23–30.
78. Flynn, M.T., C. Straight, and D. Bubenheim, *Development of an Advanced Life Support Testbed at the Amundsen-Scott South Pole Station.* Preprint.

SYED R. QASIM
The University of Texas at Arlington

STACK SAMPLING

INTRODUCTION

It is frequently necessary to determine the amount, concentration, or rate of emission of various pollutants in the exhaust streams from industrial or commercial processes. Because this generally involves the sampling and eventual analysis of the gas flowing through a stack into the outside air, it is usually called “stack emission sampling”, or more simply, “stack sampling”.

In most cases, it is not possible or practical to collect all of the gases emitted to the outside air over any reasonable time period. Therefore, it is necessary to collect only a fraction of the overall gas stream. A representative side stream is isolated from the main flow (usually by removing it from the stack altogether) and processed in some way. The gas stream may be filtered, condensed, bubbled, adsorbed, bottled, or pumped through an automatic analyzer. The equipment used for this purpose is called the sampling train. The end result of this step is usually an assessment of the contents of the stream. Meanwhile, an assessment is made about the other characteristics of the stack gas itself, such as temperature and flow rate. The information from these two assessments is then combined to produce a measure of the emissions from the stack in the desired units, such as pounds per hour, grains/dry standard cubic foot, kg/kg of fuel burned, and so on. Generally, the units are chosen to conform with applicable regulations.

Of course, the testing must include all of the pollutants that are of interest, under all of the process conditions that are needed. This must also include a wide range of checks and balances, often termed Quality Control (QC) and Quality Assurance (QA) to ensure and measure the reliability of the results. Most importantly, it all must be based on a well reasoned plan, called a Stack Test Protocol or Quality Assurance Project Plan (QAPP) that is written ahead of time and approved by all interested parties. Finally, these results must be reported in a way that fairly represents what was done, to allow regulators and/or the sources to make informed decisions.

A prime criterion for method selection is that the method must produce data useful for the purpose intended. For example, is the detection limit of the method low enough to prove compliance with the emission standard, or is the method able to measure and/or account for cyclonic flow in the stack, or is it able to differentiate between similar chemical species, or is that necessary? These sorts of questions fall into a general category called QA. They can be restated as six data quality parameters:

P	Precision	Repeatability
A	Accuracy	Bias, closeness to “correct”
R	Representativeness	Typical of actual stack gas
C	Comparability	Similar to other data
C	Completeness	Enough information
S	Sensitivity	Low enough detection levels

These so-called PARCCS Parameters are simply ways to ensure that the results of a sampling project will yield useful results. When methods published by EPA are used, the PARCCS parameters have already been determined and are built into the methods. For novel methods, the PARCCS parameters must be determined, at least qualitatively. This is well beyond the scope of this chapter but must be borne in mind should new methodology need to be developed to fit a specific circumstance.

This leads to the most important warning concerning stack testing. In all but the most dire emergencies, stack testing projects should be planned and carried out by trained, experienced stack testing teams. No one should believe that even a close reading of this chapter would provide sufficient background to plan or perform stack tests.

There are four recent advances in the stack testing field that rate special mention. The first is the increasing reliance on external quality assurance, as embodied in audit samples and devices. Several agencies and organizations, particularly the Emissions, Monitoring and Analysis Division in USEPA’s North Carolina facility, are now producing reliable Performance Evaluation (PE) audit samples that can be obtained for the purpose of checking analytical accuracy at the time of sample analyses. They are available for many parameters, as listed in the specific Test Methods.

The second major recent advance in stack testing is the development and proliferation of reliable continuous emission monitoring systems (CEMS). CEMS are generally electronic analyzers that determine and record instantaneous concentrations of given parameters continuously.

For example, a CEMS for stack gas opacity, termed a transmissometer, shines a light beam through the stack gas and measures the fraction of light transmitted, taking a reading every few seconds. A CEMS for SO_2 extracts a small sample of stack gas and sends it to a monitor outside the stack for analysis and recording. The CEMS records present an excellent picture of the continuing status of compliance of the stack gas, reveal any inconsistencies, failures of control devices, or process upsets, and allow plant operators to act immediately when any anomaly appears. It is likely that more and more CEMS will be required as they become available. CEMS are becoming more available for mercury, particulate matter and ammonia.

While detailed discussion of CEMS technology, procedures, and operation cannot be presented here, the Performance Specifications for the CEMs currently required by USEPA may be found in Appendix B to 40 CFR Part 60, immediately following the Test Methods.

A third major advance in stack testing is the introduction of the 300 series methods found in 40 CFR Part 63 Appendix A. In an effort to measure the hazardous air pollutants to demonstrate emission reductions pursuant to MACT requirements, EPA has introduced a number of 300 series methods. Some of these methods are specific for pollutants from specific sources. Method 301, which is undergoing revision in 2005, is a field validation procedure that will enable source owners to validate their own test methods in the absence of a recognized EPA method.

The fourth is the rewrite of the manual methods into a standard format. The instrumental methods are being rewritten in 2005, along with some major changes.

The following sections of this chapter will deal with the selection of sampling and analytical methods, followed by some general information on protocol and final report preparation.

Selection of a Sampling Method

The choice of stack sampling methodology is most strongly dependent on the pollutants to be measured. In many cases, similar pollutants can be measured with the same or slightly modified methods, while dissimilar pollutants may require totally different methods.

The most obvious categorization of pollutants is based on their physical state: gaseous, liquid, or solid. Gaseous pollutants are generally easier to sample and can be collected using one of a few simple train configurations. Liquid and solid pollutants, usually lumped into a single category called particulates, require a totally different collection concept.

The following sections will describe the generalized methods employed for collecting particulate and gaseous stack samples. These are followed by more detailed descriptions of methods to be used for specific compounds, or groups of compounds. When the U.S. Environmental Protection Agency (EPA) has designated a method as a Test Method, it is required by EPA and by most states for compliance determinations. For convenience, the appropriate EPA designations are indicated. When a Test Method is used in establishing an emission limit and is specified by a regulation that method is called the Reference Method.

Particulate Sampling

When the pollutant of interest is or is attached to solid particles or liquid droplets at stack conditions, it is necessary to select a method that physically traps the particles. But the first step must be the selection of a side stream that is truly representative of the stack exhaust gases.

A representative sample of the stack exhaust gas will look and behave like a small-scale version of the actual exhaust gas. It will contain the same fraction of particulates as the main stream (including the same ratios of large and small particles) and will contain a fair share of material from each part of the stack cross section. (This is necessary because gases flow faster near the center of a stack and slower near the walls due to friction).

In addition a representative sample must be taken at a location that is free of unusual flow patterns, such as cyclonic flow (in which a significant component of the flow is not along the axis of the stack) or stratified flow (in which the particulates are bunched along one side of the stack). This is because it is very difficult to figure out the actual average flow rate and particulate rate when the measurements are all skewed by the flow anomalies. In most cases, flow abnormalities are caused by recognizable disturbances, such as bends, fans, expansions, contractions, or shape changes in the duct. These disturbances, whether upstream or downstream from the sampling location, have the potential for making useful testing very difficult, or even impossible. For that reason, the first criterion for good particulate testing is to find a location that is sufficiently far from flow disturbances. Extensive testing has shown that a sampling location 8 stack diameters downstream from any disturbance and 2 diameters upstream from any disturbance is sufficiently far. In this measurement, the term stack diameter is used literally for circular stacks. For rectangular stacks, an equivalent stack diameter is calculated.

In some cases, it is impossible to find a location in the stack or in any straight duct leading to the stack that satisfies these criteria. It is possible to use a location closer to disturbances. However, other provisions must be taken to account for the possible inaccuracies introduced by the disturbed flow. All of this is described in detail in Test Method 1.

Once a sampling location is selected, it is necessary to collect samples of the gas stream that are representative of the gas flowing by that location. This is achieved by sampling for a short time at each of several points across the stack cross-section.

In practice, two or more holes, or ports, are cut in the stack wall and a sampling probe inserted. The probe is essentially a hollow tube shaped like a shepherd's crook with the short end, or nozzle, facing into the gas stream. The gas stream is then pumped by suction from the main stream through the nozzle and probe into the collection part of the sampling train located outside the stack. The probe is held in one spot, aligned into the main stream, for a specified time. It is then moved to another point and held for the same time. This process is repeated along the line between the port and the opposite wall. The process of moving the probe along this line is called a traverse, and the individual sampling points termed traverse points. The configuration of the ports and the traverse points is generally chosen according to Test Method 1.

Three additional considerations must be addressed in order for the sample to be considered representative. First, the particulate-laden gas stream entering the nozzle must be typical of the stream flowing by it. Second, the makeup of the gas stream leaving the probe and entering the rest of the sampling train must be substantially the same as it was when it entered the nozzle. Third, sampling must be conducted at a time and for sufficient duration to cover any inconsistencies in the pollution emission rate.

The first of these conditions may sound like overkill. However, the previous work was to ensure the representativeness of the location. This part concerns the representativeness of the gas collected there. This is ensured by careful design of the nozzle and control o the side stream flow rate. The opening of the nozzle is designed with sharply tapered edges. The nozzle itself is shaped to minimize deposition of particulates on the inside walls as the stream turns 90°.

The sampling stream flow rate is extremely important because of the difference in aerodynamics and inertial effects of particles. Very small particles tend to behave like gas molecules and tend to follow gas flow stream lines. For these particles, sample flow rates are not critical. Large particles, however, do not necessarily follow the gas flow streamlines. Instead, their flow is controlled more by their inertia. In other words, they tend to keep going in straight lines. Thus, if the flow rate into the sampling nozzles is different than the local gas flow rate, the gas itself and the fine particles will be skewed, either into or out of the nozzle, depending on the relative rates. The large particles, however, will continue along their straight paths. Those, and only those, in direct line with the nozzle face will enter. This can have a significant effect on the measured particulate concentrations, depending on the degrees of error in the nozzle flow rate and on the fraction of particulate mass attributable to the large particles.

Sampling at exactly the right flow rate is termed isokinetic sampling. Sampling at too great a velocity is called superisokinetic, while sampling at too low a velocity is called subisokinetic. Generally, superisokinetic sampling results in an underestimation of the actual particulate concentration (termed a low bias), while subisokinetic sampling results in an overestimation (high bias). Test Method 5 contains instructions for choosing the appropriate nozzle size and sampling flow rate to ensure isokinetic sampling. Sample flow rate and nozzle size are based on the volume of sample gas that needs to be collected and on the flow rate in the stack gas. The needed sample volume is based on the amount of particulate needed for the physical or chemical analyses to be conducted, and will be discussed in the analysis section. The stack gas flow rate is determined according to procedures described in Test Method 2. The procedure involves the measurement of linear flow rate by means of the relationship between the static and dynamic pressure in the gas stream. The static pressure is the pressure of the gas stream, as measured by a pressure tap perpendicular to the flow. The dynamic pressure is the pressure exerted by the flowing stream and is indicative of the flow velocity. It is measured by a pressure tap facing directly into the flow stream. In practice, a device called an “S-type pitot tube” is used to measure static and dynamic pressure at a single location. Standard calculations are then used to compute the flow rate. By moving the pitot tube across a stack cross section, the flow rate at each point can be determined. All of this is described in detail in Test Method 2.

Next, it is necessary to assure that the sample gas stream does not change substantially between the time it enters the nozzle and leaves the probe for the collection part of the sampling train. This is accomplished by ensuring that the construction and operation of the probe do not interfere physically or chemically with the flowing sample stream. The nozzle and probe must be made of materials, such as stainless steel, glass, or teflon, that are smooth and do not react with the stream. In addition, the probe may be heated to ensure that vapors in the stream do not condense on the walls of the probe. A stainless steel nozzle and a glass-lined probe heated to 120 ± 14°C (248 ± 25°F) will suffice for most situations. However, special construction materials and/or probe temperature settings may be required for sampling exhaust streams from certain source types of containing certain contaminants. The appropriate sections of this chapter should be consulted in detail before any decisions are made.

The third and final consideration for ensuring representative particulate sampling is that the sampling is conducted at a time and over a sufficient time period to account for variabilities in the exhaust steam. In general, it is desirable to measure the maximum possible emissions, so as to determine compliance under so-called worst case conditions. This is accomplished by first determining the stage or stages in the plant process mot likely to produce the greatest emission rate. This might require preliminary testing, or it may be specified in the applicable regulations.

Once the time for testing is selected, it is necessary to decide the duration of each sample run and the number of runs to be performed. Generally, this is specified in the applicable regulation or in the Test Methods. However, in some cases, it may be necessary to select different sampling periods. The basic reason for sampling for any given time is to account for temporal variability in the emissions. Very few processes produce emission streams that are truly constant over more than a few minutes at a time. In most cases, though, an hour or two-hour sampling period will be sufficient to smooth out any inherent variation in the exhaust stream. In practical terms, this is accomplished by pumping the sample stream through the filter, solutions, or sorbents for the full sampling period. This effectively averages the collected sample over the entire time period. Of course, it is possible that variations in the emission rate with time could mean that higher concentrations are measured at one point in a traverse, such as near one stack wall, etc. However, in most cases, these variations are not significant.

The individual regulations or Test methods usually state the required sampling duration. There are two reasons that the sampling duration might be changed. First, if the source is known to vary over a longer or shorter cycle, such that a different sampling period would be likely to yield more meaningful results. Second, with some state operating permits establishing very low emission limits, sampling times are increased to provide adequate quantification limits.

At this point, the sample gas stream should be as representative as can be collected using these methods. However, there is still the inherent variability of the methods themselves. To account for this, most testing programs include a series of three or more test runs in a single test. Depending on the underlying regulations, the results of the test runs can be used independently or averaged to reach a final emission measurement.

Once an appropriate sample gas stream has been extracted from an exhaust stream, it must be collected in some way for subsequent analysis. Most particulate sampling trains separate the particulate matter from the rest of the gas stream, saving the particulates and exhausting the particulate-free gas sample stream. This separation can be accomplished physically, chemically, or both. Examples of physical separation are filtration, inertial separation (using a "cyclone"), or condensation. Examples of chemical separation include dissolving into solution, adsorption, or chemical reaction into a solution.

One or more of these separation techniques may be appropriate for a given particulate material or under given process and/or stack conditions. The specific technique to be used and the limiting conditions may be found under the specific Test Methods described later in this Section. The essence of each is to ensure that all of the particulate matter flowing through the nozzle and probe is actually collected for subsequent analysis.

The following is a brief list of these particulate sampling principles, rearranged in an order that might be followed for an actual test.

1) Select a sampling location far from disturbances.
2) Determine traverse point locations.
3) Select appropriate nozzle, probe, heater, and so on.
4) Determine appropriate time for testing (worst case, etc.)
5) Determine isokinetic sampling rates.
6) Select sample train configuration.
7) Perform test runs.

Gaseous Sampling

When the pollutant of interest is a gas at ambient pressure, sampling is much easier and more straightforward than for particulate samples. This is because gas streams are almost always well mixed across the stack and are not subject to the internal considerations at the nozzle or in the probe. As a result, there is usually no need to worry about the sample location, traversing the stack cross-section, or isokinetic sampling. An exception is the measurement of nitrogen oxides from gas turbines using method 20. This method requires a stratification study.

The only real concern is that the probe be constructed of materials that will not react with or adsorb contaminants from the sample stream.

Sampling is usually performed by inserting a probe at a convenient location and sampling at the centroid of the cross-section. As with particulate sampling, it is important to measure the gas flow rate in the stack to allow calculations of emission rate. This is done in the same way as for particulate sampling using Test Method 2.

Protocol and Final Report Preparation

A stack sampling project, like most other investigative work, is not likely to succeed unless it is well planned and documented. The Stack Test Protocol, or Quality Assurance Project Plan, is the means used to document the planning of the project. The more detail that can be included in the protocol, the better the likelihood that the test will succeed the first time. Most EPA Regions and State Agencies have specific protocol formats and require specific types of information. Therefore, the project manager should contact the Agency well before the projected test date to obtain the format and to discuss any special conditions that need to be included, such as audit samples.

The final Stack Test Report is just as important as the Protocol. It is the means by which the testing team documents what they did and their results. If the Report does not fairly report what actually happened on the stack, in the laboratory and in the calculations, the entire test might well be wasted. Again, it is advisable for the project manager to contact the regulatory agency well before the test to obtain information of acceptable report formats and special information that might be needed. It is likely, though, that the Agency will require copies of all field data sheets, lab data sheets and print-outs, calculation procedures, examples, and results, diagrams, etc. A fourth advance has been made in the dissemination of stack testing information through EPA's electronic bulletin board. The Emission Measurement Technical Information Center (EMTIC) bulletin board is available as part of the Technology Transfer Network Bulletin Board Service (TTNBBS). The EMTIC bulletin board includes promulgated methods, proposed methods, some state test methods, papers on stack analysis, a data base on validated methods for various compounds, etc. It can be used to get answers to specific questions. Access to the TTNBBS is available through the Internet at "http://www.epa.gov//ttn"

Stack Test Guidance is available at http://www.epa.gov/Compliance/assistance/air/index.html

This his document does not address test methods. However, it does provide a good discussion of regulatory requirements for stack testing, including notifications, time frames, observation by regulatory agencies, and reporting.

TEST METHOD DESCRIPTIONS

The main body of this chapter includes brief descriptions of selected current U.S. EPA Test Methods. These are the methods approved by U.S. EPA for testing emissions from sources subject to the New Source Performance Standards (NSPS), found in 40 CFR Part 60, and the National Emissions Standards for Hazardous Air Pollutants (NESHAPS), found at 40 CFR Parts 61 and 63. The methods themselves can be found in the appendices to those regulations.

The U.S. EPA Test Methods may, of course, be used for purposes other than NSPS or NESHAPS. However, their applicability and validity may be unsure. The same is true for methods developed by the states or by others. Many methods are completely appropriate for a given circumstance for which they have been extensively verified (as have the EPA methods for NSPS and NESHAPS). However, their validity under other circumstances should always be questioned until and unless their performance can be confirmed.

The descriptions presented here are valuable for developing general understanding of the equipment and procedures. However, the methods should never be attempted without a thorough reading and understanding of the methods themselves. Stack Testing is still a very complex process that requires experience if useful results are to be obtained.

To assist in the selection of a Test Method, Table 1 lists the parameters that can be measured, along with the appropriate methods.

Test Method 1

Test Method 1 is used to determine representative traverse points for measuring solid or liquid pollutants and/or determine total volumetric flow rate from a stationary source. The procedures described in this method are used to determine the minimum number of points, the location of these points, and whether the chosen points are free from cyclonic flow. The Method contains a rote procedure for choosing point locations that are at the centroids or equal area portions of the stack cross section. This ensures equal weighting of all flows into the average flow rate determination.

The minimum number of sampling points is determined from Figure 1. This figure is applicable to both round and rectangular ducts based on the distance from the nearest disturbance, bend, exit o other obstruction which might disrupt the flow of gas through the duct. The ducts must be at least

TABLE 1

Parameters	Test methods	Conditions
Arsenic	108, 108A–C	
Beryllium	103, 104	103 Screening
Carbon Disulfide	15	
Carbon Monoxide	3, 10	
Carbon Dioxide	3, 3A, 3B, 3C, 20	3A Instrumental
Carbonyl Sulfide	15	
Condensible PM	202	
Chromium	306, 306A	Electroplating
Dioxins	23	
Dry Molecular Weight	3	
Excess Air	3	
Field Validation	301	
Flow Rate	2	Volumetric
Flow Rate	2A	Small stacks
Flow Rate	2B	Gasoline Vapor Incin
Flow Rate	2C	For small ducts
Flow Rate	2D	For small ducts
Flow Rate	2E	Landfill gas production
Fluoride (total)	13A, B, 14	AL plants
Fugitive Emissions	22	
Gasoline Vapors	27	Leaks from Tanks
Halogenated Organic	307	Vapor from solvent cleaning
Hydrogen Chloride	26	
Hydrogen Sulfide	11, 15	
Lead	12	Inorganic

(continued)

TABLE 1 *(continued)*

Parameters	Test methods	Conditions
Leaking Gasoline		
Tank Organics	27	
Leaks—Organic	21	
Mercury	101, 101A,	102, 101A From Incinerators
	102	In hydrogen
Mercury	105	In Sewage Sludge
Metals	29	
Moisture	4	
Nitrogen Oxides	20	
Nitrogen Oxides	7, 7A–E	Different methods
Nonmethane		
Organics	25	
Organics—Leaking	21	
Organics (gaseous)	18, 25A–B	A-B-different analyzers
Organics	25C	Landfill gas
Organics	25D	Waste samples
Organics	25E	Waste samples
Organics	304A–B	Biodegradation
Organics	305	Individual organics in waste
Organics	311	Hazardous air pollutants in paints
Oxygen	3, 3A, 3B, 3C, 20	3A instrumental
Particulates	5, A-I	A-I Specif Facials,
	17	In stack filter
PM_{10}	201 OR 201A, and 202	
Polonium-210	III	
Sampling Site	I	
Sulfur Dioxide	6, 6A–C, 8	A—Fossil fuel, B—daily average, C—instrumental
Sulfur Compounds	15A, 16, 16A	16A Total Reduced, 16—Semicontinuous
Sulfuric Acid Mist	8	
Surface Coatings	24, 24A	Volatiles, water, density, solids
Surface Tension	306B	Chromium electroplating
Organics (gaseous)	18	
Traverse Point	1	
Traverse Points	1A	For small ducts
Velocity	2, 2A-2H	
Vinyl Chloride	107	In Wastewater,
	107A	resin slurry
Vinyl Chloride	106	
Visible Emissions	9, 9-Alt, 303,	9-Alt-Lidar,
	303A,	
	22	fugitives
Volatile Organics		
Capture Efficiency(VOCs)	204 *et al.*	
Wood Heaters	28	Certification and auditing
Wood Heaters	28A	Air to fuel ratio

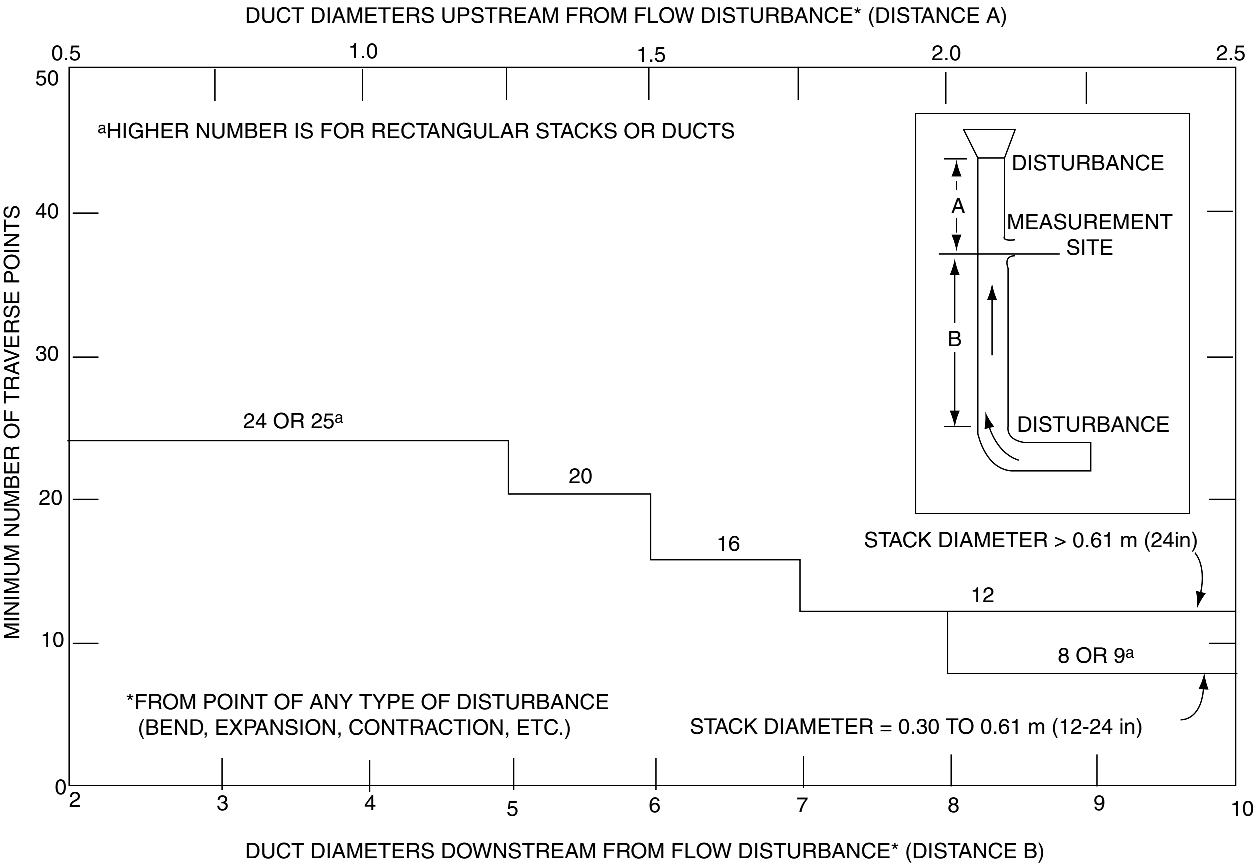


FIGURE 1 Minimum number of traverse points for particulate traverses.

12 inches (0.3 m) in diameter or 113 in^2 (0.071 m^2) cross-sectional area. Sampling ports must not be within 2 duct diameters downstream or half a diameter upstream from any flow disturbance. Sampling points are then determined by dividing the stack into equal area sections as shown in Figures 2 and 3. A table is provided in the Method which gives the percentage of the stack diameter from the inside wall to each traverse point. For stacks greater than 24 inches in diameter, no point should be closer than 1 inch from the wall; for smaller stacks, no closer than 0.5 inch.

Once these criteria are met measurement of the direction of flow is made to insure absence of significant cyclonic flow. The angle of the flow is determined by rotating a Type S pitot tube until a null or zero differential pressure is observed on the manometer. The angle of the pitot tube with the stack is then measured. This procedure is repeated for each sampling point and the average of the absolute values of the angles calculated. If the average angle is greater than 20º, the sampling site is not acceptable and a new site must be chosen, the stack extended, or straightening veins installed.

A few unusual cases have been accounted for by the method. If the duct diameter or size is smaller than that required by Method 1, Method 1A can be used. For cases where 2 equivalent stack diameters downstream or a half diameter upstream are not available, a directional velocity probe can be used to determine the absence of cyclonic flow, as described in Method 1. If the average angle is less than 20º, then the sampling site is satisfactory; however a greater number of points must be used.

Test Method 2

Test Method 2 is used to determine the average velocity in a duct by measuring the differential pressure across a Type S (Stausscheibe) pitot tube. The Type S pitot tube is preferable to the standard pitot tube when there are particles that could cause plugging of the small holes in the standard pitot tube. Measurement sites for velocity determination are chosen as described in Method 1, that is, required number of sites and absence of cyclonic or swirling flow.

The type S and standard pitot tubes are shown in Figure 4. When the Type S pitot tube has been correctly manufactured and installed on a probe as shown in Figure 5, there is no interference and calibration is not necessary. A pitot tube constant of 0.84 is assumed. If the criteria for interferences are not met, the method discusses the necessary calibration procedures.

TRAVERSE POINT	DISTANCE, % of diameter
1	4.4
2	14.6
3	29.6
4	70.4
5	85.4
6	95.6

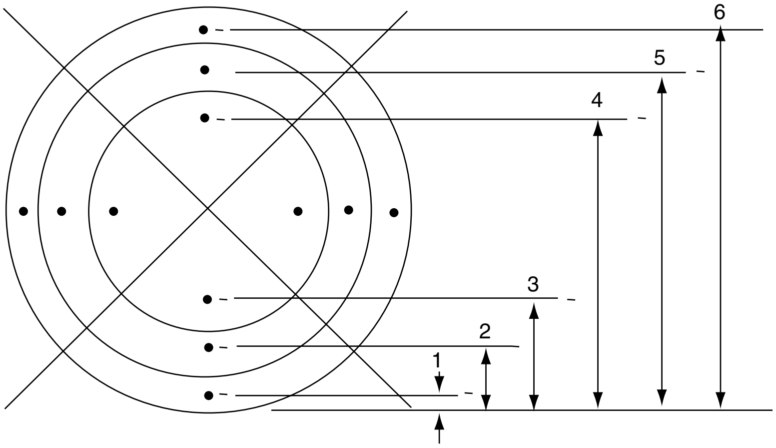


FIGURE 2 Example showing circular stack cross section divided into 12 equal areas, with location of traverse points indicated.

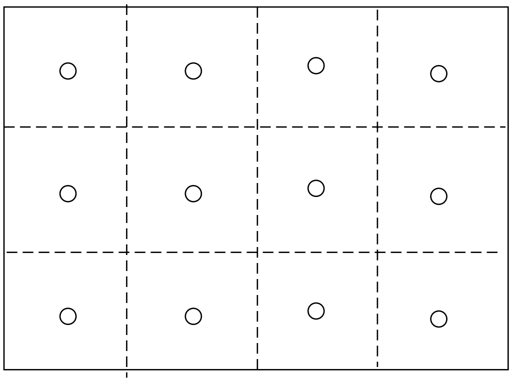


FIGURE 3 Example showing rectangular stack cross section divided into 12 equal areas, with a traverse point at centroid of each area.

Velocity, as measured with a pitot tube, is proportional to the square root of the differential pressure across the two sides of the pitot tube and the density of the stack gas. Most sampling trains use a combination inclined–vertical manometer to measure the velocity head, or Δp. These manometers usually have 0.01 inch of water subdivisions from 0–1 inch of water. The vertical section has 0.1 inch divisions from 1–10 inches of water. This type of gauge provides sufficient accuracy down to 0.05 inches; below that a more sensitive gauge should be used.

The temperature of the gases is usually measured using a type K (Chromel-Alumel) thermocouple mounted on the probe. The absolute pressure is calculated by adding the static pressure in the stack to the barometric pressure. The molecular weight of the stack gases is determined using Methods 3 and 4.

Velocity is calculated by the equation below:

$$V_s = K_p C_p (\Delta p^{0.5})_{avg} \{T_{s(avg)}/(P_s M_s)\}^{0.5}$$

where:
K_p = Velocity equation constant

$$K_p = 34.97 \frac{m}{sec} \left[\frac{(g/g\text{ - mole})(mmHg)}{(\ K)(mmH_2O)} \right]^{0.5} \text{ metric}$$

$$K_p = 85.49 \frac{ft}{sec} \left[\frac{(lb/lb\text{ - mole})(in.Hg)}{(\ R)(in.H_2O)} \right]^{0.5} \text{ English}$$

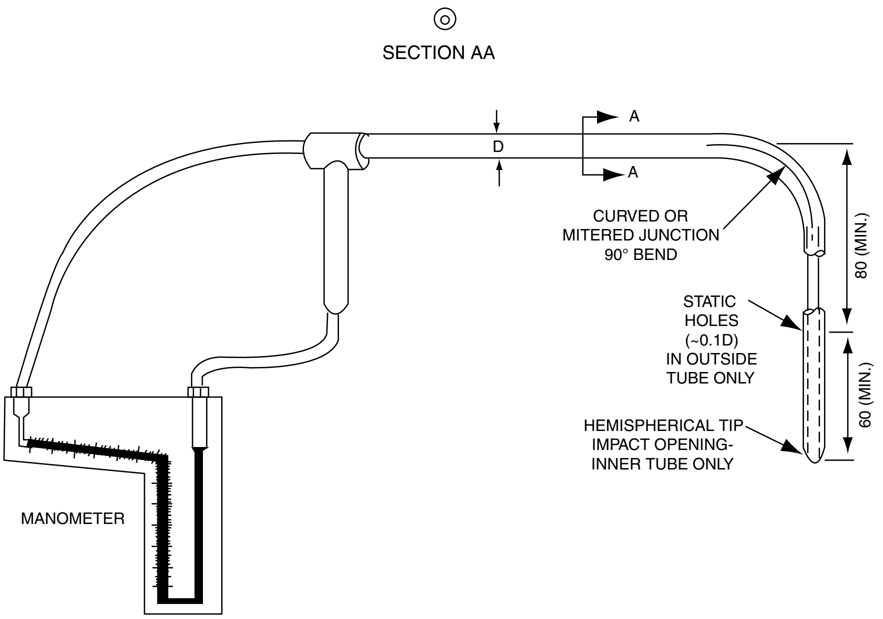
C_p = Pitot tube Coefficient (0.84 for S Type without interferences)
Δp = pressure difference across the two sides of the pitot tube (velocity head of the stack gas)
P_s = Absolute pressure of the stack, mm Hg or in. Hg
M_s = Molecular weight of the wet stack gases, g/g mole or lb/lb mole
T_s = Absolute stack temperature, ºK (273 + ºC) or ºR (460 + ºF)

The average dry volumetric stack flow is:

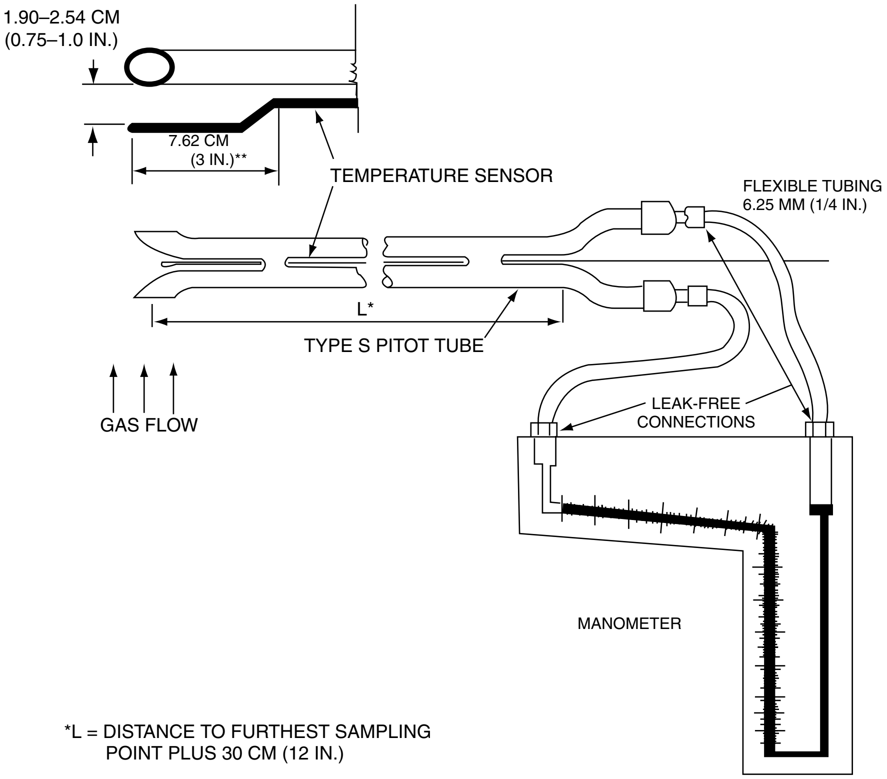
$$Q_{sd} = 3{,}600(1 - B_{ws})V_s A(T_{std}/T_{s(avg)})(P_s/P_{std})$$

where:

Q_{sd} = Average stack gas dry volumetric flow rate
B_{ws} = Water vapor in the gas stream from Method 4 or 5
V_s = Average stack gas velocity
A = Cross-sectional area of the stack
T_{std} = Standard absolute temperature 293ºK or 528ºF
P_{std} = Standard absolute pressure 760 mm Hg or 29.92 in. Hg



Standard pitot tube design specifications.

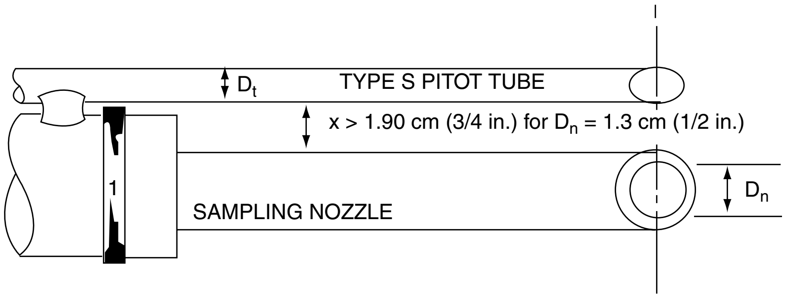


*L = DISTANCE TO FURTHEST SAMPLING POINT PLUS 30 CM (12 IN.)

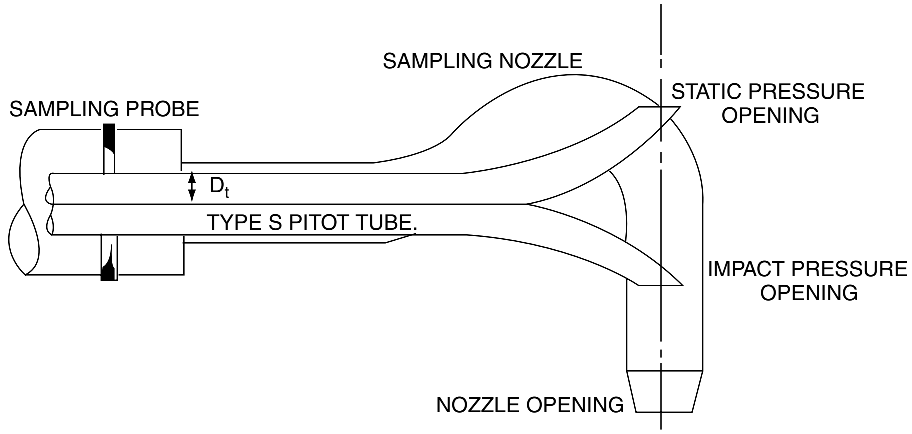
**PITOT TUBE - TEMPERATURE SENSOR SPACING

Type S pitot tube-manometer assembly.

FIGURE 4 Type S pitot tube-manometer assembly.



(a) BOTTOM VIEW: SHOWING MINIMUM PITOT-NOZZLE SEPARATION.



(b) SIDE VIEW: TO PREVENT PITOT TUBE FROM INTERFERING WITH GAS FLOW STREAMLINES APPROACHING THE NOZZLE, THE IMPACT PRESSURE OPENING PLANE OF THE PITOT TUBE SHALL BE EVEN WITH OR DOWNSTREAM FROM THE NOZZLE ENTRY PLANE

FIGURE 5 (a) Bottom view: showing minimum pitot-nozzle separation. (b) Side view: to prevent pitot tube from interfering with gas flow streamlines approaching the nozzle, the impact pressure opening plane of the pitot tube shall be even with or downstream from the nozzle entry plane.

Other methods 2A through 2H are used for specific conditions.

Test Method 3

Test Method 3 is used to determine the oxygen (O_2) and carbon dioxide (CO_2) concentration from combustion gas streams for the determination of molecular weight.

The method can also be used for other processes where compounds other than CO_2, O_2, CO or N_2 are not present in concentrations that will affect the results significantly. The O_2 and CO_2 can then be used to calculate excess air, molecular weight of the gas, or to correct emission rates as specified by various subparts of 40 CFR Part 60. Method 3 can also be used to determine carbon monoxide when concentrations are in the percent range.

Two types of analyzers can be used depending on the use intended for the data. Both analyzers depend on the absorption of components in the combustion gases by specific chemicals. The Orsat Analyzer sequentially absorbs CO_2, O_2, and CO. The change in sample volume is measured with a gas burette after each absorption step. Potassium hydroxide solution is used to absorb CO_2, forming potassium carbonate. When no further change in volume is noted, the difference from the starting volume is the amount of CO_2 present. Since the starting volume in the burette is usually 100 ml, the difference in ml is also the concentration of CO_2 in percent. The absorbent solution for O_2 is a solution of alkaline pyrogallic acid or chromous chloride. The CO absorbent is usually cuprous chloride or sulfate solution. The Fyrite type analyzers are available for either CO_2 or O_2, however they do not provide the accuracy of the Orsat Analyzer, **using Method B.**

Test Method 3A

Test Method 3A is an instrumental method for determining O_2 and CO_2.

From stationary sources when specified in the applicable regulations. Calibration procedures are similar to those discussed in Method 6C.

Test Method 3B

The Orsat analyzer is required for emission rate corrections and excess air determinations. Concentration values from 3 consecutive analyses must differ by no more than

0.3 percentage points when used for above purposes. When only the molecular weight determination is desired, the analysis is repeated until the dry molecular weights from any three analyses differ by less than 0.32 g/g-mole (lb/lb-mole). The Fyrite analyzer can be used only for determination of molecular weight.

Sampling is done in one of three methods: grab, integrated, or multi-point integrated. Grab samples are used when the source is spacially uniform and concentration as a function of time is required, or the source is considered uniform over time. Integrated samples are the most commonly used. A leak-free 30 liter plastic bag is filled using a pump and flow meter to insure that the sample is representative of the same period of time as the emission test. Bags must be analyzed within 8 hours for determination of molecular weight, or 4 hours when samples were collected for emission rate correction. For a multi-point integrated sample, the stack is traversed as described in Method 1. Samples are uniformly collected in suitable sampling bags.

Leak-checks of the analyzer and bags are required. The leak-check should also insure that the valve does not leak in the open position. To assure the data quality when both O_2 and CO_2 are measured, a fuel factor, F_0, should be calculated using the equation:

$$F_0 = \frac{20.9 - \%O_2}{\%CO_2}$$

The value of F_0 should be compared with the values given in the Test Method for the fuel used. If F_0 is not within the specified range, the source of error must be investigated.

The EPA Test Method 3 write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 3C

Test Method 3C is a gas chromatographic method used to measure CO_2, CH_4, N_2 and O_2 from landfill gases.

Test Method 4

Test Method 4 is used to determine moisture content in stack gases. In this method a sample is extracted at a constant rate from the stack; the moisture is removed and determined either gravimetrically or volumetrically. Often this measurement is made as part of particulate emission measurements.

When saturated conditions are present in the stack gases, (i.e. water droplets are present), the test method may yield questionable results. Moisture content for saturated stack gases should be based on stack temperature measurements and either phschrometric charts or vapor pressure tables. Molecular weight determinations and velocity calculations should be based on the lower of the two water concentration determinations.

The procedure described next for Method 5 is appropriate for Method 4. If particulate measurements are not required, the method can be simplified as follows:

1) The filter required in Method 5 can be replaced with a glass wool plug.
2) The probe need be heated only enough to prevent condensation.
3) The sampling rate does not have to be isokinetic, instead the source can be sampled at a constant rate within ± 10%.

The EPA Test Method 4 write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose. A table for determination of vapor pressure of water at saturation is located in section 3.3.6 page 7 of the *Quality Assurance Handbook of Air Pollution Measuring Systems: Volume III.*

Test Method 5

Test Method 5 is used to measure particulate emissions from stationary sources. Stack gases containing particulate matter are drawn isokinetically through a glass fiber filter. The volume of gas after removal of all water is measured with a dry gas meter. The particulates are measured gravimetrically.

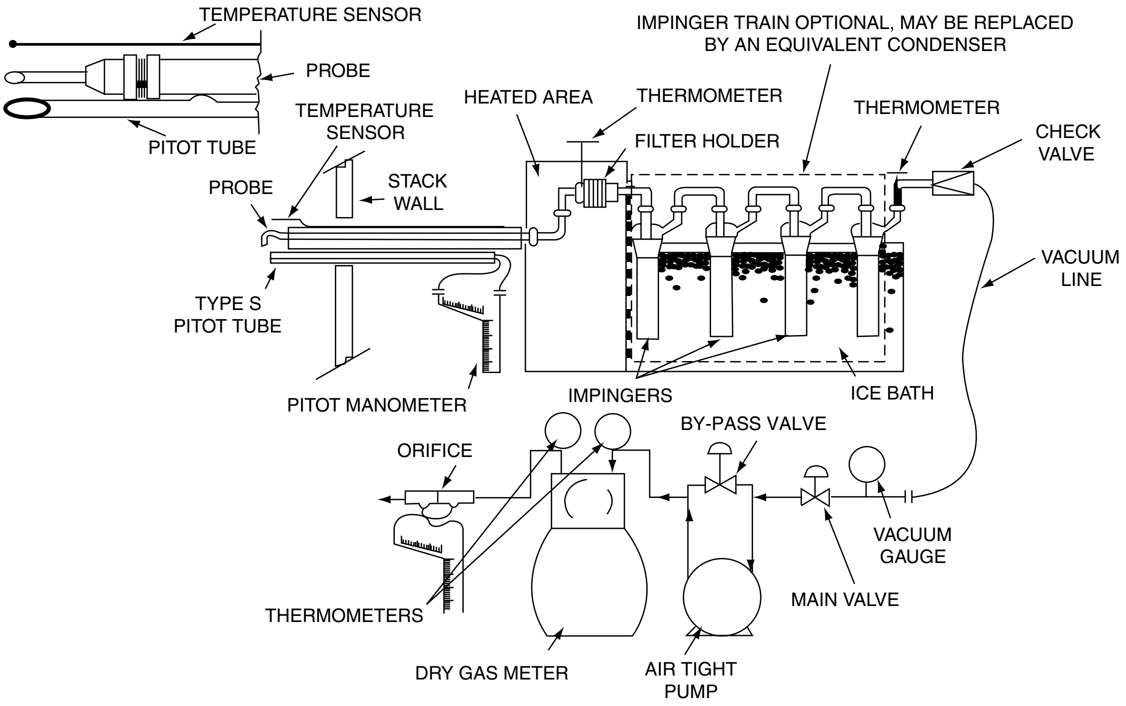
The equipment used in Method 5 forms the basis for many other emission measurements, therefore, the components of the Method 5 Train, Figure 6, will be described in some detail.

Probe Nozzle

The probe nozzles are usually stainless steel tubing in the shape of a Shepherd's Crook with a sharp tapered opening. Nozzles are available in a variety of sizes from 1/8" to 3/4" in diameter. They are calibrated by measuring the diameter in three different directions. The difference between the high and low value must be less than 0.004". The nozzle tips should be inspected before each use to ensure that the nozzle has not been damaged. During the sample recovery phase, all external particulate matter is wiped off and the opening capped with aluminum foil. When the probe has cooled sufficiently, the nozzle is removed. The nozzle is then cleaned by rinsing with reagent grade acetone and brushing with a Nylon bristle brush until further rinses with acetone are free of particulates. The acetone is saved in a bottle labeled Container 2 for later analysis.

Probe

Probe liners are usually constructed of borosilicate glass, or, if stack temperatures are expected to exceed 480ºC, quartz. Metal liners can be used subject to EPA approval. The probe liner is wrapped wit heater tape to maintain temperatures adequate to prevent condensation. The temperature of



Schematic of Method 5 sampling train.

FIGURE 6 Schematic of Method 5 sampling train.

the probe liner is regulated with a variable voltage controller. The controller is calibrated by introducing air at various temperatures in the range of the stack temperature and determining the controller setting necessary to maintain the probe temperature required by the Method (120 ± 14°C) or the specific regulation, such as a subpart of 40 CFR Part 60. The probe liner, with the heating element, is inserted in a stainless steel tube.

The probe liner is cleaned in much the same way as the nozzle. Once it has cooled it is removed from the train and the ends wiped free of silicone grease and covered. During cleaning, the probe is loosely held on an incline with utility clamps. The probe is then rinsed with acetone while slowly rotating the probe. A nylon brush, extended with teflon tubing, is used to scrub the inside of the probe liner. The acetone is saved in the bottle labeled Container 2 for later analysis.

Pitot Tube

A type S pitot tube is secured to the outside of the probe. The pitot tube was discussed in the Method 2 description. The special relationship to the nozzle is critical, in terms of both the location and the alignment. The pitot tube should be inspected before and after each run to insure that it has not been damaged. The pitot tube and the tubing connecting it to the manometer should be leak checked before and after each run. This can easily be done by slipping rubber tubing over the end of one side of the pitot, blowing gently on the rubber tubing producing at least 3 inches water pressure, then clamping the rubber tubing with a pinch clamp. No change in pressure should be observed in 15 seconds.

Filter

The filter holder is made of borosilicate glass with a glass frit filter support and silicone rubber gasket. Clamps of varying designs are used to seal the filter between the two halves of the holder.

The filter is a glass fiber filter capable of capturing at least 99.95% of 0.3 micron particles. Filters, desiccated at room temperature and ambient pressure, are weighted every six hours until the change is less than 0.5 mg. This process is called desiccating to constant weight. Alternatively the filters can be dried in an oven to constant weight. They are allowed to cool to room temperature in a desiccator before each weighing.

After the test, the filters are inspected for signs of gases passing around the edges of the filter and for water stains on the filter. Either would seriously compromise the results. The filter is carefully transferred to a petri dish. Any pieces of filter sticking to the housing are removed with forceps or a sharp instrument. The front half of the filter holder is rinsed with acetone and the acetone saved in the bottle labeled Container 2 for later analysis.

The filter holder must be maintained at 120 ± 14°C, **or as specified in the applicable regulation.**

Any heating system capable of providing this temperature around the filter holder during the run is acceptable.

Condenser

The condensation of the water vapor serves two purposes: first, it prevents condensation in the dry gas meter and the vacuum pump. Second, the collected condensate is used to determine the water vapor content in the stack gases. This collection is considered quantitative if the exit gases are maintained below 20ºC (68ºF). Four Greenburg-Smith impingers are connected in series and placed in an ice bath. The first, third, and fourth impingers are modified by removing the tip and replacing it with a straight piece of glass. One hundred ml of water is placed in the first and second impingers. The third is left empty and the fourth is filled with 200 to 300 grams of silica gel, at least some of which is the indicating type that turns from blue to pink when it is spent. The silica gel is usually put into a preweighed bottle before the start of testing. After use, it is returned to the same bottle and the weight difference is recorded.

After the sampling is complete the total water collected is determined by measuring the liquid in the first three impingers and subtracting the starting 200 ml. Any oil film or color of the water should be noted. Added to the liquid water collected is the weight gain of the silica gel. The color of the silica gel should be noted after sampling; if it all has changed to pink, all of the water passing through the train water may not have been collected.

Meter System

The metering system is used to withdraw isokinetically a measured quantity of gas from the stack. A vacuum pump is used to withdraw the sample. There is a vacuum gauge located before the vacuum pump. A dry gas meter capable of 2% accuracy is used to measure the gas sample volume. Two thermometers, one at the meter inlet and one at the outlet, are used to measure the gas temperature. To maintain isokinetic sampling rates, it is important to know the gas flow rate. This is done by measuring the pressure difference across an orifice in the exit from the dry gas meter.

The meter must be calibrated before initial use against a wet test meter. Then after each field use the calibration is checked at one point with either a wet test meter or critical orifice. The calibration check can not deviate from the initial reading by more than 5%. The metering system should be leak checked from the pump to the orifice meter prior to initial use and after each shipment.

Nomograph

The sampling rate necessary to maintain isokinetic conditions is dependent on the conditions in the stack (i.e. water content, average molecular weight of the gas, temperature, velocity, and pressure) and at the meter (i.e. temperature and pressure). The correct sampling rate is determined from the above parameters. The sampling rate is controlled by selecting the nozzle size and regulating vacuum. They serve as the course and fine adjustment of sampling rate. An initial velocity traverse is done to determine the stack conditions. A nomograph, special slide rule, or computer program is used to select an appropriate nozzle size to allow the fine adjustments of the vacuum pump to cover the expected range. This is necessary because the vacuum pump has a limited range. Because the stack velocity can change during the run, the operator must be able to rapidly recalculate the desired flow rate through the meter. The slide rule and computer can also be used to correct rapidly for changes in temperatures and pressures.

Sampling

The first step in determining particulate emissions is the selection of the sampling site and number of traverse points by Method 1. The following are the steps to perform particulate sampling:

1) Set up the equipment as shown in Figure 5.
2) Do initial traverse to determine appropriate nozzle.
3) Install nozzle.
4) Leak check from nozzle to vacuum pump.
5) Insert probe into stack and heat probe and filter.
6) Start traverse maintaining isokinetic sampling.
7) After traverse, leak check equipment.
8) After cooling, clean probe, nozzle and front of filter housing. Transfer filter to a petri dish.
9) Determine the amount of water collected.
10) Calculate the isokinetic variation. It must be within 10%.
11) Evaporate acetone from probe and nozzle washes and weigh.
12) Desiccate or oven dry the filter to constant weight.
13) Calculate the particulate emissions in the required units.

The EPA Test Method 5 write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 5A

Test Method 5A is used to determine particulate emissions from the Asphalt Processing and Asphalt Roofing Industry. This method differs from Method 5 in only two aspects:

1 The glass fiber filter is maintained at 42º ± 10ºC (108º ± 18ºF).
2 Trichloroethane is used to wash the probe and front half of the filter housing instead of acetone. This may be evaporated at 38ºC.

Test Method 5B

Test Method 5B is used to determine the non-sulfuric acid particulate matter from stationary sources. This method is very similar to Method 5 except that the filter housing is kept at 160º ± 14ºC (320º ± 25ºF) during the sample collection and dried at the same temperature for six hours. This volatilizes the sulfuric acid mist that collected on the filter. Once the acetone from the probe wash has been evaporated, the residue is also dried at high temperature to remove the sulfuric acid.

Test Method 5D

The purpose of this method is to select appropriate alternate locations and procedures for sampling emissions from positive pressure fabric filters. Many times these air pollution control devices were not designed with emission sampling in mind. This method should be consulted if a source using fabric filters does not met the criteria specified in Method 1.

Test Method 5E

Test Method 5E is used for the determination of particulate emissions from the wool fiberglass insulation industry. Method 5 has been modified to include the measurement of condensable hydrocarbons in this method. A 0.1N NaOH solution is used in place of distilled water in the impingers. The particulates condensed in this solution are analyzed with a Total Organic Carbon (TOC) analyzer. The sum of the filtered particulates and the condensed particulates is reported as the total particulate mass.

Test Method 6

Test method 6 is used for **determining the sulfur** dioxide (SO_2) emissions from stationary sources. The sulfuric acid mist and the sulfur trioxide are separated from the SO_2, and the SO_2 quantified using the barium-thorium titration method.

In this method, the use of midget impingers is recommended. However, the standard size impingers as used in Method 5 or 8 can be used if the modifications required by Method 6 are implemented. SO_2 can be determined simultaneously with moisture and particulates by making the required modifications.

The train for Method 6 is similar to Method 5 except for the size of the impingers and the use of a glass wool plug in the probe tip, replacing the particulate filter. The first impinger or bubbler (fritted glass tip) contains 15 ml of 80% isopropanol and a glass wool plug at the exit. The first impinger will collect SO_3 in the isopropanol solution and the glass wool will prevent the carry-over of sulfuric acid mist. The isopropanol should be checked for the presence of peroxides with potassium iodide. Peroxides would prematurely oxidize the SO_2 and capture it along with the SO_3. The next two midget impingers each contain 15 ml of 3% hydrogen peroxide. This will oxidize the SO_2 to sulfuric acid for analysis. The hydrogen peroxide should be prepared daily from 30% hydrogen peroxide. A drying tube containing silica gel is used to protect the vacuum pump and dry gas meter. The dry gas meter must be capable of 2% accuracy for a 20 liter sample. The vacuum pump should be a leak-free diaphragm pump. A surge tank should be used to eliminate pulsations.

Sampling for SO_2 is not done isokinetically since the gas is assumed to be uniformly dispersed in the stack. Sampling is done at one point and at a constant rate (+ 10%). Crushed ice should be added as necessary to maintain 68ºF at the last impinger.

After sampling, the train is leak checked, then the ice drained and the train purged with clean air for 15 minutes at the sampling rate. This will remove any SO_2 dissolved n the isopropanol and carry it over to the peroxide solution for oxidation and analysis. The isopropanol solution is then discarded. The peroxide solution containing the oxidized SO_2 is transferred to a graduated cylinder and diluted to 100 ml with distilled water. A 20 ml aliquot with four drops of thorium indicator is titrated with barium perchlorate. The solutions should be standardized as described in the method. The end point for this titration can be difficult to catch. It is possible, however, to get replicate titrations within 1% or 0.2 ml as required by the method. Audit samples are available through the EPA's Emissions, Monitoring and Analysis Division.

The EPA Test Method 6 write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with the equipment and titrations specified in this method.

Test Method 6A

Test Method 6A is used for the simultaneous determination of SO_2 and CO_2 from fossil fuel combustion source. Moisture may also be determined by this method. The train for Method 6A is very similar to Method 6 with the following exceptions:

1) The probe is heated to prevent moisture condensation.
2) The fourth impinger contains 25 grams of anhydrous calcium sulfate to remove the water. This is weighed after sampling to determine the moisture content.
3) In place of the drying tube in Method 6, there is a CO_2 absorber tube containing Ascarite II. This is weighed to determine the CO_2 concentration.

As with Method 6, the EPA Test Method 6A write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with the equipment and titrations specified in this method. The Method 6 audit samples are also appropriate for this method.

Test Method 6B

Test Method 6B is used for the simultaneous determination of SO_2 and CO_2 daily average emissions from fossil fuel combustion sources.

The train for Method 6B is very similar to Method 6A with the following exceptions:

1) The probe is heated to 20ºC above the source but not greater than 120ºC.

2) The isopropanol bubbler or impinger is eliminated. An empty bubbler is used for the collection of liquid droplets.
3) The stack gases are extracted from the sampling point intermittently over the specified period, usually 24 hours. An industrial timer is used to cycle the pump on for at least 12 periods of at least two minutes each. Between 25 and 60 liters of gas must be collected.

The EPA Test Method 6B write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with the equipment and titrations specified in this method. The Method 6 audit samples are also appropriate for this method.

Test Method 6C

Test Method 6C is an instrumental method for the determination of sulfur dioxide from stationary sources. No specific instrument is required by this method. Ultraviolet, nondispersive infrared, or fluorescence instruments can be used providing they meet the performance specifications and the test procedures are followed.

The following Measurement System Performance Specification must be passed by the instrument before actual environmental samples are analyzed:

1) Analyzer Calibration Error must be less than ± 2% of the span for the zero, mid-range, and high-range calibration gases.
2) Sampling System Bias must be less than ± 5% of the span for the zero, mid-range, and high-range calibration gases.
3) Zero Drift must be less than ± 3% of the span over the period of each run.
4) Calibration Drift must be less than ± 3% of the span over the period of each run.

The analytical range must be selected such that the SO_2 emission limit required of the source is not less than 30% of the instrument span. Any run in which the SO_2 concentration in the stack gas goes off-scale must be repeated.

The EPA Test Method 6C write-up contains detailed descriptions of the calibration gases required, calibration procedures, sampling procedures in addition to a list of references. It should be read in detail before the Method is attempted. The manufacturer's instructions will provide instrument specific instructions. Testing should be performed only by personnel trained and experienced with the equipment being used.

Test Method 7

Test Method 7 is used for the determination of nitrogen oxide (NO_x) emissions from stationary sources. In this method the NO_x concentration of a grab sample, collected in an evacuated two liter flask, is measured colorimetrically using the phenoldisulfonic acid procedure.

The apparatus for sample collection consists of probe, squeeze bulb, vacuum pump, manometer, and a two liter flask. The probe is heated if necessary to prevent condensation. A one way squeeze bulb is used to purge the probe before sampling. The vacuum pump is used to evacuate the flask. A manometer is used to measure the flask pressure. The two liter flask contains 25 ml of a H_2SO_4/H_2O_2 solution. This solution will absorb the NO_2. Oxygen from the source is required for the oxidation of the NO. If less than 1.5% oxygen is present in the stack gases, the steps described in the method to introduce additional oxygen must be used. Once the flask is evacuated 75 mmHg, valves are rotated to isolate the flask from the pump, and a sample is introduced into the flask. It should take less than 15 seconds for the flask to reach ambient pressure. If a longer time is required, check the probe for plugging.

After sampling, the flask should sit for at least 16 hours for the NO to oxidize. The flask is then shaken for two minutes before checking the pressure of the flask. The liquid is then transferred to a container, and the flask rinsed with distilled water. The pH of the liquid is adjusted to between 9 and 12 with 1 normal NaOH. Just prior to analysis the liquid is transferred to a 50 ml volumetric flask and diluted to 50 ml. A 25 ml aliquot is transferred to an evaporating dish and evaporated to dryness. Two ml of the phenoldisulfonic acid solution is added to the evaporating dish along with 1 ml of distilled water and four drops of concentrated sulfuric acid. The solution is again heated for three minutes followed by the addition of 20 ml of water. The solution is then adjusted to pH 10 with ammonium hydroxide. The solution is filtered if necessary and transferred to a 100 ml volumetric flask and diluted to 100ml. The absorbance is measured at the optimum wave length and compared against standards for quantification. Audit samples are available through the EPA's Emissions, Monitoring and Analysis Division.

The EPA Test Method 7 write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by personnel trained and experienced with the equipment and the spectrophotometer used in this method.

Test Method 7A

Test Method 7A is man alternative to Method 7 for the determination of nitrogen oxide (NO_x) emissions from stationary sources. The sample is collected with the same sampling train used in Method 7. However, instead of using the colorimetric phenoldisulfonic acid procedure, ion chromatography is used for quantification. Audit samples, available through the EPA's Quality Assurance Division, are required for this method.

The EPA Test Method 7A write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by personnel trained and experienced with the sampling train and the ion chromatograph used for this method.

Test Method 7B

Test Method 7B is used for the determination of nitrogen oxide (NO_x) emissions from nitric acid plants. This method is very similar to Method 7 except that in this method the NO_x concentration of the grab sample is measured using an ultraviolet spectrophotometer. Audit samples, available through EPA's Quality Assurance Division, are required for each set of compliance samples.

The EPA Test Method 7B write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by personnel trained and experienced with the equipment and spectrophotometer used for this method.

Test Method 7C

Test Method 7C is used to determine NO_x from fossil-fuel fired steam generators, electric utility plants, nitric acid plants, or other sources as specified in the regulations. In this method, an integrated sample is collected in alkaline-permanganate solution. The NO and NO_2 in the stack gas are oxidized to NO_2 and NO_3 by the permanganate. The NO_3 is then reduced by cadmium to NO_2, then the NO_2 from both starting chemicals is measured colorimetrically.

The train used for Method 7C is similar to the one used for Method 6 except that the three impingers are larger and have restricted orifices at the tips. 200 ml of $KMnO_4$/NaOH solution is placed into each of the three impingers. The probe is heated as necessary to prevent condensation. A sampling rate of 400–500 cc/minute should be used since greater rates will give low results. The sample run must be a minimum of 1 hour. Carbon dioxide should be measured during the run with either an Orsat or Fyrite analyzer.

The EPA Test Method 7C write-up contains detailed instructions for sample preparation, analysis and calibrations, along with a list of references. It should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with the equipment and titrations specified in this method. The Method 7 audit samples are appropriate for this method as well.

Test Method 7D

Test Method 7D is used to determine NO_x from stationary sources. In this method, an integrated sample is collected in alkaline-permanganate solution. The NO and NO_2 in the stack gas are oxidized to NO_3 by the permanganate. The NO_3 is analyzed by ion chromatography.

The train and sample recovery procedure used for Method 7D is identical to the one used for Method 7C. The Method describes the calibration and recommended chromatographic conditions for the analysis by ion chromatography.

The EPA Test Method 7D write-up contains detailed instructions for sample preparation, analysis and calibration along with a list of references. It should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with the equipment and ion chromatography used in this method. The Method 7 audit samples are appropriate for this method.

Test Method 7E

Test Method 7E is an instrumental method for the determination of NO_x from stationary sources. In this method, a gas sample is continuously extracted from a stack and conveyed to a chemiluminescence analyzed for determination of NO_x.

The performance specifications and test procedures in Method 6C are incorporated by reference to ensure data reliability.

The chemiluminescence analyzer is based on the reaction of NO with ozone to produce NO_2 and a photon of light. This light is measured with a photomultiplier. If total NO^x is required, the NO_2 is thermally converted to NO before analysis. When the converter is used NOx is measured.

The EPA Test Method 7E write-up contains instructions for calibrations, along with a list of references. The manufacturer's instructions should be followed for set up of the instrument. Testing should be performed only by trained personnel.

Test Method 8

Test Method 8 will provide data on both the sulfur dioxide and sulfuric acid mist from stationary sources. This method is similar to Method 6 except that the sulfuric acid in the isopropanol solution is also measured by the barium–thorin titration method.

Instead of using the midget impingers as used in Method 6, the full size impingers, as in Method 5, are used. A filter is installed between the isopropanol impinger and the first hydrogen peroxide impinger to catch any sulfuric acid carry over.

The EPA Test Method 8 write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 9

Test Method 9 is used for visually determining the opacity of emissions from stacks or ducts. The specified procedures are to be followed by human observers trained and certified by procedures also contained in the method.

The training and certification portion of the method is based on teaching observers to identify smoke of given opacity between zero and 100% in increments of 5%. This is accomplished using a smoke generator machine capable of producing white and black smoke separately and identifying the opacity with a transmissometer, or smoke meter calibrated accurately to within ±2%. Once the observers have "calibrated their eyes" to identify emissions of different opacities, they are given a test consisting of 25 black smoke readings and 25 white smoke readings. An observer first takes a classroom training session then passes the two sets of smoke observations and is thus certified to "read smoke." To pass, the average error for each color smoke must be no greater than 7.5% and no individual reading more than 15% from the correct value. Observers must be recertified every six months, needing to pass the test but not take the classroom training.

The remainder of the method describes the procedure to be followed by the observer in observing actual smoke emissions and for recording them. The basic idea for performing visible emission (VE) observations is to try and minimize the influence of those factors that might bias the results of the observations or might cause them to be unreliable. The four major variables that can be controlled by the observer all relate to position. They are: distance from the emission point, the viewing angle relative to the direction of the emission, the angle relative to the sun, and the angle relative to the wind. The goal is to keep the sun to your back, the wind to your side, the emission at about eye level, and at a sufficient distance to allow comfortable viewing.

VE observations are momentary glances at the densest part of the plume every 15 seconds. They are recorded in 5% increments on an observational record sheet that also contains a diagram of the observer/source/conditions and other relevant information. Normally, 24 consecutive readings constitute a minimum data set, representing the average emissions over that 6-minute period. However, more or fewer observations may be required or allowed by a specific regulation.

More detailed instructions for operating a certification program and for making VE observations is included in the method. However, in almost all cases, a certification course would be offered by an experienced organization. In the case of this method, the admonition about experienced personnel does not apply. Anyone who can pass the test can perform VE observations.

Test Method 9, Alt 1

Alternative 1 to Test Method 9 allows for the remote observation of emission plume opacity by a mobile lidar system, rather than by human observers, as specified in Method 9. Lidar is a type of laser radar (Light Detection and Ranging) that operates by measuring the backscatter of its pulsed laser light source from the plume. It operates equally well in daylight or at night.

The lidar unit shall be calibrated at least once a year and shall be subjected to one of two routine verification procedures every four hours during observations. The calibration shall be performed either on the emissions from a Method 9 smoke generator or using screen material of known opacity. The system is considered to be in calibration if the lidar's average reading is within ± 5% for the smoke generator, or within ±3% for the screens. The routine verification procedures require either the use of neutral density filers or an optical generator. In either case, average readings within ±3% are required.

The actual operation of the lidar system is conceptually straightforward but technically complex. The unit is positioned with an unobstructed view of a single emission, the backscatter recorded, and the plume opacity calculated. However, the actual procedures for accomplishing this must be performed only by personnel experienced at operating lidar units and interpreting their results. Detailed criterion are provided in the method for both.

Test Method 10

Test Method 10 is used for the determination of carbon monoxide (CO) emissions from stationary sources employing a nondispersive infrared analyzer (NDIR). Either an integrated or continuous sample can be extracted from the source for analysis by this method.

The analyzer compares the infrared energy transmitted through a reference cell to the same length cell containing a sample of the stack gases. A filter cell is used to minimize the effects of CO_2 and hydrocarbons. The use of ascarite traps further reduces the interference of carbon dioxide. Silica gel is used to remove water which would also interfere with the measurement of CO. Both the ascarite and silica gel traps are placed in an ice bath.

The EPA Test Method 10 write-up contains detailed instructions for calibrations, leak check procedure, and sampling, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and analyzer designed for this purpose.

Test Method 11

Test Method 11 is used to determine the hydrogen sulfide (H_2 S) content of fuel gas streams in petroleum refineries. The H_2S is absorbed by cadmium sulfate solution to form cadmium sulfide.

The sampling train is similar to the one used for Method 6. Five midget impingers are used. The first contains hydrogen peroxide to absorb SO_2, which would interfere with the analysis. The second impinger is empty to prevent carry-over of the peroxide into the cadmium sulfate absorbing solution in the last three impingers. If the gas to be sampled is pressurized, a needle valve is used to regulate the flow. If the pressure is not sufficient, a diaphragm pump is used to draw the sample through the train. A sampling rate of 1 (±10%) liter per minute is maintained for at least 10 minutes. As with Method 6, the train is purged after sampling is complete.

After sampling, the peroxide solution, containing the SO_2 trapped as H_2SO_4, is discarded. The contents of the third, fourth, and fifth impingers are quantitatively transferred to a 500 ml flask. Excess acidic iodine solution is added to the flask. After allowing 30 minutes for the sulfide to react with the iodine, the amount of excess iodine is determined by titration with either 0.01 N sodium thiosulfate or phenylarsine oxide.

The EPA Test Method 11 write-up contains detailed instructions for calibrations, leak check procedure, sampling, and calculations, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and the titrations required by this method.

Test Method 12

Test Method 12 is used for the determination of inorganic lead (Pb) emissions from stationary sources. Both particulate and gaseous forms of lead are measured by this method.

The sampling train is identical to the Method 5 train. The water in the first two impingers is replaced by dilute nitric acid to collect any gaseous lead emissions, converting them to the nitrate. The filter is not weighed. Instead it is treated with acid to dissolve the lead compounds, which are then analyzed using atomic absorption spectroscopy. The impinger liquid is evaporated to 15 ml and digested according to the instructions in the method. Once the sample from the impingers is prepared, it is also analyzed by atomic absorption spectroscopy.

The EPA Test Method 12 write-up contains detailed instructions for calibration, leak check procedure, sampling, and analysis, along with a list of references. It should be read in detail before the Method is attempted. Analysis of audit samples, prepared and distributed by EPA's Emissions, Monitoring and Analysis Division, is required. As with all of these methods, testing should be performed only by trained and experienced personnel using the equipment and atomic absorption spectrometer required for this analysis.

Test Method 13A

Test Method 13A is used for the determination of inorganic fluoride from stationary sources as specified in the regulations. It does not measure organic fluorine compounds such as the chlorofluorocarbons (freons).

The sampling train for Method 13A is similar to the Method 5 train, except that the filter can be located either in the standard position or placed behind the third and fourth impingers. If the filter is placed before the impingers, it should be either paper or organic membrane such that it will withstand temperatures up to 135ºC and retain at least 95% of 0.3 μm particles. Distilled water is used in clean-up instead of the acetone used in Method 5.

The total fluoride concentration of the combined filter, probe wash, and impinger contents is determined using the SPADNS Zirconium Lake Colorimetric Method. The sample preparation procedure includes breaking up of the filter, evaporating of the water and fusing with NAOH at 600ºC. The remaining ash is acidified with H_2SO_4, then distilled at 175ºC. SPADNS reagent is added to a suitable aliquot of the distillate and the absorbance at 570 nm compared with standard solution.

The EPA Test Method 13A write-up contains detailed instructions for sample preparation, calibrations, sample analysis, blanks analysis, and calculations, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by personnel trained and experienced with the equipment, chemicals, and the analytical procedures required by this method.

Test Method 13B

Test Method 13B is used for the determination of fluoride emissions from stationary sources. It uses the same sampling and sample preparation procedures as Method 13A. Analysis is with specific ion electrode method instead of the spectrophotometric method used there.

The EPA Test Method 13B write-up contains instructions for calibration along with a list of references. The manual for the specific ion electrode and the meter should be consulted before using this equipment. As with all of these methods, testing should be performed only by personnel trained and experienced with the equipment, chemicals, and the analytical procedures required by this method.

Test Method 14

Test method 14 provides guidance for setting up a sampling system for fluoride emissions from potroom roofs at primary aluminum plants. Gaseous and particulate emissions are drawn into a permanent sampling manifold through several large nozzles. The sample is then transported from the sampling manifold to ground level where it is sampled and analyzed by either Method 13A or 13B.

Anemometers are required for velocity measurements in the roof monitors (the large roof vents at such places). These anemometers are required to meet the specifications in the method for construction material, accuracy, etc. One anemometer is required for every 85 meters of roof monitor length. For roof monitors less than 130 meters, two anemometers are required.

An industrial exhaust fan is attached to the sample duct. The fan capacity must be adjustable and have sufficient capacity to permit isokinetic sampling. The fan is adjusted to draw a volumetric flow rate such that the entrance velocity into each manifold nozzle approximates the average effluent velocity in the roof monitor. A standard pitot tube is used for the velocity measurement of the air entering the nozzles because the standard pitot obstructs less of the cross section than a Type S pitot tube.

The EPA Test Method 14 write-up contains detailed instructions for the set-up, calibration, and use of permanent manifolds for sampling emissions from potroom roof monitors at aluminum plants. It should be read in detail before considering the installation of this type of system.

Test Method 15

Test Method 15 uses gas chromatography for the determination of hydrogen sulfide (H_2S), carbonyl sulfide (COS), and carbon disulfide (CS_2) emissions from stationary sources, such as tail gas control units for sulfur recovery plants. Any gas chromatographic system with a flame photometric detector (FDP) shown to be capable of resolving the compounds of interest and meeting the specifications in the method is acceptable.

A stainless steel or glass probe equipped with a particulate filter is used to extract stack gas continually. Condensed water, carbon dioxide, carbon monoxide, and elemental sulfur can interfere with the analysis. The analyst must show that these substances will not affect the determination. A dilution system is usually employed to reduce the effects of the CO and CO_2. The water condensation problem is minimized by heating the sampling lines by the dilution system. Heating the sampling lines will help to prevent buildup of sulfur in the lines. However, the probe and sampling lines should be inspected after each run and cleaned if necessary. If the probe is observed to be clogged during a run, the run must be repeated. The performance tests for calibration precision and calibration drift must fall within the stated limits. Calibration procedures and leak checks in the method must be followed. Aliquots of the diluted, heated sample stream are withdrawn periodically and injected directly into the GC-FPD for analysis. A sample run consists of a minimum of 16 injections over a period of not less than three hours or more than six hours.

The EPA Test Method 15 write-up contains performance specifications, instructions for sampling and calibration along with a list of references. The manual for the gas chromatograph should be consulted before using this equipment. As with all of these methods, testing should be performed only by personnel trained and experienced with the sampling procedure and gas chromatography required by this method.

Test Method 15A

Test Method 15A is used to determine total reduced sulfur emissions from sulfur recovery plants in petroleum refineries.

An integrated gas sample is extracted from the stack. A measured amount of air is added to the stack gas during sampling. This oxygen enriched mixture is passed through an electrically heated quartz tube at 1100 ± 50°C, and the sulfur compounds are oxidized to sulfur dioxide (SO_2). The remainder of the train is identical to that used for Method 6. The SO_2 collected in the hydrogen peroxide solution is analyzed using the barium-thorin titration technique.

The EPA Test Method 15A write-up contains detailed instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 16

Test Method 16 is a semi-continuous method for the determination of total reduced sulfur (TRS) emissions from stationary sources such as recovery furnaces, lime kilns, and smelt dissolving tanks at kraft pulp mills. Total reduced sulfur includes hydrogen sulfide (H_2S), methyl mercaptan (MeSH), dimethyl sulfide (DMS), and dimethyl disulfide (DMDS). These compounds are separated by gas chromatography and measured with a flame photometric detector.

The interferences, performance specifications, sampling and analysis procedures are very similar to those of Test Method 15.

The EPA Test Method 16 write-up contains instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 16A

This method is applicable to the determination of total reduced sulfur (TRS) emissions including hydrogen sulfide (H_2S), carbonyl sulfide (COS), dimethyl sulfide (DMS), and dimethyl disulfide (DMDS).

The sampling train for Method 16A is similar to the Method 15A train except that there is a citric acid scrubber before the oxidation tube to remove sulfur dioxide. Because the sources that would use this method have sufficient oxygen in their exhausts, additional air is not added before the oxidation step. The oxidized sulfur compounds are collected and analyzed as in Method 6. Sampling runs are either three hours, or three samples collected for one hour each. This provides data that are comparable with Method 16, which requires runs to be from three to six hours.

A system performance check is done to validate the sampling train components and procedure before the test and to validate the results after the test. Audit samples for Method 6 are used with this method.

The EPA Test Method 16A write-up contains instructions, along with a list of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 16B

This method is applicable to the determination of total reduced sulfur (TRS) emissions, i.e., hydrogen sulfide (H_2S), carbonyl sulfide (COS), dimethyl sulfide (DMS), and dimethyl disulfide (DMDS).

The sampling and analytical equipment and procedures train for Test Method 16B is similar to those used in Method 16A, except that gas chromatography is used to determine the SO_2 formed by the oxidation of the TRS, instead of the Method 6 procedure. There must be at least 1% oxygen in the stack gas for the oxidation procedure to work properly. As with Method 16, 16 injections are required per run over a period of not less than three hours nor more than six.

A system performance check is done to validate the sampling train components and procedure before the test and to validate the results after the test.

The EPA Test Method 16B and the other methods incorporated by reference contain instructions, along with a list of references. These should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 17

Test Method 17 is used for the measurement of particulate emissions from stationary source using an in-stack filter. In this method, unlike Method 5 in which the filter housing is maintained at 250 ± 25°F, there is no control of the filter temperature. This is important since particulate matter is not an absolute quantity, but dependent on the temperature at which it is measured. For example, some hydrocarbons are solids at the Method 5 standard temperature, but would not be solid at a stack temperature of 400°F. However, controlling the temperature of the filter can be difficult. Therefore, when particular matter is known to be independent of temperature, it is desirable to eliminate the glass probe liner and heated filter system necessary for Method 5 and to sample at the stack temperature using an in-stack filter. This method is intended to be used only when specified by the applicable regulation and only within the applicable temperature range. Except for the filter, this method is very similar to Method 5.

The EPA Test Method 17 and 5 write-ups contain detailed instructions, along with a list of references. They should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 18

Test Method 18 uses gas chromatography (GC) to determine the identity and concentration of organic compounds emitted from stationary sources. This method is based on the ability of a material in a GC column to separate a mixture of gaseous phase organic mixtures into its component compounds. An appropriate detector must be chosen to provide a response when a compound of interest passes through it. A chart recorder provides a plot of detector response versus time. When a compound separated by the column passes through the detector, the signal increases from the baseline to a maximum then returns to the baseline; this is called a peak. The time from when the mixed sample was injected to the time of maximum peak height is the retention time. The compounds are tentatively identified by comparison of the retention time with that of known compounds. The components are then quantified by comparison of the peak height with that of known concentrations of the identified compound.

There are many variables in gas chromatography, making the GC a very versatile tool. However, these parameters require careful selection to provide the necessary separation and quantification. They include: column material, column temperature, column packing material, carrier gas glow rate, injection port temperature, detector type, detector temperature, etc.

A presurvey is necessary before the actual sampling to determine the VOCs present and their approximate concentration. A presurvey sample is collected in an evacuated glass flask, purged through a glass flask, or collected in a tedlar bag. The sample containers are then heated to duct temperature to vaporize any condensed compounds. The optimum chromatographic conditions for separation are determined. If any peaks are not identified by comparison with known samples, other techniques such as GC/mass spectroscopy can be used for identification.

Based on the presurvey analysis, calibration standards are prepared. There should be at least three standard concentrations bracketing the expected concentration for each compound found during the presurvey. A calibration standard may contain more than one compound.

The final sampling can take four forms: integrated bag sample, direct injection, diluted interface sampling, and the adsorption tube procedure. For the integrated bag sample, a tedlar bag is placed in a rigid container then filled by evacuating the container. This sucks sample gas into the bag, eliminating the possibility of contamination or absorption by a sampling pump. The bag should be heated to the source temperature before analysis. The contents of the bag are flushed through a heated gas sample loop on the GC. An automated valve injects the contents of the loop onto the chromatographic column. The resulting peaks are identified by retention time comparison and quantified against the prepared standards. Bags should be analyzed within two hours of collection.

The direct injection technique does not permit integrated sampling, however it does eliminate the possibility of adsorption or contamination by the bags. All sampling lines must be maintained at stack temperatures to prevent condensation. The sample is sucked directly through the gas sample loop. Analysis is the same as for bag samples.

The dilution interface sampling and analysis procedure is appropriate when the source concentration is too high for direct injection. The stack gases are diluted by a factor of 100:1 or 10:1 by the addition of nitrogen or clean dry air.

Adsorption tubes can be used to collect organic compounds from stack gases. The selection of the adsorben is based on the chemicals present in the stack gas. Once a known volume of gas has been drawn through the tube the tube can be taken back to the laboratory for analysis. Tubes can generally be stored up to a week by refrigerating a sample. Once back at the laboratory, the adsorbent is extracted with suitable solvent and the solvent analyzed by GC. There are many variables that affect the efficiency of the tube for collecting representative samples. The quantitative recovery percentage of each organic compound from the adsorbent material must be known. When the adsorbent capacity of the tube is exceeded, material will break through the tube and not be collected. This is dependent on the sample matrix, i.e. the amount of moisture and the effect of other compounds competing for the adsorbent.

The EPA Test Method 18 write-up and the QA Handbook Section 3.16 contain detailed instructions for sampling, analysis and calibrations, along with a list of references for many organic compounds of interest. Both should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with source sampling and gas chromatography. Audit samples are available for many compounds through EPA's Emissions, Monitoring and Analysis Division.

Test Method 19

Test Method 19 is not a stack test method. Instead, it contains procedures for measuring the sulfur content of fossil fuels. It also contains extensive procedures for calculating emission rates for particulates, nitrogen oxides (NO_x) and sulfur dioxide (SO_2) and for calculating SO_2 removal efficiency, all for electric utility steam generators. These calculations are based on sulfur in fuel measured by this method, on stack gas concentrations measured by the appropriate methods, and by ultimate fuel analyses (determination of the major elemental constituents in samples).

The sulfur content in fuel, either before or after pretreatment, is determined using ASTM procedures specified in this method. The primary consideration is the representativeness of the sampling procedure. The SO_2 removal efficiency of the pretreatment is calculated by the standard (in-out)(in) procedure, in which the 5% values are corrected for the gross caloric value of the fuel to compensate for changes in mass related to the S removal. SO_2 removal attributed to the flue gas desulfurization unit is calculated in the same way, but without need for correction. Overall percent removal is calculated from them.

Many regulations prefer particulate, SO_2, and NO_x emission rates to the heat input to the boiler, as "pounds/million Btu." In order to calculate these values, the emission rates must be corrected by so-called F factors to account for the different heat content of the fuel and the excess combustion air. This method includes detailed procedures or calculating F factors from ultimate analyses and average F factors for various fuels. It also includes procedures for correcting for moisture and for combining various results together.

Finally, Method 19 describes procedures for combining the SO_2 emission rates calculated before and after a control device for determining removal efficiency.

Test Method 20

Test Method 20 is used for determining the nitrogen oxide (NO_x) emission rate from stationary gas turbines. Since measurement of oxygen (O_2) or carbon dioxide (CO_2) is needed in order to calculate NO_x emissions, this method also includes procedures for their determination. Finally, the method includes provisions for calculating sulfur dioxide (SO_2) emissions based on Method 6 measurements of SO_2 and diluent measurements from this method.

The basic principles of this method is that a gas sample is extracted from eight or more traverse points, filtered, stripped of moisture, and fed to automatic NO_x and diluent analyzers. The results of the instrument readings are then used to calculate the NO_x and/or SO_2 emission rates.

The sampling train begins with a heated probe, a calibration valve (where calibration gases can be added instead of stack gas), a moisture trap, a heated pump, an automatic analyzer for either O_2 or CO_2 (to indicate the amount of excess air in the exhaust stream), a NO_2 to NO converter (to ensure the inclusion of all NO_x molecules in the analysis), and a NO_x analyzer.

The method includes detailed specifications for the calibration gases and for the analyzers. All of the calibration gases must be either traceable to National Institute of Standards and Technology Standard Reference Materials (NIST SRMs) or must be certified in a manner specified in the method. The analyzers must be able to pass a calibration check and must be subjected to a response time determination. The NO_x monitor must pass an interference response check (to ensure that high levels of CO, CO_2, O_2, and/or SO_2 will not interfere in its performance). In addition, the NO_2 to NO converter must be checked with calibration gases to ensure that the conversion will be stable at peak loading.

Before conducting the actual NO_x and diluent measurements, a sampling rate and site must be selected and a preliminary diluent traverse performed. Site selection is both important and difficult. The method presents some factors to be considered, such as turbine geometry and baffling, but leaves to the testing team the selection of a representative location. The guidance provided in section 6.1.1. of the method should be read carefully before a site is selected. The preliminary traverse is then performed, sampling for a period of one minute plus the instrument response time at each of at least eight points (The number of points, between eight and 49, is determined by a calculation procedure included in the method) for the purpose of selecting the eight sampling points exhibiting the lowest O_2 concentration or the highest CO_2 concentration. Those are the points expected to show the highest NO_x concentrations and are the eight to be sampled for both NO_x and diluent.

Three test runs constitute a test at each load condition specified in the regulation. A run consists of sampling at each designated traverse point for at least one minute plus the instrument response time.

The measured CO_2 and NO_x values are then averaged algebraically and corrected to dry conditions using data from a Method 4 traverse. The CO_2 concentration is corrected to account for fuel heat input by using an F factor obtained from Method 19 (either from the Table there or by calculation based on ultimate analysis). Both NO_x and diluent concentrations are then corrected to 15% O_2 (to ensure consistent comparison with regulatory standards). Finally, the emission rates are calculated using the corrected concentrations. The most sensitive portion of Method 20 is the calibration and operation of the diluent analyzer, as the diluent concentration strongly affects the NO_x corrections. All of the procedures should be attempted only by trained personnel using appropriate equipment.

Test Method 21

Test Method 21 is used for determining Volatile Organic Compound (VOC) leaks from process equipment. The method actually presents three different ways for determining leaks. The first and most common way is to place the nozzle of a portable instrument right against the potential leak source. If the instrument reads above an arbitrarily defined value, a leak is found. The second way is for screening prior for conducting a survey with the portable instrument. This screening is performed by spreading a soap solution on the flange, for example, and looking for bubbles. The third way to look for leaks is to look for the absence of leaks. This is performed by measuring background concentration in the vicinity of the process equipment with the portable instrument and then determining whether the concentration at the equipment is significantly higher than background. The choice of procedure is dictated entirely by the prevailing regulation.

The second factor dictated by the applicable regulation is the definition of a leak. Method 21 defines all characteristics of a leak but one. As stated above, Method 21 procedures determine that a piece of equipment is leaking when the instrument is placed against it and reads above an arbitrarily defined value. The regulation must supply the value. Typical levels are 10,000 ppmv or 25,000 ppmv. The definition of no detectable emission is then defined as a reading less than 5% of the leak value above background. Thus, if background is 500 ppmv and the leak definition is 25000 ppmv, no detectable emission is reported at less than $0.05 \times 25000 + 500 = 1750$ ppmv.

Method 21 contains detailed specifications for the selection calibration, and operation of the instrument, without specifying the detector. Flame ionization detectors and photoionization detectors are the most common, but others, such as infrared absorption, have also worked successfully for this purpose. One of the most important criteria for selecting an instrument is its ability to detect the compounds of interest at the necessary concentration levels.

In order to allow consistent instrument selection and calibration, a single compound is generally specified in the regulation as the reference compound. All calibrations and instrument checks are then performed using calibration gases that are mixtures of that compound in air. A two point calibration is performed, using zero air, at one end, and a concentration close to the leak definition at the other.

Method 21 also contains detailed instructions for conducting leak detection surveys, describing acceptable procedures for measuring leaks at various types of equipment. These surveys may be conducted by personnel who have been trained to do so but who may not know the details of instrument troubleshooting, etc. However, someone at each facility performing the survey should be qualified to work on the instruments.

Test Method 22

Test Method 22 is used for determining fugitive emissions and smoke emissions from flares. Like Method 9, it is a visual method performed by trained observers. However, this method is not a quantitative method. This method involves only the timing of the observed emission and the calculation of the fraction of time that the emission is observed.

Fugitive emissions include all emissions that are not emitted directly from a stack or duct. They include emissions from doors, windows, cracks or holes as well as emissions from piles or directly from process equipment or operations.

In order to perform valid readings, the observer finds a suitable location more than 15 feet from the source, starts a stopwatch to record elapsed observation time, and begins observing the source. The observer continuously observes the source, starting a second stopwatch when emissions are noted, stopping it when they stop. This process continues until either the regulatory requirement has been met or 15 to 20 minutes have elapsed. At that point, the observer must take a 5 to 10 minute break. The clock time of all observations and the time and duration of emission observation should be recorded on an appropriate form, along with a suitable description and drawing of the observation position, etc.

Observers need not pass the certification exam of Method 9 but must have training equivalent to the classroom part of Method 9 certification.

Test Method 24

Test Method 24 is not a stack test method. Instead, it is used to determine the amount of volatile organic solvent used to determine the amount of volatile organic solvent water content, density, volume of solids, and weight of solids of surface coatings. This method applies to paints varnish, lacquer, and other related coatings. American Society of Testing Methods (ASTM) procedures have been incorporated by reference into this test method.

Test Method 24A

Test Method 24A is not a stack test method. Instead, it is used to determine the volatile organic content (VOC) and density of printing inks and related coatings.

The amount of VOC is determined by measuring the weight loss of a one to three gram sample heated to 120 ± 2°C at an absolute pressure of 510 ± 51 mm Hg for 4 hours or heated in a forced draft oven at atmospheric pressure for 24 hours. The coating density and solvent density are determined using ASTM D 1475–60, which is incorporated to this method by reference.

The EPA Test Method 24A write-up and the ASTM method contain detailed instructions for sample preparation and analysis. These should be read before attempting this method.

Test Method 25

This method is used to determine the emissions of total gaseous non-methane organics (TGNMOs) as carbon. It is used mainly for testing sources such as combustion sources for which the organic constituents of the emission are unknown and/or complex. The collected sample is injected onto a gas chromatographic column where the TGNMOs are separated from CO_2 and methane. Using an unusual backflush procedure, the TGNMOs are then oxidized to CO_2, catalytically reduced to methane, and analyzed using a flame ionization detector. By measuring the organics in this way, the difference in flame ionization response factors for various organic compounds that may be present is eliminated.

There are some limitations of this method, which the analyst should be familiar with.

1) Organic particulate matter will interfere with the analysis, giving higher values than actual. This can be eliminated by using a heated filter.
2) The minimum detection limit is 50 ppm as carbon.
3) When both water and CO_2 are present in the stack gas there can be a high bias in the concentration of the sample. This is due to the inability of the chromatographic column to separate the carbon dioxide from the hydrocarbons when large amounts of CO_2 and water are present in the sample gas. When the product of the percent water times the percent CO_2 is greater than 100, this bias is considered to be significant. In such cases, other methods should be used.

An emission sample is drawn at a constant rate through a heated filter and then through a condensate trap cooled with dry ice into an evacuated sample tank. The probe is heated to maintain an exit gas temperature of at least 129ºC (266ºF). The filter is heated in a chamber capable of maintaining a gas temperature of 121 ± 3ºC (250 ± 5ºF). The sample tank must be leak checked and cleaned by the procedures described in the method. The pressure of the sample tank is measured before and after sampling, and this information is used to determine the amount of sample collected. After sampling the sample tank valve is closed and the condensate trap is capped.

The condensate trap will contain water, CO_2, and condensed organic compounds. The CO2 will interfere with the analysis of the organics in the trap. CO2 is removed by purging the trap with clean air into the original sampling tank. The method describes this procedure in more detail and includes a figure showing the recommended equipment. After the purge is complete, the sample tank is pressurized to 1060 mm Hg absolute. The condensate tube is then connected to an oxidation catalyst and heated while purging with air or oxygen. The water collected with the sample and that produced by the oxidation of the hydrocarbons is removed with a cold trap. The CO_2 produced from the oxidation of the condensed organic compounds is collected in an intermediate collection vessel.

The analysis of the gas in the sample collection tank and the intermediate collection vessel both use gas chromatographic separation. The gas chromatograph column and operating conditions are chosen to separate CO, CH_4, and CO_2. These compounds are first oxidized then reduced to methane that is quantified by the flame ionization detector. The non-methane organics are retained on the column. A valve then reverses the flow of carrier gas through the column, back flushing the organics off the column through the oxidation then reduction catalyst before going to the detector. The TGNMO concentration is the sum of the non-methane organics and CO_2 from the intermediate sampling vessel.

The Method requires extensive quality assurance and quality control measures to insure that valid data are produced. The analysis of two audits gases provided by the Emissions, Monitoring and Analysis Division at Research Triangle Park, NC is required in the Method. This method should be performed only by personnel familiar with the sampling method and the gas chromatography necessary for the analysis. There have been recent changes in this Method; therefore it should be read carefully by all persons involved in its use.

Test Method 25A

Test Method 25A uses a Flame Ionization Analyzer (FIA) that has a flame ionization detector to directly measure that total gaseous organic compounds from a stationary source. A gas sample is extracted from the center of the stack. The sample is filtered if necessary to remove particulates which could give high readings or clog the detector. The concentration is expressed in terms of propane or the concentration as carbon is calculated.

This method is much simpler to perform than Method 25. However, there are two major limitations that must be understood.

1) The FIA will not separate or identify the organic compounds present.
2) The response of the detector is different for each compound.

Method 25A is the method of choice for the following situations.

1) When only one compound is present.
2) When the organic compounds present consist of only hydrogen and carbon, in which case the response factor will be about the same for all.
3) When the relative percentages of the compounds are known, in which case proper calibration standards can be prepared.
4) When a consistent mixture of compounds is present before and after an emission control device, and only the relative efficiency is to be determined.
5) When the FIA can be calibrated against mass standards of the compounds emitted.

This method will not distinguish between methane and nonmethane hydrocarbons. If this is required, Method 25 should be used or methane should be determined using Method 18 and subtracted.

The following Measurement System Performance Specification must be passed by the instrument before actual environmental samples are analyzed:

1) Analyzer Calibration Error must be less than ± 5% of the span for the zero, mid-range, and high-range calibration gases.
2) Zero Drift must be less than ± 3% of the span over the period of each run.
3) Calibration Drift must be less than ± 5 of the span over the period of each run.

The analytical range must be selected based on the applicable regulation and is usually 1.5 to 2.5 times the emission limit. The EPA Test Method 25A write-up contains detailed descriptions of the calibration gases required, calibration procedures, sampling procedures in addition to a list of references. It should be read in detail before the Method is attempted. The manufacturer's instructions will provide instrument specific instructions. Testing should be performed only by personnel trained and experienced with the equipment being used.

Test Method 27

Test Method 27 is used for determining leaks from gasoline delivery tank trucks or rail cars. It does not involve actual measurements of gasoline emissions. Instead, it involves the pressurization and/or evacuation of the tank and the subsequent measurement of the pressure and/or evacuation of the tank and the subsequent measurement of the pressure and/or vacuum after a given number of minutes. The pressure, time span, and allowable pressure change are all specified in the applicable regulation. Typically, the initial pressure is 450 mm water, with an allowable pressure loss of 75 mm water after five minutes.

Prior to conducting this test the tank must be emptied of both gasoline liquid and gasoline vapor and must be thermally stabilized. This is best accomplished by flushing the tank with a non-volatile liquid such as heating oil. In addition, care must be exercised to protect against fire by ensuring proper electrical grounding of the tank.

A suitable pump is attached to the tank and pressure or vacuum applied until the specified level is reached, as indicated on a liquid manometer or other instrument capable of reading up to 500 mm water gauge pressure, with ± 2.5 mm precision. The valve is then shut and the reading taken again after the specified period.

This test method should be performed only by people experienced in dealing with gasoline delivery equipment and operation.

Test Method 101

Test Method 101 is used to determine the emissions of gaseous and particulate mercury (Hg) when the gas stream is predominantly air. It is used mainly at chloralkali facilities that produce chlorine gas from salt in mercury-based galvanic cells. The gas stream is extracted from the stack isokinetically using a sampling train like the standard Method 5 train.

The main differences are that the impingers contain an iodine monochlorine solution (ICI), and no filter is employed. The probe must be lined with borosilicate or quartz glass tubing to prevent reactions with the mercury.

The method relies on the reaction of both particulate and gaseous mercury with the ICI to form $HgCl_2$, which remains in the impinger solutions. During subsequent analysis, the $HgCl_2$ is reduced to elemental mercury with a solution of HCI and $SnCl_2$, forming H_2SnCl_6. The mercury is then aerated into an optical cell where it is measured by atomic absorption spectrophotometry (AA).

Sample train preparation is about the same as for Method 5, with a few exceptions. First, care must be taken in selecting the nozzle size to allow the use of a single nozzle for each entire test run. Second, the 0.1 M ICI solution must be prepared, used to clean the impingers, and added to them. Third, it may be necessary to break a 2 hour sampling run into two or more subruns if high Hg or SO_2 concentrations lead to liberation of free iodine (evidenced by reddening of the first impinger solution). Finally, an empty impinger may be used as a knock-out chamber prior to the silica gel to remove excess moisture.

Calibration of the sampling train and actual sampling proceed exactly as with Method 5, as do calculations of percent isokinetic conditions. Sample recovery is essentially the same as for method 5 except that 0.1 M ICI solution is used as the rinse solution, to ensure capture of all mercury from the probe walls, etc.

The analytical system of this Method is designed to free the Hg from solution and to allow the Hg vapor to flow into an optical cell connected to an AA spectrophotometer that records the absorption at 253.7 nm light, the characteristic wavelength of Hg. This is accomplished by mixing a stannous chloride solution with aliquots of the recovered sample in a closed container then aerating with nitrogen. The nitrogen carrier-gas and the Hg then flow through the optical cell.

The flow rate through the aeration cell is calibrated with a bubble flowmeter or a wet test meter. The heating system for the optical cell, needed to prevent condensation on the cell windows, is calibrated with a variable transformer. The spectrophotometer itself is calibrated by using six different aliquots of a 200 ng/ml working solution of mercury, repeating each analysis until consecutive pairs agree to within ±3% of their average. Either peak area or peak height may be used. The aliquots are added to the aeration cell and N_2 bubbled through as during cell calibration, except without the flowmeter. A straight line is then fitted through the five points; it must pass within ±2 percent of full scale of the origin.

Analysis follows the same procedures as calibration except that aliquots of diluted sample are substituted for the working standard. Again, the aerated sample from each aliquot is analyzed until consecutive pairs agree to within 3%. The average is then compared to the calibration line and used in the succeeding calculations for determining the mass of mercury in the original sample solution. This value is then combined with the stack and flow volume factors in the same way as for Method 5 to arrive at the total mercury emission rate, in grams/day. Isokinetic variation and test acceptability are also determined according to Method 5 criteria.

The EPA Method 101 write-up contains detailed instruction for each of these steps, along with estimates of range, sensitivity, accuracy, precision, possible interferences and a list of references. It should be read in detail before the Method is attempted. Also, as with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose.

Test Method 101A

Test Method 101A is used to determine total particulate and gaseous mercury (Hg) emissions from sewage sludge incinerators. It is almost identical to Method 101. The gas stream is sampled isokinetically using, essentially, a Method 5 sampling train and bubbled through an acidic potassium permanganate ($KMnO_4$) solution where the mercury is collected and held in solution. During subsequent analysis, the mercury is reduced to elemental form, and aerated into an optical cell where it is analyzed by an atomic absorption spectrophotometer.

The main differences from Method 101 are as follows:

- There are two main differences in the sampling train. First, the impingers are filled with an acidic solution of 4 percent $KMnO_4$ solution instead of ICl. Second, a filter may be used prior to the impingers if the exhaust gas stream is expected to contain high concentrations of particulate matter.
- Analysis, including calibration, is also essentially the same as for Method 101 except for modifications related to the change from ICl to $KMnO_4$ as the oxidizing agent. In this case, the reducing solution is an acidic solution of $SnCl_2$, as in Method 101 plus sodium chloride hydroxylamine. The rest of the procedure is generally the same.

As for Method 101, the EPA Method 101A write-up should be read in detail, and the method should be attempted only by experienced personnel.

Test Method 102

The Method 102 is used to determine the emissions of particulate and gaseous mercury (Hg) when the gas stream is predominantly hydrogen. Such streams are common at chlor-alkali facilities that produce chlorine gas from salt in mercury-based galvanic cells. The equipment, procedures, and calculations are identical to those employed in Method 101 except for special safety precautions related to the flammability and explosiveness of hydrogen streams or related to hydrogen's low molecular weight.

For example, probe heaters, fans, and timers are used. Also, venting provisions are more elaborate. Finally, meter box calibrations must be performed with hydrogen or other special gases.

As with all other procedures, the EPA Method should be read in detail and attempted only by trained personnel using proper equipment. This is even more important than normal because of the serious risk of explosion.

Test Method 103

Test Method 103 is used as a screening method for determining approximate emissions of beryllium (Be). The Method has not been extensively verified and is generally used to produce order to magnitude estimates. If the results of a Method 103 test show the Be emissions to be within a factor of 10 or so of the level of interest (such as a regulatory emission standard), the test would normally be repeated using Method 104, which is much more reliable, if more expensive.

Method 103 uses a rough isokinetic sampling procedure, in which a sample probe is placed at only three locations along a stack diameter. The sample train consists of a nozzle and probe connected to a filter and a meter-pump system. No impingers are used.

Sample site location and train operation are the same as for Method 5 except, of course, for the reduction in the number of points and the absence of impingers. Points are chosen at 25, 50, and 75 percent of the stack diameter at the selected location. Sample recovery with acetone is also essentially the same.

Test Method 104

Test Method 104 is used to measure beryllium (Be) emissions at Be extraction plants, machine shops, ceramic plants, etc. The gas stream is extracted from the stack isokinetically using a Method 5 sampling train and Method 5 calibration and sampling procedures with a few minor exceptions. The sample is also recovered according the Test Method 5. The sample is then digested in acid and analyzed for Be by atomic adsorption spectrophotometry (AA). Once the sample has been collected and recovered, the solution is digested by addition of concentrated HNO_3, heating until light brown fumes indicate destruction of all organic matter, cooling, and subsequent addition of concentrated H_2SO_4 and concentrated $HClO_4$. The resulting solution is evaporated to dryness and redissolved in HCl. During this process, extreme care must be taken to avoid emanation of dangerous perchlorates. Therefore, all work should be done under an appropriate hood.

Analysis is then performed using a nitrous oxide/acetylene flame AA at 234.8 nm. If aluminum, silicon or certain other materials may be present, interferences may exist. They can be eliminated by following procedures cited in the Test Method. Calibration follows the AA manufacturer's specifications with specific provisions for dealing with cases in which the concentration of the sample is outside the normal calibration range.

The EPA Test Method 104 write-up contains detailed instructions, along with a lists of references. It should be read in detail before the Method is attempted. As with all of these methods, testing should be performed only by trained and experienced personnel using equipment and materials designed for this purpose. This is especially true for this Method, considering the hazard associated with Be exposure, as well as the perchlorate danger.

Test Method 105

Test Method 105 is not a stack testing method. Instead, it is the method used by EPA for estimating the maximum possible emissions of mercury from sewage sludge treatment plants. This is accomplished by measuring the mercury content of the feed sludge, multiplying this by the maximum feed rate to calculate the maximum emissions if all of the mercury in the sludge were to go up the stack. Method 105 thus serves as an inexpensive screening method for use at plants that expect their emissions to be well below the emission standard.

The Method requires the collection of about one liter of sludge each half hour for 8 hours. The samples are then combined into one, weighed, digested in potassium permanganate, and analyzed as in Method 101A, using atomic spectrophotometry.

Details, including the range, limitations, reagents, and equipment, may be found in the method write-up. Even though the field portion of this test is relatively straightforward, it should be performed only by trained personnel.

Test Method 106

Test Method 106 is used to measure the emissions of vinyl chloride monomer (VCM) in stack gas. It does not measure VCM in particulate matter. Stack gas is withdrawn from the centroid of the stack into a tedlar bag using the bag-in-a-box technique that isolates the sample from the pump. The sample is then analyzed directly using a gas chromatograph-flame ionization detector (GC-FID).

The sampling probe for this method is a standard stainless steel probe. The rest of the sampling train is unusual, however, because most of it is actually never seen by the sample. The probe is connected directly by new teflon tubing to an empty tedlar bag (usually between 50 and 100 liters capacity). The bag is sealed in a large box that is, in turn, connected to a needle valve, pump, carbon tube, and rotameter for measuring the flow. As the pump evacuates the air in the box, the bag sucks gas from the stack. Through the use of this indirect pumping procedure, there is no need for concern for contamination from the pump, or loss of VCM in the pump. While elaborate box configurations are available, a 55 gallon drum with two fittings and a sealable top works fine.

Prior to sampling, the bags and the box should be leak checked. Just before sampling, the bag should be partially filled with stack gas to condition the bag and purge the sample lines. It is then emptied by switching the pump to the bag fitting. Sampling then proceeds at a fixed rate proportional to the stack flow rate.

The GC-FID is calibrated by injection of 3 VCM standard gases prepared from a single certified 99.9% cylinder of VCM gas or cylinder standards appropriately certified by the manufacturer. During calibration and subsequent sample analysis, the VCM peaks must be of sufficient size and must not overlap with interfering peaks (such as acetadehyde). Procedures for calculating overlap are provided in Appendix C to 40 CFR Part 61 (right after the NESHAPS Test Methods). Procedures for minimizing overlap and interference are mentioned in the Method but are not specified. This is why an experienced GC operator is required for this analysis.

Immediately before analyzing samples, the analyst must analyse 2 audit gas cylinders supplied by EPA or another independent party. These audit cylinders contain concentration of VCM in nitrogen in 2 ranges to ensure the accuracy of the analytical procedures.

The sample is then injected directly into the GC-FID. The ratio of the peak height to peak area for the VCM peak is compared to the ratio for the nearest peak standard peak. If they differ by more than 10%, interference is probably present and an alternate GC column is needed. It is strongly recommended that the laboratory be given actual samples of stack gas for analysis at least a day before on official test is to take place. This will allow time to select appropriate levels for the standard gas concentrations and to determine whether alternate GC columns are required.

The concentration of VCM in the bag is then calculated using the GC calibration curve and the temperature, pressure, and humidity in the stack. This process is usually repeated for three test runs.

After the sample is extracted from the bag, the bag must be leak checked by filling it with air and connecting it to a water manometer. Any displacement of the manometer after 10 minutes indicates a leak. The box must also be checked for leaks. This is accomplished by placing the bag in the box-and evacuating the bag with a rotameter in line before the pump. Any displacement of the rotameter after the bag appears empty indicates a leak.

During the entire sampling and analytical procedure, extreme can must be taken to avoid exposure to the VCM gas, which is carcinogenic. The carbon tube included in the sampling train is for the purpose of absorbing VCM. Nevertheless, the tubing existing the pump should be aimed well away from the sampling team. Similarly, all laboratory work should be performed in appropriate vented or hooded areas. This is one more reason why this test for VCM should be performed only by experienced personnel who are extremely familiar with their equipment and with the procedure.

Test Method 107

Test Method 107 is not a stack testing method. It is used for determining the Vinyl Chloride Monometer (VCM) content of water or slurry samples associated with the manufacture of polyvinyl chloride (PCV) from VCM. Such sampling is often required as part of a comprehensive program for measuring the total VCM emitted from a PCV facility, to ensure that VCM removed from exhaust gas streams is not simply transferred to a water stream and discharged into the environment.

The method involves the collection of a sample of the wastewater stream or a stream of slurry containing PCV resin in a 60 ml vial, quickly capping the vial without trapping any air, and refrigerating. The vial is then conditioned at 90ºC for an hour. During that time the VCM dissolved in the water or remaining in the resin (termed Residual VCM, or RVCM) will reach equilibrium with the material that vaporizes, forming the so-called headspace in the vial.

The headspace is subsequently sampled by syringe and injected into a Gas Chromatograph-Flame Ionization Detector (GC-FID). The GC-FID is calibrated using vials filled with standard gas mixtures bracketing the expected range. The VCM concentration in the samples is calculated from the GC-FID response factor.

As with the other VCM Test Methods, this procedure should be performed only by trained individuals taking special precautions to avoid exposure to the carcinogenic VCM vapors.

Test Method 107A

Test Method 107A is not a stack testing method. It is a method used for determining the Vinyl Chloride Monomer (VCM) content in solvent solutions or in polyvinyl Chloride (PCV) resin slurry samples. This procedure supplements stack testing at PCV manufacturing facilities for determining the overall emissions of VCM from exhaust gas streams as well waste water and fugitive sources.

The method involves the collection of liquid or slurry samples in tightly capped 20ml glass vials. A small amount of the sample material is subsequently withdrawn by syringe and injected into a Gas Chromatograph (GC), where the sample vaporizes and is analyzed for VCM content. The GC is calibrated using standard solutions of VCM in an appropriate solvent. VCM content in the sample is calculated based on the response factor of the GC.

As with other VCM Test Methods, this method should be attempted only by trained personnel who should take special precautions to avoid exposure to the carcinogenic VCM vapors.

Test Method 108

Test Method 108 is used to determine stack gas emissions of particulate and gaseous forms of inorganic arsenic (As) from stationary sources. The principle is that a side stream is withdrawn isokinetically and both the particulate and gaseous forms are collected on a filter and in water solution. The arsenic is then concentrated, combined into one sample, and analyzed by flame atomic absorption (AA) spectrophotometry.

Sample train construction and operation are almost exactly ad described in Method 5 to ensure isokinetic sampling. Two of the impingers need to be modified slightly, the filter box is heated, and the sample flow rate is reduced.

The filter sample is recovered by digestion. Each portion of the sample (the impinger and filter catches) is then prepared for analysis by combining with nitric acid and boiling to dryness.

If sample concentrations are sufficiently low, special procedures must be followed to the detection limit of the AA. Either a Vapor Generator Procedure or a Graphite Furnace Procedure may be employed, following steps detailed in the Method and in the manufacturer's instructions.

Calibration standard solutions are prepared from stock arsenic solutions. The stock solutions are prepared by dissolving As_2O_3 in NaOH, adding nitric acid, heating to dryness and reconstituting with water. The calibration standards are then prepared by diluting aliquots of the stock solution with nitric acid. The sampling train is calibrated in much the same way as for Method 5.

Standard absorbances are then determined against blank levels. Samples are then run in the same way, with concentrations determined form the resulting calibration curve.

Concurrent with analysis of samples, special Quality Assurance Audit Samples must be analyzed. These samples are prepared and supplied through EPA's Emissions, Monitoring and Analysis Division in North Carolina. To be effective in evaluating the analyst's technique and the standards preparation, the Audit Samples must be analyzed by the same people and using the same equipment that will be used throughout the test program. EPA or the State agency must be notified at least 30 days in advance to ensure timely delivery of the Audit Samples.

The As concentration in the original stack gas is then computed with reference to the gas flow rate, moisture content, and measured As content.

As with all of these Test Methods, Method 108 should only be attempted by properly trained personnel using appropriate equipment.

Test Method 108A

Test Method 108A is not a stack sampling method. Instead, it is a procedure for measuring the Arsenic (As) content of As ore samples, particularly form nonferrous foundries. It is used primarily for determining the As feed rate, often needed to calculate As emission rate relative to feed rate.

In this method, the ore sample is finely pulverized, digested in hot acid, and filtered. Depending on the concentration of the sample, it is then analyzed either by flame atomic absorption spectrophotometry or by the graphite furnace procedure. In either case, the manufacturer's instructions are followed.

There are also two mandatory Quality Assurance checks associated with this procedure. First, there is a check for matrix effects, identical to the one described in Test Method 12. Second, it is required that all laboratories analyzing samples according to this method first acquire audit samples prepared and distributed by the Emissions, Monitoring and Analysis Division of US EPA, as with Method 12. These are to be analyzed along with the actual samples. The address for obtaining these samples is included in the Method.

As with all of these Test Methods, they should only be attempted by properly trained individuals, using appropriate equipment.

Test Method 111

Test Method 111 is used for measuring the emissions of Polonium (Po)-210 in stack gases. The sampling is exactly as in Method 5. In the analysis, the particulate Po-210 is dissolved in solution and deposited on a silver disc. The radioactive disintegration rate is measured using an alpha spectrometry system.

After sampling using the Method 5 train and techniques, the filter containing the sample is prepared for analysis. Care must be taken to complete analysis within 30 days of sampling to ensure that the results are not biased high as the result of decay of lead 210.

Sample preparation consists of repeated dissolution of the sample filter in concentrated hydrofluoric acid and heating to dryness until the filter is gone. The sample is then dissolved in a mixed acid solution and prepared for screening. The screening is designed to determine the approximate radioactivity of the sample, allowing the selection of an appropriate sample concentration to protect the detector. Once the aliquot size is chosen, an amount of Pb-209 is added in approximately equal activity. This will be used to assess Po recovery in the spectrometry system.

A specially prepared 3.8 cm silver disc, with one side painted to prevent deposition, is suspended in the heated sample solution for 3 hours. After rinsing in distilled water, the disc is placed in the alpha spectrometry system and its emissions counted for 1000 minutes. The picocuries of Po-210 per filter are then calculated from the counts in the Po-210 region, the counts of Po-209, and the Po recovery, based on the Po-209 counts.

The emission rate of Po-210 is then calculated based on the stack flow rate, as measured by Method 2. In most cases, the emission rate desired would be for rock processing plants, such as for phosphate rock. In that case, the processing rate is placed in the denominator or the rate equation. Detailed procedures for all typical calculations are included in the Method.

Several quality assurance steps are required by this method. First, the Alpha Spectrometry System must be standardized. This is accomplished by filtering a standardized solution of a different alpha-emitting actinide element and exposing the detector to the dried filter. Second a standardized Po-209 solution must be filtered and exposed to the detector in the same way. Next, each sample is analyzed in duplicate, with the difference between pairs required to be below stated limits. Finally, an independent analyst, under the direction of the Quality Assurance Officer for the testing project, should prepare a performance evaluation sample of Po, unknown to the analyst, for every 10 samples analyzed.

As with all of the Test Methods, Method III should not be attempted by anyone not trained to perform stack sampling. In particular, the analysis should only be performed by individuals experienced in dealing with radioactive sample analyses and hydrofluoric acid digestion.

Method 0010 (Modified Method 5)

Modified Method 5, or MM5, as it is usually called is used for Resource Conservation and Recovery Act (RCRA) trial burns to determine semi-volatile organic emissions from hazardous waste incinerators. This method, under the designation method 0010, can be found in SW-846.

In many ways, the equipment and operation of this train are similar to Test Method 5. A sample is extracted isokinetically from the stack and filtered to collect particulate material. In addition, the gases are cooled in a condenser and then trapped with Amberlite XAD-2 resin. A methylene chloride/methanol mixture is used for the probe wash instead of acetone.

The semi-volatile compounds on the filter and XAD-2 resin are extracted using an appropriate continuous extractor, and are concentrated for analysis. The condensate in the first impinger and, if required, the contents of the other impinger, are extracted using a separatory funnel and concentrated for analysis. The solvent rinses from the train are concentrated for analysis.

Analysis of these samples is done using Method 8270 which is also found in SW-846. The discussion of this method is beyond the scope of the next.

Modified Method 5 contains detailed descriptions of the quality control, calibration procedures, and sampling procedures in addition to a list of references. It should be read in detail before the Method is attempted. Testing should be performed only by personnel trained and experienced with the equipment being used.

Test Method 0030 Volatile Organic Sampling Train (VOST)

The Volatile Organic Sampling Train (VOST) is used to collect volatile organic compounds during resource Conservation and Recovery Act (RCRA) trial burns. This method can be found in SW-846 as Method 0030. It provides for the collection of gas samples in a pair of sorbent traps with subsequent desorption, concentration, and analysis of the contents of each tube. This method has been validated for many organic compounds boiling between 30°C and 100°C. This method should not be used for low boiling chlorofluorocarbons. Compounds which are water soluble may not be purged from the condensate completely and therefore should not be analyzed by this method. This method has been successfully used for some compounds boiling higher than 100°C.

In this method, a 20 liter gas sample is extracted from the stack with a glass lined probe at the rate of either 1 liter/minute or 0.5 liter/minute. Samples are collected from the center of the stack, not isokinetically. The sample gas is cooled with a condenser that has ice water pumped through it. The objective is to cool the gas to below 20ºC before the organic components are adsorbed onto a pair of sorbent resin traps in series. Any condensate present is collected between the two resin traps. The first trap contains approximately 1.6 grams of 2,6-diphenylene oxide polymer (Tenax). The backup trap contains approximately 1 gram each of Tenax and carbon. Both traps are cleaned by purging with a nitrogen while heating the traps to 190ºC. The method provides detailed instruction for the preparation of the sorbent tubes. The gases are dried before they are pumped by a vacuum pump through a dry gas meter.

Because of the relatively short time that each pair of tubes may collect samples, and the need to sample over at least an hour, a total of three pairs of tubes must be analyzed for each run. To account for the possibility of tube breakage during transport and analysis, between 4 and 6 pairs are actually collected. Each tube is analyzed by first heating the trap to desorb the compounds that are then collected on a smaller analytical trap. This tube is heated quickly and the compounds separated by gas chromatography and quantified with a suitable detector. The back-up traps are analyzed separately. When the traps are desorbed, the second, or back-up, trap should contain less than 30% of each organic compound. If this is not the case, there is a possibility of significant breakthrough of that compound through the traps. The condensate should be analyzed to show that there are no organic compounds present. To increase the detection limit, all three of each type of tube can be analyzed together.

This method requires extensive quality assurance and quality control measures to insure that valid data is produced. The majority of problems with this technique are usually from contamination of the sorbent traps in the field or laboratory. The analysis of audit gas provided by the Quality Assurance Division at Research Triangle Park, NC is required in the Method. This method should be performed only by personnel familiar with the sampling method and the gas chromatography necessary for the analysis.

Test Methods 204 et al.

EPA has published a number of Test Methods, Protocols, Procedures, and Regulations all aimed at determining the effectiveness of a control device to capture potential emissions of volatile organic compounds (VOCs). As a prelude to testing, most of the Procedures require the source to construct and verify a total enclosure, either permanent or temporary. Method 204 and Methods 204 A through F (published in Appendix M of 40 CFR Part 51) are to be used for measuring the VOC content in various process and emission streams, and for verifying the total enclosure. Then, additional Procedures from Part 52 are used to determine the actual Capture Efficiency.

DONALD G. WRIGHT

MARCUS E. KANTZ
U.S. Environmental Protection Agency

STATISTICAL METHODS FOR ENVIRONMENTAL SCIENCE

All measurement involves error. Any field which uses empirical methods must therefore be concerned about variability in its data. Sometimes this concern may be limited to errors of direct measurement. The physicist who wishes to determine the speed of light is looking for the best approximation to a constant which is assumed to have a single, fixed true value.

Far more often, however, the investigator views his data as samples from a larger population, to which he wishes to apply his results. The scientist who analyzes water samples from a lake is concerned with more than the accuracy of the tests he makes upon his samples. Equally crucial is the extent to which these samples are representative of the lake from which they were drawn. Problems of inference from sampled data to some more general population are omnipresent in the environmental field.

A vast body of statistical theory and procedure has been developed to deal with such problems. This paper will concentrate on the basic concepts which underlie the use of these procedures.

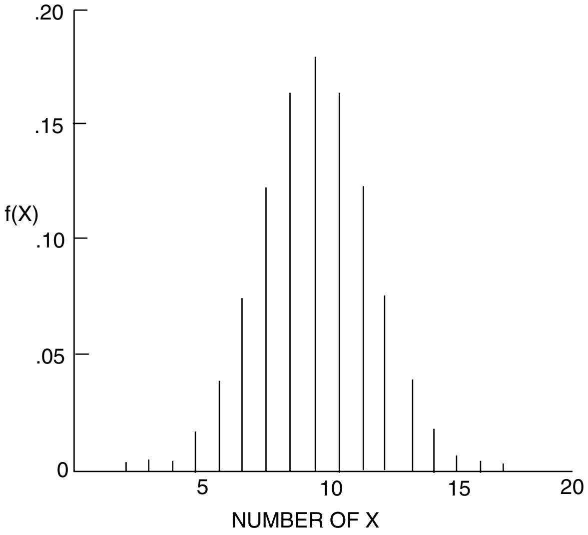


FIGURE 1

DISTRIBUTIONS

Discrete Distributions

A fundamental concept in statistical analysis is the probability of an event. For any actual observation situation (or experiment) there are several possible observations or outcomes. The set of all possible outcomes is the sample space. Some outcomes may occur more often than others. The relative frequency of a given outcome is its probability; a suitable set of probabilities associated with the points in a sample space yield a probability measure. A function *x,* defined over a sample space with a probability measure, is called a random variable, and its distribution will be described by the probability measure.

Many discrete probability distributions have been studied. Perhaps the more familiar of these is the binomial distribution. In this case there are only two possible events; for example, heads and tails in coin flipping. The probability of obtaining x of one of the events in a series of n trials is described for the binomial distribution by where θ is the probability of obtaining the selected event on a given trial. The binomial probability distribution is shown graphically in Figure 1 for $\theta = 0.5$, $n = 20$.

$$f(x;\theta,n) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}, \qquad (1)$$

It often happens that we are less concerned with the probability of an event than in the probability of an event *and* all less probable events. In this case, a useful function is the cumulative distribution which, as its name implies gives for any value of the random variable, the probability for that and all lesser values of the random variable. The cumulative distribution for the binomial distribution is

$$F(x;\theta,n) = \sum_{i=0}^{x} f(x;\theta,n). \qquad (2)$$

It is shown graphically in Figure 2 for $\theta = 0.5$, $n = 20$.

An important concept associated with the distribution is that of the moment. The moments of a distribution are defined as

$$\mu_k = \sum_{i=1}^{n} x_i^k f(x_i) \qquad (3)$$

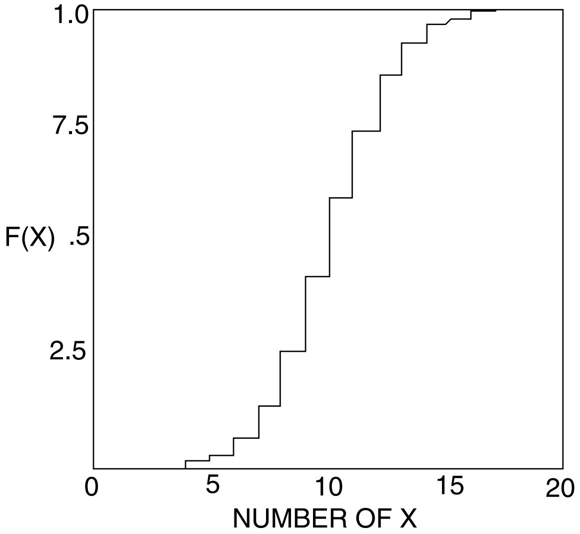


FIGURE 2

for the first, second, third, etc. moment, where $f(x_i)$ is the probability function of the variable x. Moments need not be taken around the mean of the distribution.

However, this is the most important practical case. The first and second moments of a distribution are especially important. The mean itself is the first moment and is the most commonly used measure of central tendency for a distribution. The second moment about the mean is known as the variance. Its positive square root, the standard deviation, is a common measure of dispersion for most distributions. For the binomial distribution the first moment is given by

$$\mu = n\theta \tag{4}$$

and the second moment is given by

$$\sigma^2 = n\theta(1-\theta). \tag{5}$$

The assumptions underlying the binomial distribution are that the value of θ is constant over trials, and that the trials are independent; the outcome of one trial is not affected by the outcome of another trial. Such trials are called Bernoulli trials. The binomial distribution applies in the case of sampling with replacement. Where sampling is without replacement, the hypergeometric distribution is appropriate. A generalization of the binomial, the multinomial, applies when more than two outcomes are possible for a single trial.

The Poisson distribution can be regarded as the limiting case of the binomial where n is very large and θ is very small, such that $n\theta$ is constant. The Poisson distribution is important in environmental work. Its probability function is given by

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \tag{6}$$

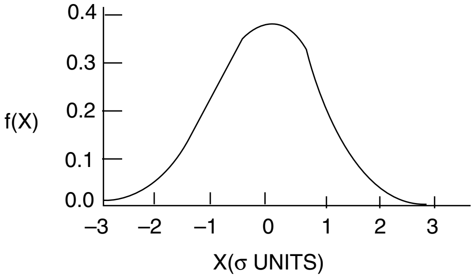


FIGURE 3

where $\lambda = n\theta$ remains constant.
Its first and second moments are

$$\mu = \lambda \tag{7}$$

$$\sigma^2 = \lambda. \tag{8}$$

The Poisson distribution describes events such as the probability of cyclones in a given area for given periods of time, or the distribution of traffic accidents for fixed periods of time. In general, it is appropriate for infrequent events, with a fixed but small probability of occurrence in a given period. Discussions of discrete probability distributions can be found in Freund among others. For a more extensive discussion, see Feller.

Continuous Distributions

The distributions mentioned in the previous section are all discrete distributions; that is, they describe the distribution of random variables which can be taken on only discrete values.

Not all variables of interest take on discrete values; very commonly, such variables are continuous. The analogous function to the probability function of a discrete distribution is the probability density function. The probability density function for the standard normal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{9}$$

It is shown in Figure 3. Its first and second moments are given by

$$\mu = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} \mathrm{d}x = 0 \tag{10}$$

and

$$\sigma^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \mathrm{d}x = 1. \tag{11}$$

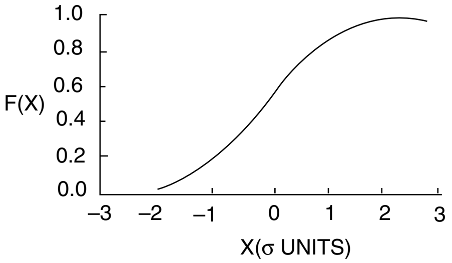


FIGURE 4

The distribution function for the normal distribution is given by

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,dt. \tag{12}$$

It is shown in Figure 4.

The normal distribution is of great importance for any field which uses statistics. For one thing, it applies where the distribution is assumed to be the result of a very large number of independent variable, summed together. This is a common assumption for errors of measurement, and it is often made for any variables affected by a large number of random factors, a common situation in the environmental field.

There are also practical considerations involved in the use of normal statistics. Normal statistics have been the most extensively developed for continuous random variables; analyses involving nonnormal assumptions are apt to be cumbersome. This fact is also a motivating factor in the search for transformations to reduce variables which are described by nonnormal distributions to forms to which the normal distribution can be applied. Caution is advisable, however. The normal distribution should not be assumed as a matter of convenience, or by default, in case of ignorance. The use of statistics assuming normality in the case of variables which are not normally distributed can result in serious errors of interpretation. In particular, it will often result in the finding of apparent significant differences in hypothesis testing when in fact no true differences exists.

The equation which describes the density function of the normal distribution is often found to arise in environmental work in situations other than those explicitly concerned with the use of statistical tests. This is especially likely to occur in connection with the description of the relationship between variables when the value of one or more of the variables may be affected by a variety of other factors which cannot be explicitly incorporated into the functional relationship. For example, the concentration of emissions from a smokestack under conditions where the vertical distribution has become uniform is given by Panofsky as

$$C = \frac{Q}{\sqrt{2\pi V D \sigma_y}} e^{-y^2/2\sigma y^2}, \tag{13}$$

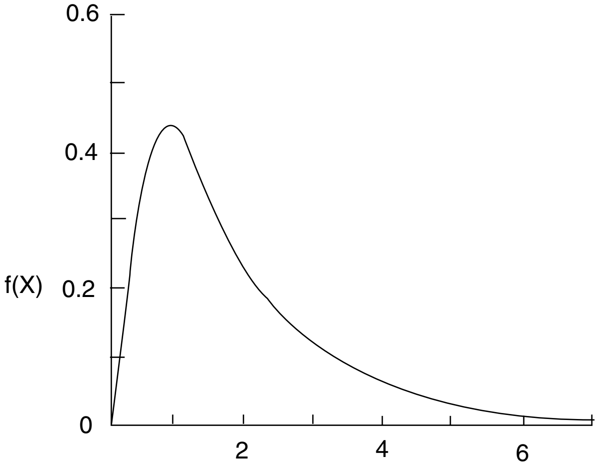


FIGURE 5

where y is the distance from the stack, Q is the emission rate from the stack, D is the height of the inversion layer, and V is the average wind velocity. The classical diffusion equation was found to be unsatisfactory to describe this process because of the large number of factors which can affect it.

The lognormal distribution is an important non-normal continuous distribution. It can be arrived at by considering a theory of elementary errors combined by a multiplicative process, just as the normal distribution arises out of a theory of errors combined additively. The probability density function for the lognormal is given by

$$\begin{aligned} f(x) &= 0 \text{ for } x \le 0 \\ f(x) &= \frac{1}{\sqrt{2\pi\sigma x}} e^{-(1nx-\mu)^2/2\sigma^2} \text{ for } x > 0. \end{aligned} \tag{14}$$

The shape of the lognormal distribution depends on the values of μ and σ^2. Its density function is shown graphically in Figure 5 for $\mu = 0$, $\sigma = 0.5$. The positive skew shown is characteristic of the lognormal distribution.

The lognormal distribution is likely to arise in situations in which there is a lower limit on the value which the random variable can assume, but no upper limit. Time measurements, which may extend from zero to infinity, are often described by the lognormal distribution. It has been applied to the distribution of income sizes, to the relative abundance of different species of animals, and has been assumed as the underlying distribution for various discrete counts in biology. As its name implies, it can be normalized by transforming the variable by the use of logarithms. See Aitchison and Brown (1957) for a further discussion of the lognormal distribution.

Many other continuous distributions have been studied. Some of these, such as the uniform distribution, are of minor importance in environmental work. Others are encountered occasionally, such as the exponential distribution, which has been used to compute probabilities in connection with the expected failure rate of equipment. The distribution of times between occurrences of events in Poisson processes are described by the exponential distribution and it is important in the theory of such stochastic processes (Parzen, 1962). Further discussion of continuous distributions may be found in Freund (1962) or most other standard statistical texts.

A special distribution problem often encountered in environmental work is concerned with the occurrence of extreme values of variables described by any one of several distributions. For example, in forecasting floods in connection with planning of construction, or droughts in connection with such problems as stream pollution, concern is with the most extreme values to be expected. To deal with such problems, the asymptotic theory of extreme values of a statistical variable has been developed. Special tables have been developed for estimating the expected extreme values for several distributions which are unlimited in the range of values which can be taken on by their extremes. Some information is also available for distributions with restricted ranges. An interesting application of this theory to prediction of the occurrence of unusually high tides may be found in Pfafflin (1970) and the Delta Commission Report (1960) Further discussion may be found in Gumbel.

HYPOTHESIS TESTING

Sampling Considerations

A basic consideration in the application of statistical procedures is the selection of the data. In parameter estimation and hypothesis testing sample data are used to make inferences to some larger population. The data are assumed to be a random sample from this population. By random we mean that the sample has been selected in such a way that the probability of obtaining any particular sample value is the same as its probability in the sampled population. When the data are taken care must be used to insure that the data are a random sample from the population of interest, and make sure that there must be no biases in the selective process which would make the samples unrepresentative. Otherwise, valid inferences cannot be made from the sample to the sampled population.

The procedures necessary to insure that these conditions are met will depend in part upon the particular problem being studied. A basic principle, however, which applies in all experimental work is that of randomization. Randomization means that the sample is taken in such a way that any uncontrolled variables which might affect the results have an equal chance of affecting any of the samples. For example, in agricultural studies when plots of land are being selected, the assignment of different experimental conditions to the plots of land should be done randomly, by the use of a table of random numbers or some other randomizing process. Thus, any differences which arise between the sample values as a result of differences in soil conditions will have an equal chance of affecting each of the samples.

Randomization avoids error due to bias, but it does nothing about uncontrolled variability. Variability can be reduced by holding constant other parameters which may affect the experimental results. In a study comparing the smog-producing effects of natural and artificial light, other variables, such as temperature, chamber dilution, and so on, were held constant (Laity, 1971) Note, however, that such control also restricts generalization of the results to the conditions used in the test.

Special sampling techniques may be used in some cases to reduce variability. For example, suppose that in an agricultural experiment, plots of land must be chosen from three different fields. These fields may then be incorporated explicitly into the design of the experiment and used as control variables. Comparisons of interest would be arranged so that they can be made within each field, if possible. It should be noted that the use of control variables is not a departure from randomization. Randomization should still be used in assigning conditions within levels of a control variable. Randomization is necessary to prevent bias from variables which are not explicitly controlled in the design of the experiment.

Considerations of random sampling and the selection of appropriate control variables to increase precision of the experiment and insure a more accurate sample selection can arise in connection with all areas using statistical methods. They are particularly important in certain environmental areas, however. In human population studies great care must be taken in the sampling procedures to insure representativeness of the samples. Simple random sampling techniques are seldom adequate and more complex procedures, have been developed. For further discussion of this kind of sampling, see Kish (1965) and Yates (1965). Sampling problems arise in connection with inferences from cloud seeding experiments which may affect the generality of the results (Bernier, 1967). Since most environmental experiments involve variables which are affected by a wise variety of other variables, sampling problems, especially the question of generalization from experimental results, is a very common problem. The specific randomization procedures, control variables and limitations on generalization of results will depend upon the particular field in question, but any experiment in this area should be designed with these problems in mind.

Parameter Estimation

A common problem encountered in environmental work is the estimation of population parameters from sample values. Examples of such estimation questions are: What is the "best" estimate of the mean of a population: Within what range of values can the mean safely be assumed to lie?

In order to answer such questions, we must decide what is meant by a "best" estimate. Probably the most widely used method of estimation is that of maximum likelihood, developed by Fisher (1958). A maximum likelihood estimate is one which selects that parameter value for a distribution describing a population which maximizes the probability of obtaining the observed set of sample values, assuming random sampling. It has the advantages of yielding estimates which fully utilize the information in the sample, if such estimates exist, and which are less variable under certain conditions for large samples than other estimates.

The method consists of taking the equation for the probability, or probability density function, finding its maximum value, either directly or by maximizing the natural logarithm of the function, which has a maximum for the same parameter values, and solving for these parameter values. The sample mean, $\hat{\mu} = (\Sigma_{n}^{i=1} x_i)/N\theta$, is a maximum likelihood estimate of the true mean of the distribution for a number of distributions. The variance, $\hat{\sigma}^2$, calculated from the sample by $\hat{\sigma}^2 = (\Sigma_{i=1}^{n} (x_i - \hat{\mu})^2$, is a maximum likelihood estimate of the population σ^2 for the normal distribution.

Note that such estimates may not be the best in some other sense. In particular, they may not be unbiased. An unbiased estimate is one whose value will, on the average, equal that of the parameter for which it is an estimate, for repeated sampling. In other words, the expected value of an unbiased estimate is equal to the value of the parameter being estimated. The variance is, in fact, biased. To obtain an unbiased estimate of the population variance it is necessary to multiply σ^2 by $n/(n-1)$, to yield s^2, the sample variance, and $s, (+\sqrt{s^2})$ the sample standard deviation.

There are other situations in which the maximum likelihood estimate may not be "best" for the purposes of the investigator. If a distribution is badly skewed, use of the mean as a measure of central tendency may be quite misleading. It is common in this case to use the median, which may be defined as the value of the variable which divides the distribution into two equal parts. Income statistics, which are strongly skewed positively, commonly use the median rather than the mean for this reason.

If a distribution is very irregular, any measure of central tendency which attempts to base itself on the entire range of scores may be misleading. In this case, it may be more useful to examine the maximum points of $f(x)$; these are known as modes. A distribution may have 1, 2 or more modes; it will then be referred to as unimodal, bimodal, or multimodal, respectively.

Other measures of dispersion may be used besides the standard deviation. The probable error, p.e., has often been used in engineering practice. It is a number such that

$$\int_{\mu-p.e.}^{\mu+p.e.} f(x)dx = 0.5. \tag{15}$$

The p.e. is seldom used today, having been largely replaced by σ^2.

The interquartile range may sometimes be used for a set of observations whose true distribution is unknown. It consists of the limits of the range of values which include the middle half of sample values. The interquartile range is less sensitive than the standard deviation to the presence of a few very deviant data values.

The sample mean and standard deviation may be used to describe the most likely true value of these parameters, and to place confidence limits on that value. The standard error of the mean is given by $s/\sqrt{n}$ (n = sample-size). The standard error of the mean can be used to make a statement about the probability that a range of values will include the true mean. For example, assuming normality, the range of values defined by the observed mean $1.96s/\sqrt{n}$ will be expected to include the value of the true mean in 95% of all samples.

A more general approach to estimation problems can be found in Bayseian decision theory (Pratt *et al.*, 1965). It is possible to appeal to decision theory to work out specific answers to the "best estimate" problem for a variety of decision criteria in specific situations. This approach is well described in Weiss (1961). Although the method is not often applied in routine statistical applications, it has received attention in systems analysis problems and has been applied to such environmentally relevant problems as resource allocation.

Frequency Data

The analysis of frequency data is a problem which often arises in environmental work. Frequency data for a hypothetical experiment in genetics are shown in Table 1. In this example, the expected frequencies are assumed to be known independently of the observed frequencies. The chi-square statistic, χ^2, is defined as

$$\chi^2 = \sum \frac{(E-O)^2}{E^2} \tag{16}$$

where E is the expected frequency and O is the observed frequency. It can be applied to frequency tables, such as that shown in Table 1. Note that an important assumption of the chi-square test is that the observations be independent. The same samples or individuals must not appear in more than one cell.

In the example given above, the expected frequencies were assumed to be known. In practice this is very often not the case; the experimenter will have several sets

TABLE 1

Hypothetical data on the frequency of plants producing red, pink and white flowers in the first generation of an experiment in which red and white parent plants were crossed, assuming single gene inheritance, neither gene dominant of observed frequencies, and will wish to determine whether or not they represent samples from one population, but will not know the expected frequency for samples from that population.

		Flower color		
		Red	Pink	White
Number of plants	expected	25	50	25
	observed	28	48	24

In situations where a two-way categorization of the data exists, the expected values may be estimated from the marginals. For example, the formula for chi-square for the fourfold contingency table shown below is

	Classification II	
Classification I	A	B
	C	D

$$\chi^2 = \frac{N\left(|AD - BC| - \frac{N}{2}\right)^2}{A\cdot B\cdot C\cdot D}. \tag{17}$$

Observe that instead of having independent expected values, we are now estimating these parameters from the marginal distributions of the data. The result is a loss in the degrees of freedom for the estimate. A chi-square with four independently obtained expected values would have four degrees of freedom; the fourfold table above has only one. The concept of degrees of freedom is a very general one in statistical analysis. It is related to the number of observations which can vary independently of each other. When expected values for chi-square are computed from the marginals, not all of the $O - E$ differences in a row or column are independent, for their discrepancies must sum to zero. Calculation of means from sample data imposes a similar restriction; since the deviations from the mean must sum to zero, not all of the observations in the sample can be regarded as freely varying. It is important to have the correct number of degrees of freedom for an estimate in order to determine the proper level of significance; many statistical tables require this information explicitly, and it is implicit in any comparison. Calculation of the proper degrees of freedom for a comparison can become complicated in specific cases, especially that of analysis of variance. The basic principle to remember, however, is that any linear independent constraints placed on the data will reduce the degrees of freedom. Tables for value of the χ^2 distribution for various degrees of freedom are readily available. For a further discussion of the use of chi-square, see Snedecor.

Difference between Two Samples

Another common situation arises when two samples are taken, and the experimenter wishes to know whether or not they are samples from populations with the same parameter values. If the populations can be presumed to be normal, then the significance of the differences of the two means can be tested by

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{18}$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the sample means, s_1^2 and s_1^2 are the sample variances, N_1 and N_2 are the sample sizes. and the population variances are assumed to be equal. This is the **t**-test, for two samples. The **t**-test can also be used to test the significance of the difference between one sample mean and a theoretical value. Tables for the significance of the **t**-test may be found in most statistical texts.

The theory underlying the **t**-test is that the measures of dispersion estimated from the observations within a sample provide estimates of the expected variability. If the means are close together, relative to that variability, then it is unlikely that the populations differ in their true values. However, if the means vary widely, then it is unlikely that the samples come from distributions with the same underlying distributions. This situation is diagrammed in Figure 6. The **t**-test permits an exact statement of how unlikely the null hypothesis (assumption of no difference) is. If it is sufficiently unlikely, it can be rejected. It is common to assume the null hypothesis unless it can be rejected in at least 95% of the cases, though more stringent criteria (99% or more) may be adopted if more certainty is needed.

The more stringent the criterion, of course, the more likely it is that the null hypothesis will be accepted when, in fact, it is false. The probability of falsely rejecting the null hypothesis is known as a type I error. Accepting the null hypothesis when it should be rejected is known as a type II error. For a given type I error, the probability of correctly rejecting the null hypothesis for a given true difference is known as the power of the test for detecting the difference. The function of these probabilities for various true differences in the parameter under test is known as the power function of the test. Statistical tests differ in their power and power functions are useful in the comparison of different tests.

Note that type I and type II errors are necessarily related; for an experiment of a given level of precision, decreasing the probability of a type I error raises the probability of a type II error, and vice versa. Thus, increasing the stringency of one's criterion does not decrease the overall probability of an erroneous conclusion; it merely changes the type of error which is most likely to be made. To decrease the overall error, the experiment must be made more precise, either by increasing the number of observations, or by reducing the error in the individual observations.

Many other tests of mean difference exist besides the **t**-test. The appropriate choice of a test will depend on the assumptions made about the distribution underlying the observations. In theory, the **t**-test applies only for variables which are continuous, range from ± infinity in value, and

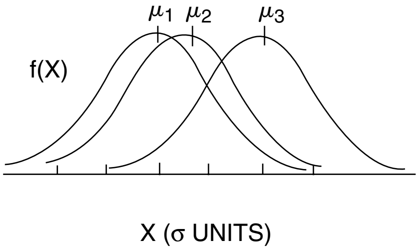


FIGURE 6

are normally distributed with equal variance assumed for the underlying population. In practice, it is often applied to variables of a more restricted range, and in some cases where the observed values of a variable are inherently discontinuous. However, when the assumptions of the test are violated, or distribution information is unavailable, it may be safer to use nonparametric tests, which do not depend on assumptions about the shape of the underlying distribution. While nonparametric tests are less powerful than parametric tests such as the **t**-test, when the assumptions of the parametric tests are met, and therefore will be less likely to reject the null hypothesis, in practice they yield results close to the **t**-test unless the assumptions of the **t**-test are seriously violated. Nonparametric tests have been used in meteorological studies because of nonnormality in the distribution of rainfall samples. (Decker and Schickedanz, 1967). For further discussions of hypothesis testing, see Hoel (1962) and Lehmann (1959). Discussions of nonparametric tests may be found in Pierce (1970) and Siegel (1956).

Analysis of Variance (ANOVA)

The **t**-test applies to the comparison of two means. The concepts underlying the **t**-test may be generalized to the testing of more than two means. The result is known as the analysis of variance. Suppose that one has several samples. A number of variances may be estimated. The variance of each sample can be computed around the mean for the sample. The variance of the sample means around the grand mean of all the scores gives another variance. Finally, one can ignore the grouping of the data and complete the variance for all scores around the grand mean. It can be shown that this “total” variance can be regarded as made up of two independent parts, the variance of the scores about their sample means, and the variance of these means about the grand mean. If all these samples are indeed from the same population, then estimates of the population variance obtained from within the individual groups will be approximately the same as that estimated from the variance of sample means around the grand mean. If, however, they come from populations which are normally distributed and have the same standard deviations, but different means, then the variance estimated from the sample means will exceed the variance are estimated from the within sample estimates.

The formal test of the hypothesis is known as the F-test. It is made by forming the F-ratio.

$$F = \frac{\text{MSE}_{(1)}}{\text{MSE}_{(2)}} \tag{19}$$

Mean square estimates (MSE) are obtained from variance estimates by division by the appropriate degrees of freedom. The mean square estimate in the numerator is that for the hypothesis to be tested. The mean square estimate in the denominator is the error estimate; it derives from some source which is presumed to be affected by all sources of variance which affect the numerator, except those arising from the hypothesis under test. The two estimates must also be independent of each other. In the example above, the within group MSE is used as the error estimate; however, this is often not the case for more complex experimental designs. The appropriate error estimate must be determined from examination of the particular experimental design, and from considerations about the nature of the independent variables whose effect is being tested; independent variables whose values are fixed may require different error estimates than in the case of independent variables whose values are to be regarded as samples from a larger set. Determination of degrees of freedom for analysis of variance goes beyond the scope of this paper, but the basic principle is the same as previously discussed; each parameter estimated from the data (usually means, for (ANOVA) in computing an estimator will reduce the degrees of freedom for that estimate.

The linear model for such an experiment is given by

$$X_{ij} = \mu + G_i + e_{ij,} \tag{20}$$

Where X_{ij} is a particular observation, μ is the mean, G_i is the effect the Gth experimental condition and e_{ij} is the error uniquely associated with that observation. The e_{ij} are assumed to be independent random samples from normal distributions with zero mean and the same variances. The analysis of variance thus tests whether various components making up a score are significantly different from zero.

More complicated components may be presumed. For example, in the case of a two-way table, the assumed model might be

$$X_{ijk} = \mu + R_{\mathrm{i}} + C_j + R_{cij} + \mathrm{e}_{ijk} \tag{21}$$

In addition to having another condition, or main effect, there is a term RC_{ij} which is associated with that particular combination of levels of the main effects. Such effects are known as interaction effects.

Basic assumptions of the analysis of variance are normality and homogeneity of variance. The **F**-test however, has been shown to be relatively “robust” as far as deviations from the strict assumption of normality go. Violations of the assumption of homogeneity of variance may be more serious. Tests have been developed which can be applied where violations of this assumption are suspected. See Scheffé (1959; ch.10) for further discussion of this problem.

Innumerable variations on the basic models are possible. For a more detailed discussion, see Cochran and Cox (1957) or Scheffé (1959). It should be noted, especially, that a significant F-ratio does not assure that all the conditions which entered into the comparison differ significantly from each other. To determine which mean differences are significantly different, additional tests must be made. The problem of multiple comparisons among several means has been approached in three main ways; Scheffé’s method for post-hoc comparisons; Tukey’s gap test; and Duncan’s multiple range test. For further discussion of such testing, see Kirk (1968).

Computational formulas for ANOVA can be found in standard texts covering this topic. However, hand calculation becomes cumbersome for problems of any complexity, and a number of computer programs are available for analyzing various designs. The Biomedical Statistical Programs (Ed. by Dixon 1967) are frequently used for this purpose. A method recently developed by Fowlkes (1969) permits a particularly simple specification of the design problem and has the flexibility to handle a wide variety of experimental designs.

SPECIAL ESTIMATION PROBLEMS

The estimation problems we have considered so far have involved single experiments, or sets of data. In environmental work, the problem of arriving at an estimate by combining the results of a series of tests often arises. Consider, for example, the problem of estimating the coliform bacteria population size in a specimen of water from a series of dilution tests. Samples from the water specimen are diluted by known amounts. At some point, the dilution becomes so great that the lactose broth brilliant green bile test for the presence of coliform bacteria becomes negative (Fair and Geyer, 1954). From the amount of dilution necessary to obtain a negative test, plus the assumption that one organism is enough to yield a positive response, it is possible to estimate the original population size in the water specimen.

In making such an estimate, it is unsatisfactory simply to use the first negative test to estimate the population size. Since the diluted samples may differ from one another, it is possible to get a negative test followed by one or more positive tests. It is desirable, rather, to estimate the population from the entire series of tests. This can be done by setting up a combined hypothesis based on the joint probabilities of all the obtained results, and using likelihood estimation procedures to arrive at the most likely value for the population parameter, which is known as the Most Probable Number (MPN) (Fair and Geyer, 1954). Tables have been prepared for estimating the MPN for such tests on this principle, and similar procedures can be used to arrive at the results of a set of tests in other situations.

Sequential testing is a problem that sometimes arises in environmental work. So far, we have assumed that a constant amount of data is available. However, very often, the experimenter is making a series of tests, and wishes to know whether he has enough data to make a decision at a given level of reliability, or whether he should consider taking additional data. Such estimation problems are common in quality control, for example, and may arise in connection with monitoring the effluent from various industrial processes. Statistical procedures have been developed to deal with such questions. They are discussed in Wald.

CORRELATION AND RELATED TOPICS

So far we have discussed situations involving a single variable. However, it is common to have more than one type of measure available on the experimental units. The simplest case arises where values for two variables have been obtained, and the experimenter wishes to know how these variables relate to one another.

Curve Fitting

One problem which frequently arises in environmental work is the fitting of various functions to bivariate data. The simplest situation involves fitting a linear function to the data when all of the variability is assumed to be in the *Y* variable. The most commonly used criterion for fitting such a function is the minimization of the squared deviations from the line, referred to as the least squares criterion. The application of this criterion yields the following simultaneous equations:

$$\sum_{i=1}^{n} Y_i = nA + \sum_{i=1}^{n} X_i \tag{22}$$

and

$$\sum_{i=1}^{n} X_i Y_i = A\sum_{i=1}^{n} X_i + B\sum_{i=1}^{n} X_i^2 . \tag{22}$$

These equations can be solved for *A* and *B,* the intercept and slope of the best fit line. More complicated functions may also be fitted, using the least squares criterion, and it may be generalized to the case of more than two variables. Discussion of these procedures may be found in Daniel and Wood.

Correlation and Regression

Another method of analysis often applied to such data is that of correlation. Suppose that our two variables are both normally distributed. In addition to investigating their individual distributions, we may wish to consider their joint occurrence. In this situation, we may choose to compute the Pearson product moment correlation between the two variables, which is given by

$$r_{xy} = \frac{\operatorname{cov}(x_i y_i)}{\sigma_x \sigma_y} \tag{23}$$

where $\operatorname{cov}(x_i\, y_i)$ the covariance of *x* and *y,* is defined as

$$\sum_{i=1}^{n} \frac{(x_i - \mu_x)(y_i - \mu_y)}{n}. \tag{24}$$

It is the most common measure of correlation. The square of *r* gives the proportion of the variance associated with one of the variables which can be predicted from knowledge of the other variables. This correlation coefficient is appropriate whenever the assumption of a normal distribution can be made for both variables.

Another way of looking at correlation is by considering the regression of one variable on another. Figure 7 shows the relation between two variables, for two sets of bivariate data, one with a 0.0 correlation, the other with a correlation of 0.75. Obviously, estimates of type value of one variable based on values of the other are better in the case of the higher correlation. The formula for the regression of *y* on *x* is given by

$$\frac{y-\hat{\mu}_y}{\hat{\sigma}_y}=\frac{r_{xy}(x-\hat{\mu}_x)}{(\hat{\sigma}_x)}. \qquad (25)$$

A similar equation exists for the regression of *x* on *y.*

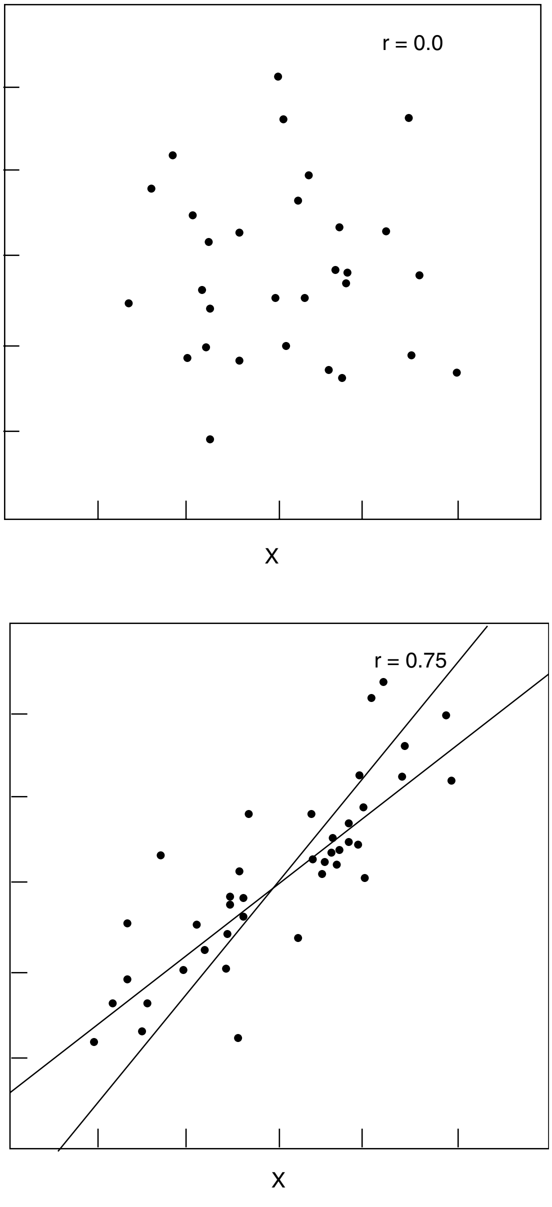


FIGURE 7

A number of other correlation measures are available. For ranked data, the Spearman correlation coefficient, or Kendall's tau, are often used. Measures of correlation appropriate for frequency data also exist. See Siegel.

MULTIVARIATE ANALYSIS

Measurements may be available on more than two variables for each experiment. The environmental field is one which offers great potential for multivariate measurement. In areas of environmental concern such as water quality, population studies, or the study of the effects of pollutants on organisms, to name only a few, there are often several variables which are of interest. The prediction of phenomena of environmental interest, or such as rainfall, or floods, typically involves the consideration of many variables. This section will be concerned with some problems in the analysis of multivariate data.

Multivariate Distributions

In considering multivariate distributions, it is useful to define the *n*-dimensional random variable *X* as the vector

$$X'=[X_1,X_2,\mathrm{K},X_n]. \qquad (26)$$

The elements of this vector will be assumed to be continuous unidimensional random variables, with density functions $f_1(x_1)$,F$f_2(x_2)$K,$f_n(x_n)$ and distribution functions $F_1(x_1)$,$F_2(x_2)$K,$F_n(x_n)$ Such a vector also has a joint distribution function.

$$F(x_1,x_2,\mathrm{K},x_n)=P(X_1\le x_1,\mathrm{K},X_n\le x_n) \qquad (27)$$

where *P* refers to the probability of all the stated conditions occurring simultaneously.

The concepts considered previously in regard to univariate distribution may be generalized to multivariate distributions. Thus, the expected value of the random vector, *X,* analogous to the mean of the univariate distribution, is

$$E(X')=[E(X_1),E(X_2),\mathrm{K}E(X_n)], \qquad (28)$$

where the $E(X_i)$ are the expected values, or means, for the univariate distributions.

Generalization of the concept of variance is more complicated. Let us start by considering the covariance between two variables,

$$\sigma_{ij}=E[X_i-E(X_i)][X_j-E(X_j)]. \qquad (29)$$

The covariances between each of the elements of the vector *X* can be computed; the covariances of the *i*th and *j*th elements will be designed as σ_{ij} If $i = j$ the covariance is the variance of X_i, and will be designed as σ_{ij} The generalization of the concept of variance to a multidimensional variable then becomes the matrix of variances and covariances. This matrix will be called the covariance matrix. The covariance matrix for the population is given as

$$\Sigma = \begin{bmatrix} \sigma_{11}\sigma_{12}\mathrm{K}\sigma_{1n} \\ \sigma_{21}\sigma_{22}\mathrm{K}\sigma_{2n} \\ \mathrm{K\ K\ K\ K\ K} \\ \sigma_{n1}\sigma_{2n2}\ldots\sigma_{nn} \end{bmatrix}. \tag{30}$$

A second useful matrix is the matrix of correlations

$$\begin{bmatrix} r_{11}\Lambda\, r_{1n} \\ r_{21}\Lambda\, r_{2n} \\ \Lambda\ \Lambda\ \Lambda \\ r_{n1}\ldots r_{nn} \end{bmatrix} \tag{31}$$

If the assumption is made that each of the individual variables is described by a normal distribution, then the distribution of X may be described by the multivariate normal distribution. This assumption will be made in subsequent discussion, except where noted to the contrary.

Tests on Means

Suppose that measures have been obtained on several variables for a sample, and it is desired to determine whether that sample came from some known population. Or there may be two samples; for example, suppose data have been gathered on physiological effects of two concentrations of SO_2 for several measures of physiological functioning and the investigator wishes to know if they should be regarded as samples from the same population. In such situations, instead of using **t**-tests to determine the significance of each individual difference separately, it would be desirable to be able to perform one test, analogous to the **t**-test, on the vectors of the means.

A test, known as Hotelling's $\mathbf{T}^2$ test, has been developed for this purpose. The test does not require that the population covariance matrix be known. It does, however, require that samples to be compared come from populations with the same covariance matrix, an assumption analogous to the constant variance requirement of the **t**-test.

To understand the nature of $\mathbf{T}^2$ in the single sample case, consider a single random variable made up of any linear combination of the n variables in the vector X (all of the variables must enter into the combination, that is, none of the coefficients may be zero). This variable will have a normal distribution, since it is a sum of normal variables, and it can be compared with a linear combination of elements from the vector for the population with the same coefficients, by means of a **t**-test. We then adopt the decision rule that the null hypothesis will be accepted only if it is true for all possible linear combinations of the variables. This is equivalent to saying that it is true for the largest value of **t** as a function of the linear combinations. By maximizing $\mathbf{t}^2$ as a function of the linear combinations, it is possible to derive $\mathbf{T}^2$. Similar arguments can be used to derive $\mathbf{T}^2$ for two samples.

A related function of the mean is known as the linear discriminant function. The linear discriminant function is defined as the linear compound which generates the largest $\mathbf{T}^2$ value. The coefficients used in this compound provide the best weighting of the variables of a multivariate observation for the purpose of deciding which population gave rise to an observation. A limitation on the use of the linear discriminant function, often ignored in practice, is that it requires that the parameters of the population be known, or at least be estimated from large samples. This statistic has been used in analysis of data from monitoring stations to determine whether pollution concentrations exceed certain criterion values.

Other statistical procedures employing mean vectors are useful in certain circumstances. See Morrison for a further discussion of this question.

Multivariate Analysis of Variance (MANOVA)

Just as the concepts underlying the **t**-test could be generalized to the comparison of more than two means, the concepts underlying the comparison of two mean vectors can be generalized to the comparison of several vectors of means.

The nature of this generalization can be understood in terms of the linear model, considered previously in connection with analysis of variance. In the multivariate situation, however, instead of having a single observation which is hypothesized to be made up of several components combined additively, the observations are replaced by vectors of observations, and the components by vectors of components. The motivation behind this generalization is similar to that for Hotelling's $\mathbf{T}^2$ test: it permits a test of the null hypothesis for all of the variables considered simultaneously.

Unlike the case of Hotelling's $\mathbf{T}^2$, however, various methods of test construction do not converge on one test statistic, comparable to the F test for analysis of variance. At least three test statistics have been developed for MANOVA, and the powers of the various tests in relation to each other are very incompletely known.

Other problems associated with MANOVA are similar in principle to those associated with ANOVA, though computationally they are more complex. For example, the problem of multiple comparison of means has its analogous problem in MANOVA, that of determining which combinations of mean vectors are responsible for significant test statistics. The number and type of possible linear models can also ramify considerably, just as in the case of ANOVA. For further discussion of MANOVA, see Morrison (1967) or Seal.

Extensions of Correlation Analysis

In a number of situations, where multivariate measurements are taken, the concern of the investigator centers on the prediction of one of the variables. When rainfall measurements are taken in conjunction with a number of other variables, such as temperature, pressure, and so on, for example, the purpose is usually to predict the rainfall as a function of the other variables. Thus, it is possible to view one variable as the dependent variable for *a prior* reasons, even though the data do not require such a view.

In these situations, the investigator very often has one of two aims. He may wish to predict one of the variables from all of the other variables. Or he may wish to consider one variable as a function of another variable with the effect of all the other variables partialled out. The first situation calls for the use of multiple correlation. In the second, the appropriate statistic is the partial correlation coefficient.

Multiple correlation coefficients are used in an effort to improve prediction by combining a number of variables to predict the variable of interest. The formula for three variables is

$$r_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}. \tag{32}$$

Generalizations are available for larger numbers of variables. If the variables are relatively independent of each other, multiple correlation may improve prediction. However, it should be obvious that this process reaches an upper limit since additional variables, if they are to be of any value, must show a reasonable correlation with the variable of interest, and the total amount of variance to be predicted is fixed. Each additional variable can therefore only have a limited effect.

Partial correlation is used to partial out the effect of one of more variables on the correlation between two other variables. For example, suppose it is desired to study the relationship between body weight and running speed, independent of the effect of height. Since height and weight are correlated, simply doing a standard correlation between running speed and weight will not solve the problem. However, computing a partial correlation, with the contribution of height partialled out, will do so. The partial correlation formula for three variables is

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{1 - r_{13}^2}\sqrt{1 - r_{23}^2}}, \tag{33}$$

where $r_{12.3}$ gives the correlation of variables 1 and 2, with the contribution of variable 3 held constant. This formula may also be extended to partial out the effect of additional variables.

Let us return for a moment to a consideration of the population correlation matrix, *p*. It may be that the investigator has same *a priori* reason for believing that certain relationships exist among the correlations in this matrix. Suppose, for example there is a reason to believe that several variables are heavily dependent on wind velocity and that another set of variables are dependent on temperature. Such a pattern of underlying relations would result in systematic patterns of high and low correlations in the population matrix, which should be reflected in the observed correlation matrix. If the obtained correlation matrix is partitioned into sets in accordance with the *a priori* hypothesis, test for the independence of the sets will indicate whether or not the hypothesis should be rejected. Procedures have been developed to deal with this situation, and also to obtain coefficients reflecting the correlation between sets of correlations. The latter procedure is known as canonical correlation. Further information about these procedures may be found in Morrison.

Other Analyses of Covariance and Correlation Matrices

In the analyses discussed so far, there have been *a priori* considerations guiding the direction of the analysis. The situation may arise, however, in which the investigator wishes to study the patterns in an obtained correlation or covariance matrix without any appeal to *a priori* considerations. Let us suppose, for example, that a large number of measurements relevant to weather prediction have been taken, and the investigator wishes to look for patterns among the variables. Or suppose that a large number of demographic variables have been measured on a human population. Again, it is reasonable to ask if certain of these variables show a tendency to be more closely related than others, in the absence of any knowledge about their actual relations. Such analyses may be useful in situations where large numbers of variables are known to be related to a single problem, but the relationships among the variables are not well understood. An investigation of the correlation patterns may reveal consistencies in the data which will serve as clues to the underlying process.

The classic case for the application of such techniques has been the study of the human intellect. In this case, correlations among performances on a very large number of tasks have been obtained and analyzed, and many theories about the underlying skills necessary for intellectual function have been derived from such studies. The usefulness of the techniques are by no means limited to psychology, however. Increasingly, they are being applied in other fields, as diverse as biology (Fisher and Yates, 1964) and archaeology (Chenhall, 1968). Principal component analysis, a closely related technique, has been used in hydrology.

One of the more extensively developed techniques for the analysis of correlation matrices is that of factor analysis. To introduce the concepts underlying factor analysis, imagine a correlation matrix in which the first x variables and the last $n - x$ variables are all highly correlated with each other, but the correlation between any of the first x and any of the second $n - x$ variables is very low. One might suspect that there is some underlying factor which influences the first set of variables, and another which influences the second set of variables, and that these two factors are relatively independent statistically, since the variables which they influence are not highly correlated. The conceptual simplification is obvious; instead of worrying about the relationships among n variables as reflected in their $n(n - 1)/2$ correlations, the investigator can attempt to identify and measure the factors directly.

Factor analysis uses techniques from matrix algebra to accomplish mathematically the process we have outlined intuitively above. It attempts to determine the number of factors, and also the extent to which each of these factors influences the measured variables. Since unique solutions to this problem do not exist, the technique has been the subject of considerable debate, especially on the question of how to determine the best set of factors. Nevertheless, it can be useful in any situation where the relationships among a large set of variables is not well understood.

ADDITIONAL PROCEDURES

Multidimensional Scaling and Clustering

There are a group of techniques whose use is motivated by considerations similar to those underlying the analysis of correlation matrices, but which are applied directly to matrices of the distances, or similarities, between various stimuli. Suppose, for example, that people have been asked to judge the similarity of various countries. These judgments may be scaled by multidimensional techniques to discover how many dimensions underlie the judgments. Do people make such judgments along a single dimension? Or are several dimensions involved? An interesting example of this sort was recently analyzed by Wish (1972). Sophisticated techniques have been worked out for such procedures.

Multidimensional scaling has been most extensively used in psychology, where the structure underlying similarity or distance measurements may not be at all obvious without such procedures. Some of these applications are of potential importance in the environmental field, especially in areas such as urban planning, where decisions must take into account human reactions. They are not limited to such situations however, and some intriguing applications have been made in other fields.

A technique related in some ways to multidimensional analysis is that of cluster analysis. Clustering techniques can be applied to the same sort of data as multidimensional scaling procedures. However, the aim is somewhat different. Instead of looking for dimensions assumed to underlie the data, clustering techniques try to define related clusters of stimuli. Larger clusters may then be identified, until a hierarchical structure is defined. If the data are sufficiently structured, a “Tree” may be derived.

A wide variety of clustering techniques have been explored, and interest seems on the increase (Johnson, 1967). The procedures used depend upon the principles used to define the clusters. Clustering techniques have been applied in a number of different fields. Biologists have used them to study the relationships among various animals; for example, a kind of numerical taxonomy.

The requirements which the data must meet for multidimensional scaling and clustering procedures to apply are usually somewhat less stringent than in the case of the multivariate procedures discussed previously. Multidimensional scaling in psychology is often done on data for which an interval scale of measurement cannot be assumed. Distance measures for clustering may be obtained from the clustering judgments of a number of individuals which lack an ordinal scale. This relative freedom is also useful in many applications where the order of items is known, but the equivalence of the distances between items measured at different points is questionable.

Stochastic Processes

A stochastic or random process is any process which includes a random element in its description. The term stochastic process is frequently also used to describe the mathematical description of any actual stochastic process. Stochastic models have been developed in a number of areas of environmental concern.

Many stochastic processes involve space or time as a primary variable. Bartlett (1960) in his discussion of ecological frequency distributions begins with the application of the Poisson distribution to animal populations whose density is assumed to be homogeneous over space, and then goes on to develop the consequences of assuming heterogeneous distributions, which are shown to lead to other distributions, such as the negative binomial. The Sutton equation for the diffusion of gases applied to stack effluents, a simplification of which was given earlier for a single dimension (Strom, 1968) is another example of a situation in which statistical considerations about the physical process lead to a spatial model, in this case, one involving two dimensions.

Time is an important variable in many stochastic models. A number of techniques have been developed for the analysis of time series. Many of the concepts we have already considered, such as the mean and variance, can be generalized to time series. The autocorrelation function, which consists of the correlation of a function with itself for various time lags, is often applied to time series data. This function is useful in revealing periodicities in the data, which show up as peaks in the function. Various modifications of this concept have been developed to deal with data which are distributed in discrete steps over time. Time series data, especially discrete time series data, often arise in such areas as hydrology, and the study of air pollution, where sampling is done over time. Such sampling is often combined with spatial sampling, as when meterological measurements are made at a number of stations.

An important consideration in connection with time series is whether the series is stationary or non-stationary. Stationarity of a time series implies that the behavior of the random variables involved does not depend on the time at which observation of the series is begun. The assumption of stationarity simplifies the statistical treatment of time series. Unfortunately, it is often difficult to justify for environmental measurements, especially those taken over long time periods. Examination of time series for evidence of non-stationarity can be a useful procedure, however; for example, it may be useful in determining whether long term climatic changes are occurring (Quick, 1992). For further discussion of time series analysis, see Anderson.

Stochastic models of environmental interest are often multivariate. Mathematical models applied to air pollution may deal with the concentrations of a number of pollutants, as well as such variables as temperature, pressure, precipitation, and wind direction. Special problems arise in the evaluation of such models because of the large numbers of variables involved, the large amounts of data which must be processed for each variable, and the fact that the distributions of the variables are often nonnormal, or not well known. Instead of using analytic methods to obtain solutions, it may be necessary to seek approximate solutions; for example, by extensive tabulation of data for selected sets of conditions, as has been one in connection with models for urban air pollution.

The development of computer technology to deal with the very large amounts of data processing often required has made such approaches feasible today. Nevertheless, caution with regard to many stochastic models should be observed. It is common to find articles describing such models which state that a number of simplifying assumptions were necessary in order to arrive at a model for which computation was feasible, and which then go on to add that even with these assumptions the computational limits of available facilities were nearly exceeded, a combination which raises the possibility that excessive simplification may have been introduced. In these circumstances, less ambitious treatment of the data might prove more satisfactory. Despite these comments, however, it is clear that the environmental field presents many problems to which the techniques of stochastic modelling can be usefully applied.

ADDITIONAL CONSIDERATIONS

The methods outlined in the previous sections represent a brief introduction to the statistics used in environmental studies. It appears that the importance of some of these statistical methods, particularly analysis of variance, multivariate procedures and the use of stochastic modelling will increase. The impact of computer techniques has been great on statistical computations in environmental fields. Large amounts of data may be collected and processed by computer methods.

ACKNOWLEDGMENTS

The author is greatly indebted to D.E.B. Fowlkes for his many valuable suggestions and comments regarding this paper and to Dr. J. M. Chambers for his critical reading of sections of the paper.

REFERENCES

Aitchison, J. and J. A. Brown *The Lognormal Distribution,* Cambridge Univ. Press, 1957.

Anderson, T. W., *The Statistical Analysis of Time Series,* John Wiley and Sons, Inc., New York, 1971.

Bailey, N. T. J., *The Elements of Stochastic Processes: With Applications to the Natural Sciences,* John Wiley and Sons, Inc., New York, 1964.

Bartlett, M. S., *An Introduction to Stochastic Processes,* 2nd Ed., Cambridge Univ. Press, 1966.

Bartlett, M. S., *Stochastic Population Models in Ecology and Epidemiology,* Methuen and Co., Ltd, London, John Wiley and Sons, Inc., New York, 1960.

Bernier, J., On the design and evaluation of cloud seeding experiments performed by Electricite de France, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* **5,** Lecam, L. M. and J. Neyman, Eds., University of California Press, Berkeley, 1967, p. 35.

Castillo, E., *Extreme Value Theory in Engineering,* Academic Press, London, 1988.

Chernall, R. G., The impact of computers on archaeological theory, *Computers and the Humanities,* **3,** 1968. p. 15.

Cochran, W. G. and G. M. Cox, *Experimental Designs,* 2nd Ed., John Wiley and Sons, Inc., New York, 1957.

Coles, S. G., *An Introduction to Statistical Modelling of Extremes,* Springer, 2001.

Computational Laboratory Staff, *Tables on the Cumulative Binomial Probability Distribution,* Harvard University Press, Cambridge, MA., 1955.

Cooley, William W. and Paul R. Lohnes, *Multivariate Data Analysis,* John Wiley and Sons, Inc., New York, 1971.

Cox, B., J. Hunt, P. Mason, H. Wheater and P. Wolf, Eds., *Flood Risk in a Changing Climate,* Phil. Trans. of the Royal Society, 2002.

Cox, D. R., and D. V. Hinkley, A note on the efficiency of least squares estimates, *J. R. Statis. Soc.,* **B30,** 284–289.

Cramer, H., *The Elements of Probability Theory,* John Wiley and Sons, Inc., New York, 1955.

Cramer, H., *Mathematical Methods of Statistics,* Princeton University Press, Princeton, NJ, 1946.

Daniel, C. D. and F. S. Wood, *Fitting Equations to Data,* John Wiley and Sons, Inc., New York, 1971.

Decker, Wayne L. and Paul T. Schickedanz, The evaluation of rainfall records from a five year cloud seeding experiment in Missouri, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,* **5,** Lecam, L. M. and J. Neyman, Eds., Univ. of California Press, Berkeley, 1967, p. 55.

Rapport Delta Commissie, *Beschouwingen over Stormvloeden en Getijbeweging,* III 1 S Bigdragen Rijkwaterstaat, The Hague, 1960.

Dixon, W. J., Ed., *Biomedical Computer Programs,* Univ. of California Publications in Automatic Computation No. 2, Univ. of California Press, Berkeley, 1967.

Fair, G. M. and J. C. Geyer, *Water Supply and Wastewater Disposal,* John Wiley and Sons, Inc., New York, 1954.

Feller, W., *An Introduction to Probability Theory and Its Applications,* 3rd Ed., **1,** John Wiley and Sons, Inc., New York, 1968.

Fisher, N. I., *Statistical Analysis for Circular Data,* Cambridge University Press, 1973.

Fisher, R. A., *Statistical Methods for Research Workers,*13th Ed., Hafner, New York, 1958.

Fisher, R. A. and L. H.C. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proc. Camb. Phil. Soc., **24,** 180–190. 1928.

Fisher, R. A. and F. Yates, *Statistical Tables for Biological, Agricultural and Medical Research,* 6th Ed., Hafner, New York, 1964.

Fowlkes, E. B., Some operators for ANOVA calculations, *Technometrics,* **11,** 1969, p. 511.

Freund, J. E., *Mathematical Statistics,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1962.

General Electric Company, defense Systems Department, *Tables of the Individual and Cumulative Terms of Poisson Distribution,* Van Nostrand, Princeton, NJ, 1962.

Gumbel, E. J., *Statistics of Extremes,* Columbia University Press, New York, 1958.

Gumbel, E. J., *Statistical Theory of Extreme Values and Some Practical Applications,* Applied Mathematics Series No. 3, National Bureau of Standards, US Government Printing Office, Washington, DC, 1954.

Harman, H. H., *Modern Factor Analysis,* 2nd Ed., University of Chicago Press, Chicago, 1967.

Hoel, P. G., *Introduction to Mathematical Statistics,* 3rd Ed., John Wiley and Sons, Inc., New York, 1962.

Institute of Hydrology, *Flood Estimation Handbook,* Wallingford, UK, 1999.

Johnson, S. C., Hierarchical clustering schemes, *Pschometrika,* **32,** 1967, p. 241.

Jolicoeur, P. and J. E. Mosimann, Size and shape variation in the painted turtle: A principal component analysis, *Growth,* **24,** 1960, p. 339.

Kirk, R. E., *Experimental Design: Procedures for the Behavioral Sciences,* Brooks Cole Belmont, CA, 1968.

Kish, Leslie, *Survey Sampling,* John Wiley and Sons, Inc., Belmont, CA, 1965.

Laity, J. L., A smog chamber study comparing blacklight fluorescent lamps with natural sunlight, *Environmental Science and Technology,* **5,** 1971, p. 1218.

Lehmann, E. L., *Testing Statistical Hypotheses,* John Wiley and Sons, Inc., 1959.

Lieberman, G. J. and D. B. Owen, *Tables of the Hypergeometric Probability Distribution,* Stanford University Press, Stanford, CA, 1961.

Liebesman, B. S., *Decision Theory Applied to a Class of Resource Allocation Problems,* Doctoral dissertation, School of Engineering and Science, New York University, 1970.

MacInnis, C. and J. R. Pfafflin, Municipal wastewater, *The Encyclopedia of Environmental Science and Engineering,* **2,** 4th Ed., Gordon and Breach Science Publishers, 1998.

Molina, E. C., *Poisson's Exponential Binomial Limit,* Van Nostrand Company, Inc., New York, 1942.

Mardia, K. V. and P. E. Jupp, *Directional Statistics,* John Wiley and Sons, Inc., New York, 2000.

Morrison, D. F., *Multivariate Statistical Methods,* McGraw-Hill Book Co., New York, 1967.

Moses, H., Urban air pollution modeling, *The Encyclopedia of Environmental Science and Engineering,* **2,** 4th Ed., Gordon and Breach Science Publishers, New York, 1998.

Panofsky, Hans, Meteorology of air pollution, *The Encyclopedia of Environmental Science and Engineering, 1***,** 4th Ed., Gordon and Breach Science Publishers, New York, 1998.

Parzen, Emmanuel, *Stochastic Processes,* Holden-Day, Inc., San Francisco, 1962.

Pfafflin, J. R., A statistical approach to prediction of recurrence intervals of abnormally high tides, *Ocean Engineering,* **2,** Pergamon Press, Great Britain, 1970.

Pierce, A., *Fundamentals of Nonparametric Statistics,* Dickenson Pub. Co., Belmont, CA, 1970.

Pratt, J. W., H. Raiffa and R. Schlaifer, *Introduction to Statistical Decision Theory,* McGraw-Hill Book Co., New York, 1965.

Quick, Michael C., Hydrology, *The Encyclopedia of Environmental Science and Engineering,* **1,** 4th Ed., Gordon and Breach Science Publishers, New York, 1998.

Romig, H. G., *Binomial Tables,* John Wiley and Sons, Inc., New York, 1953.

Scheffer, H., *The Analysis of Variance,* John Wiley and Sons, Inc., 1959.

Scheffer, H., A method for judging all contrasts in the analysis of variance, *Biometrika,* **40,** 1953, p. 87.

Seal, Hilary L., *Multivariate Statistical Analysis for the Biologist,* Methuen and Co., London, 1964.

Siegel, S., *Nonparametric Statistics for the Behavioral Sciences,* McGraw-Hill, New York, 1956.

Snedecor, G. W. and W. G. Cochran, *Statistical Methods,* 6th Ed., The Iowa State University Press, Ames, 1967.

Strom, Gordon H., Atmospheric dispersion of stack effluents, *Air Pollution,* **1,** Stern, A. C., ed., 2nd Ed., Academic Press, New York, 1968.

Tables of the Binomial Probability Distribution, National Bureau of Standards, Applied Mathematics Series No. 6, US Govt. Printing Office, Washington, DC, 1950.

Wald, A., *Sequential Analysis,* John Wiley and Sons, Inc., New York, 1947.

Weiss, L., *Statistical Decision Theory,* McGraw-Hill Book Co., New York, 1961.

Wilks, D. S., *Statistical Methods in the Atmospheric Sciences,* International Geophysics Series, **59,** p. 467, Academic Press.

Wish, Myron, Nine dimensions of nation perception: Cross-culture and intertask variation, *Proceedings 80th Annual Convention, APA,* 1972.

Yates, Frank, *Sampling Methods for Censuses and Surveys,* 3rd. Ed., Hafner, New York, 1965.

SHEILA M. PFAFFLIN
AT&T

T

THE TERRESTRIAL SYSTEM

Space-vehicle Earth was so superbly well designed, equipped and supplied as to have been able to sustain human life aboard it for at least two million years—despite humanity's comprehensive ignorance which permitted a fearfully opinionated assumption of an infinitely extensive planar *World.* Humanity also spontaneously misassumed that its local pollution could be dispelled by the world's infinite extensiveness. Humanity also misassumed that an infinite succession of new and pleasing varieties of abundant, vital resources would be disclosed progressively as man exhausted first one and bespoiled another of the as yet known valuable, because vital, resources. Man must learn in a spontaneously self-enlightening manner to discard many, if not most, of yesterday's false premises, and axioms only believingly accepted; and he must, on his own, discard false premises and learn that only the non-sense Universe is reliable and that a lunatic is not a crazy man but one so sane, well informed, well coordinated, self-disciplined cooperative and fearless as to be the first Earthian human to have been ferried to a physical landing upon the Moon and thereafter to have been returned safely to reboard his mother space vehicle "Earth".

Long, long ago—little bands of humans seeking fish and fruits, or following animals, frequently became permanently lost and separated from one another. Endowed with the procreative urge, those of the few males and females surviving in company inbred for generations in their respective remotenesses utterly unaware of one another's tribes and separate tribal evolution—and thus evolved a plurality of superficial differences in appearance through special chromosomic concentration brought about by the special characteristics of the survival adaptation process. Thus have developed hundreds of only superficially different types, some very numerous and powerful, some successfully monopolizing specific land areas and others as yet wandering. In their ignorance, all of humanity's national governments assume, misinformedly, that there is not and never will be enough of the vital resources to support all or even a large number of humans, *ergo,* that they must automatically fight one another to the death to discover which government might survive. Often to encourage their respective peoples, political leaders evolve partly expedient and partly idealistic ideologies suitable to their viewpoints, but all the ideologies misassume an only-you-or-me—not both—survival premise as having no axiomatic alternative. Because of the invisibility of 99.9% of the source information that contradicts the assumption of a fundamental inadequacy of resources, the probability is that if man is left exclusively to political contriving he will become embroiled in approximately total self-destruction. While the top speed of the intercontinental ballistic rocket is many times that of a bullet, its 20,000 mph is as nothing beside radar-sight's speed of 70,000,000 mph. The speed of information is now so swift that for the first time in history the lethal missile is no longer faster than man's ability to apprehend both its coming and its specific course—twenty minutes before it can reach him. But the ability to see it coming does not confer the capability to dodge it. Now every one of the opposed political systems' swiftest rocketry attacks can be detected so far in advance that each and every side can dispatch, retaliatorily, not only its full arsenal of atomic warheads but also all its rocket-borne chemical and biological warfare missiles. All opposed sides can and will retaliate automatically *in toto,* thus bringing about approximately total human destruction of vast millions immediately, with the balance to be destroyed soon thereafter by the radiational, biological and chemical contamination.

As in our industrio-social age we now design everything *except* the astro-vehicle paraphernalia, all the metals that have ever been mined and put to use have an average quarter century recycling, invention-to-obsolescence periodicity which includes the scrap, melt, redesign and re-use cycling time. All the metals ever mined and so put to use are now invested in structures and machines that, if operated at full capacity, could take care of only 44% of humanity. The rate at which we have been finding and mining new metals is far slower than the rate of increase of human population. This means that if we freeze the world's design standards at their present levels, which are far below the standards of the astro-vehicle technology, 56% of humanity, which means humanity's political majority, is doomed to premature demise, and to want and suffering en route to that early death. There is nothing that politics, per se, can do to alter that condition;

only a design revolution—such as that which is already "made to order" in the potentially thousand-fold performance per pounds, minutes and kilowatts advancement to be realized by the astro-vehicle industry—can change those fundamental conditions of humanity overnight from failure to comprehensive world-around, human success.

Between 1900 and 1969 our space-vehicle Earth's passengers have experienced an increase of from less than 1% of its total population to 41% of total world population now enjoying a standard of living superior to that either experienced or dreamed of by any monarch before 1900. During that time the material resources per each world man were continually decreasing so that the advancement was not accomplished by exploiting more resources. This historical forty-folding of the percentage of humanity's "haves" can only be explained as the fallout of ever higher performance per pound technology as developed for the ships of the world's water and air oceans. That an over-night advancement from 40 to 100% is possible can be understood when we realize that the technological fall-out into our domestic economy of ships of extraterrestrial astrogation have not had time to have important effect on the standard of living because their technological fallout has not yet had time to occur.

It seems eminently clear that we not only must put our space programs on highest priority of attention and resource investment but that all humanity must be accredited and financed to enter into a new re-educational system that is geared to develop our most prominent awareness, that we indeed *are in space* and that all of our concern is with the fact that our space-vehicle Earth and its life-energy-giving Sun, and the tide-pumping Moon can provide ample sustenance and power for all humanity's needs to be derived from our direct energy income without further robbing our fossil fuels energy savings account. In reality, the Sun, the Earth and the Moon are nothing else than a most fantastically welldesigned and space-programmed team of vehicles. All of us *are,* always have been, and so long as we exist, *always will be—nothing else but—astronauts.* Let's pull our heads out of the brain benumbing, mind frustrating misinformedly conditioned reflexes. If it is going to be "All ashore who's going ashore," once more intent to return to nonspace DOWN HERE ON EARTH, humanity is doomed.

But there is hope in sight. The young! While the university students are intuitively skeptical of the validity of any and all evolution-blocking establishments, *ergo,* negatives, the high school age youth thinks spontaneously and positively in astro and electromagnetic technology and their realistic uses. The young of all age levels abhor hypocrisy. They are bored with obsolete UP and DOWN dancing, with bureaucratic inertia, bias of any kind or fear-built security. They disdain white, gray, black and blue lies. The students and school children around the world have idealistic compassion for all humanity. There is a good possibility that they may take over and successfully operate SPACESHIP EARTH. How may we use our intellectual capability to higher advantage? Our muscle is very meager as compared to the muscles of many animals. Our integral muscles are as nothing compared to the power of a tornado or the atom bomb which society contrived—in fear—out of the intellect's fearless discoveries of generalized principles governing the fundamental energy behaviors of physical universe.

In organizing our grand strategy we must first discover where we are now; that is, what our present navigational position in the universal scheme of evolution is. To begin our position-fixing aboard our Spaceship Earth we must first acknowledge that the abundance of immediately consumable, obviously desirable or utterly essential resources have been sufficient until now to allow us to carry on despite our ignorance. Being eventually exhaustible and spoilable, they have been adequate only up to this critical moment. This cushion-for-error of humanity's survival and growth up to now was apparently provided just as a bird inside of the egg is provided with liquid nutriment to develop it to a certain point. But then by design the nutriment is exhausted at just the time when the chick is large enough to be able to locomote on its own legs. And so as the chick pecks at the shell seeking more nutriment it inadvertently breaks open the shell. Stepping forth from its initial sanctuary, the young bird must now forage on its own legs and wings to discover the next phase of its regenerative sustenance.

My own picture of humanity today finds us just about to step out from amongst the pieces of our just one-second-ago broken eggshell. Our innocent, trial-and-error-sustaining nutriment is exhausted. We are faced with an entirely new relationship to the universe. We are going to have to spread our wings of intellect and fly or perish; that is, we must dare immediately to fly by the generalized principles governing universe and not by the ground rules of yesterday's superstitious and erroneously conditioned reflexes. And as we attempt competent thinking we immediately begin to reemploy our innate drive for comprehensive understanding.

The architects and planners, particularly the planners, though rated as specialists, have a little wider focus than do the other professions. Also as human beings they battle the narrow views of specialists—in particular, their patrons—the politicians, and the financial and other legal, but no longer comprehensively effective, heirs to the great pirates'—now only ghostly—prerogatives. At least the planners are allowed to look at *all* of Philadelphia, and not just to peek through a hole at one house or through one door at one room in that house. So I think it's appropriate that we assume the role of planners and begin to do the largest scale comprehensive thinking of which we are capable.

We begin by eschewing the role of specialists who deal only in parts. Becoming deliberately expansive instead of contractive, we ask, "*How* do we think in terms of *wholes*?" if it is true that the bigger the thinking becomes the more lastingly effective it is, we must ask, "How big can we think?"

One of the modern tools of high intellectual advantage is the development of what is called general systems theory. Employing it we begin to think of the largest and most comprehensive systems, and try to do so scientifically. We start by inventorying all the important, known variables that are operative in the problem. But if we don't really know how big "big" is, we may not start big enough, and are thus likely to leave unknown, but critical, variables outside the system which will continue to plague us. Interaction of the unknown variables inside and outside the arbitrarily chosen limits of the system are probably going to generate misleading or outrightly wrong answers. If we are to be effective, we are going to have to think in both the biggest and most minutely-incisive ways permitted by intellect and by the information thus far won through experience.

Can we think of, and state adequately and incisively, what we mean by universe? For universe is, inferentially, the biggest system. If we could start with universe, we would automatically avoid leaving out any strategically critical variables. We find no record as yet of man having successfully defined the universe—scientifically and comprehensively—to include the nonsimultaneous and only partially overlapping, micro-macro, always and everywhere transforming, physical and metaphysical, omni-complementary but nonidentical events.

Man has failed thus far, as a specialist, to define the microcosmic limits of divisibility of the nucleus of the atom, but, epochally, as accomplished by Einstein, has been able to define successfully the physical universe but not the metaphysical universe; nor has he, as yet, defined total universe itself as combining both the physical and metaphysical. The scientist was able to define physical universe by virtue of the experimentally-verified discovery that energy can neither be created nor lost and, therefore, that energy is conserved and is therefore finite. That means it is equatable. Einstein successfully equated the physical universe as $E = Mc^2$. His definition was only a hypothetical venture until fission proved it to be true. The physical universe of associative and dissociative energy was found to be a closed, but nonsimultaneously occurring, system—its separately occurring events being mathematically measurable; i.e., weighable and equatable. But the finite physical universe did not include the metaphysical weightless experiences of universe. All the unweighables, such as any and all our thoughts and all the abstract mathematics, are weightless. The metaphysical aspects of universe have been thought by the physical scientists to defy "closed system's" analysis. I have found, however, as we shall soon witness, the total universe including both its physical and metaphysical behaviors and aspects are scientifically definable.

Einstein and others have spoken exclusively about the physical department of universe in words which may be integrated and digested as *the aggregate of nonsimultaneous and only partially overlapping, nonidentical, but always complementary, omni-transforming, and weighable energy events.* Eddington defines science as "the earnest attempt to set in order the facts of experience." Einstein and many other first-rank scientists noted that science is concerned exclusively with "facts of experience."

Holding to the scientists' experiences as all important, I define universe, including both the physical and metaphysical, as follows: *The universe is the aggregate of all of humanity's consciously-apprehended and communicated experience with the nonsimultaneous, nonidentical, and only partially overlapping, always complementary, weighable and unweighable, ever omni-transforming, event sequences.*

Each experience begins and ends—ergo, is finite. Because our apprehending is packaged, both physically and metaphysically into time increments of alternate awakeness and asleepness as well as into separate finite conceptions such as the discrete energy quanta and the atomic nucleus components of the fundamental physical discontinuity, all experiences are finite. Physical experiments have found no solids, no continuous surfaces or lines—only discontinuous constellations of individual events. An aggregate of finites in finite. *Therefore, universe as experimentally defined, including both the physical and metaphysical, is finite.*

It is therefore possible to initiate our general systems formulation at the all inclusive level of universe whereby no strategic variables will be omitted. Thee is an operational grand strategy of General Systems Analysis that proceeds from here. It is played somewhat like the game of "Twenty Questions," but GSA is more efficient—that is, is more economical—in reaching its answers. It is the same procedural strategy that is used by the computer to weed out all the wrong answers until only the right answer is left.

Having adequately defined the whole system we may proceed to subdivide, progressively. This is accomplished through progressive division into two parts—one of which, by definition could not contain the answer—and discarding of the sterile part. Each progressively retained life part is called a "bit" because of its being produced by the progressive binary "yes" or "no" bi-section of the previously residual live part. The magnitude of such weeding operations is determined by the number of successive bits necessary to isolate the answer.

How many "bi-secting bits" does it take to get rid of all the irrelevancies and leave in lucid isolation that specific information you are seeking? We find that the first subdividing of the concept of universe—bit one—is into what we call a system. *A system subdivides universe into all the universe outside the system (macrocosm) and all the rest of the universe which is inside the system (microcosm) with the exception of the minor fraction of universe which constitutes the system itself.* The system divides universe not only into macrocosm and microcosm but also coincidentally into typical conceptual and nonconceptual aspects of universe—that is, an overlappingly-associable consideration, on the one hand, and, on the other hand, all the nonassociable, nonoverlappingly-considerable, nonsimultaneously-transforming events of nonsynchronizable disparate wave frequency rate ranges.

A thought is a system, and is inherently conceptual—though often only dimply and confusedly conceptual at the moment of first awareness of the as yet only vaguely describable thinking activity. Because total universe is nonsimultaneous it is not conceptual. Conceptuality is produced by isolation, such as in the instance of one single, static picture held out from a moving-picture film's continuity, or scenario. Universe is an evolutionary-process scenario without beginning or end, because the shown part is continually transformed chemically into fresh film and re-exposed to the ever self-reorganizing process of latest thought realizations which must continually introduce new significance into the freshly written description of the ever-transforming events before splicing the film in again for its next projection phase.

Heisenberg's principle of "indeterminism" which recognized the experimental discovery that the act of measuring always alters that which was being measured turns experience into a continuous and never-repeatable evolutionary scenario. One picture of the scenario about the caterpillar phase does not communicate its transformation into the butterfly phase, etc. The question, "I wonder what is outside the outside-of-universe?" is a request for a single picture description of a scenario of transformations and is an inherently invalid question. It is the same as looking at a dictionary and saying, "Which word is the dictionary?" It is a meaningless question.

It is characteristic of "all" thinking—of all system's conceptioning—that all the lines of thought interrelationships must return cyclically upon themselves in a plurality of directions, as do various great circles around spheres. Thus may we interrelatedly comprehend the constellation—or system—of experiences under consideration. Thus may we comprehend how the special-case economy demonstrated by the particular system considered also discloses the generalized law of energy conservation of physical universe.

To hit a duck in flight a hunter does not fire his gun at the bird where the gunner sees him but ahead of the bird, so that the bird and the bullet will meet each other at a point not in line between the gunner and the bird at time of firing. Gravity and wind also pull the bullet in two different directions which altogether impart a mild corkscrew trajectory to the bullet. Two airplanes in nighttime dogfights of World War II firing at each other with tracer bullets and photographed by a third plane show clearly the corkscrew trajectories as one hits the other. Einstein and Reiman, the Hindu mathematician, gave the name *geodesic lines* to these curvilinear and *most economical lines of interrelationship between two independently moving "events"*—the events in this case being the two airplanes.

A great circle is a line formed on a sphere's surface by a plane going through the sphere's center. Lesser circles are formed on the surfaces of spheres by planes cutting through spheres but not passing through the sphere's center. When a lesser circle is superimposed on a great circle it cuts across the latter at two points, *A* and *B*. It is a shorter distance between *A* and *B* on the great circle's shortest are than it is on the lesser circle's shortest arc. Great circles are geodesic lines because they provide the most economical (energy, effort) distances between any two points on a spherical system's surface; therefore, nature, which always employs only the most economical realizations must use those great circles which, unlike spiral lines, return upon themselves in the most economical manner. All the system's paths must be topologically and circularly interrelated for conceptually definitive, locally transformable, polyhedronal understanding to be attained in our spontaneous—ergo, most economical—geodesicly structured thoughts.

Thinking itself consists of self-disciplined dismissal of both the macrocosmic and microcosmic irrelevancies which leaves only the lucidly-relevant considerations. The macrocosmic irrelevancies are all the events too large and too infrequent to be synchronizably tuneable in any possible way without consideration (a beautiful word meaning putting stars together). The microcosmic irrelevancies are all the events which are obviously too small and too frequent to be differentially resolved in any way or to be synchronizably-tuneable within the lucidly-relevant wave-frequency limits of the system we are considering.

How many stages of dismissal of irrelevancies does it take—that is, proceeding from "universe" as I defined it, how many bits does it take—lucidly to isolate all the geodesic interrelations of all the "star" identities in the constellation under consideration? The answer is the formula $(N^2 - N)/2$ where N is the number of stars in the thought-discerned constellation of focal point entities comprising the problem.

"Comprehension" means identifying all the most uniquely economical inter-relationships of the focal point entities involved. We may say then that:

$$\text{Comprehension} = \frac{N^2 - N}{2}.$$

This is the way in which thought processes operate with mathematical logic. The mathematics involved consist of topology, combined with vectorial geometry, which combination I call "synergetics"—which word I will define while clarifying its use. By questioning many audiences, I have discovered that only about one in three hundred are familiar with synergy. The word is obviously not a popular word. Synergy is the only word in our language that means *behavior of whole systems unpredicted by the separately observed behaviors of any of the system's separate parts of any subassembly of the system's parts.* There is nothing in the chemistry of a toenail that predicts the existence of a human being.

I once asked an audience of the National Honors Society in chemistry, "How many of you are familiar with the word, synergy?" and all hands went up. Synergy is the essence of chemistry. The tensile strength of chrome-nickel, steel, which is approximately 350,000 pounds per square inch, is 100,000 PSI greater than the sum of the tensile strengths of all of each of its alloyed together, component, metallic elements. Here is a "chain" that is 50% stronger than the sum of the strengths of all links. We think popularly only in the terms of a chain being no stronger than its weakest link, which concept fails to consider, for instance, the case of an endlessly interlinked chain of atomically self-renewing links of omni-equal strength or of an omni-directionally interlinked chain matrix of ever renewed atomic links in which one broken link would be, only momentarily, a local cavern within the whole mass having no weakening effect on the whole, for every link within the matrix is a high frequency, recurring, break-and-make restructuring of the system.

Since synergy is the only word in our language meaning behavior of wholes unpredicted by behavior of their parts, it is clear that society does not think there are behaviors of whole systems unpredicted by their separate parts. This means that society's formally-accredited thoughts and ways of accrediting others are grossly inadequate in comprehending the nonconceptual qualities of the scenario "universal evolution."

There is nothing about an electron alone that forecasts the proton, nor is there anything about the Earth or the Moon that forecasts the coexistence of the sun. The solar system is synergetic—unpredicted by its separate parts. But the interplay of Sun as supply ship of Earth and the Moon's gravitationally produced tidal pulsations on Earth all interact to produce the biosphere's chemical conditions which permit but do not cause the regeneration of life on Spaceship Earth. This is all synergetic. There is nothing about the gases given off respiratorily by Earth's green vegetation that predicts that those gases will be essential to the life support of all mammals aboard Spaceship Earth, and nothing about the mammals that predicts that the gases which they give off respiratorily are essential to the support of the vegetation aboard our Spaceship Earth. Universe is synergetic. Life is synergetic.

Summarizing synergetically I may conclude that since my experimental interrogation of more than one hundred audiences all around the world has shown that less than one in three hundred university students has ever heard of the word synergy, and since it is the only word that has that meaning it is obvious that the world has not thought there are any behaviors of whole systems unpredictable by their parts. This is partially the consequence of overspecialization and of leaving the business of the whole to the old pirates to be visibly conducted by their stooges, the feudal kings or local politicians.

There is a corollary of synergy which says that the known behavior of the whole and the known behavior of a minimum of known parts often makes possible the discovery of the values of the remaining parts as does the known sum of the angles of a triangle plus the known behavior of three of its six parts make possible evaluating the others. Topology provides the synergetic means of ascertaining the values of any system of experiences.

Topology is the science of fundamental pattern and structural relationships of even constellations. It was discovered and developed by the mathematician Euler. He discovered that all patterns can be reduced to three prime conceptual characteristics: to *lines; points* where two lines cross or the same line crosses itself; and *areas,* bound by lines. He found that there is a constant relative abundance of these three fundamentally unique and no further reducible aspects of all patterning

$$P + A = L + 2.$$

This reads: the number of points plus the number of areas always equals the number of lines plus the number constant two. There are times when one area happens to coincide with others. When the faces of polyhedra coincide illusionarily the congruently hidden faces must be accounted arithmetically in formula.

Thus man has developed an externalized metabolic regeneration organism involving the whole of Spaceship Earth and all its resources. Any human being can physically employ that organism, whereas only one human can employ the organically integral craft tool. All 91 of the 92 chemical elements thus far found aboard our spaceship are completely involved in the world-around industrial network. The full family of chemical elements is unevenly distributed, and therefore our total planet is at all times involved in the industrial integration of the unique physical behaviors of each of all the elements. Paradoxically, at the present moment our Spaceship Earth is in the perilous condition of having the Russians sitting at one set of the co-pilot's flying controls while the Americans sit at the other. France controls the starboard engines, and the Chinese control the port engines, while the United Nations controls the passenger operation. The result is an increasing number of UFO hallucinations of sovereign states darting backwards and forwards and around in circles, getting nowhere, at an incredibly accelerating rate of speed.

All of humanity's tool extensions are divisible into two main groups: the craft and the industrial tools. I define the craft tools as all those tools which could be invented by one man starting all alone, naked in the wilderness, using only his own experience and his own integral facilities. Under these isolated conditions he could and did invent spears, slings, bows, and arrows, etc. By industrial tools I mean all the tools that cannot be produced by one man, as for instance the SS *Queen Mary.* With this definition, we find that the spoken word, which took a minimum of two humans to develop, was the first industrial tool. It brought about the progressive integration of all individual generation-to-generation experiences and thoughts of all humanity everywhere and everywhen. The Bible says, "In the beginning was the word"; I say to you, "In the beginning of industrialization was the spoken word." With the graphic writing of the words and ideas we have the beginning of the computer, for the computer *stores* and retrieves information. The written word, dictionary and the book were the first information storing and retrieving systems.

The craft tools are used initially by man to make the first industrial tools. Man is using his hands today most informatively and expertly only to press the buttons that set in action the further action of the tools which reproduce other tools which may be used informatively to make other tools. In the craft economies craftsman artists make only end or consumer-products. In the industrial economy the craftsman artists make the tools and the tools make the end or consumer-products. In this industrial development the mechanical advantages of men are pyramided rapidly and synergetically into invisible magnitudes of ever more incisive and inclusive tooling which produces ever more with ever less resource investment per each unit of end-product, or service, performance.

As we study industrialization, we see that we cannot have mass production unless we have mass consumption. This was effected evolutionarily by the great social struggles of labor to increase wages and spread the benefits and prevent reduction of the numbers of workers employed. The labor movement made possible mass purchasing; ergo, mass production ergo, low prices on vastly improved products and services, which have altogether established entirely new and higher standards of humanity's living.

Our labor world and all salaried workers, including school teachers and college professors, are now, at least subconsciously if not consciously, afraid that automation will take away their jobs. They are afraid they won't be able to do what is called "earning a living," which is short for earning the right to live. This term implies that normally we are supposed to die prematurely and that it is abnormal to be able to earn a living. It is paradoxical that only the abnormal or exceptional are entitled to prosper. Yesterday the term even inferred that success was so very abnormal that only divinely ordained kings and nobles were entitled to eat fairly regularly.

It is easy to demonstrate to those who will take the time and the trouble to unbias their thoughts that automation swiftly can multiply the physical energy part of wealth much more rapidly and profusely than can man's muscle and brain-reflexed—manually-controlled production. On the other hand humans alone can foresee, integrate, and anticipate the new tasks to be done by the progressively automated wealth-producing machinery. To take advantage of the fabulous magnitudes of real wealth waiting to be employed intelligently by humans and unblock automation's postponement by organized labor we must give each human who is or becomes unemployed a life fellowship in research and development or in just simple thinking. Man must be able to dare to think truthfully and to act accordingly without fear of losing his franchise to live. The use of mind fellowships will permit humans comprehensively to expand and accelerate scientific exploration and experimental prototype development. For every 100,000 employed in research and development, or just plain thinking, one probably will make a breakthrough that will more than pay for the other 99,999 fellowships. Thus, production will no longer be impeded by humans trying to do what machines can do better. Contrariwise, omni-automated and inanimately powered production will unleash humanity's unique capability—its metaphysical capability. Historically speaking, these steps will be taken within the next decade. There is no doubt about it. But not without much social crisis and consequent educational experience and discovery concerning the nature of our unlimited wealth. Through the universal research and development fellowships, we're going to start emancipating humanity from being muscle and reflex machines. We're going to give everybody a chance to develop their most powerful mental and intuitive faculties. Given their research and development fellowship, many who have been frustrated during their younger years may feel like going fishing. Fishing provides an excellent opportunity to think clearly; to review one's life; to recall one's earlier frustrated and abandoned longings and curiosities. What we want everybody to do is to *think* clearly.

We soon will begin to generate wealth so rapidly that we can do very great things. I would like you to think what this may do realistically for living without spoiling the landscape, or the antiquities or the trails of humanity throughout the ages, or despoiling the integrity of romance, vision, and harmonic creativity. All the great office buildings will be emptied of earned living workers, and the automated office-processing of information will be centralized in the basements of a few buildings. This will permit all the modernly mechanized office buildings to be used as dwelling facilities.

When we approach our problems on a universal, general systems basis and progressively eliminate the irrelevancies, somewhat as we peel petals from an artichoke, at each move we leave in full visibility the next most important layer of factors with which we must deal. We gradually uncover *you* and *me* in the heart of now. But evolution requires that we comprehend each layer in order to unpeel it. We have now updated our definitions of universe by conforming them with the most recent and erudite scientific findings such as those of Einstein and Planck. Earlier in our thinking we discovered man's function in universe to be that of the most effective metaphysical capability experimentally evidenced thus far within our locally observable phases and time zones of universe. We have also discovered that it is humanity's task to comprehend and set in order the special case facts of human experience and to win therefrom knowledge of the *a priori* existence of a complex of generalized, abstract principles which apparently altogether govern all physically evolving phenomena of universe.

We have learned that only and exclusively through use of his mind can man inventively employ the generalized principles further to conserve the locally available physical energy of the only universally unlimited supply. Only thus can man put to orderly advantage the various, local, and otherwise disorderly behaviors of the entropic, physical universe. Man can and may metaphysically comprehend, anticipate, shunt, and meteringly introduce the evolutionarily organized environment events in the magnitudes and frequencies that best synchronize with the patterns of his successful and metaphysical metabolic regeneration while ever increasing the degrees of humanity's space and time freedoms from yesterday's ignorance sustaining survival procedure chores and their personal time capital wasting.

Now we have comprehended and peeled off the layers of petals which disclosed not only that physical energy is conserved but also that it is ever increasingly deposited as a fossil-fuel savings account aboard our Spaceship Earth through photosynthesis and progressive, complex, topsoil fossilization buried ever deeper within Earth's crust by frost, wind, flood, volcanoes, and earthquake upheavals. We have thus discovered also that we can make all of humanity successful through science's worldengulfing industrial evolution provided that we are not so foolish as to continue to exhaust in a split second of astronomical history the orderly energy savings of billions of years' energy conservation aboard our Spaceship Earth. These energy savings have been put into our Spaceship's life-regeneration-guaranteeing bank account for use only in self-starter functions.

The fossil fuel deposits of our Spaceship Earth correspond to our automobile's storage battery which must be conserved to turn over our main engine's self-starter. Thereafter, our "main engine," the life regenerating processes, must operate exclusively on our vast daily energy income from the powers of wind, tide, water, and the direct Sun radiation energy. The fossil-fuel savings account has been put aboard Spaceship Earth for the exclusive function of getting the new machinery built with which to support life and humanity at ever more effective standards of vital physical energy and reinspiring metaphysical sustenance to be sustained exclusively on our Sun radiation's and Moon pull gravity's tidal, wind, and rainfall generated pulsating and therefore harnessable energies. The daily income energies are excessively adequate for the operation of our main industrial engines and their automated productions. The energy expended in one minute of a tropical hurricane equals the combined energy of all the USA and USSR nuclear weapons. Only by understanding this scheme may we continue for all time ahead to enjoy and explore universe as we progressively harness evermore of the celestially generated tidal and storm generated wind, water, and electrical power concentrations. We cannot afford to expend our fossil fuels faster than we are "recharging our battery," which means precisely the rate at which the fossil fuels are being continually deposited within Earth's spherical crust.

We have discovered that it is highly feasible for all the human passengers aboard Spaceship Earth to enjoy the whole ship without any individual interfering with another and without any individual being advantaged at the expense of another, provided that we are not so foolish as to burn up our ship and its operating equipment by powering our prime operations exclusively on atomic reactor generated energy. The too-shortsighted and debilitating exploitation of fossil fuels and atomic energy are similar to running out automobiles only on the self-starters and batteries and as the latter become exhausted replenishing the batteries only by starting the chain reaction consumption of the atoms with which the automobiles are constituted.

We have discovered also why we were given our intellectual faculties and physical extension facilities. We have discovered that we have the inherent capability and inferentially the responsibility of making humanity comprehensively and sustainably successful. We have learned the difference between brain and mind capabilities. We have learned of the superstitions and inferiority complexes built into all humanity through all of history's yesterdays of slavish survival under conditions of abysmal illiteracy and ignorance wherein only the most ruthless, shrewd, and eventually brutish could sustain existence, and then for no more than a third of its known potential life span.

This all brings us to a realization of the enormous educational task which must be successfully accomplished right now in a hurry in order to convert man's spin-dive towards oblivion into an intellectually mastered power pullout into safe and level flight of physical and metaphysical success, whereafter he may turn his Spaceship Earth's occupancy into a universe exploring advantage. If it comprehends and reacts effectively, humanity will open an entirely new chapter of the experiences and the thoughts and drives thereby stimulated.

Most importantly we have learned that from here on it is success for all or for none, for it is experimentally proven by physics that "unity is plural and at minimum two"—the complementary but not mirror-imaged proton and neutron. You and I are inherently different and complementary. Together we average as zero—that is, as eternity. Now having attained that cosmic degree of orbital conceptioning we will use our retro-rocket controls to negotiate our reentry of our Spaceship Earth's atmosphere and return to our omnibefuddled present. Here we find ourselves maintaining the fiction that our cross-breeding World Man consists fundamentally of innately different nations and races which are the antithesis of that cross-breeding. Nations are products of many generations of local inbreeding in a myriad of remote human enclaves. With grandfather chiefs often marrying incestuously the gene concentrations brought about hybrid nationally-unique physiological characteristics which in the extreme northern hibernations bleached out the human skin and in the equatorial casting off of all clothing inbred darkly tanned pigmentation. All are the consequence only of unique local environment conditions and super inbreeding.

The crossbreeding world people on the North American continent consists of two separate *input* sets. The first era input set consists of those who came with the prevailing winds and ocean currents eastward to the North, South, and Central Americas by raft and by boat from across the Pacific, primarily during an age which started at least thirty thousand years ago, possibly millions of years ago, and terminated 300 years ago. The eastbound trans-Pacific migration peopled the west coast of both South and North America and migrated inland towards the two continents' middle ground in Central America and Mexico. In Mexico today will be found every type of human characteristic and every known physiognomy, each of which occur in such a variety of skin shades from black to white that they do not permit the ignorance-invented "race" distinctions predicted only superficially on extreme limits of skin color. The second or west-bound input era set of crossbreeding world man now peopling the Americas consists of the gradual and slower migration around the world from the Pacific Ocean westward into the wind, "following the sun," and travelling both by sea through Malaysia, across the Indian Ocean up the Persian Gulf into Mesopotamia and overland into the Mediterranean, up the Nile from East Africa into the South and North Atlantic to America—or over the Chinese, Mongolian, Siberian, and European hinterlands to the Atlantic and to the Americas.

Now both east and westbound era sets are crossbreeding with one another in ever-accelerating degree on America's continental middleground. This omni reintegration of world man from all the diverse hybrids is producing a crossbred people on the Pacific Coast of North America. Here with its aerospace and oceans penetrating capabilities, a world type of humanity is taking the springboard into all of the hitherto hostile environments of universe into the ocean depths and into the sky and all around the earth. Returning you again to our omni-befuddled present, we realize that reorganization of humanity's economic accounting system and its implementation of the total commonwealth capability by total world society, aided by the computer's vast memory and high speed recall comes first of all of the first-things-first that we must attend to make our space vehicle Earth a successful man operation. We may now raise our sights, in fact must raise our sights, to take the initiative in planning the world-around industrial retooling revolution. We must undertake to increase the performance per pound of the world's resources until they provide all of humanity a high standard of living. We can no longer wait to see whose biased political system should prevail over the world.

You may not feel very confident about how you are going to earn your right to live under such world-around patronless conditions. But I say to you the sooner you do the better chance we have of pulling out of humanity's otherwise fatal nose dive into oblivion. As the world political economic emergencies increase, remember that we have discovered a way to make the total world work. It must be initiated and in strong momentum before we pass the point of no return. You may gain great confidence from that fact that your fellow men, some of them your great labor leaders, are already aware and eager to educate their own rank and file on the fallacy of opposition to automation.

I have visited more than 300 universities and colleges around the world as an invited and appointed professor and have found an increasing number of students who understand all that we have been reviewing. They are comprehending increasingly that elimination of war can only be realized through a design and invention revolution. When it is realized by society that wealth is as much everybody's as is the air and sunlight, it no longer will be rated as a personal handout for anyone to accept a high standard of living in the form of an annual research and development fellowship.

I have owned successively, since boyhood, fifty-four automobiles. I will never own another. I have now given up driving. I began to leave my cars at airports—never or only infrequently getting back to them. My new pattern requires renting new cars at the airports as needed. I am progressively ceasing to own things, not on a political-schism basis, as for instance Henry George's ideology, but simply on a practical basis. Possession is becoming progressively burdensome and wasteful and therefore obsolete. Why accumulate mementos of far away places when you are much more frequently in those places than at your yesterday's home, state, city and street identified residences, as required for passport, taxing, and voting functions? Why not completely restore the great cities and buildings of antiquity and send back to them all their fragmented treasures now deployed in the world's museums? Thus, may whole eras be reinhabited and experienced by an ever increasingly interested, well-informed, and inspired humanity. Thus, may all the world regain or retain its regenerative metaphysical mysteries.

I travel between Southern and Northern Hemispheres and around the world so frequently that I no longer have any so-called normal winter and summer, nor normal night and day, for I fly in and out of the shaded or sunflooded areas of the spinning, orbiting Earth with ever-increased frequency. I wear three watches to tell me what time it is at my "home" office, so that I can call them by long distance telephone. One is set for the time of day in the place to which I am next going, and one is set temporarily for the locality in which I happen to be. I now see the Earth realistically as a sphere and think of it as a spaceship. It is big, but it is comprehensible. I no longer think in terms of "weeks" except as I stumble over their antiquated stop-and-go habits. Nature has no "weeks." Quite clearly the peak traffic patterns exploited by businessmen who are eager to make the most profit in order to prove their right to live causes everybody to go in and out of the airport during two short moments in the twenty-four hours with all the main facilities shut down two-thirds of the time. All our beds around the world are empty for two-thirds of the time. Our living rooms are empty seven-eighths of the time.

The population explosion is a myth. As we industrialize, down goes the annual birth rate. We observed[4] that, by 1997, the whole world became industrialized, and, as with the United States, and as with all Europe and China and Japan today, the birth rate is dwindling, and the bulge in population will be recognized as accounted for exclusively by those who are living longer.[4]

When world realization of its unlimited wealth has been established there as yet will be room for the whole of humanity to stand indoors in greater New York City, with more room for each human than at an average cocktail party.

We will oscillate progressively between social concentrations in cultural centers and in multideployment in greater areas of our Spaceship Earth's as yet very ample accommodations. The same humans will increasingly converge for metaphysical intercourse and deploy for physical experiences.

Each of our four billion humans' shares of the Spaceship Earth's resources as yet today amount to two-hundred billion tons.

It is also to be remembered that despite the fact that you are accustomed to thinking only in dots and lines and a little bit in areas does not defeat the fact that we live in omnidirectional space-time and that a four dimensional universe provides ample individual freedoms for any contingencies.

So, planners, architects, and engineers, take the initiative. Go to work, and above all co-operate and don't hold back on one another or try to gain at the expense of another. Any success in such lopsidedness will be increasingly short-lived. These are the synergetic rules that evolution is employing and trying to make clear to us. They are not man-made laws. They are the infinitely accommodative laws of the intellectual integrity governing universe.[2]

The preceding material has been excerpted from the work of Dr. R. Buckminster Fuller[1,2] with permission of the author.

As a fitting postscript to this article by the late Professor Fuller, the reader is referred to the declaration on climate change adopted at the 1989 Ministerial Conference held in the Netherlands[3] which is presented in Appendices (Table 11 of this Encyclopedia).

This was an important landmark in an ongoing series of international meetings in the field of climate change policy at the political level.

Consultations and discussions prior to, and during, the Conference culminated in the adoption of a Declaration, by consensus of all parties present (This included 67 countries). The Noordwijk Declaration's unique new concepts and targets are addressed as follows:

- CO_2-emission stabilization and future reductions;
- a global forest stock balance and future net forest growth;
- funding mechanisms for both existing and additional funds;
- elements of a climate change convention;
- the principle of shared responsibility and the particular responsibilities of both developed and developing countries.

Also, there is an agreement to strengthen the amendments of the Montreal Protocol to phase out chlorofluorocarbons in a more timely fashion (i.e. by the year 2000).

The Kyoto accord,[6] negotiated in 1997 in Kyoto, Japan, requires industrial nations—with varying targets—to reduce their emissions of greenhouse gases which trap heat and result in global warming below their 1990 levels, in the five years from 2008 to 2012. As of Feb. 16, 2005, the date the agreement took effect, 35 nations planned to cut their emissions of carbon dioxide and other greenhouse gases.

However, in the United States, which generates a fifth of the world's greenhouse gases and which formally rejected the Kyoto pact in 2001, a growing number of companies regard mandatory reductions as inevitable. In keeping with the spirit of Kyoto, Michael G. Morris, chief executive of American Electric Power, the largest electricity generator in the United States and a top emitter of CO_2, has pledged a 10% reduction in those emissions from its plants by 2006, rather than resisting the agreement's philosophy. Nations with rapidly growing economies, like China and India, which approved the agreement, are not required to reduce greenhouse gas emissions in Kyoto Phase One—even though together, they already account for 14 percent of the world's total. "For the European Union, the target is an 8 percent reduction below emissions levels in 1990. But the Germans went beyond that and agreed to a more ambitious target of 21 percent because they expected windfall gains by shutting down polluting, coal-fired power plants in the former East Germany. (It now seems likely to fall somewhat short of that.)"[5] The Kyoto Protocol encourages trading carbon dioxide emissions credits, some of which may be earned from reforestation projects which absorb CO_2 from the atmosphere and others from the use of cleaner technologies.[5]

REFERENCES

1. American Scholar, Spring, 1968.
2. *Operating Manual for Spaceship Earth,* S.I.U. Press, Carbondale, I11. 1969.
3. P. Vellinga, Declaration of the Ministerial Conference on Atmospheric and Climatic Change—Noordwijk, The Netherlands, Nov. 6 and 7 (1989).
4. Editorial Note (1997)—R.B.F. Prediction.
5. M. Landler, *The New York Times,* Mixed Feelings as Treaty on Greenhouse Gases Takes Effects, February 16, 2005.
6. See Appendix, Table 12, of this volume.

R. BUCKMINSTER FULLER (DECEASED)
Southern Illinois University

TEACHING OF ENVIRONMENTAL STUDIES: see ENVIRONMENTAL EDUCATION

THERMAL EFFECTS ON FISH ECOLOGY

Of all environmental factors that influence aquatic organisms, temperature is the most all-pervasive. There is always an environmental temperature while other factors may or may not be present to exert their effects. Fish are, for all practical purposes, thermal conformers, or obligate poikilotherms. That is, they are able to exert little significant influence on maintaining a certain body temperature by specialized metabolic or behavioral means. Their body temperature thus fluctuates nearly in concert with the temperature of their aquatic medium (although particularly large, actively-moving fish such as tuna have deep muscle temperatures slightly higher than the water). Intimate contact at the gills of body fluids with the outside water and the high specific heat of water provide a very efficient heat exchanger that insures this near identity of internal and external temperatures.

Every response of fish, from incubation of the egg to feeding activity, digestive and metabolic processes, reproduction, geographic distribution, and even survival, proceeds within a thermal range dictated by the immediate environment. As human activities change this thermal environment, such as through deforestation, damming or thermal discharges from power stations, the activities of indigenous fish species must also change. Depending upon the magnitude and rates of the thermal changes, there may be minor readjustments of the rates of metabolism and growth, or major changes in the distribution of species and of the functioning of the affected aquatic ecosystems.

In our recent environmental awareness, we have coined the phrase “thermal pollution” for extensive thermal changes to natural aquatic environments that are believed to be detrimental to desired fish populations. The key to controlling “thermal pollution” is a firm understanding of how temperature affects fish, and of the circumstances that truly constitute pollution.

The subject of thermal effects on fishes has been given critical scientific review periodically especially over the years (e.g. Fry, 1947; Bullock, 1955; Brett, 1956; Fry, 1964; Fry, 1967 and Brett, 1970). Scientific knowledge as a basis for controlling pollution is clearly more advanced in this area than for almost any other environmental factor. This knowledge has been applied to the context of thermal modifications by electricity generating stations in two symposium volumes (Parker and Krenkel, 1969; Krenkel and Parker, 1969) and by Cairns (1968), Clark (1969), Parker and Krenkel (1969) and Countant (1970 and 1972). The voluminous scientific literature on temperature effects on fishes may be easily searched for specific information in bibliographies by Kennedy and Mihursky (1967), Raney and Menzel (1969) and annual literature reviews by Coutant (1968, 1969, 1970, 1971) and Coutant and Goodyear (1972). Readers seeking more than a general review are advised to read these materials. (*See also* Alabaster 1986).

While fish must conform to water temperature, they have evolved mechanisms other than body temperature regulation to deal with vicissitudes of temperature fluctuations that occur geographically, seasonally and daily. That such mechanisms exist became apparent when fish physiologists realized that at any one temperature a fish may survive or die, be hyperactive or be numbed into activity, be stimulated to migrate or be passive, be sexually mature or immature, all depending upon the state of previous temperature exposures. Temperature affects organisms not only by absolute level (as in physics and chemistry) but also by change. Like light, temperature can exert effects through daily or seasonal patterns that exhibit a special quality beyond that of absolute level[†].

The functional properties of temperature acting on fish can be summarized as follows: Temperature can act as a lethal agent that kills the fish directly, as a stressing agent that destroys the fish indirectly, as a controlling factor that sets the pace of metabolism and development, as a limiting factor that restricts activity and distribution, as a limiting factor that restricts activity and distribution, as a masking factor that interacts with other environmental factors by blocking or altering their potential expression, and as a directing agent in gradients that stimulate sensory perception and orient activity. Each of these properties can be visualized as acting on two levels—on the individual fish and on the population of any one fish species.

TEMPERATURE AS A LETHAL AGENT

Mass mortalities of fish in nature have often been reported, but usually the causes are obscure. Fish rarely die in places and at times when proper field instrumentation is operating or when trained observers are at hand. Many deaths probably go unnoticed, for scavengers may act quickly or water

---

[†] Clear distinction must be made between *heat* which is a quantitative measure of energy of molecular motion that is dependent upon the mass of an object or body of water and *temperature* which is a measure (unrelated to mass) of energy intensity. Organisms respond to temperature, not to heat.

currents disperse carcasses (particularly of small fishes). The most common reports are of cold kills brought about by particularly severe winters or rapid drops in temperature (e.g. summaries by Brett, 1970). It is well known among fishery biologists that the abundance of a species reproduced in any one year varies tremendously, a fact that many scientists have attributed in part to deaths from unfavorable temperatures at early life stages where the fish are too small to be recognized as constituting a "fish kill".

Studies of temperature tolerance in fishes began in the last century. The early method of determining the lethal end-point (generally the cessation of opercular movements) by slow heating or cooling was generally supplanted in the 1940s by a more precise method of direct transfer to a series of preset temperatures in which the rates of dying of individual fish and the statistical variation among many individuals could be obtained. These experiments demonstrated the importance of recent past history of the fish, both the controlled holding temperature imposed in the laboratory prior to testing acclimation and the seasonal environmental temperature when fish were tested directly from field collections (acclimatization).

These experiments also showed that each species of fish (and often each distinct life stage of one species) has a characteristic range of temperature that it will tolerate that is established by internal biochemical adjustments made while at the previous holding temperature (Figure 1). Ordinarily (for purposes of comparison) the upper and lower ends of this range are defined by survival of 50% of a sample of individuals similar in size, health and other factors, for a specified length of time, often one week. The tolerance range is shifted upward by long-term holding (acclimation) in warmer water, and downward by acclimation to cooler water. This accommodation is limited, however, at the lower end by freezing point of water (for species in temperate latitudes) and at the upper end by an ultimate lethal threshold. The graphic representation (Figure 1) is a geometric figure for which an area can be computed. The areas (as degrees squared) provide convenient measures of the relative overall sensitivity of tolerance among different species and life stages (a small area or zone on the graph signified high thermal sensitivity).

It is not surprising that rough species such as carp and goldfish were found to have large thermal tolerance zones.

Outside the thermal tolerance zone, premature death is inevitable and its onset is a function of both temperature and time of exposure (thermal resistance). Death occurs more rapidly the farther the temperature is from the threshold (Figure 2), an attribute common to the action of toxicants, pharmaceuticals, and radiation. The duration of survival of half of a test population of fish at extreme temperature can be expressed as an equation based on experimental data for each acclimation temperature:

$$\log \text{survival time}_{(\text{min})} = a + b\,(\text{Temp}_{(^\circ\text{C})}),$$

in which a and b are intercept and slope of the linear regression lines in Figure 2. In some cases the time-temperature relationship is more complex than this semi-logarithmic model, but this expression is the most generally applicable and is the one most generally accepted by the scientific community. The equation defines the average rate of dying at any extreme temperature.

The thermal resistance equations allow prediction of fish survival (or death) in zones where human activity induces

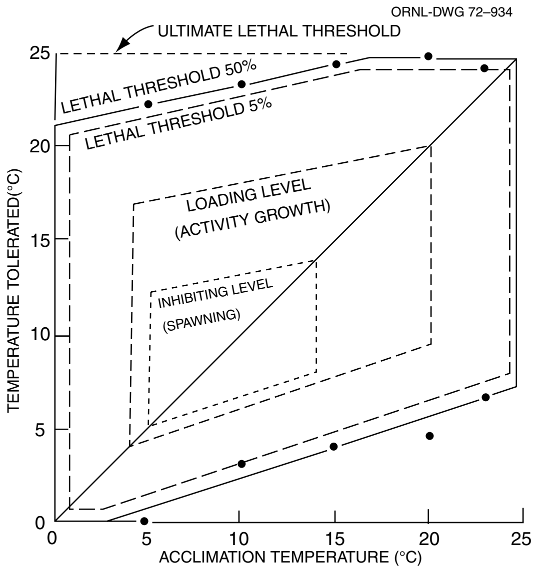


FIGURE 1 Upper and lower lethal temperatures for young sockeye salmon with various acclimation temperatures, plotted to show the ranges of tolerance, and within these ranges more restrictive requirements for activity, growth or spawning. (Reproduced by permission from Coutant, 1972.)

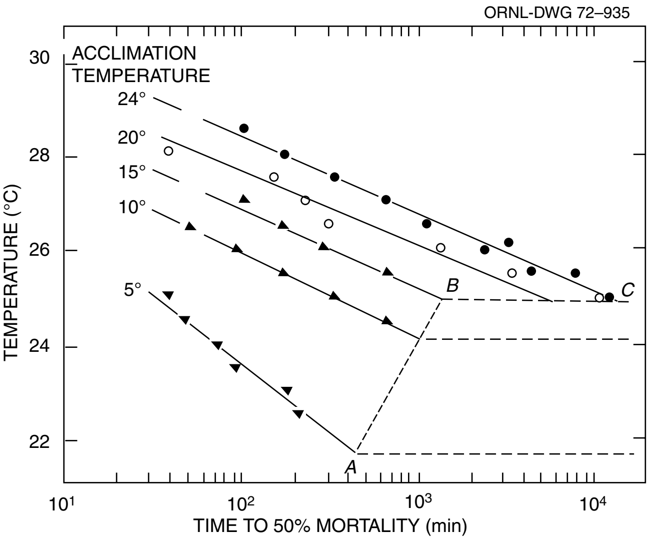


FIGURE 2 Median resistance times to high temperatures among young chinook salmon acclimated to the temperatures indicated. Line A-B denotes rising lethal threshold levels with increasing acclimation temperature. This rise ceases at higher acclimation temperatures. (Reproduced by permission from Coutant, 1972.)

extreme high temperatures. For example, juvenile salmon and trout were found to pass through warm mixing zones of thermal discharges to the Columbia River during their seaward migration (Becker *et al.,* 1971). The thermal exposure was a complex pattern of rapid temperature rise (often to temperatures beyond the tolerance zone) followed by a slow decline as the heated effluent mixed with the cooler river. By using the equation-expressed rates of dying at each of the temperatures briefly experienced, and the length of time the fish were exposed to each incremental temperature, the ability of the fish to survive the exposure was estimated and compared with actual field exposures. Similar predictions can be made for proposed thermal discharges, and corrective engineering can be selected before the project is constructed. Similar predictions can be made for circumstances where fish may become acclimated to warm water (e.g. in a discharge canal) and then be cooled rapidly and face a potential cold kill. This predictive methodology is further described by Coutant (1972).

TEMPERATURE AS A STRESSING FACTOR

Death need not come to fish directly from temperature or its change. In natural ecological systems death often comes as the result of a secondary agent acting upon a fish weakened by some stress such as temperature. This secondary agent is often disease or predator. A potentially lethal high temperature will, for example, induce loss of equilibrium before the physiological death point is reached, and equilibrium loss (going "belly-up") in a natural environment is an open invitation to predators. In fact, ongoing research indicates that stress from relatively small temperature changes (both up and down) will induce selective predation on the stressed fish. The effect appears to follow a time-temperature pattern similar to that for death, with stress appearing after shorter exposures and lower temperatures than required for death directly. The predictability developed for lethal responses can be applied to these stressing conditions as well, if we wish to prevent "ecological death."

TEMPERATURE AS A CONTROLLING FACTOR

Metabolism

Within the zone of thermal tolerance of any species (Figure 1), the most important contributor to survival and success in nature is the dynamic cycle of energy intake, conversion and utilization for activity, development (the differentiation of cells) and growth (multiplication of cells and storage of energy reserves). Since the time that Fry (1947) observed that environmental temperature controls energy metabolism, there has been extensive research in this area of fish physiology and biochemistry. This research has yielded important generalizations about the temperature responses of fish, and the physiological and biochemical "reasons" for these responses.

Metabolic processes are basically chemical in character. Among the most significant vital chemical reactions are the actions of the living catalysts (enzymes) which control the oxidation of organic food materials. Most enzymes show an optimum temperature at which they reach a maximum rate of catalytic activity. This is sometimes higher than the upper lethal threshold for the whole fish. The aggregate of many metabolic reactions also exhibits a temperature optimum, or point of maximum rate, which is often remarkably similar for various functions involved, for example digestion, development and locomotion (Figure 3). Through genetic selection, the optimum has become different for any two species. Below the optimum, the maximum rate possible is controlled by water temperature. These rates can be quite different for various functions. It should be noted that the optimum temperature

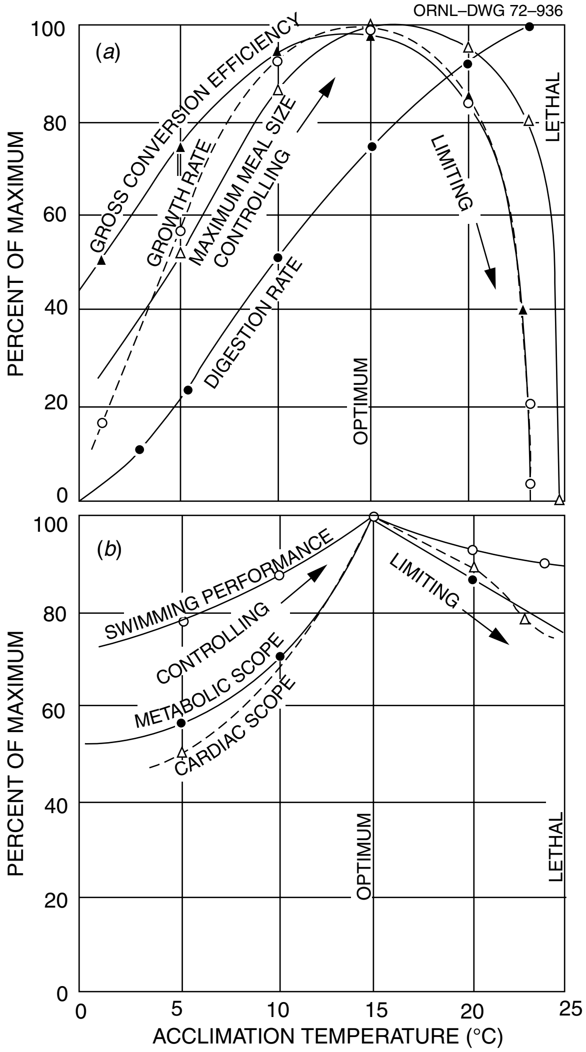


FIGURE 3 Performance of sockeye salmon in relation to acclimation temperature. There are three characteristic type responses; two have coinciding optima. (Reproduced by permission from Coutant, 1972.)

and the maximum metabolic rates at any given temperature may be quite different during embryonic development and during the lifetime of the fully-developed fish.

Of the various methods that have been used to measure metabolic rates (see Brett, 1971), the most often measured has been the rate of oxygen consumption. This provides an instantaneous measure of enzyme activity so long as no oxygen debt, or delayed oxidation of certain chemical compounds, is accumulated. Three levels of metabolic rates have been commonly recognized for fish: (1) *Standard metabolic rate,* representing that fraction which is just necessary to maintain vital functions of a resting fish, (2) *routine metabolic rate,* which also includes the energy demands of routine, spontaneous activity, and (3) *active metabolic rate,* which represents the maximum level of oxygen consumed by a working (swimming) fish. The amount of energy available for active work (or growth) is termed the *metabolic scope* for activity, and it is the difference between active and standard metabolic rates. Each of these is related to temperature in a different way. The most important measure for a fish's ability to cope with the overall environmental demands is the metabolic scope, which has an optimum temperature (Figure 3).

Activity

As temperature controls the metabolic rate which provides energy for activity, that activity, then, is also controlled. The literature contains many references to increases in fish activity with temperature rise, particularly swimming performance. This increase in activity ceases at an optimum temperature that appears to coincide with the temperature of maximum metabolic scope (Figure 3).

Growth

Temperature is one of the principal environmental factors controlling growth of fishes, others being light and salinity. There recently has been a considerable amount of laboratory experimentation to separate these often-correlated influences on growth.

Whenever there is abundant food, increasing temperature enhances growth rate up to an optimum (Figure 3) above which there is a decline. Low temperatures generally retard growth, although organisms residing habitually in cold areas such as the arctic have evolved metabolic compensations that allow good growth even at low extremes. Optimum growth appears to occur at about the same temperature as maximum metabolic scope. Restriction of food generally forces the optimum growth temperature toward cooler levels and restricts the maximum amount of growth attainable (Brett *et al.,* 1969).

TEMPERATURE AS A LIMITING FACTOR

As the previous discussion implied, there comes a point (the optimum) on a rising temperature scale at which increased temperature no longer speeds processes but begins to limit them. In contrast to the gradual increase in performance with temperature rise exhibited at suboptimum temperatures, the responses at levels above optimum often show a precipitous decline (Figure 3). Performance is often reduced to zero several degrees below temperatures which would be directly lethal in the relatively short period of one week. One of the most significant of thermal limitations from the standpoint of a fish's overall success in this environment is upon set growth rate for the population. If a majority of individuals of the species cannot sustain positive growth, then the population is likely to succumb. While it is probably unnecessary for populations to grow at maximum rates, there must be a thermal maximum for prolonged exposures of any fish species that is less than the established lethal levels at which growth limitation becomes critical for continued population survival. The requirement for sustained growth may be one of the most important mechanisms of geographic limitations of species. Intensive research in this area is needed to establish rational upper temperature standards for water bodies.

TEMPERATURE AS A MASKING FACTOR

All other environmental factors, such as light, current, or chemical toxins, act upon fish simultaneously within a temperature regime. With so much of a fish's metabolic activity dependent upon temperature, both immediate and previous, it is little wonder that responses to other environmental factors change with differing temperature. The interactions are seemingly infinite, and the general impression that one obtains is that temperature is masking a clear-cut definition of the response pattern to any other environmental parameter.

This pessimism overstates the case, however. Two-factor experimentation is routine today, and interactions of temperature and a variety of pollutants are now becoming clear. For instance, research in Britain has shown that the effect of increased temperature on the toxicity of poisons to fish is generally to reduce their time of survival in relatively high lethal concentrations, but median *threshold* concentrations for death may not be markedly changed, or may even be increased (Ministry of Technology, 1968). An increase in temperature of 8°C reduced the 48 hr LC_{50} (median lethal concentration) to rainbow trout by a factor of 1.8 for zinc (i.e. increased toxicity) but increased it (i.e. reduced toxicity by about 1.2 for phenol, by 2.0 for undissociated ammonia, and by 2.5 for cyanide. The effect of temperature on ammonia toxicity is further expressed by changing the dissociation of ammonia in water and thus the percentage of actively toxic ammonia available. For estuarine and marine fishes temperature-salinity interactions are of special importance, and are receiving increased research attention.

TEMPERATURE AS A DIRECTING AGENT

Gradient responses

Numerous observations of fish in horizontal and vertical thermal gradients both in the laboratory and under field conditions have demonstrated preferred or selected temperatures. There are wide differences among species, and some differences among life stages of any one species. The preferred temperature is dependent upon recent prior thermal history, but continuous exposure to a gradient (in which metabolic acclimation gradually takes place) results in a "final preferendum". Preferred ranges have been shown to coincide with the species-specific optimum temperature for maximum metabolic scope for activity, and thus the directive mechanism would appear to have survival value.

Many fish have a delicate sense for temperature discrimination. The threshold for teleosts (bony fish) appears to be on the order of ±0.05°C, although elasmobranches (sharks, rays) have a threshold quite a bit higher (about ±0.8°C). Orientation responses have generally been elicited by differences of about 0.5°C (Brett, 1971). Many fish are very capable of detecting undesirable temperatures and of avoiding water masses that are potentially detrimental to them.

Directive cues

A mechanistic response to temperature gradients is often overridden by seasonal influences and special behavior patterns involving temperature-oriented activities such as migration. The seasonal response to a specific temperature has been shown to have great importance for reproductive activity of a large number of fishes.

The sequence of events relating to gonad maturation, spawning migration, courting behavior, release of gametes, and subsequent development of egg and embryo represents one of the most complex phenomena in nature. While temperature cues appear critical in many cases, the interactions with other factors such as seasonal light intensity are still not clearly understood. Advance or retardation of reproduction has been closely related to temperature of the months preceding spawning in such fish as the cod *Gadus morhua.* The difference in the effect of temperature governing a rate phenomenon (controlling or limiting) and temperature acting as a releasing factor is clearly shown in cases where falling temperatures induce spawning, as in the Pacific salmon.

Temperature appears to confine spawning to a narrower range than most other functions. The average range for spawning of marine fish is one-quarter to one-third that of the lethal range (Brett, 1971).

SUMMARY

From this brief introduction, we can see that temperature is probably the preeminent master factor in the lives of fish. No study of fish in relation to their environment ("fish ecology") would be meaningful without consideration of thermal relationships. This review can direct the curious to more comprehensive treatises. From a different perspective, there are few environmental modifications that man could make to aquatic systems that would be so assured to causing some ecological change as temperature. Within limits, fish possess effective mechanisms for adapting to thermal changes, for such changes are a normal part of their existence. Man must be careful not to exceed these limits, however, if he wishes to preserve a productive commercial and recreational fishery.

REFERENCES

1. Abrams, P.W., M. Tranter, T.P. Davis and I.L. Blackwood, 1989, Geochemical studies in a remote Scottish upland catchment; II. Streamwater chemistry during snowmelt, Water, Air and Soil Pollution, 43, 3/4.
2. Alabaster, J.S., 1986, Habitat modification and freshwater fisheries, Butterworth, Stoneham, MA.
3. Becker, C.D., C.C. Coutant and E.F. Prentice, 1971. Experimental drifts of juvenile salmonids through effluent discharges at Hanford, Part II. 1969 Drifts and conclusions USAEC Rept., BNWL-1529, Batelle Northwest, Richland, Washington.
4. Brett J.R., 1956, Some principles in the thermal requirements of fishes, *Quarterly Review of Biology* **31**(2), 75–87.
5. Brett, J.R., 1970, Temperature—animals—fishes, O. Kinne, Ed., in *Marine Ecology,* **1,** Environmental Factors, Part 1, pp. 515–560.
6. Brett, J.R., 1971, Energetic responses of salmon to temperature, a study of some thermal relations in the physiology and freshwater ecology of sockeye salmon (*Oncorhynchus nerka*). *American Zoologist* **11,** 99–113.
7. Brett, J.R., J.E. Shelbourn and C.T. Shoop, 1969, Growth rate and body composition of fingerling sockeye salmon, *Oncorhynchus merka,* in relation to temperature and ration size, *J. Fish. Res. Bd. Canada* **26,** 2363–2394.
8. Bullock, T.H., 1955, Compensation for temperature in the metabolism and activity of poikilotherms, *Biol. Rev.* **30**(3), 311–342.
9. Cairns, John, Jr., 1968, We're in hot water, *Scientist and Citizen* **10**(8), 187–198.
10. Clark, J.R., 1969, Thermal pollution and aquatic life, *Sci. Amer.,* **220**(3), 18–27.
11. Coutant, C.C., 1968, Thermal pollution—Biological effects a review of the literature of 1967, *J. Water Poll. Cont. Fed.* **40**(6), 1047–1052.
12. Coutant, C.C., 1969, Thermal pollution—Biological effects a review of the literature of 1968, Battelle-Northwest, Richland, Wash.; BNWL-SA-2376, *J. Warer Poll.* Cont. Fed. **41**(6), 1036–1053.
13. Coutant, C.C., 1970, Thermal Pollution—Biological effects a review of the literature of 1969, Battelle-Northwest, Richland, Wash.; BNWL-SA-3255, *J. Water Poll. Cont. Fed.* **42**(6), 1025–1057.
14. Coutant, C.C., 1970, Biological aspects of thermal pollution. I. Entrainment and discharge land effects, CRC Critical Reviews in *Environmental Control* **1**(3), 341–381.
15. Coutant, C.C., 1971, Thermal pollution—Biological effects, in A review of literature of 1970 on wastewater and water pollution control, *J. Water Poll. Cont. Fed.* **43**(6), 1292–1334.
16. Coutant, C.C., 1972, Biological aspects of thermal pollution, II. Scientific basis for water temperature standards at power plants, CRC Crit. Rev. in *Envir. Control.*
17. Coutant, C.C. and C.P. Goodyear, 1972, Thermal effects, in A review of the literature of 1971 on wastewater and water pollution control, *J. Water Poll. Cont. Fed.* **44**(6), 1250–1294.
18. Fry, F.E.J., 1947, Effects of the environment on animal activity, Univ. Toronto Stud. Biol. Ser. No. 55. Publ. Ont. Fish. Res. Lab. No. 68, 1–62.
19. Fry, F.E.J., 1964, Animals in aquatic environments: Fishes (Chap. 44), Handbook of physiology, Section 4: Adaptation to the environment, Amer. Physiol. Soc., Wash. D.C.
20. Fry, F.E.J., 1967. Responses of vertebrate poikilotherms to temperature, in *Thermobiology,* A.H. Rose (ed.) Academic Press, London, pp. 375–409.
21. Kennedy, V.S. and J.A. Mihursky, 1967, Bibliography on the effects of temperature in the aquatic environment, Univ. of Maryland, Nat. Res. Inst. Cont. No. 326, 89 p.
22. Krenkel, P.A. and F.L. Parker, 1969, *Biological Aspects of Thermal Pollution,* Vanderbilt Univ. Press, Nashville, Tennessee.
23. Ministry of Technology, UK, 1968, Water pollution research 1967, p. 63.

24. Parker, F.L. and P.A. Krenkel, 1969b, Thermal pollution: Status of the art, Dept. of Envir. and Water Res. Eng., Vanderbilt Univ. Rept. No. 3.
25. Parker, F.L. and P.A. Krenkel 1969a. *Engineering Aspects of Thermal Pollution,* Vanderbilt Univ. Press, Nashville, Tennessee.
26. Raney, E.C. and B.W. Menzel, 1969, Heated effluents and effects on aquatic life with emphasis on fishes: A bibliography, Ichthyological Associates Bull. No. 2 prepared with Cornell Univ. Water Resources and Marine Sciences Center and Philadelphia Electric Company, 470 p.

CHARLES C. COUTANT
Oak Ridge National Laboratory

TOXIC EFFECTS: see AIR POLLUTANT EFFECTS; EFFECTS OF CHEMICALS

TOXICOLOGY

The terms *toxicology, toxicity,* or *toxic substance* (*toxicant*) are used daily in the scientific and general literature. Review of almost any daily newspaper will reveal one or more articles on the toxic effects of a substance, most of which when released into the environment are called *pollutants.* Today there are scientific journals devoted to the subject of toxicity, illustrating the importance of this topic. However, many do not understand the term *toxicology* or have an understanding of its concepts. So what is a good definition of toxicity? It can be best defined as the science of poisons. Of course, this brings us to the question of what a poison is: any substance that can result in a detrimental effect when the concentration is increased. An increased response as compared to increasing concentration has been called a "dose-response curve," which will be discussed later.

When using the definition of toxicity provided above, most will consider poisoning of animals and humans; however, this definition can be extended to all life forms, including microbes (Thomulka et al., 1996) and plants (Haarmann and Lange, 2000). In the broadest term, toxic insult can be evaluated from an ecological viewpoint and can encompass effects to an ecosystem. This is what is commonly considered when looking at poisoning in an industrial environment. However, in today's changing environment, the viewpoint from an industrial perspective is changing to include the entire environment. The scope of toxicology is ever-increasing, and from the point of view of an engineer, especially an environmental engineer, should not be limited. Depending on the focus, toxicity can also be viewed from global impact (e.g., mercury release from burning fossil fuels) to that which affects single-celled organisms in a local pond.

Public awareness has raised the term *toxicity* to an everyday usage, although most do not understand how to properly apply this term. Most consider that when something is listed as toxic it means an effect from an exposure has occurred. Certainly in the most general sense this is true. Forgotten for the term *toxicity* is that every substance is toxic, at least in the right dose. So what can be added to the concept of a poison is that the dose makes the poison.

For engineers, often the terms *hazardous substance* or *waste* are used as substitutes for *toxicity.* This in the strict definition is not correct, in that a hazardous waste may not act as a poison, but rather result in a physical effect (e.g., a burn). However, even a substance capable of causing a burn will do so in proportion to the concentration applied. Thus, even for these types of substances, there is a dose-response effect. If any effect from a substance is considered a toxic response, then *hazardous waste* is another name for *toxicity.* In most cases a hazardous waste is a mixture of substances and/or chemicals at a site, and its release was uncontrolled or unregulated. Regardless, this mixture will have its own dose-response, while the individual chemicals or substances will exhibit separate responses (a differing dose-response curve).

What is of importance to many engineers when examining toxicity is the use of standard references. Table 1 lists number of textbooks and governmental sources that contain various numerical values for toxicity and basic information on chemicals. These sources are a very good staring point to obtain basic information about a chemical, its regulatory limits, and general information on the hazards associated with the substance.

AREAS OF TOXICOLOGY

Toxicology can be divided into a variety of subareas. These areas can be categorized by organ systems, chemicals (substances), or discipline. Examples of categorization are shown in Table 2, along with a brief description. For the most part, engineers will work in the general areas of environmental and occupational toxicology, although some will venture into others as well. In special cases, engineers will venture into areas such as forensic toxicology. What needs to be kept in mind is that toxicology is an area that borrows from other basic fields of science, such as chemistry, physics, biology, and mathematics.

ENDPOINTS OF TOXICITY

Historically, toxicology was associated with the response of animals when exposed to an agent or agents. Mostly this has been performed using small rodents such as mice and rats. However, for engineers, animal toxicity data are only one part, especially for work that relates to the environmental areas. For example, evaluation of a hazardous-waste site can involve the toxic effects to plants, invertebrates, microbes, and aquatic organisms. Commonly, toxicity of a substance or toxicant is often referred to a single organism. In the environmental area, as well as in others, there may be many different types of organisms affected, along with different effects among these organisms. Use of a single value will not likely represent toxicity to entire groups or a system. Thus, representation of toxicity as a single value may be misleading. Toxicity endpoints for a chemical can vary by logarithmic orders, even for the same organism. This is illustrated by the chemical copper for *Strongylocentrotus purpuratus* using the endpoint EC_{50}, which is the median effective concentration (where 50% of the organisms are affected at a given period of time). ED_{50} is the median exposure dose, which is the concentration in air or water. The other commonly used endpoint of measure for industrial (occupational) toxicology is the median lethal dose (LD_{50}; again, this is a value where 50% of the organisms die at the given concentration, assuming that the mean and median values are equal, as in a normal curve, although used in more studies to refer to the median concentration). Obviously the LD_{50} is not useful in setting occupational-exposure limits, but provides a relative comparison for different chemicals. Similar in nature to the LD_{50} is the EC_{50}. Here the concentration has to be in some unit of air or liquid (water) for the endpoint to be measured. The variability for a chemical as related to effective endpoints (dose) can be illustrated using copper in aquatic organisms (Table 3). The LD_{50} of copper for the various organisms listed have a large variation (log order). This variation is commonly observed when evaluating a chemical among different organisms and even the same organism between laboratories.

A toxic response can be reported as any endpoint measurement that is reproducible. This can include death, as represented by an LD_{50} or another, such as a behavior endpoint measurement, which could be an EC_{50}. When evaluating

TABLE 1

Some common references on environmental and occupational toxicology

Klaassen CD (1996), Casarett and Doulls Toxicology: the basic science of poisons	An excellent reference on toxicology, although generally written at the graduate level.
ACGIH® (2004), TLV's and BEI's	Threshold limit values (TLVs) and biological exposure indices (BEIs) values, which provide the upper exposure limit for many chemicals.
Hathaway et al. (1991), Proctor and Hughes' chemical hazards of the workplace	Provides information on many chemicals—including regulatory exposure limits and basic information on the chemical.
OSHA (29 CFR 1910. 1000)	Permissible exposure limits (PELs), which are the maximum exposure limit set by the U.S. government.
NIOSH Criteria Documents	Information on a specific chemical as provided by NIOSH. However, these reports are not updates and some that are older will not have the most up-to-date information.
Niesink et al. (1995), Toxicology: principles and applications	General toxicology reference that focuses on the occupational environment.
Lippmann (1992), Environmental toxicants: human exposures and their health effects	Provides information through chapters on specific topics that relate to both environmental and occupational toxicology.
Rand and Petrocelli (1985), Fundamentals of aquatic toxicology: methods and applications	A good basic textbook on aquatic toxicology.
NIOSH (1994), NIOSH pocket guide to hazardous chemicals	Provides exposure values, physical properties, and keywords on health hazards for many chemicals of industrial interest.

ACGIH®—American Conference of Governmental Industrial Hygienists

OSHA—U.S. Occupational Safety and Health Administration

NIOSH—National Institute for Occupational Safety and Health (an example of these documents is NIOSH, Criteria for recommended standard occupational exposure to hydrogen fluoride, Department of Health and Human Services (DHHS) (NIOSH) Pub Nos. 76–141)

TABLE 2

Some areas of toxicology

Environmental	Concerned with effects on the environment, which can be considered pollution. This can be further divided into air, soil, and water systems. There can also be a measurement on a species as well.
Forensic	The occurrence of toxic effects on people and possibly other organisms, such as livestock, that is in relation to a crime.
Occupational	Effects of chemicals or substances on those in the working environment and industry.
Regulatory	Effects of chemicals (may also be drugs) in regard to the risk associated with the purposes or in some cases prevention of that chemical's use. This is often associated with some regulation or law, like the U.S. Clean Air Act (CAA).
Mechanistic	Evaluates a chemical's mechanism of action and how this action causes a toxic effect on the organism.

TABLE 3

Aquatic toxicology values of various organisms for copper

Organism	LD_{50}	Reference
Mesocyclops peheiensis	75 μg/l	Wong and Pak, 2004
Tilapia zillii	6.1 mg/l	Zyadah and Abdel-Baky, 2000
Mysis sp. (from Nile River)	2.89 mg/1	Zyadah and Abdel-Baky, 2000
Mugil cephalus	5.3 mg/1	Zyadah and Abdel-Baky, 2000
Photobacterium phosphoreum	>100 mg/1	Thomulka et al., 1993
Strongylocentrotus purpuratus	15.3 μg/1$^+$	Phillips et al., 2003
Penaes merguiensis	0.38 mg/1	Ahsanullah and Ying, 1995

$^+$ An EC_{50}.

data, the endpoint must be identified, especially when looking at nonlethal measurements such as EC_{50}'s.

There are three general routes of exposure: inhalation, dermal (skin), and ingestion (oral). A fourth route, which is more related to medical situations, is injection. Depending on the chemical and the activity employed, one or more of these will have a great deal of importance in the toxic outcome. Occupationally, the most important route is inhalation, since it generally results in the most severe health consequences. Dermal effects are the most numerous, but in most cases are of minor importance. Most dermal effects are related to irritation of the skin and related allergic reactions. As a general rule in occupational toxicology, skin problems are the most common, although effects such as cancer of various organs can also be of concern (Lange, 2003). Using cement as an example, epidemiological studies have reported this agent to cause cancer in a variety of organs. The organs or systems of carcinogenic concern include the skin, bladder, stomach, and lungs (Smailyte et al., 2004), although the most common problem reported in occupations using this building material is dermatological (skin) (Winder and Carmondy, 2002; Lange, 2003), which is a noncarcinogenic occupational hazard. This illustrates that a chemical can have multiple toxic endpoints for different organs.

Most toxicologists divide the exposure to humans and organisms into four categories: acute, subacute, subchronic, and chronic. *Acute* is commonly defined as a single or repeated exposure that occurs over a 24-hour period that results in a measurable effect. Although this definition is not perfect, it tells us that acute cases are generally of short duration and high concentration. *Subacute,* on the other hand, is exposure that occurs over about a 1-month time period and in this case is generally lower in concentration, and the effect requires a longer period of time to occur in comparison to a true acute exposure. It is not uncommon to report acute effects as case studies. In the case report by Dote et al. (2003), an industrial worker accidentally exposed (sprayed) himself with the agent hydrogen fluoride (HF), or hydrofluoric acid. HF is a highly corrosive agent that can result in serous chemical burns, and in this case the burns occurred on the face of the industrial worker. As a result of this exposure, the worker died within a half hour as a result of acute respiratory failure. In the case of HF, this substance would be considered a hazard to both the respiratory and dermal systems, in this case inhalation being the main route of exposure that resulted in death. To put HF exposure in perspective, Hathaway et al. (1991) reported that the LD_{50} for a 5-minute exposure is between 500 and 800 parts per million (ppm).

Chronic toxicology is defined as an effect resulting from an exposure that occurs over a long period of time, like years. Certainly the time period of measurement also depends on the length of an organism's life history as well. *Subchronic,* as compared to chronic, is of shorter duration with a higher concentration and can be considered to occur within a time period of 1 to 3 months for people. Although these terms are discussed for an occupational setting, the terms are also applied to environmental toxicology. Historically, acute exposure was a key factor in exposure prevention. As industrial exposures are becoming better controlled, there has been a change in focus to chronic conditions, at least in the developed countries.

Since inhalation is the most important route of exposure in the occupational (industrial) environment, most reported limits of acceptable exposure are for this route. However, in other systems, such as aquatic or terrestrial, dermal contact or ingestion may be the most important routes of exposure.

OCCUPATIONAL EXPOSURE LIMIT VALUES

For occupational exposure, established upper limits have been published by governmental and private agencies or groups. These values are: permissible exposure limit (PEL), threshold limit value (TLV), and recommended exposure limit (REL). PELs are established by the U.S. Occupational Safety and Health Administration (OSHA) and are the legal standard for the maximum exposure level. OSHA PELs are published in the Code of Federal Regulations (CFR) at 29 CFR 1910.1000. It should be noted that these exposure concentrations are mostly for inhalation, as previously mentioned, and the levels represented are somewhat out of date, since they have to go through a regulatory process for updating. TLVs are established by the American Conference of Governmental Industrial Hygienists (ACGIH), which is considered a consensus organization. Many consider these values to be the most up-to-date, although they are, like most decision-making processes, subject to industry pressure and other political factors when being established. Generally, TLVs are lower in concentration than PELs, although there are exceptions to this statement. It can be considered that the PELs, as they change, are also subject to industry and political considerations as well. Both the PELs and TLVs are established for an 8-hour time-weighted average (TWA). This average is an arithmetic mean of all the exposures collected in that workday. The formula for making a TWA is shown below.

$$\text{TWA} = (C_1 \times T_1) + (C_2 \times T_2) + \ldots + (C_n \times T_n)/ \\ (T_1) + (T_2) + \ldots + (T_n)$$

C—concentration
T—time

The maximum and ideal time of sample (exposure) collection is 8 hours, although this is not usually feasible. Most consider that to obtain a TWA the sample should be collected for at least 6.5 hours of the 8-hour work shift. The remaining 1.5 hours would be included as a 0 exposure level. The REL is a 10-hour TWA exposure limit and is set by the National Institute of Occupational Safety and Health (NIOSH) as a value to be considered by OSHA in the rule-making process. For all the values (PEL, TLV, and REL), they are established for a 40-hour workweek.

When evaluating exposure limits, exceedance can be considered for a single measurement or summation of measurements (Letters to the Editor, 1998). There has been considerable discussion of the correct evaluation for exposure. For those chemicals that are considered to be chronic in nature, disease appears to follow the arithmetic mean of exposure, suggesting that summation exposure values best represent potential health effects (Lange, 2002).

A short-term exposure limit (STEL) has also been established for many chemicals. STELs are for 15-minute periods with at least 2 hours of exposure below the PEL, as an example, with no more than four exposure periods (STELs) occurring per day. When applying STELs, the PEL should not be exceeded when these values are included in the TWA. If there is an exceedance of the PEL, appropriate personal protective equipment is then required.

Exposure limit values (TLV-TWA) are established using three general criteria. First, in order of importance, are epidemiological data. Occupational and in some cases environmental epidemiology studies provide the most important information on the hazards from a chemical. Since there are different types of epidemiological studies, those of the greatest strength, in order, are: cohort, case-control, cross-sectional, and ecological. Next is animal experimentation in identifying hazards, and last are case studies or reports. The ACGIH publishes documentation summarizing the basis for establishing and setting TLVs and is often useful as a general reference. Another good reference that provides summary information on chemicals is Hathaway et al. (1991).

Exposure levels are given in units of mg/m^3, ppm, and fibers per cubic centimeter (f/cc). In most cases these values are for inhalation, but there are some listed for skin (e.g., decaborane).

Another value that is of importance to toxicologists in the industrial environment is IDLH (immediately dangerous to life and health). The problem with IDLH is that it has two different definitions (NIOSH/OSHA/USCG/EPA, 1985). The Mine Safety and Health Administration (MSHA) (30 CFR 11.3[t]) defines IDLH as the concentration that will cause immediate death of permanent injury. However, NIOSH, in the *Pocket Guide* (1994; see Table 1), defines this as the maximum concentration where one can escape within 30 minutes without irreversible health effects. So care must be taken when using IDLH values, as each source has completely different criteria.

DOSE-RESPONSE

In toxicity there exists an increased response to a chemical with the chemical's increasing concentration. This is known as the "dose-response effect" and is fundamental to toxicology. In general, it can be said that every chemical has a dose-response effect. The response is any repeatable indicator or measurement that is used to evaluate the response of an organism to the chemical. At some point the concentration becomes high enough that the response is 100%. Figure 1 shows time of exposure to various concentrations of the chemical sodium bisulfate (Haarmann and Lange, 2000). As the concentration of each chemical varies there is a reduction in root length after a given period of time. In many cases the curve would appear reversed, where there would be no inhibition at the lower concentrations and inhibition at the higher levels. However, here, for the root length, which was for radish-seed elongation, the highest length is at the lower concentration of chemical. The shape of the dose-response curve can provide information on the effect of a chemical, and data extracted from this relationship is often used in risk-assessment analysis. LD_{50} and related values are extracted from dose-response curves. Different formulas can be used to obtain this information as well (Thomulka et al., 1996).

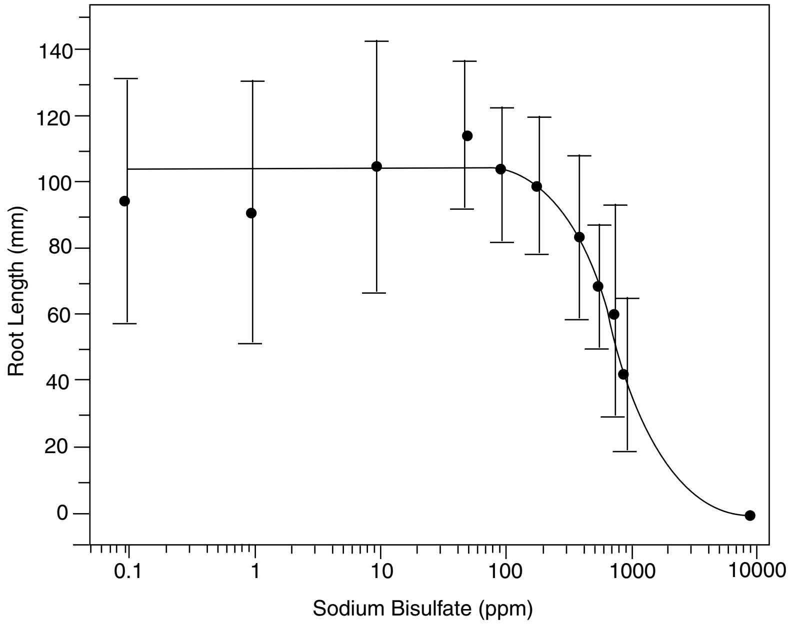


FIGURE 1 Dose-response curve for sodium bisulfate in Lake Erie water (from Haarmann and Lange, 2000; with permission from Parlar Scientific Publications).

Dose-response curves are often used to provide information on a chemical as well as comparison to other chemicals. Potency is one factor that can be derived from the dose-response. This term refers to the concentrations that result in an increasing response to the chemical. Two chemicals can have the same slope on a dose-response curve, but have different potencies. Thus, various information can be extracted from dose-response curves.

EXPOSURE

Exposure can be considered to be at the heart of toxicology. Just because you are exposed does not mean that there will be an effect or even that the chemical will be taken up by the organism. There are a number of factors that influence the cause and effect, including absorption, distribution, excretion, and biotransformation. To understand exposure, a brief discussion of each will be presented.

A toxicant is often called a *xenobiotic,* which means a foreign substance, and these terms are often used interchangeably in texts. In some cases, a xenobiotic may not be foreign to the organism (e.g., selenium), but exist in a higher or lower concentration that results in a disease state. Of importance to environmental and occupational toxicology is that a lower concentration may also result in disease or an undesired event, which for the purposes of this chapter will be considered a toxic action. In some unusual cases increased occupational exposure has been reported to result in beneficial effects. This has been illustrated by the exposure of organic dust that appears to reduce lung cancer (Lange, 2000; Lange et al., 2003). However, it needs to be noted that exposure to organic dust (like cotton dust, in the textile industry) also results in severe respiratory diseases (e.g., bysinosis), which outweigh any benefits of reduced lung cancer, as in this case.

Absorption

Absorption is the process where a xenobiotic crosses a membrane or barrier (skin) and enters the organism, most commonly into the blood. As previously mentioned, the major routes of absorption are ingestion (the gastrointestinal [GI] system), inhalation (lungs), and dermal (skin). Oral intake is not a common route of occupational exposure, but one of major importance environmentally. Transport across barriers occur as passive transport, active transport, facilitated diffusion, or specialized transport. Transport can occur in the uptake and excretion of chemicals. Passive transport, which is simple diffusion, follows Frick's Law and does not require energy. Here a concentration gradient exists, and molecules move from the higher to the lower concentration. As a rule, for biological systems, the more nonionized the form of a molecule, the better it is transported across lipid membranes. The membranes of cells are composed of a lipid bilayer, thus favoring nonionized compounds. Active transport involves the movement of a chemical against a gradient and requires the use of energy. This requires a transporter molecule to facilitate the movement and would be subject to saturation of the system. Facilitated transport is similar to active transport, except it does not work against a gradient and does not require energy. There are other specialized forms of transport, such as phagocytosis by macrophages. These various transport mechanisms are also used to bring essential substances and xenobiotics into the organisms.

Absorption in the GI tract can occur anywhere from the mouth to the rectum, although there are some generalizations that can be made. If the chemical is an organic acid or base, it will most likely be absorbed in locations where it exists in its most lipid-soluble form. The Henderson-Hasselbalch equation can be used to determine at what pH a chemical exists as lipid-soluble (nonionized) as compared to ionized. As a general rule, ionized forms of a chemical are not easily absorbed across biological membranes.

For the lungs, gases, vapors, and particles can be absorbed. In the lungs, ionization of a chemical is not as important as it is for the GI tract. This is due to the rapid absorption of chemicals and the thinness of the separation of alveolar cells (air in the lungs and blood system) with the body fluids (blood). Ionized molecules are also generally nonvolatile and are therefore usually not in high concentration in the air. Particles are separated as they travel the pulmonary system. The larger ones (say, greater than 10 μm in size) are removed early in the pulmonary system, like in the nasal area, whereas the smaller ones (say, 1 μm) enter the alveolar region. As a general rule, it can be said that particles around 5 to 10 μm are deposited in the nasopharyngeal area, those 2 to 5 μm in the tracheobronchial area, and those less than 1 to 2 μm in the alveolar region. The alveolar region is where air is exchanged with the blood system, oxygen is taken up, and waste gases (carbon dioxide) are returned to the atmosphere. Particles that are deposited into the alveolar region have been termed "respirable dust" (Reist, 1993). Distribution of particles described is not exact, but provides a generalization of particle distribution for lungs. Some chemicals, like those that are highly water-soluble (e.g., formaldehyde), can be scrubbed out at various locations of the respiratory tract. Here, formaldehyde is removed by the nose, and in general this is a site of its toxic action, irritation, and nasal cancer (Hansen and Olsen, 1995).

Skin is generally not highly penetrable and is a good overall protective barrier. This protection is a result of the multiple layers of tissue associated with the skin. However, the primary layer of protection is the stratum corneum. This is the top layer of cells on the skin; it is dead and can vary in thickness. On the hands and feet this cell layer can be 400 to 600 μm thick, while on the legs it can be 8 to 15 μm. Some chemicals can disrupt the skin's protection and allow chemicals to pass more easily. An example of this is dimethyl sulfoxide (DMSO), which can de-fat the skin and allow better penetration of chemicals.

Distribution

After a chemical enters the organism, it is usually distributed rapidly. This distribution is commonly achieved by the blood system. Many chemicals have locations in the organism where

TABLE 4
Locations or organs of toxic action by classes of chemical compound

Class of Chemical/Substance	Location or Organ (Example)
Metals	Kidney, bone, immune
Solvents	Liver
Pesticides	Nervous
Radiation	Blood

they concentrate (e.g., lead in bone). It is often important to know where a chemical is concentrated or its organ of toxicity. Some generalities, although not complete, can be made for different classes of compounds (Table 4). However, when evaluating toxicity it is necessary to obtain specific information on the compound because there are many exceptions to general rules of site of toxic action. It is not uncommon that one chemical will have multiple organs or locations of toxicity. A good example of this is the metal arsenic. Arsenic can be both an environmental and occupational poison. Ingestion of arsenic in drinking water, at elevated concentrations, has been shown to result in skin cancer (which has been referred to as Blackfoot disease) as well as other forms of cancer (e.g., lung; Bhamra and Costa, 1992) and noncancer diseases (e.g., dermatological; Lange, 2004a). Environmental problems associated with arsenic exposure (via water) can be most acute and are well illustrated in a well-water problem for Bangladesh (Murshed et al., 2004). Here water wells were established to provide safe drinking-water sources (free of microbial contaminates). However, at the time these wells were placed it was not known that the soil contained high levels of arsenic. This resulted in drinking-water sources being contaminated with this metal. Subsequently, there has been a high rate of arsenic-related diseases (e.g., bladder, liver, and lung cancer; Chen and Ahsan, 2004) as a direct result of using these water sources. Arsenic does not only result in cancer, it also causes many environmentally related noncancer diseases (Milton et al., 2003). As mentioned, there are also occupational diseases from this metal (Bhamra and Costa, 1992; Lange, 2004a). For example, workers in smelting plants that use arsenic have been shown to exhibit elevated levels of lung cancer, and from these types of studies arsenic has been identified as a lung carcinogen. Although arsenic has been reported to cause detrimental effects, it should be noted that it is also an essential trace element. Deficiency in arsenic has been reported to result in various health problems as well as increased mortality (Bhamra and Costa, 1992). Thus, many chemicals can have a dual role in causing and preventing disease. It has even been suggested that some chemicals and substances can have a protective effect in the occupational environmental (Lange, 2000; Lange et al., 2003).

Chemicals can also be identified individually with a site or organ system being affected. Examples of chemicals and their general site of action are shown in Table 5. Certainly this list is not comprehensive, but provides the range of organ systems a single chemical can influence in the disease process. Effects can be both acute and chronic along with many having both carcinogenic and noncarcinogenic properties (e.g., benzene).

TABLE 5
Specific chemicals and some of their general organs or sites of action

Chemical	Location or Organ (Example)
Aluminum	Endrocrine, kidney, lung
Arsenic	Bladder, skin, heart, liver, lung, nervous
Benzene	Blood, liver
Cadmium	Kidney, reproductive
Carbon monoxide	Blood
Coke oven gases	Lung
Cotton dust	Lung
Ethanol	Liver
Formaldehyde	Lung (respiratory)
Fungus (*Fusarium moniliforme*)	Liver
Lead	Bone, blood, heart, kidney, nervous
Mercury	Kidney, nervous, heart
Methyl ethyl ketone	Heart
Paraquat	G1, heart, lung
Phenol	Liver, skin
Polyaromatic hydrocarbons	Immune, liver, reproductive
Polychlorinated biphenys	Immune
Rotenone	Endrocrine, eye, lung, skin
Tetrachloroethylene	Kidney
Thallium	Eye

"Heart" includes the vascular system as a general group.

Excretion

Toxicants that are taken up by an organism must be eliminated in some way. There are three major routes of excretion (urine, feces, and air [exhalation]) and several minor routes (hair, nails, saliva, skin, milk, and sweat). Many compounds are biotransformed before being excreted. This biotransformation results in xenobiotics being more water-soluble. As will be mentioned later, biotransformation involves a two-step process known as Phase I and Phase II biotransformation. Generally, substances with the greatest toxicity are those that do not completely undergo the biotransformation process.

Urinary excretion involves elimination through the kidney and is commonly considered the most important route of excretion. The kidney receives about 25% of the cardiac output. Toxic agents are generally excreted by being filtered out through the glomeruli or tubules in the kidney.

Fecal excretion can involve both the GI tract and liver/gallbladder. Some toxicants pass through the alimentary system (GI tract) unabsorbed or modified by bacteria or other processes in this system. Biliary excretion involves removal of toxicants from the blood by the liver and their subsequent elimination through a fecal route. Here a xenobiotic is biotransformed by the lever and transported to the gallbladder, which then excretes the chemical into the GI tract for elimination. There are some cases where a chemical eliminated by this route is then reabsorbed by the intestine into the body, resulting in a long half-life for this substance. This process is known as the "enterohepatic cycle." Ideally chemicals are metabolized into a polar form, making these poorly reabsorbable. However, microbes in the intestine can transform these compounds into a more lipid-soluble compound, which favors reabsorption.

Exhalation

Substances that exists in a gas phase are mostly eliminated through the lungs. These chemicals are mostly eliminated through simple diffusion, with elimination generally related to the inverse proportion of their rate of absorption. Thus, chemicals with low blood-solubility are rapidly eliminated, while others with high solubility are eliminated slowly.

Other Routes

Several other routes of excretion have been mentioned. Overall, these other routes are of minor importance in elimination of toxicants. However, they can be used to test the existence and concentration of various toxicants in the organism. This is commonly known as "biological monitoring." For example, hair can be used to test where a person has suffered from previous exposure to and possible toxicity of heavy metals, like arsenic. Thus, these minor excretion routes can be important for specific areas of toxicology (e.g., forensic). It should be noted that the major routes can also be used for biological monitoring, with urine and blood being the most important, particularly clinically and occupationally.

Biological Monitoring

Biological monitoring has become a common method for evaluating absorption of chemicals and drugs. It has been used for such activities as drug and alcohol testing. Methods have been established to determine the absorbed dose of a chemical, which are therefore important in many areas of toxicology, including clinical, forensic, and occupational toxicology. The ACGIH has established BEI values for some chemicals as one measure of monitoring risk to industrial populations. This allows evaluation of exposure from all routes, including occupational and nonoccupational. In many cases, only one route of exposure is evaluated, airborne levels, while exposure from other routes (e.g., dermal) contributes to the absorbed and toxic dose. Biological monitoring can be used for both major and minor routes of excretion. As noted, hair and nails can be used to evaluate exposure to heavy metals. An example of biological monitoring in the occupational environment is for methyl ketone (MEK), which has been suggested to be measured at the end of a work shift using urine as the biological fluid. The ACGIH BEI for MEK is 2 mg/l.

Biological monitoring is also used as part of medical evaluations and in environmental toxicology as well. A good example of its use in medical evaluations is for lead-abatement workers. Blood lead levels (BLL) for workers in this industry or exposure category have been established by OSHA. Here workers having a BLL over 40 μg/dl (deciliter of whole blood—100 ml of blood) are required to undergo an annual medical examination. Workers over 50 μg/dl are required to be removed from the work area (removal from exposure) until the BLL (two connective readings) is below 40 μg/dl. This illustrates the use of biological monitoring in prevention of occupational disease and its incorporation in regulatory toxicology.

Environmentally, lead is often monitored in children since it can cause harm in a number of organ systems and with effects that are characterized with a developing organism. The Centers for Disease Control and Prevention (CDC) suggest that children below the age of 6 not have a BLL that exceeds 10 μg/dl. This is the lowest level that has been suggested to have biological effects for humans. Biological concentrations of chemicals have also been used to evaluate exposure and toxic effects in organisms other than man.

Monitoring of biological fluids and tissue in environmental toxicology is a common practice (Pip and Mesa, 2002). Both plants (Pip and Mesa, 2002) and animals (Madenjian and O'Connor, 2004) are used for evaluating the distribution and uptake of toxicants from polluted environments. Monitoring can also be extended to abiotic conditions that influence toxicity to organisms (Mendez et al., 2004). The use of biological systems for monitoring can include effects on metabolism and other systems as well (Lange and Thomulka, 1996). Thus, biological monitoring is commonly used in both environmental and occupational settings as well as other areas of toxicology. Monitoring of this nature has even been extended to ecosystems as a methodology for evaluating health.

Biotransformation

Xenobiotic substances that are taken up by an organism must eventually be eliminated. To eliminate many of these chemicals, they must be transformed into a water-soluble product. This transformation is called "biotransformation." In many vertebrates, this transformation occurs in the liver, although other tissues and organs (e.g., the kidney) are also involved. Generally, chemicals are absorbed as lipid compounds and excreted as water-soluble (hydrophilic) compounds. Hydrophilic compounds can be easily passed along with the urine and feces. In the lungs, volatile compounds are favored for excretion in the exhaled gas, while those that are nonvolatile are generally retained. If chemicals were not biotransformed, their rate of excretion as lipid-soluble compounds would be very long, and this would result in buildup of xenobiotics. The rate at which a chemical is metabolized or excreted is called its half-life ($t_{1/2}$). Half-lives can be very short (as in minutes) or long (as in years).

Biotransformation and *metabolism* are often used as synonymous terms. In general they can be used interchangeably, although here *biotransformation* is used in describing the metabolism of xenobiotics that are not part of normal metabolism or at concentrations related to pollutant or toxicant exposure.

Some chemicals are able to actually increase or stimulate the biotransformation of other compounds. This is known as "induction." Induction can occur for a variety of compounds.

As previously mentioned, biotransformation is generally divided into two categories, Phase I and Phase II. Phase I reactions involve oxidation, reduction, and hydrolysis, which prepare the compound to undergo a Phase II reaction. Phase II involves conjugation. Commonly the most toxic products of a chemical are those from Phase I. If the system becomes saturated, Phase I compounds will seek alternative routes of metabolism, and this may result in more toxic intermediates. If this occurs, it is said that the metabolic system has become saturated.

MIXTURE TOXICITY

Most toxicology studies involve the use of a single compound; however, rarely in the real world does exposure occur to only a single substance. Although single-exposure events do occur, they generally result in acute toxicity, while multiple exposures are more frequently associated with chronic events. Certainly there are numerous exceptions to this rule, like asbestos and mesothelioma, but even with asbestos there are mixtures associated with this substance. One of the best illustrations for a mixture is asbestos and smoking in the case of lung cancer. Here smoking magnifies the potential effect of inhaled asbestos, resulting in a higher-than-expected rate of lung cancer than would occur for either alone. Most exposures in the industrial environmental focus on a single predominant toxicant associated with that activity, or at the most the top two or three chemicals, and generally concerns are identified with acute events. Both PEL and TLV are established with nonexposure time periods between exposures and often have an emphasis on acute occurrences. In environmental toxicology this is not always the case, since most regulatory standards have been established to protect against chronic events, considering most organisms spend their entire life in a single media. This is also true for humans as related to air and water pollution.

Mixture toxicity or interaction studies can be generally categorized by several terms (Table 6). *Additivity* is when two chemicals together exhibit equal toxicity with each having the same additive response. So if chemicals A and B were mixed and have an effect of ten, by adding five units of each, than adding ten units of A alone or B alone would have the effect of ten as well. *Synergism* is where the combination of the two chemicals magnify the outcome, as in asbestos and smoking. Asbestos may cause 1 cancer in 1000 and smoking 200 cases in 1000, but when together this may rise to 700 cases out of 1000. *Antagonism* is when one chemical reduces the effect caused when combined with another. *Potentiation* is when one chemical allows another to have its full toxic potential. This can be illustrated when the barrier of the skin is disrupted, as with DMSO, and a chemical that would not previously pass through the skin now enters easily. Generally, most chemical combinations exhibit additivity.

TABLE 6
Terms used for identifying mixture interactions

Term	Example by Numerical Value
Additivity	$5 + 5 = 10$
Synergism	$5 + 5 = 40$
Antagonism	$5 + 5 = 7$
Potentiation	$1 + 5 = 12$

Unfortunately, little information exists on chemical combinations (Lange and Thomulka, 1997). The lack of information is often due to the complexity and costs associated with these studies. However, recent advances in using bacterial systems (Lange and Thomulka, 1997) for evaluating mixtures does provide a more cost-effective and convenient way of testing more than one chemical.

There have been a number of methods published, excluding statistical comparisons, for evaluating two chemicals in combination. One of the early methods was a graphic representation of the two chemicals together, called an "isobole plot" (Lange and Thomulka, 1997). Here chemical combinations at some set value (like each chemical's LD_{50}) are plotted. Usually combinations of 100% of A, 80(A)/20(B)%, 60/40%, 20/80%, and 100% of B are used in making the plot. When this graph is represented in proportions, it is called an isobologram (Lange et al., 1997).

Another method that employs a formula is called the additive index (AI) (Lange and Thomulka, 1997). Here two chemicals using the same endpoint value (like LD_{50}) are evaluated, and these results are incorporated into the formula to obtain the AI. The AI is shown below:

$$S = \mathrm{A}m/\mathrm{A}i + \mathrm{B}m/\mathrm{B}i$$

S is sum of activity
A and B are chemicals
i is individual chemical and m is mixture of toxicities (LD_{50})

for S 1.0, the AI $= 1/S - 1.0$
for S 1.0, the AI $= S(-1) + 1$

For the AI, a negative number (sum of activity, S) suggests that the chemicals are less than additive (antagonistic), with zero being additive and a positive value synergistic. Certainly in these calculations the numbers are not exact, so confidence intervals (CIs) are often incorporated to reflect the range of these mixture interactions. In using CI values, at 95%, the upper and lower CIs are used to determine the range. If the CI range includes zero, then this mixture is considered to be additive.

Mixture toxicity is a commonly discussed topic, but as mentioned, it is not well understood. One basis for synergism is related to inhibition of detoxification pathways; however, as noted, most chemical mixtures are additive, which is probably due to few chemicals, at least in mixtures, using exactly the same metabolic pathways. Other methods exist for evaluating mixtures (e.g., the mixture-toxicity index; Lange et al., 1997). Determination of interactions for more than one chemical can in many ways be identified as an art (Marking, 1985). However, as science develops, better methods are being developed to evaluate combinations.

CARCINOGENICITY

The existence of cancer-causing chemicals has been known for thousands of years. However, it was not until recently that a direct relationship between environment or occupation and cancer was established. One of the early examples of an occupational relationship was provided by the English physician Percival Pott around 1775. Pott observed a high number of cases of scrotum cancer in chimney sweeps. He concluded that this cancer was a result of soot exposure in this occupational group. Later Japanese investigators (Yamagawa and Ichikawa, 1915) determined that coal tar (a common component of which is polyaromatic hydrocarbons), a component in soot, exhibited carcinogenic effects on animals, providing a basic animal model to support the occupational observations of Pott.

Cancer in its simplest term is the unregulated or uncontrolled growth of cells in an organism. Cancer, or neoplasm, can be either benign or malignant. Those that are benign generally occupy a given space and do not spread to other parts of the body. If the cancer is said to be malignant, it is then metastatic and can spread and form secondary sites at various other locations within the body.

Probably the best-known cancer-causing agent is cigarette smoke. Studies have shown that direct and indirect use of this product can result in cancer. Doll and Hill (1954) demonstrated that cigarette smoking was a major cause of lung cancer. Although this was an epidemiological study, these types of investigations opened up a new era of investigation into cancer-causing agents.

A cancer-causing agent generally has two processes in the causation of a tumor: initiation and promotion. This has resulted in chemicals being identified as either initiators or promoters, although there are some, known as "complete carcinogens," that exhibit both properties. This concept was developed by painting chemicals on the skin of mice at different time periods and observing whether tumor formation occurred. It was discovered that for some chemical combinations, the initiator had to be applied before the promoter. When the promoter was applied first, a time period waited, and then the initiator applied, no tumor formation occurred. Cancer can also be caused by other nonchemical factors such as heredity and viruses.

There has been considerable debate as to the amount of cancer caused by environmental pollutants and exposures in the occupational environment. However, it is known that there are a large number of agents capable of causing cancer in both the environmental and occupational settings. A list of a few occupationally associated carcinogens is shown in Table 7.

TABLE 7
Some carcinogenic chemicals (substances) and the cancers they cause

Substance	Cancer Caused
Aniline	Bladder
Arsenic	Skin, lung
Benzene	Leukemia
Cadmium oxide	Prostate, lung
Carbon tetrachloride	Liver
Cement dust (Portland cement)	Lung, stomach
Coke oven gases	Lung, kidney
Ethylene oxide	Leukemia
Lead arsenate	Lung, skin
Mustard gas	Larynx, lung
Nickel	Lung, nasal
Styrene oxide	Stomach
Toxaphene	Liver

This list is not complete but demonstrates the large variety and locations of cancers. Most environmental engineers look at the agents capable of or identified as causing cancer when evaluating a situation; however, this is usually done for simplicity, in that cancer is an endpoint of clarity—it exists or it does not exist. It must be kept in mind that there are other endpoints of interest as well that are noncarcinogenic (e.g., kidney toxicity).

To classify carcinogens, several agencies list chemicals or substances according to their degree of carcinogenicity. One of the most frequently cited agencies is the International Agency for Research on Cancer (IARC). The IARC is located in Lyon, France, and is part of the World Health Organization. As part of this agency's charter, it publishes monographs for various substances and is considered by many an excellent reference on information on carcinogens. This agency classifies cancer-causing agents into five different groups (Table 8). These grouping are based on data from epidemiological and animal studies. Many consider the IARC to be the best source of information and classification for carcinogens.

Group 1 indicates that there is sufficient epidemiological data that the substance is a human carcinogen. This is the highest level of classification, and as noted in Table 8, an example is arsenic. Group 2 has two classifications, A and B. Group 2A represents limited epidemiological evidence but sufficient animal evidence that the substance is a carcinogen, while with group 2B there is sufficient animal evidence, but epidemiological data are lacking or of poor quality. With Group 3 there is inadequate evidence for classifying a chemical or substance as a carcinogen. Group 4 evidence supports that it is not a carcinogen.

The IARC is not the only agency that classifies carcinogens. The National Toxicity Program (NTP) provides a classification scheme. Here carcinogens are listed as known human carcinogens or as reasonably anticipated to be a human carcinogen.

TABLE 8
IARC classification groups for carcinogenic substances

Group 1: Carcinogenic to humans (common called "known") carcinogen (examples: asbestos, arsenic) (evidence supports the chemical or substance as a human carcinogen)

Group 2A: Probably carcinogenic to humans (examples: diethyl sulfate, vinyl bromide) (limited evidence in humans and sufficient evidence in experimental animals)

Group 2B: Possibly carcinogenic to humans (examples: bracken fern, chlordane) (limited evidence in humans and less than sufficient evidence in experimental animals)

Group 3: Unclassified or not classified as carcinogenic to humans (examples: aldrin, aniline) (inadequate evidence in humans and inadequate or limited evidence in animals)

Group 4: Probably not carcinogenic to humans (example: caprolactam) (evidence suggesting lack of carcinogenicity in humans and animals)

For a chemical to be classified as a known human carcinogen there must be sufficient evidence to support a causal relationship between its exposure and the occurrence of human cancer. For substances to be listed as an anticipated human carcinogen there must exist sufficient evidence that it is a carcinogen, but alternative explanations exist or there is insufficient evidence supporting classification, experimentally and epidemiologically, as a carcinogen.

Regardless of the classification used, evidence of carcinogenicity for a substance requires epidemiological and experimental evidence. Epidemiological information is considered to be the strongest in establishing a substance as a carcinogen.

REGULATORY TOXICOLOGY

Regulatory toxicology is the area that interrelates toxicology with regulatory standards. The purpose of this area is to establish standards to provide protection against a specific chemical or group of chemicals. In many cases, standards are established before the full knowledge of a chemical is complete. Some identify this type of decision making to be part of risk assessment. In the United States, regulations related to toxicology can be generally divided into the major agencies that promulgate criteria for chemicals. These are the Food and Drug Administration (FDA), the Environmental Protection Agency (EPA), OSHA, MSHA, and the Consumer Product Safety Commission (CPSC). There are other agencies (e.g., the Department of Transportation), but for the purposes of this section they are considered minor. The agencies that are important for environmental engineers are the EPA and OSHA. However, those with the mining industry will also consider MSHA of great importance. The EPA, in general, establishes standards for environmental protection, and OSHA for protection related to those in the occupational environment. For consumer substances and products, the CPSC regulates toxicity.

OSHA came into existence with the passage of the Occupational Safety and Health Act on December 29, 1970 (effective April 28, 1971). OSHA as well as MSHA are part of the U.S. Department of Labor. OSHA has five major parts, with each regulating different industrial groups (Table 9).

TABLE 9
Sections of the CFR related to OSHA standards

29 CFR 1910—General industry
29 CFR 1915—Shipyards
29 CFR 1917—Marine terminals
29 CFR 1918—Longshoring
29 CFR 126—Construction

TABLE 10
Some environmental acts of importance

Clean Air Act
Clean Water Act
Toxic Substance Control Act
Resource Conservation and Recovery Act
National Environmental Policy Act
Comprehensive Environmental Response, Compensation, and Liability Act Emergency Planning and Right to Know Act

The OSHA act sets out two primary duties for employers, which are for them to maintain a workplace free of recognized hazards and to comply with OSHA regulations. The act also requires that employees comply with the act, although clarification of this requirement is often lacking. Requirements of the employer are called the General Duty Clause. States can have their own OSHA plan and enforce OSHA as a state provision if they meet certain requirements. Currently, there are 23 state plans.

Commonly, environmental engineers will be required to interact with OSHA inspectors. OSHA often conducts inspections as a random process, or more frequently does so as a result of complaint. When an inspection occurs, the inspector will present identification to the management of the facility. If a labor union exists, the inspector must also notify the labor-union representative. Usually there is then an examination of the OSHA records, usually materials safety data sheets (MSDS) and the OSHA 200 form. Lack of MSDS, which is part of the Hazard Communication Plan, is one of the most frequently cited violations. A walkthrough is then conducted, which may include the collection of samples. At the end of this process there is a closing conference. At this time alleged violations are discussed. If citations are issued they can consist of one of three types: imminent danger, serious violations, and willful violations.

Employers can contest citations. This is usually initiated through an informal hearing. If the employer then decides to contest the citation, there is a specific process that must be undertaken. OSHA has an independent review commission as part of the Department of Labor to hear contested citations. To contest the citation, the employer must file notice within 15 working days by certified mail. There is no formal application, but the notice must clearly indicate that the citation is being contested. The employer must also post notice of this contestation. The notice is sent to the area director of OSHA.

The EPA was not created, as OSHA was, through legislative action, but rather by presidential order. This agency establishes regulations for environmental protection. There are a number of environmental acts and provisions (Table 10).

Some of the regulations established by the EPA overlap OSHA and other agencies. There is often confusion about the interpretation of the various regulations when they cross regulatory boundaries. Since most environmental regulations involve control of toxicants, toxicology is an important aspect of EPA regulations.

REFERENCES

Ahsanullah M. and Ying W. (1995). Toxic effects of dissolved copper in *Penaes merguiensis* and *Penaes monodin. Bulletin of Environmental Contamination and Toxicology.* 55:81–88.

American Conference of Governmental Industrial Hygienists. (2004). TLV's and BEI's. Cincinnati, OH.

Bhamra R.K. and Costa M. (1992). Trace elements aluminum, arsenic, cadmium, mercury, and nickel. In *Environmental toxicants: human exposures and their health effects* (Lippmann M. ed.). Van Nostrand Reinhold, New York.

Chen Y. and Ahsan H. (2004). Cancer burden from arsenic in drinking water in Bangladesh. *American Journal of Public Health.* 94:741–44.

Doll R. and Hill A.B. (1954). The mortality of doctors in relation to their smoking habits. *British Medical Journal.* ii:1451–55.

Dote T., Kono K., Usuda K., Shimizu H., Kawasaki T., Dote E. (2003). Lethal inhalation exposure during maintenance operations of a hydrogen fluoride liquefying tank. *Toxicology and Industrial Health.* 19:51–54.

Haarmann F. and Lange J.H. (2000). Toxicity of sodium bisulfate, sodium oxalate and copper sulfate with three different water sources using a root elongation test. *Fresenius Environmental Bulletin.* 9:172–78.

Hansen J. and Olsen J.H. (1995). Formaldehyde and cancer morbidity among male employees in Denmark. *Cancer Causes and Control.* 6:354–60.

Hathaway G.J., Proctor N.H., Hughes J.P., Fischman M.L. (1991). *Proctor and Hughes' chemical Hazards of the Workplace.* Van Nostrand Reinhold, New York.

Klaassen C.D. (1996). *Casarett and Doulls toxicology: The Basic Science of Poisons.* McGraw-Hill (5th edition), New York.

Lange J.H. (2000). Reduced cancer rates in agricultural workers: a benefit of environmental and occupational endotoxin exposure. *Medical Hypotheses.* 55:383–85.

Lange J.H. (2002). Airborne exposure and soil levels associated with lead abatement of a steel tank. *Toxicology and Industrial Health.* 18:28–38.

Lange J.H. (2004a) Arsenic exposure levels during cleanup of flyash and dermatitis in an air sampling technician. *Bulletin of Environmental Contamination and Toxicology.* 72:1098–100.

Lange J.H. (2003). Cement: a common cancer agent? (Letter to the Editor.) *Toxicology and Industrial Health* 18:183.

Lange J.H., Mastrangelo G., Fedeli U., Fadda E., Rylander R., Lee E. (2003). Endotoxin exposure and lung cancer mortality by type of farming: is there a hidden dose-response relationship? *Annuals of Agriculture and Environmental Medicine.* 10:229–32.

Lange J.H. and Thomulka K.W. (1996). Evaluation of mixture toxicity for nitrobenzene and trinitrobenzene at various equitoxic concentrations using the *Vibrio harveyi* bioluminescence toxicity test. *Fresenius Environmental Bulletin.* 7:444–51.

Letters to the Editor. (1998). A log-normal distribution-based exposure assessment method for unbalanced bias. *Annuals of Occupational Hygiene.* 42:413–22.

Lippmann M. (1992). *Environmental Toxicants: Human Exposures and Their Health Effects.* Van Nostrand Reinhold, New York.

Madenjian C.P. and O'Connor D.V. (2004). Tropic transfer efficiency of DDT to lake trout (Salvenlinus namaycush) from their prey. *Bulletin of Environmental Contamination and Toxicology.* 72:1219–25.

Marking L.L. (1985). Toxicity of chemical mixtures. In *Fundamentals of Aquatic Toxicology: Methods and Applications* (Rand G.M., Petrocelli S.R., eds.). Taylor and Francis, Bristol, PA.

Mendez L., Acosta B., Arreola-Lizarraga A., Padilla G. (2004). Anomalous levels of heavy metals in sediments from Guaymas Bay, Mexico. *Bulletin of Environmental Contamination and Toxicology.* 72:1101–6.

Milton A.H., Hasan Z., Rahman A., Rahman M. (2003). Non-cancer effects of chronic arsenicosis in Bangladesh: preliminary results. *Journal of Environmental Science and Health: Toxic Hazardous Substances and Environmental Engineering.* 38:301–5.

Murshed R., Douglas R.M., Ranmuthugala G., Caldwell B. (2004). Clinicians' role in management of arsenicosis in Bangladesh: interview study. *British Medical Journal.* 328:493–94.

National Institute of Occupational Safety and Health. (1994). NIOSH Pocket Guide to Chemical Hazards. DHHS (NIOSH) Pub Nos. 94–116, Washington, DC.

Niesink R.J.M., de Vries J., Hollinger M.A. (1995). *Toxicology: Principles and Applications.* CRC Press, New York.

NIOSH/OSHA/USCG/EPA. (1985). Occupational Safety and Health Guidance Manual for Hazardous Waste Activities. DHHS (NIOSH) Pub Nos. 85–115, Washington, DC.

Phillips B.M., Nicely P.A., Hunt J.W., Anderson B.S., Tjeerdema R.A., Palmer S.E., Palmer F.H., Puckett H.M. (2003). Toxicity of cadmium-copper-zinc-nickel-zinc mixture to larval purple sea urchins *(Strongylocentroutus purpuratus). Bulletin of Environmental Contamination and Toxicology.* 70:592–99.

Pip E. and Mesa C. (2002). Cadmium, copper, and lead in two species of *Artemisia* (Compositae) in southern Manitoba, Canada. *Bulletin of Environmental Contamination and Toxicology.* 69:644–48.

Rand G.M. and Petrocelli S.R. (1985). *Fundamentals of Aquatic Toxicology: Methods and Applications.* Taylor and Francis, Bristol, PA.

Reist P.C. (1993). *Aerosol Science and Technology.* McGraw-Hill, New York.

Smailyte G., Kurtinaitis J., Anderson A. (2004). Mortality and cancer incidence among Lithuanian cement producing workers. *Occupational and Environmental Medicine.* 61:529–34.

Thomulka K.W., Abbas C.G., Young D.A., Lange J.H. (1996). A method of evaluating median effective concentrations of chemicals with bioluminescent bacteria. *Bulletin of Environmental Contamination and Toxicology.* 56:446–52.

Thomulka K.W., McGee D.J., Lange J.H. (1993). Use of the bioluminescent bacterium *Photobacterium phosphoreum* to detect potentially biohazardous materials in water. *Bulletin of Environmental Contamination and Toxicology.* 51:538–44.

Winder C. and Carmondy M. (2002). The dermal toxicity of cement. *Toxicology and Industrial Health.* 18:321–31.

Wong C.K. and Pak A.P. (2004). Acute and subacute toxicity of the heavy metals copper, chromium, nickel, and zinc, individually and in mixtures to the fresh water copepod *Mesocyclops peheiensis. Bulletin of Environmental Contamination and Toxicology.* 73:190–96.

Zyadah M.A. and Abdel-Baky T.E. (2000). Toxicity and bioaccumulation of copper, zinc, and cadmium in some aquatic organisms. *Bulletin of Environmental Contamination and Toxicology.* 64:740–47.

J.H. LANGE
Envirosafe Training and Consultants

U

URBAN AIR POLLUTION MODELING

INTRODUCTION

Urban air pollution models permit the quantitative estimation of air pollutant concentrations by relating changes in the rate of emission of pollutants from different sources and meteorological conditions to observed concentrations of these pollutants. Many models are used to evaluate the attainment and maintenance of air quality standards, urban planning, impact analysis of existing or new sources, and forecasting of air pollution episodes in urban areas.

A mathematical air pollution model may serve to gain insight into the relation between meteorological elements and air pollution. It may be likened to a transfer function where the input consists of both the combination of weather conditions and the total emission from sources of pollution, and the output is the level of pollutant concentration observed in time and space. The mathematical model takes into consideration not only the nature of the source (whether distributed or point sources) and concentrations at the receptors, but also the atmospheric processes that take place in transforming the concentrations at the source of emission into those observed at the receptor or monitoring station. Among such processes are: photochemical action, adsorption both on aerosols and ground objects, and of course, eddy diffusion.

There are a number of areas in which a valid and practical model may be of considerable value. For example, the operators of an industrial plant that will emit sulfur dioxide want to locate it in a particular community. Knowing the emission rate as a function of time; the distribution of wind speeds, wind direction, and atmospheric stability; the location of SO_2-sensitive industrial plants; and the spatial distribution of residential areas, it is possible to calculate the effect the new plant will have on the community.

In large cities, such as Chicago, Los Angeles, or New York, during strong anticyclonic conditions with light winds and low dispersion rates, pollution levels may rise to a point where health becomes affected; hospital admissions for respiratory ailments increase, and in some cases even deaths occur. To minimize the effects of air pollution episodes, advisories or warnings are issued by government officials. Tools for determining, even only a few hours in advance, that unusually severe air pollution conditions will arise are invaluable. The availability of a workable urban air pollution model plus a forecast of the wind and stability conditions could provide the necessary information.

In long-range planning for an expanding community it may be desirable to zone some areas for industrial activity and others for residential use in order to minimize the effects of air pollution. Not only the average-sized community, but also the larger megalopolis could profitably utilize the ability to compute concentrations resulting from given emissions using a model and suitable weather data. In addition, the establishment of an air pollution climatology for a city or state, which can be used in the application of a model, would represent a step forward in assuring clean air.

For all these reasons, a number of groups have been devoting their attention to the development of mathematical models for determining how the atmosphere disperses materials. This chapter focuses on the efforts made, the necessary tools and parameters, and the models used to improve living conditions in urban areas.

COMPONENTS OF AN URBAN AIR POLLUTION MODEL

A mathematical urban air pollution model comprises four essential components. The first is the source inventory. One must know the materials, their quantities, and from what location and at what rate they are being injected into the atmosphere, as well as the amounts being brought into a community across the boundaries. The second involves the measurement of contaminant concentration at representative parts of the city, sampled properly in time as well as space. The third is the meteorological network, and the fourth is the meteorological algorithm or mathematical formula that describes how the source input is transformed into observed values of concentration at the receptors (see Figure 1). The difference between what is actually happening in the atmosphere and what we think happens, based on our measured

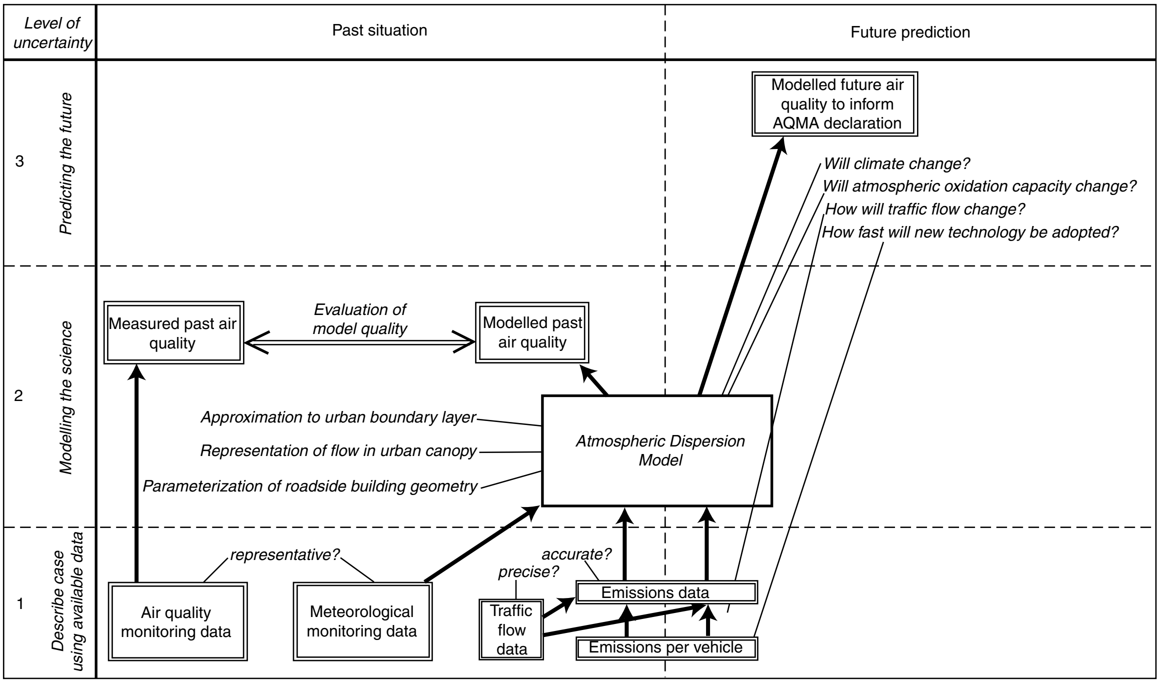


FIGURE 1 Schematic diagram showing flow of data into and out of the atmospheric dispersion model, and three categories of uncertainty that can be introduced (From Colvile et al., 2002, with permission from Elsevier).

sources and imperfect mathematical formulations as well as our imperfect sampling of air pollution levels, causes discrepancies between the observed and calculated values. This makes the verification procedure a very important step in the development of an urban air pollution model. The remainder of this chapter is devoted to these four components, the verification procedures, and recent research in urban air pollution modeling.

Accounts may be found in the literature of a number of investigations that do not have the four components of the mathematical urban air pollution model mentioned above, namely the source inventory, the mathematical algorithm, the meteorological network, and the monitoring network. Some of these have one or more of the components missing. An example of this kind is the theoretical investigation, such as that of Lucas (1958), who developed a mathematical technique for determining the pollution levels of sulfur dioxide produced by the thousands of domestic fires in a large city. No measurements are presented to support this study. Another is that of Slade (1967), which discusses a megalopolis model. Smith (1961) also presented a theoretical model, which is essentially an urban box model. Another is that of Bouman and Schmidt (1961) on the growth of pollutant concentrations in the cities during stable conditions. Three case studies, each based on data from a different city, are presented to support these theoretical results. Studies relevant to the urban air pollution problem are the pollution surveys such as the London survey (Commins and Waller, 1967), the Japanese survey (Canno et al., 1959), and that of the capital region in Connecticut (Yocum et al., 1967). In these studies, analyses are made of pollution measurements, and in some cases meteorological as well as source inventory information are available, but in most cases, the mathematical algorithm for predicting pollution is absent. Another study of this type is one on suspended particulate and iron concentrations in Windsor, Canada, by Munn et al. (1969). Early work on forecasting urban pollution is described in two papers: one by Scott (1954) for Cleveland, Ohio, and the other by Kauper et al. (1961) for Los Angeles, California. A comparison of urban models has been made by Wanta (1967) in his refreshing article that discusses the relation between meteorology and air pollution.

THE SOURCE INVENTORY

In the development of an urban air pollution model two types of sources are considered: (1) individual point sources, and (2) distributed sources. The individual point sources are often large power-generating station stacks or the stacks of large buildings. Any chimney stack may serve as a point source, but some investigators have placed lower limits on the emission rate of a stack to be considered a point source in the model. Fortak (1966), for example, considers a source an individual point source if it emits 1 kg of SO_2 per hour, while Koogler et al. (1967) use a 10-kg-per-hour criterion. In addition, when ground concentrations are calculated from the emission of an elevated point source, the effective stack height must be determined, i.e., the actual stack height plus the additional height due to plume rise.

Information concerning emission rates, emission schedules, or pollutant concentrations is customarily obtained by means of a source-inventory questionnaire. A municipality with licensing power, however, has the advantage of being able to force disclosure of information provided by a source-inventory questionnaire, since the license may be withheld until the desired information is furnished. Merely the awareness of this capability is sufficient to result in gratifying cooperation. The city of Chicago has received a very high percentage of returns from those to whom a source-inventory questionnaire was submitted.

Information on distributed sources may be obtained in part from questionnaires and in part from an estimate of the population density. Population-density data may be derived from census figures or from an area survey employing aerial photography.

In addition to knowing where the sources are, one must have information on the rate of emission as a function of time. Information on the emission for each hour would be ideal, but nearly always one must settle for much cruder data. Usually one has available for use in the calculations only annual or monthly emission rates. Corrections for diurnal patterns may be applied—i.e., more fuel is burned in the morning when people arise than during the latter part of the evening when most retire. Roberts et al. (1970) have referred to the relationship describing fuel consumption (for domestic or commercial heating) as a function of time—e.g., the hourly variation of coal use—as the "janitor function." Consideration of changes in hourly emission patterns with season is, of course, also essential.

In addition to the classification involving point sources and distributed sources, the source-inventory information is often stratified according to broad general categories to serve as a basis for estimating source strengths. The nature of the pollutants—e.g., whether sulfur dioxide or lead—influences the grouping. Frenkiel (1956) described his sources as those due to: (1) automobiles, (2) oil and gas heating, (3) incinerators, and (4) industry; Turner (1964) used these categories: (1) residential, (2) commercial, and (3) industrial; the Connecticut model (Hilst et al., 1967) considers these classes: (1) automobiles, (2) home heating, (3) public services, (4) industrial, and (5) electric power generally. (Actually, the Connecticut model had a number of subgroups within these categories.) In general, each investigator used a classification tailored to his needs and one that facilitated estimating the magnitude of the distributed sources. Although source-inventory information could be difficult to acquire to the necessary level of accuracy, it forms an important component of the urban air pollution model.

MATHEMATICAL EQUATIONS

The mathematical equations of urban air pollution models describe the processes by which pollutants released to the atmosphere are dispersed. The mathematical algorithm, the backbone of any air pollution model, can be conveniently divided into three major components: (1) the source-emissions subroutine, (2) the chemical-kinetics subroutine, and (3) the diffusion subroutine, which includes meteorological parameters or models. Although each of these components may be treated as an independent entity for the analysis of an existing model, their inferred relations must be considered when the model is constructed. For example, an exceedingly rich and complex chemical-kinetic subroutine when combined with a similarly complex diffusion program may lead to a system of nonlinear differential equations so large as to preclude a numerical solution on even the largest of computer systems. Consequently, in the development of the model, one must "size" the various components and general subroutines of compatible complexity and precision.

In the most general case, the system to be solved consists of equations of continuity and a mass balance for each specific chemical species to be considered in the model. For a concise description of such a system and a cogent development of the general solution, see Lamb and Neiburger (1971).

The mathematical formulation used to describe the atmospheric diffusion process that enjoys the widest use is a form of the Gaussian equation, also referred to as the modified Sutton equation. In its simplest form for a continuous ground-level point source, it may be expressed as

$$\frac{\chi}{Q} = \frac{1}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2} \right) \tag{1}$$

where

χ: concentration (g/m^3)

Q: source strength (g/sec)

u: wind speed at the emission point (m/sec)

σ_y: perpendicular distance in meters from the centerline of the plume in the horizontal direction to the point where the concentration falls to 0.61 times the centerline value

σ_z: perpendicular distance in meters from the centerline of the plume in the vertical direction to the point where the concentration falls to 0.61 times the centerline value

x, y, z: spatial coordinates downwind, cross-origin at the point source

Any consistent system of units may be used.

From an examination of the variables it is readily seen that several kinds of meteorological measurements are necessary. The wind speed, u, appears explicitly in the equation; the wind direction is necessary for determining the direction of pollutant transport from source to receptor.

Further, the values of σ_y and σ_z depend upon atmospheric stability, which in turn depends upon the variation of temperature with height, another meteorological parameter. At the present time, data on atmospheric stability over large urban areas are uncommon. Several authors have proposed diagrams or equations to determine these values.

The temperature variation with height may be obtained by means of thermal elements mounted on radio or television towers. Tethered or free balloons carrying suitable sensors may also be used. Helicopter soundings of temperature have been used for this purpose in New York City; Cincinnati, Ohio; and elsewhere. There is little doubt that as additional effort is devoted to the development of urban air pollution models, adequate stability measurements will become available. In a complete study, measurements of precipitation, solar radiation, and net radiation flux may be used to advantage. Another meteorological variable of importance is the hourly temperature for hour-to-hour predictions, or the average daily temperature for 24-hour calculations. The source strength, Q, when applied to an area source consisting of residential units burning coal for space heating, is a direct function of the number of degree-hours or degree-days. The number of degree-days is defined as the difference between the average temperature for the day and 65°. If the average temperature exceeds 65°, the degree-day value is considered zero. An analogous definition applies for the degree-hour. Turner (1968) points out that in St. Louis the degree-day or degree-hour values explain nearly all the variance of the output of gas as well as of steam produced by public utilities.

THE USE OF GRIDS

In the development of a mathematical urban air pollution model, two different grids may be used: one based on existing pollution sources and the other on the location of the instruments that form the monitoring network.

The Pollution-Source Grid

In the United States, grid squares 1 mile on a side are frequently used, such as was done by Davidson, Koogler, and Turner. Fortak, of West Germany, used a square 100 × 100 m. The Connecticut model is based on a 5000-ft grid, and Clarke's Cincinnati model on sectors of a circle. Sources of pollution may be either point sources, such as the stacks of a public utility, or distributed sources, such as the sources representing the emission of many small homes in a residential area.

The Monitoring Grid

In testing the model, one resorts to measurements obtained by instruments at monitoring stations. Such monitoring stations may also be located on a grid. Furthermore, this grid may be used in the computation of concentrations by means of the mathematical equation—e.g., concentrations are calculated for the midpoints of the grid squares. The emission grid and monitoring grid may be identical or they may be different. For example, Turner used a source grid of 17 × 16 miles, but a measurement grid of 9 × 11 miles. In the Connecticut model, the source grid covers the entire state, and calculations based on the model also cover the entire state. Fortak used 480 × 800-m rectangles.

TYPES OF URBAN AIR POLLUTION MODELS

Source-Oriented Models

In applying the mathematical algorithm, one may proceed by determining the source strength for a given point source and then calculating the isopleths of concentration downwind arising from this source. The calculation is repeated for each area source and point source. Contributions made by each of the sources at a selected point downwind are then summed to determine the calculated value of the concentration. Isopleths of concentration may then be drawn to provide a computed distribution of the pollutants.

In the source-oriented model, detailed information is needed both on the strength and on the time variations of the source emissions. The Turner model (1964) is a good example of a source-oriented model.

It must be emphasized that each urban area must be "calibrated" to account for the peculiar characteristics of the terrain, buildings, forestation, and the like. Further, local phenomena such as lake or sea breezes and mountain-valley effects may markedly influence the resulting concentrations; for example, Knipping and Abdub (2003) included sea-salt aerosol in their model to predict urban ozone formation. Specifically, one would have to determine such relations as the variations of σ_y and σ_z with distance or the magnitude of the effective stack heights. A network of pollution-monitoring stations is necessary for this purpose. The use of an algorithm without such a calibration is likely to lead to disappointing results.

Receptor-Oriented Models

Several types of receptor-oriented models have been developed. Among these are: the Clarke model, the regression model, the Argonne tabulation prediction scheme, and the Martin model.

The Clarke Model

In the Clarke model (Clarke, 1964), one of the most well known, the receptor or monitoring station is located at the center of concentric circles having radii of 1, 4, 10, and 20 km respectively. These circles are divided into 16 equal sectors of 22 1/2°. A source inventory is obtained for each of the 64 (16 × 4) annular sectors. Also, for the 1-km-radius circle and for each of the annular rings, a chart is prepared relating x/Q (the concentration per unit source strength) and wind speed for various stability classes and for various mixing heights. In refining his model, Clarke (1967) considers separately the contributions to the concentration levels made by transportation, industry and commerce, space heating, and strong-point sources such as utility stacks. The following equations are then used to calculate the pollutant concentration.

$$\chi_T = \sum_{i=1}^{4} (\chi/Q)_{Ti} Q_{Ti}$$

$$\chi_I = \sum_{i=1}^{4} (\chi/Q)_{Ii} Q_{Ii}$$

$$\chi_S = \sum_{i=1}^{4} (\chi/Q)_{Si} Q_{Si}$$

$$\chi_{\text{Total}} = a\chi_T + b\chi_I + c\chi_S + \sum_{i=1}^{4} k_i \chi_p$$

where

χ: concentration (g/m^3)
Q: source strength (g/sec)
T: subscript to denote transportation sources
I: subscript to denote industrial and commercial sources
S: subscript to denote space-heating sources
p: subscript to denote point sources
i: refers to the annular sectors

The above equations with some modification are taken from Clarke's report (1967). Values of the constants *a, b,* and *c* can be determined from information concerning the diurnal variation of transportation, industrial and commercial, and space-heating sources. The coefficient k_i represents a calibration factor applied to the point sources.

The Linear Regression-Type Model

A second example of the receptor-oriented model is one developed by Roberts and Croke (Roberts et al., 1970) using regression techniques. Here,

$$\chi = C_0 + C_1Q_1 + C_2Q_2 + \sum_{i=1}^{n} k_iQ_i$$

In applying this equation, it is necessary first to stratify the data by wind direction, wind speed, and time of day. C_0 represents the background level of the pollutant; Q_1 represents one type of source, such as commercial and industrial emissions; and Q_2 may represent contributions due to large individual point sources. It is assumed that there are *n* point sources. The coefficients C_1 and C_2 and k_i represent the $1/\sigma_y\sigma_z$ term as well as the contribution of the exponential factor of the Gaussian-type diffusion equation (see Equation 1).

Multiple discriminant analysis techniques for individual monitoring stations may be used to determine the probability that pollutant concentrations fall within a given range or that they exceed a given critical value. Meteorological variables, such as temperature, wind speed, and stability, are used as the independent variable in the discriminant function.

The Martin Model

A diffusion model specifically suited to the estimation of long-term average values of air quality was developed by Martin (1971). The basic equation of the model is the Gaussian diffusion equation for a continuous point source. It is modified to allow for a multiplicity of point sources and a variety of meteorological conditions.

The model is receptor-oriented. The equations for the ground-level concentration within a given 22 1/2° sector at the receptor for a given set of meteorological conditions (i.e., wind speed and atmospheric stability) and a specified source are listed in his work. The assumption is made that all wind directions within a 22 1/2° sector corresponding to a 16-point compass occur with equal probability.

In order to estimate long-term air quality, the single-point-source equations cited above are evaluated to determine the contribution from a given source at the receptor for each possible combination of wind speed and atmospheric stability. Then, using Martin's notation, the long-term average is given by

$$\chi = \sum_{N}\sum_{L}\sum_{S} F(D_n, L, S)\chi(\rho_n, L, S)$$

where D_n indicates the wind-direction sector in which transport from a particular source (n) to the receptor occurs; ρ_n is the distance from a particular source to the receptor; $F(D_n, L, S)$ denotes the relative frequency of winds blowing into the given wind-direction sector (D_n) for a given wind-speed class (S) and atmospheric stability class (L); and N is the total number of sources. The joint frequency distribution $F(D_n, L, S)$ is determined by the use of hourly meteorological data.

A system of modified average mixing heights based on tabulated climatological values is developed for the model. In addition, adjustments are made in the values of some mixing heights to take into account the urban influence. Martin has also incorporated the exponential time decay of pollutant concentrations, since he compared his calculations with measured sulfur-dioxide concentrations for St. Louis, Missouri.

The Tabulation Prediction Scheme

This method, developed at the Argonne National Laboratory, consists of developing an ordered set of combinations of relevant meteorological variables and presenting the percentile distribution of SO$_2$ concentrations for each element in the set. In this table, the independent variables are wind direction, hour of day, wind speed, temperature, and stability. The 10, 50, 75, 90, 98, and 99 percentile values are presented as well as the minimum and the maximum values. Also presented are the interquartile range and the 75 to 95 percentile ranges to provide measures of dispersion and skewness, respectively. Since the meteorological variables are ordered, it is possible to look up any combination of meteorological variables just as one would look up a name in a telephone book or a word in a dictionary. This method, of course, can be applied only as long as the source distribution and terrain have not changed appreciably. For continued use of this method, one must be cognizant of changes in the sources as well as changes in the terrain due to new construction.

In preparing the tabulation, the data are first stratified by season and also by the presence or absence of precipitation. Further, appropriate group intervals must be selected for the meteorological variables to assure that within each grouping the pollution values are not sensitive to changes in that variable. For example, of the spatial distribution of the sources, one finds that the pollution concentration at a station varies markedly with changes in wind direction. If one plots percentile isopleths for concentration versus wind direction, one may choose sectors in which the SO_2 concentrations are relatively insensitive to direction change. With the exception of wind direction and hour of day, the meteorological variables of the table vary monotonically with SO_2 concentration. The tabulation prediction method has advantages over other receptor-oriented technique in that (1) it is easier to use, (2) it provides predictions of pollution concentrations more rapidly, (3) it provides the entire percentile distribution of pollutant concentration to allow a forecaster to fine-tune his prediction based on synoptic conditions, and (4) it takes into account nonlinearities in the relationships of the meteorological variables and SO_2 concentrations. In a sense, one may consider the tabulation as representing a nonlinear regression hypersurface passing through the data that represents points plotted in n-dimensional space. The analytic form of the hypersurface need not be determined in the use of this method.

The disadvantages of this method are that (1) at least 2 years of meteorological data are necessary, (2) changes in the emission sources degrade the method, and (3) the model could not predict the effect of adding, removing, or modifying important pollution sources; however, it can be designed to do so.

Where a network of stations is available such as exists in New York City, Los Angeles, or Chicago, then the receptor-oriented technique may be applied to each of the stations to obtain isopleths or concentration similar to that obtained in the source-oriented model. It would be ideal to have a source-oriented model that could be applied to any city, given the source inventory. Unfortunately, the nature of the terrain, general inaccuracies in source-strength information, and the influence of factors such as synoptic effect or the peculiar geometries of the buildings produce substantial errors. Similarly, a receptor-oriented model, such as the Clarke model or one based on regression techniques, must be tailored to the location. Every urban area must therefore be calibrated, whether one desires to apply a source-oriented model or a tabulation prediction scheme. The tabulation prediction scheme, however, does not require detailed information on the distribution and strength of emission sources.

Perhaps the optimum system would be one that would make use of the advantages of both the source-oriented model, with its prediction capability concerning the effects of changes in the sources, and the tabulation prediction scheme, which could provide the probability distributions of pollutant concentrations. It appears possible to develop a hybrid system by developing means for appropriately modifying the percentile entries when sources are modified, added, or removed. The techniques for constructing such a system would, of course, have general applicability.

The Fixed-Volume Trajectory Model

In the trajectory model, the path of a parcel of air is predicted as it is acted upon by the wind. The parcel is usually considered as a fixed-volume chemical reactor with pollutant inputs only from sources along its path; in addition, various mathematical constraints placed on mass transport into and out of the cell make the problem tractable. Examples of this technique are discussed by Worley (1971). In this model, derived pollution concentrations are known only along the path of the parcel considered. Consequently, its use is limited to the "strategy planning" problem. Also, initial concentrations at the origin of the trajectory and meteorological variables along it must be well known, since input errors along the path are not averageable but, in fact, are propagated.

The Basic Approach

Attempts have been made to solve the entire system of three-dimensional time-dependent continuity equations. The ever-increasing capability of computer systems to handle such complex problems easily has generally renewed interest in this approach. One very ambitious treatment is that of Lamb and Neiburger (1971), who have applied their model to carbon-monoxide concentrations in the Los Angeles basin. However, chemical reactions, although allowed for in their general formulation, are not considered because of the relative inertness of CO. Nevertheless, the validity of the diffusion and emission subroutines is still tested by this procedure.

The model of Friedlander and Seinfeld (1969) also considers the general equation of diffusion and chemical reaction. These authors extend the Lagrangian similarity hypothesis to reacting species and develop, as a result, a set of ordinary differential equations describing a variable-volume chemical reactor. By limiting their chemical system to a single irreversible bimolecular reaction of the form $A + B + C$, they obtain analytical solutions for the ground-level concentration of the product as a function of the mean position of the pollution cloud above ground level. These solutions are also functions of the appropriate meteorological variables, namely solar radiation, temperature, wind conditions, and atmospheric stability.

ADAPTATION OF THE BASIC EQUATION TO URBAN AIR POLLUTION MODELS

The basic equation, (1), is the continuous point-source equation with the source located at the ground. It is obvious that the sources of an urban complex are for the most part located above the ground. The basic equation must, therefore, be modified to represent the actual conditions. Various authors have proposed mathematical algorithms that include appropriate modifications of Equation (1). In addition, a source-oriented model developed by Roberts et al. (1970) to allow for time-varying sources of emission is discussed below; see the section "Time-Dependent Emissions (the Roberts Model)."

Chemical Kinetics: Removal or Transformation of Pollutants

In the chemical-kinetics portion of the model, many different approaches, ranging in order from the extremely simple to the very complex, have been tried. Obviously the simplest approach is to assume no chemical reactions are occurring at all. Although this assumption may seem contradictory to our intent and an oversimplification, it applies to any pollutant that has a long residence time in the atmosphere. For example, the reaction of carbon monoxide with other constituents of the urban atmosphere is so small that it can be considered inert over the time scale of the dispersion process, for which the model is valid (at most a few hours).

Considerable simplification of the general problem can be effected if chemical reactions are not included and all variables and parameters are assumed to be time-independent (steady-state solution). In this instance, a solution is obtained that forms the basis for most diffusion models: the use of the normal bivariate or Gaussian distribution for the downwind diffusion of effluents from a continuous point source. Its use allows steady-state concentrations to be calculated both at the ground and at any altitude. Many modifications to the basic equation to account for plume rise, elevated sources, area sources, inversion layers, and variations in chimney heights have been proposed and used. Further discussion of these topics is deferred to the following four sections.

The second level of pseudo-kinetic complexity assumes first-order or pseudo-first-order reactions are responsible for the removal of a particular pollutant; as a result, its concentration decays exponentially with time. In this case, a characteristic residence time or half-life describes the temporal behavior of the pollutant. Often, the removal of pollutants by chemical reaction is included in the Gaussian diffusion model by simply multiplying the appropriate diffusion equation by an exponential term of the form $\exp(-t/T)$, where T represents the half-life of the pollutant under consideration. Equations employing this procedure are developed below. The interaction of sulfur dioxide with other atmospheric constituents has been treated in this way by many investigators; for examples, see Roberts et al. (1970) and Martin (1971). Chemical reactions are not the only removal mechanism for pollutant. Some other processes contributing to their disappearance may be absorption by plants, soil-bacteria action, impact or adsorption on surfaces, and washout (for example, see Figure 2). To the extent that these processes are simulated by or can be fitted to an exponential decay, the above approximation proves useful and valid.

These three reactions appear in almost every chemical-kinetic model. On the other hand, many different sets of equations describing the subsequent reactions have been proposed. For example, Hecht and Seinfeld (1972) recently studied the propylene-NO-air system and list some 81 reactions that can occur. Any attempt to find an analytical solution for a model utilizing all these reactions and even a simple diffusion submodel will almost certainly fail. Consequently, the number of equations in the chemical-kinetic subroutine is often reduced by resorting to a "lumped parameter" stratagem. Here, three general types of chemical processes are identified: (1) a chain-initiating process involving the inorganic reactions shown above as well as subsequent interactions of product oxidants with source and product hydrocarbons, to yield (2) chain-propagating reactions in which free radicals are produced; these free radicals in turn react with the hydrocarbon mix to produce other free radicals and organic compounds to oxide NO to NO_2, and to participate in (3) chain-terminating reactions; here, nonreactive end products (for example, peroxyacetylnitrate) and aerosol production serve to terminate the chain. In the lumped-parameter representation, reaction-rate equations typical of these three categories (and usually selected from the rate-determining reactions of each category) are employed, with adjusted rate constants determined from appropriate smog-chamber data. An attempt is usually made to minimize the number of equations needed to fit well a large sample of smog-chamber data. See, for examples, the studies of Friedlander and Seinfeld (1969) and Hecht and Seinfeld (1972). Lumped parameter subroutines are primarily designed to simulate atmospheric conditions with a simplified chemical-kinetic scheme in order to reduce computing time when used with an atmospheric diffusion model.

Elevated Sources and Plume Rise

When hot gases leave a stack, the plume rises to a certain height dependent upon its exit velocity, temperature, wind speed at the stack height, and atmospheric stability. There are several equations used to determine the total or virtual height at which the model considers the pollutants to be emitted. The most commonly used is Holland's equation:

$$\Delta H = \frac{v_s}{u}\left[1.5 + \left(2.68 \times 10^{-2}(P)\left(\frac{T_s - T_a}{T_a}\right)d\right)\right]$$

where

ΔH: plume rise
v_s: stack velocity (m/sec)
d: stack diameter (m)
u: wind speed (m/sec)
P: pressure (kPa)
T_s: gas exit temperature (K)
T_a: air temperature (K)

The virtual or effective stack height is

$$H = h + \Delta H$$

where

H: effective stack height
h: physical stack height

With the origin of the coordinate system at the ground, but the source at a height *H,* Equation (2) becomes

$$\frac{\chi}{Q} = \frac{1}{\pi u \sigma_y \sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2} - \frac{(z-H)^2}{2\sigma_z^2} \right) \exp\left(-\frac{0.693t}{T_{1/2}} \right) \quad (3)$$

Mixing of Pollutants under an Inversion Lid

When the lapse rate in the lowermost layer, i.e., from the ground to about 200 m, is near adiabatic, but a pronounced inversion exists above this layer, the inversion is believed to act as a lid preventing the upward diffusion of pollutants. The pollutants below the lid are assumed to be uniformly mixed. By integrating Equation (3) with respect to z and distributing the pollutants uniformly over a height *H,* one obtains

$$\frac{\chi}{Q} = \frac{1}{\pi u \sigma_y H} \exp\left(-\frac{y^2}{2\sigma_y^2} \right) \exp\left(-\frac{0.693t}{T_{1/2}} \right)$$

Those few measurements of concentration with height that do exist do not support the assumption that the concentration is uniform in the lowermost layer. One is tempted to say that the mixing-layer thickness, *H,* may be determined by the height of the inversion; however, during transitional conditions, i.e., at dawn and dusk, the thickness of the layer containing high concentrations of pollutants may differ from that of the layer from the ground to the inversion base.

The thermal structure of the lower layer as well as pollutant concentration as a function of height may be determined by helicopter or balloon soundings.

The Area Source

When pollution arises from many small point sources such as small dwellings, one may consider the region as an area source. Preliminary work on the Chicago model indicates that contribution to observed SO_2 levels in the lowest tens of feet is substantially from dwellings and exceeds that emanating from tall stacks, such as power-generating stacks. For a rigorous treatment, one should consider the emission Q as the emission in units per unit area per second, and then integrate Q along x and along y for the length of the square. Downwind, beyond the area-source square, the plume may be treated as originating from a point source. This point source is considered to be at a virtual origin upwind of the area-source square. As pointed out by Turner, the approximate equation for an area source can be calculated as

$$\frac{\chi}{Q} = \frac{\exp\left(-\frac{y^2}{2\left[\sigma_y^2\left(x_{y0}+x\right)\right]^2} - \frac{(z-h)^2}{2\sigma_z^2} \right) \exp\left(-\frac{0.693t}{T_{1/2}} \right)}{\pi u \left[\sigma_y\left(x_{y0}+x\right)\right]\sigma_z}$$

where $\sigma_y(x_{y0} + x)$ represents the standard deviation of the horizontal crosswind concentration as a function of the distance $x_{y0} + x$ from the virtual origin. Since the plume is considered to extend to the point where the concentration falls to 0.1 that of the centerline concentration, $\sigma_y(x_{y0}) = S/403$ where $\sigma_y(x_{y0})$ is the standard deviation of the concentration at the downwind side of the square of side length S. The distance x_{y0} from the virtual origin to the downwind side of the grid square may be determined, and is that distance for which $\sigma_y(x_{y0}) = S/403$. The distance x is measured from the downwind side of the grid square. Other symbols have been previously defined.

Correction for Variation in Chimney Heights for Area Sources

In any given area, chimneys are likely to vary in height above ground, and the plume rises vary as well. The variation of effective stack height may be taken into account in a manner similar to the handling of the area source. To illustrate, visualize the points representing the effective stack height projected onto a plane perpendicular to the ground and parallel both to two opposite sides of the given grid square and to the horizontal component of the wind vector. The distribution of the points on this projection plane would be similar to the distribution of the sources on a horizontal plane.

Based on Turner's discussion (1967), the equation for an area source and for a source having a Gaussian distribution of effective chimney heights may be written as

$$\frac{\chi}{Q} = \frac{\exp\left(-\frac{y^2}{2\left[\sigma_y\left(x_{y0}+x\right)\right]^2} - \frac{(z-h)^2}{2\left[\sigma_z\left(x_{z0}+x\right)\right]^2} \right) \exp\left(-\frac{0.693t}{T_{1/2}} \right)}{\pi u \left[\sigma_y\left(x_{y0}+x\right)\right]\left[\sigma_z\left(x_{z0}+x\right)\right]}$$

where $\sigma_z(x_{z0} + x)$ represents the standard deviation of the vertical crosswind concentration as a function of the distance $x_{z0} + x$ from the virtual origin. The value of $\sigma_z(x_{z0})$ is arbitrarily chosen after examining the distribution of effective chimney heights, and the distance x_{z0} represents the distance from the virtual origin to the downwind side of the grid square. The value x_{z0} may be determined and represents the distance corresponding to the value for $\sigma_z(x_{z0})$. The value of x_{y0} usually differs from that of x_{z0}. The other symbols retain their previous definition.

In determining the values of $\sigma_y(x_{y0} + x)$ and $\sigma_z(x_{z0} + x)$, one must know the distance from the source to the point in question or the receptor. If the wind direction changes within the averaging interval, or if there is a change of wind direction due to local terrain effects, the trajectories are curved. There are several ways of handling curved trajectories. In the Connecticut model, for example, analytic forms for the trajectories were developed. The selection of appropriate trajectory or streamline equations (steady state was assumed) was based on the wind and stability conditions. In the St. Louis model, Turner developed a computer program using the available winds to provide pollutant trajectories. Distances obtained from the trajectories are then used in the Pasquill diagrams or equations to determine the values of $\sigma_y(x_{y0} + x)$ and $\sigma_z(x_{z0} + x)$.

Time-Dependent Emissions (The Roberts Model)

The integrated puff transport algorithm of Roberts et al. (1970), a source-oriented model, uses a three-dimensional Gaussian puff kernel as a basis. It is designed to simulate the time-dependent or transient emissions from a single source. Concentrations are calculated by assuming that dispersion occurs from Gaussian diffusion of a puff whose centroid moves with the mean wind. Time-varying source emissions as well as variable wind speeds and directions are approximated by a time series of piecewise continuous emission and meteorological parameters. In addition, chemical reactions are modeled by the inclusion of a removal process described by an exponential decay with time.

The usual approximation for inversion lids of constant height, namely uniform mixing arising from the superposition of an infinite number of multiple source reflections, is made. Additionally, treatments for lids that are steadily rising or steadily falling and the fumigation phenomenon are incorporated.

The output consists of calculated concentrations for a given source for each hour of a 24-hour period. The concentrations can be obtained for a given receptor or for a uniform horizontal or vertical grid up to 1000 points.

The preceding model also forms the basis for two other models, one whose specific aim is the design of optimal control strategies, and a second that repetitively applies the single-source algorithm to each point and area source in the model region.

METEOROLOGICAL MEASUREMENTS

Wind speed and direction data measured by weather bureaus are used by most investigators, even though some have a number of stations and towers of their own. Pollutants are measured for periods of 1 hour, 2 hours, 12 hours, or 24 hours. 12- and 24-hour samples of pollutants such as SO_2 leave much to be desired, since many features of their variations with time are obscured. Furthermore, one often has difficulty in determining a representative wind direction or even a representative wind speed for such a long period.

The total amount of data available varies considerably in the reviewed studies. Frenkiel's study (1956) was based on data for 1 month only. A comparatively large amount of data was gathered by Davidson (1967), but even these in truth represent a small sample. One of the most extensive studies is the one carried out by the Argonne National Laboratory and the city of Chicago in which 15-minute readings of SO_2 for 8 stations and wind speed and direction for at least 13 stations are available for a 3-year period.

In the application of the mathematical equations, one is required to make numerous arbitrary decisions: for example, one must choose the way to handle the vertical variation of wind with height when a high stack, about 500 ft, is used as a point source; or how to test changes in wind direction or stability when a change occurs halfway through the 1-hour or 2-hour measuring period. In the case of an elevated point source, Turner in his St. Louis model treated the plume as one originating from the point source up to the time of a change in wind direction and as a combination of an instantaneous line puff and a continuous point source thereafter. The occurrence of precipitation presents serious problems, since adequate diffusion measurements under these conditions are lacking. Furthermore, the chemical and physical effects of precipitation on pollutants are only poorly understood. In carrying forward a pollutant from a source, one must decide on how long to apply the calculations. For example, if a 2-mph wind is present over the measuring grid and a source is 10 miles away, one must take account of the transport for a total of 5 hours.

Determining a representative wind speed and wind direction over an urban complex with its variety of buildings and other obstructions to the flow is frequently difficult, since the horizontal wind field is quite heterogeneous. This is so for light winds, especially during daytime when convective processes are taking place. With light-wind conditions, the wind direction may differ by 180° within a distance of 1 mile. Numerous land stations are necessary to depict the true wind field. With high winds, those on the order of 20 mph, the wind direction is quite uniform over a large area, so that fewer stations are necessary.

METHODS FOR EVALUATING URBAN AIR POLLUTION MODELS

To determine the effectiveness of a mathematical model, validation tests must be applied. These usually include a comparison of observed and calculated values. Validation tests are necessary not only for updating the model because of changes in the source configuration or modification in terrain characteristics due to new construction, but also for comparing the effectiveness of the model with any other that may be suggested. Of course, the primary objective is to see how good the model really is, both for incident control as well as for long-range planning.

Scatter Plots and Correlation Measures

Of the validation techniques appearing in the literature, the most common involves the preparation of a scatter diagram relating observed and calculated values (Y_{obs} vs. Y_{calc}). The degree of scatter about the $Y_{obs} = Y_{calc}$ line provides a measure of the effectiveness of the model. At times, one finds that a majority of the points lies either above the line or below the line, indicating systematic errors.

It is useful to determine whether the model is equally effective at all concentration levels. To test this, the calculated scale may be divided into uniform bandwidths and the mean square of the deviations abou t the $Y_{\text{obs}} = Y_{\text{calc}}$ line calculated for each bandwidth. Another test for systematic error as a function of bandwidth consists of an examination of the mean of the difference between calculated and observed values for $Y_{\text{calc}} < Y_{\text{obs}}$ and similarly for $Y_{\text{calc}} > Y_{\text{obs}}$.

The square of the linear correlation coefficient between calculated and observed values or the square of the correlation ratio for nonlinear relationships represent measures of the effectiveness of the mathematical equation. For a linear relationship between the dependent variable, e.g., pollutant concentration, and the independent variables,

$$R^2 = 1 - \frac{S_y^2}{\sigma_y^2} = 1 - \frac{\text{unexplained variance}}{\text{total variance}} = \frac{\sigma_y^2 - S_y^2}{\sigma_y^2}$$

$$= \frac{\text{explained variance}}{\text{total variance}}$$

where

R^2: square of the correlation coefficient between observed and calculated values

S_y^2: average of the square of the deviations about the regression line, plane, or hyperplane

σ_y^2: variance of the observed values

Statistical Analysis

Several statistical parameters can be calculated to evaluate the performance of a model. Among those commonly used for air pollution models are Kukkonen, Partanen, Karppinen, Walden, et al. (2003); Lanzani and Tamponi (1995):

The index of agreement

$$\text{IA} = 1 - \frac{\overline{(C_p - C_o)^2}}{[|C_p - \overline{C_o}| + |C_o - \overline{C_o}|]^2}$$

R

$$R = \overline{\left[\frac{(C_o - \overline{C_o})(C_p - \bar{C}_p)}{\sigma_o \sigma_p}\right]}$$

The bias

$$\text{Bias} = \frac{\bar{C}_p - \overline{C_o}}{\overline{C_o}}$$

The fractional bias

$$\text{FB} = \frac{\bar{C}_p - \overline{C_o}}{0.5(\bar{C}_p + \overline{C_o})}$$

The normalized mean of the square of the error

$$\text{NMSE} = \frac{\overline{(C_p - \overline{C_o})^2}}{\bar{C}_p \overline{C_o}}$$

where

C_p: predicted concentrations

C_o: predicted observed concentrations

σ_o: standard deviation of the observations

σ_p: standard deviation of the predictions

The overbar concentrations refer to the average overall values.

The parameters IA and R^2 are measures of the correlation of two time series of values, the bias is a measurement of the overall tendency of the model, the FB is a measure of the agreement of the mean values, and the NMSE is a normalized estimation of the deviation in absolute value.

The IA varies from 0.0 to 1.0 (perfect agreement between the observed and predicted values). A value of 0 for the bias, FB, or NMSE indicates perfect agreement between the model and the data.

Thus there are a number of ways of presenting the results of a comparison between observed and calculated values and of calculating measures of merit. In the last analysis the effectiveness of the model must be judged by how well it works to provide the needed information, whether it will be used for day-to-day control, incident alerts, or long-range planning.

RECENT RESEARCH IN URBAN AIR POLLUTION MODELING

With advances in computer technology and the advent of new mathematical tools for system modeling, the field of urban air pollution modeling is undergoing an ever-increasing level of complexity and accuracy. The main focus of recent research is on particles, ozone, hydrocarbons, and other substances rather than the classic sulfur and nitrogen compounds. This is due to the advances in technology for pollution reduction at the source. A lot of attention is being devoted to air pollution models for the purpose of urban planning and regulatory-standards implementation. Simply, a model can tell if a certain highway should be constructed without increasing pollution levels beyond the regulatory maxima or if a new regulatory value can be feasibly obtained in the time frame allowed. Figure 2 shows an example of the distribution of particulate matter (PM_{10}) in a city. As can be inferred, the presence of particulate matter of this size is obviously a traffic-related pollutant.

Also, some modern air pollution models include meteorological forecasting to overcome one of the main obstacles that simpler models have: the assumption of average wind speeds, direction, and temperatures.

At street level, the main characteristic of the flow is the creation of a vortex that increases concentration of pollutants on the canyon side opposite to the wind direction, as

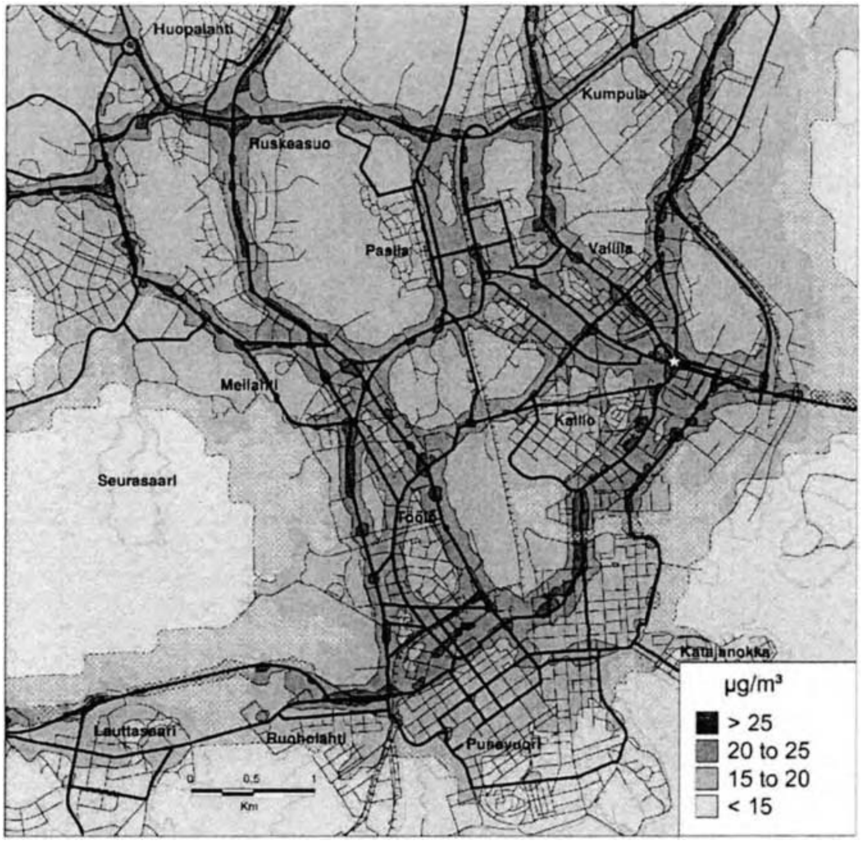


FIGURE 2 Predicted spatial distribution of the yearly means of PM_{10} in central Helsinki in 1998 (mg/m^3). The white star indicates the predicted maximum concentration in the area (40.4 mg/m^3) (From Kukkonen et al., 2001, with permission from Elsevier).

shown in Figure 3 (Berkowicz, 2000a). The Danish operational street pollution model (OSPM) has been used by several researchers to model dispersion of pollutants at street level. Several studies have assessed the validity of the model by using data for different cities in Europe.

New techniques such as Fuzzy Logic and Neural Networks have been used with great results (Kukkonen, Partanen, Karppinen, Ruuskanen, et al., 2003; Viotti et al., 2002, Pokrovskya et al., 2002; Pelliccioni and Poli, 2002). Schlink and Volta (2000) used grey box stochastic models and extended autoregressive moving average models to predict ozone formation. Other approaches have been time-series analysis, regression analysis, and statistical modeling, among others.

For additional information, the reader is referred to the references at the end of the chapter.

CONCLUDING REMARKS

Inadequacies and shortcomings exist in our assessment of each of the components of the mathematical urban air pollution model. In this section these difficulties are discussed.

The Source Inventory

For large metropolitan areas, one finds that the inventory obtained by the usual methods, such as questionnaires, is often out of date upon its completion. Continuous updating is necessary. However, in a receptor-oriented model, the requirement for a detailed source inventory is relaxed. Further, by developing a receptor-oriented "anomaly" model, one may further reduce the error resulting from inadequate source information. In the anomaly-type model, changes in the dependent variable over a given time period are calculated. This interval may be 1, 2, 4, or 6 hours.

Initial Mixing

According to Schroeder and Lane (1988), "Initial mixing refers to the physical processes that act on pollutants immediately after their release from an emission source. The nature and extent of the initial interaction between pollutants and the ambient air depend on the actual configuration of the source in terms of its area, its height above the surrounding terrain, and the initial buoyancy conditions."

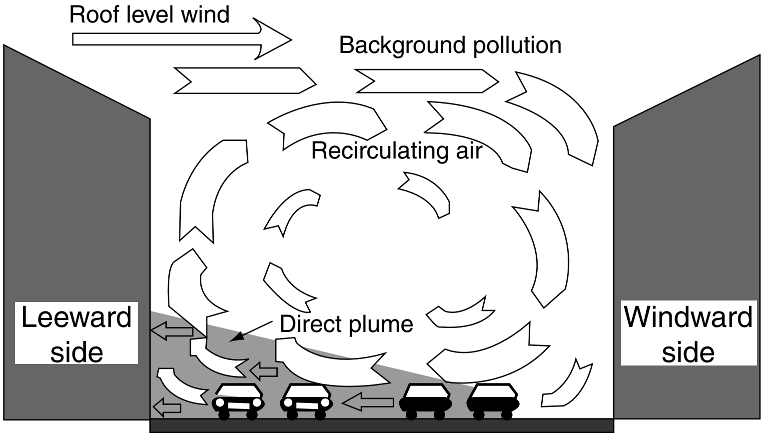


FIGURE 3 Schematic illustration of flow and dispersion of pollutants in street canyons (From Berkowicz, 2000a, with permission of Springer Science and Business media).

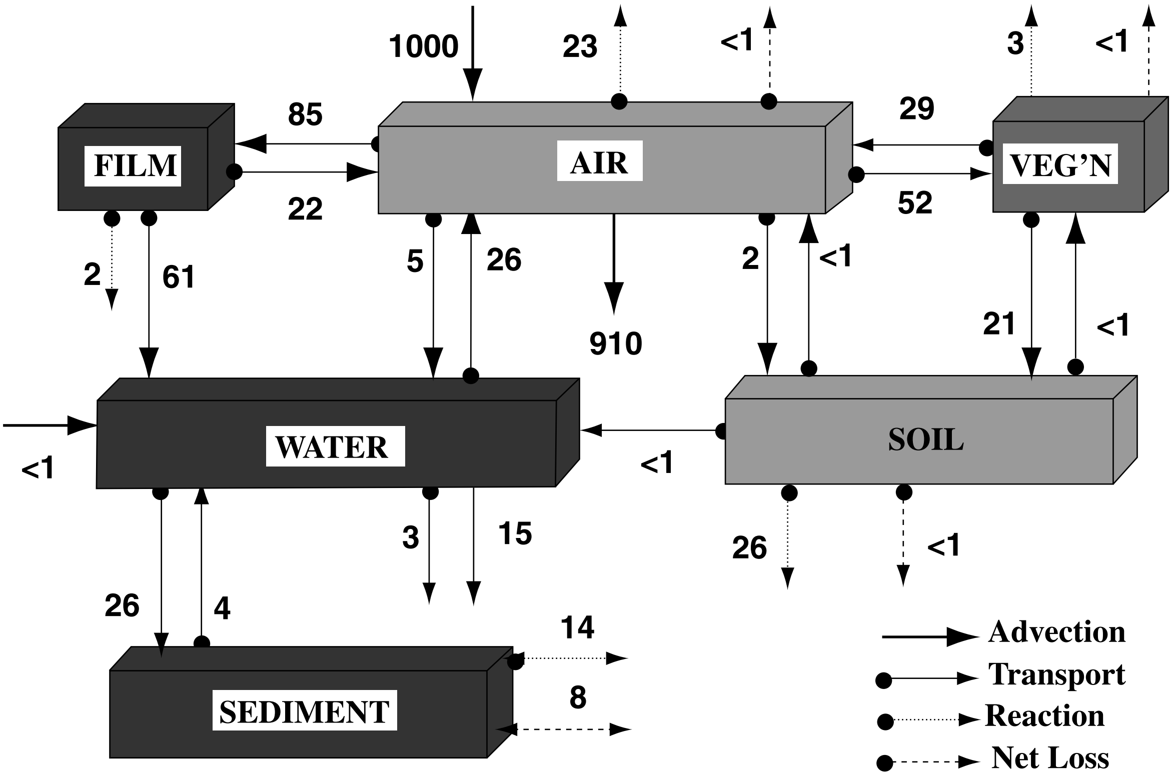


FIGURE 4 Estimated rates of chemical movement and transformation for 2,3,7,8-CL$_4$DD based on an emission of 1 mil/hour into air. Numbers shown are transport rate in mmol/hour (From Diamond et al., 2001, with permission from Elsevier).

The mixing layer is the lower region of the troposphere in which pollutants are relatively free to circulate and disperse vertically as well as horizontally because of the preponderance of small-scale turbulence. It may extend to a height as small as 50 m or as great as 5 km above the surface (Deardorff, 1975). More typically, it extends about 1 to 2 km during the day and a few hundred meters at night, although day-to-day variations can be quite large (Smith, 1982). This turbulence promotes intimate contact between vapor-phase and aerosol-associated pollutants.

Such direct contact is an important step in the chain of events that ultimately results in chemical transformations of pollutants near their source before extensive dilution has occurred and while their air concentrations are still relatively high. The consequences of limited vertical mixing may be exacerbated at northern latitudes, where air pollutants released close to the ground may disperse only to a very limited extent because of the extreme stability of air brought about by inversion layers characteristic of the Arctic, especially in winter. This situation can give rise to elevated ambient-air concentration of noxious contaminants in those regions.

Pollutant Measurements

It is necessary to remember that the distribution of any contaminants is a function of space and time. With a heterogeneous distribution of sources in space and the pronounced variation in source strength with time, one can hardly expect a few stations to describe adequately the distribution of the contaminant sources. Also, the averaging time is important. A 1-hour sample or even a 2-hour sample will bring out the diurnal effects quite well; however, 24-hour samples do not.

Photochemical Transformations

Schroeder and Lane (1988) have also discussed the reactions occurring in the atmosphere thus: "During transport and diffusion through the atmosphere, all but the most inert toxic pollutants are likely to participate in complex chemical or photochemical reactions. These processes can transform a pollutant from its primary state (the physical and chemical form in which it first enters the atmosphere) to another state that may have similar or very different characteristics."

Transformation products can differ from their precursors in chemical stabilities, toxic properties, and various other characteristics. For example, pyrene, a nontoxic, noncarcinogenic organic molecule, can react with NO_x and nitric acid in the air to form various nitropyrenes, which are highly potent, direct-acting mutagens. Secondary pollutants may be removed from the atmosphere in a manner different from that of their parent substances as a result of characteristic chemical and photochemical degradation or physical-removal mechanisms. It is difficult to formulate general statements regarding atmospheric transformations of Toxic Air Pollutants because the contributing chemical processes are numerous and complex.

The Earth's atmosphere is an efficient oxidizing medium even though most of its mass is composed of either relatively inert molecules or chemically reducing gases such as N_2, H_2, and CH_4. Nevertheless, the atmosphere acts as an oxidative system because of its overall composition and the relative chemical reactivity of natural atmospheric constituents or contaminants. Some of the more chemically reactive species known to be present in ambient air are atomic oxygen, ozone, hydroxyl and other free radicals (HO_2, CH_3O_2), peroxides (H_2O_2, CH_3O_2H), nitrogen oxides, sulfur oxides, and a wide variety of acidic and basic species. Consequently, contaminants of environmental interest, once emitted into ambient air, are converted at various rates into substances characterized by higher chemical oxidation states than their parent substances.

FUTURE CONSIDERATIONS

If mathematical modeling is to be effective, it is essential that information be available on the space-time distribution of both the pollutant and the necessary meteorological variables. By considering concentration changes with time and using the receptor-oriented approach, one may minimize the influence of the source inventory, but in the last analysis, source-inventory information is necessary for any model. In the validation procedure, one must consider the occurrence of systematic errors, i.e., readings consistently too high or too low compared to the calculated values. Similarly, large deviations between calculated and observed values should be carefully investigated. The selection of an appropriate sampling time is very important.

In conclusion, although the efforts made to date on the problem are commendable, the results can still be improved. Continued effort in the development of an urban air pollution model is necessary, and hopefully will provide the needed tools for handling urban air pollution problems.

REFERENCES

Addison, P.S., Currie, J.I., Low, D.J., and McCann, J.M. (2000), An integrated approach to street canyon pollution modelling, *Environmental Monitoring and Assessment,* 65, pp. 333–342.

Affum, J.K., Brown, A.L., and Chan, Y.C. (2003), Integrating air pollution modeling with scenario testing in road transport planning: the TRAEMS approach, *Science of the Total Environment,* 312, pp. 1–14.

Aloyan, A.E., Arutyunyan, V., Haymet, A.D., He, J.W., Kuznetsov, Y., and Lubertino, G. (2003), Air quality modeling for Houston-Galveston-Brazoria area, *Environment International,* 29, pp. 377–383.

Altshuller, A.P. (1989), Nonmethane organic compound to nitrogen oxide ratios and organic composition in cities and rural areas, *Journal of the Air and Waste Management Association,* 39, pp. 936–943.

Barna, M.G. and Gimson, N.R. (2002), Dispersion modeling of a wintertime particulate pollution episode in Christchurch, New Zealand, *Atmospheric Environment,* 36, pp. 3531–3544.

Bauer, G., Deistler, M., and Scherrer, W. (2001), Time series models for short term forecasting of ozone in the eastern part of Austria, *Environmetrics,* 12, pp. 117–130.

Beiruti, A.A. and Al-Omishy, H.K. (1985), Traffic atmospheric diffusion model, *Atmospheric Environment,* 19, No. 9, pp. 1519–1524.

Berge, E., Walker, S., Sorteberg, A., Lenkopane, M., Eastwood, S., Jablonska, H.I., and Køltzow, M.Ø. (2002), A real-time operational forecast model for meteorology and air quality during peak air pollution episodes in Oslo, Norway, *Water, Air, and Soil Pollution: Focus,* 2, pp. 745–757.

Berkowicz, R. (2000a), OSPM—a parameterised street pollution model, *Environmental Monitoring and Assessment,* 65, pp. 323–331.

Berkowicz, R. (2000b), A simple model for urban background pollution, *Environmental Monitoring and Assessment,* 65, pp. 259–267.

Borrego, C., Tchepel, O., Costa, A.M., Amorim, J.H., and Miranda, A.I. (2003), Emission and dispersion modeling of Lisbon air quality at local scale, *Atmospheric Environment,* 37, pp. 5197–5205.

Borrego, C., Tchepel, O., Monteiro, A., Barros, N., and Miranda, A. (2002), Influence of traffic emissions estimation variability on urban air quality modelling, *Water, Air, and Soil Pollution: Focus,* 2, pp. 487–499.

Bouhamra, W.S., and Elkilani, A.S. (1994), Prediction of long term NH_3 concentration using the superstack model, *Journal of Hazardous Materials,* 36, pp. 275–292.

Bouman, D.J. and Schmidt, F.H. (1961), On the growth of ground concentrations of atmospheric pollution in cities during stable conditions, *Beitrage zur Physik der Atmosphere,* 33, No. 3/4.

Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., and Smallbone, K., (2000), A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments, *Science of the Total Environment,* 253, pp. 151–167.

Buttazzoni, C., Lavagnini, I., Marani, A., Grandi, F.Z., and Del Turco, A. (1986), Probability model for atmospheric sulphur dioxide concentrations in the area of Venice, *Journal of the Air Pollution Control Association,* 36, No. 9, pp. 1026–1030.

Canno, D., Fukui, S., Ikada, H., and Ono, Y. (1959), Atmospheric SO_2 concentrations observed in Keihin Industrial Center and their relation to meteorological elements, *International Journal of Air and Water Pollution,* 1, No. 3, pp. 234–240.

Carrington, D.B. and Pepper, D.W. (2002), Predicting winds and air quality within the Las Vegas Valley, *Communications in Numerical Methods in Engineering,* 18, pp. 195–201.

Chatterton, T., Dorling, S., Lovett, A., and Stephenson, M. (2000), Air quality in Norwich, UK: multi-scale modelling to assess the significance of city, county and regional pollution sources, *Environmental Monitoring and Assessment,* 65, pp. 425–433.

Chelani, A.B., Gajghate, D.G., Tamhane, S.M., and Hasan, M.Z. (2001), Statistical modeling of ambient air pollutants in Delhi, *Water, Air, and Soil Pollution,* 132, pp. 315–331.

Chen, K.S., Ho, Y.T., Lai, C.H., and Chou, Y.M. (2003), Photochemical modeling and analysis of meteorological parameters during ozone episodes in Kaohsiung, Taiwan, *Atmospheric Environment,* 37, pp. 1811–1823.

Clarke, J.F. (1964), A simple diffusion model for calculating point concentrations from multiple sources, *Journal of the Air Pollution Control Association,* 14, No. 9, pp. 347–353.

Clarke, J.F. (1967), Forecasting Pollutant Concentrations, U.S. Weather Bureau Forecasters Training Course, Silver Spring, Maryland.

Cogliani, E. (2001), Air pollution forecast in cities by an air pollution index highly correlated with meteorological variables, *Atmospheric Environment,* 35, pp. 2871–2877.

Colvile, R.N., Gomez-Perales, J.E., and Nieuwenhuijsen, M.J. (2003), Use of dispersion modeling to assess road-user exposure to PM2.5 and its source apportionment, *Atmospheric Environment,* 37, pp. 2773–2782.

Colvile, R.N., Woodfield, N.K., Carruthers, D.J., Fisher, B.E.A., Rickard, A., Neville, S., and Hughes, A. (2002), Uncertainty in dispersion modeling and urban air quality mapping, *Environmental Science and Policy,* 5, pp. 207–220.

Commins, B.T. and Waller, R.E. (1967), Observations from a ten-year study of pollution at a site in the city of London, *Atmospheric Environment,* 1, pp. 49–68.

Cox, W.M. and Tikvart, J.A. (1990), A statistical procedure for determining the best performing air quality simulation model, *Atmospheric Environment,* 24A, No. 9, pp. 2387–2395.

Croke, E.J., Carson, J.E., Gatz, D.F., Moses, H., Clark, F.L., Kennedy, A.S., Gregory, J.A., Roberts, J.J., Carter, R.P., and Turner, D.B. (1968), Chicago Air Pollution System Model, Second Quarterly Progress Report, Argonne National Laboratory, ANUES-CC-002, May.

Croke, E.J., Carson, J.E., Gatz, D.F., Moses, H., Kennedy, A.S., Gregory, J.A., Roberts, J.J., Croke, K., Anderson, J., Parsons, D., Ash, J., Norco, J., and Carter, R.P. (1968), Chicago Air Pollution System Model, Third Quarterly Progress Report, Argonne National Laboratory, ANUES-CC-003, October.

Davidson, B. (1967), A summary of the New York urban air pollution dynamics research program, *Journal of the Air Pollution Control Association,* 17, No. 3, pp. 154–158.

Deardorff, J.W. (1975), Background Papers for a Workshop on the Tropospheric Transport of Pollutants to the Ocean, National Academy of Sciences, Washington, DC, pp. 219–22.

Diamond, M.L., Priemer, D.A., and Law, N.L. (2001), Developing a multimedia model of chemical dynamics in an urban area, *Chemosphere,* 44, pp. 1655–1667

Eerens, H.C., Sliggers, C.J., and van den Hout, K.D. (1993), The car model: the Dutch method to determine city street air quality, *Atmospheric Environment,* 27B, No. 4, pp. 389–399.

Elbir, T. (2003), Comparison of model predictions with the data of an urban air quality monitoring network in Izmir, Turkey, *Atmospheric Environment,* 37, pp. 2149–2157.

Fenger, J. (1999), Urban air quality, *Atmospheric Environment,* 33, pp. 4877–4900.

Fortak, H. (1966), Rechnerische Ermittlung der SO_2-Grundbelastung aus Emissiondaten, Anwendung auf die Verhaltnisse des Stadtgebietes von Bremen mit Abbildungsteil, Institut fur theoretische Meterologie der freien Universitat Berlin.

Frenkiel, F.N. (1956), Atmospheric Pollution in Growing Communities, Smithsonian Report for 1956, pp. 269–299.

Friedlander, S.K. and Seinfeld, J.H. (1969), A dynamic model of photochemical smog, *Environmental Science and Technology,* 3, pp. 1175–1181.

Frohn, L.M., Christensen, J.H., and Brandt, J. (2002), Development of a high-resolution nested air pollution model: the numerical approach, *Journal of Computational Physics,* 179, pp. 68–94.

Gryning, S. and Lyck, E. (1984), Atmospheric dispersion from elevated sources in an urban area: comparison between tracer experiments and model calculations, *Journal of Climate and Applied Meteorology,* 23, No. 4, pp. 651–660.

Hanssen-Bauer, I., (1985), A simple model for diffusion of SO_2 in Bergen, *Atmospheric Environment,* 19, No. 3, pp. 415–422.

Hecht, T.A. and Seinfeld, J.H. (1972). Development and validation of a generalized mechanism for photochemical smog, *Environmental Science and Technology,* 6, pp. 45–57.

Hess, G.D. (1989), Simulation of photochemical smog in the Melbourne airshed: worst case studies, *Atmospheric Environment,* 23, No. 3, pp. 661–669.

Hilst, G.R.F., Badgley, F.I., Yocum, J.B., and Browne, N.E. (1967), The Development of a Simulation Model for Air Pollution over Connecticut, a Final Report to the Connecticut Research Commission, I and II, Travelers Research Center, Hartford, CT, October.

Hilst, G.R. and Bowne, N.E. (1966), A Study of the Diffusion of Aerosols Released from Aerial Line Sources Upwind of an Urban Complex, a Final Report to the US Army Dugway Proving Ground, I, Travelers Research Center, Hartford, CT, July.

Hodabonsi, R.F. and Brodkey, R. (1984), Evaluation of RAM Model for Cleveland, Ohio, part II, *Journal of the Air Pollution Control Association,* 31, No. 4, pp. 371–374.

Hodabonsi, R.F. and Peters, L.K. (1981), Evaluation of RAM Model for Cleveland, Ohio, *Journal of the Air Pollution Control Association,* 31, No. 3, pp. 253–255.

Hopke, P.K. (2003), Recent developments in receptor modeling, *Journal of Chemometrics,* 17, pp. 255–265.

Hov, Ø. and Larssen, S. (1984), Street canyon concentrations of nitrogen dioxide in Oslo: measurements and model calculations, *Environmental Science and technology,* 18, No. 2, pp. 82–87.

Hurley, P., Manins, P., Lee, S., Boyle, R., Ng, Y.L., and Dewundege, P. (2003), Year-long, high-resolution, urban airshed modeling: verification of TAPM predictions of smog and particles in Melbourne, Australia, *Atmospheric Environment,* 37, pp 1899–1910.

Jicha, M., Pospisil, J., and Katolicky, J. (2000), Dispersion of pollutants in street canyon under traffic induced flow and turbulence, *Environmental Monitoring and Assessment,* 65, pp. 343–351.

Jones, S.G., Fisher, B.E.A., Gonzalez-Flesca, N., and Sokhi, R. (2000), The use of measurement programmes and models to assess concentrations next to major roads in urban areas, *Environmental Monitoring and Assessment,* 64, pp. 531–547.

Jorquera, H. (2002a), Air quality at Santiago, Chile: a box modeling approach—I: carbon monoxide, nitrogen oxides and sulfur dioxide, *Atmospheric Environment,* 36, pp. 315–330.

Jorquera, H. (2002b), Air quality at Santiago, Chile: a box modeling approach—II: $PM_{2.5}$, coarse and PM_{10} particulate matter fractions, *Atmospheric Environment,* 36, pp. 331–344.

Karppinen, A., Kukkonen, J., Elolahde, T., Konttinen, M., and Koskentalo, T. (2000), A modeling system for predicting urban air pollution: comparison of model predictions with the data of an urban measurement network in Helsinki, *Atmospheric Environment,* 34, pp. 3735–3743.

Karppinen, A., Kukkonen, J., Elolahde, T., Konttinen, M., Koskentalo, T., and Rantakrans, E. (2000), A modeling system for predicting urban air pollution: model description and applications in the Helsinki metropolitan area, *Atmospheric Environment,* 34, pp. 3723–3733.

Kastner-Klein, P., Fedorovich, E., Sini, J., and Mestayer, P.G. (2000), Experimental and numerical verification of similarity concept for dispersion of car exhaust gases in urban street canyons, *Environmental Monitoring and Assessment,* 65, pp. 353–361.

Kauper, E.J., Hartman, D.F., and Hopper, C.J. (1961), Smog forecasting in the Los Angeles Basin, Air Quality Report No. 37, Air Pollution Control District, County of Los Angeles.

Ketzel, M., Berkowicz, R., and Lohmeyer, A. (2000), Comparison of numerical street dispersion models with results from wind tunnel and field measurements, *Environmental Monitoring and Assessment,* 65, pp. 363–370.

Ketzel, M., Louka, P., Sahm, P., Guilloteau, E., Sini, J.F., and Moussiopoulos, N. (2002), Intercomparison of numerical urban dispersion models—part II: street canyon in Hannover, Germany, *Water, Air, and Soil Pollution: Focus,* 2, pp. 603–613.

Kitwiroon, N., Sokhi, R.S., Luhana, L., and Teeuw, R.M. (2002), Improvements in air quality modelling by using surface boundary layer parameters derived from satellite land cover data, *Water, Air, and Soil Pollution: Focus,* 2, pp. 29–41.

Knipping, E.M. and Abdub, D.D. (2003), Impact of chlorine emissions from sea-salt aerosol on coastal urban ozone, *Environmental Science and Technology,* 37, pp. 275–284.

Kono, H. and Ito, S. (1990), A comparison of concentration estimates by the OMG volume-source dispersion model with three line source dispersion models, *Atmospheric Environment,* 24B, No. 2, pp. 253–260.

Koogler, J.B., Sholtes, R.S., Danis, A.L., and Harding, C.I. (1967), A multivariate model for atmospheric, dispersion prediction, *Journal of the Air Pollution Control Association,* 17, No. 4, pp. 711–714.

Kousa, A., Kukkonen, J., Karppinen, A., Aarnio, P., Koskentalo, T. (2001), Statistical and diagnostic evaluation of a new-generation urban dispersion modeling system against an extensive dataset in the Helsinki area, *Atmospheric Environment,* 35, pp. 4617–4628.

Kukkonen, J., Harkonen, J., Karppinen, A., Pohjola, M., Pietarila, H., and Koskentalo, T. (2001), A semi-empirical model for urban PM10 concentrations, and its evaluation against data from an urban measurement network, *Atmospheric Environment,* 35, pp. 4433–4442.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., Chatterton, T., Foxall, R., and Cawley, G. (2003), Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modeling system and measurements in central Helsinki, *Atmospheric Environment,* 37, pp. 4539–4550.

Kukkonen, J., Partanen, L., Karppinen, A., Walden, J., Kartastenpaa, R., Aarnio, P., Koskentalo, T., and Berkowicz, R. (2003), Evaluation of the OSPM model combined with an urban background model against the data measured in 1997 in Runeberg Street, Helsinki, *Atmospheric Environment,* 37, pp. 1101–1112.

Kukkonen, J., Valkonen, E., Walden, J., Koskentalo, T., Karppinen, A., Berkowicz, R., and Kartastenpaa, R. (2000), Measurements and modelling of air pollution in a street canyon in Helsinki, *Environmental Monitoring and Assessment,* 65, pp. 371–379.

Kussmann, A.S. (1967), *Status of SO_2 Concentration Forecasts and Computations,* WBO, New York (unpublished manuscript).

Kussmann, A.S. (1968), *Atmospheric diffusion model for metropolitan areas as applied to New York City* (unpublished manuscript).

Lamb, R.G. and Neiburger, M. (1971), An interim version of a generalized urban air pollution model, *Atmospheric Environment*, 2, pp. 239–264.

Lanzani, G. and Tamponi, M. (1995), A microscale Lagrangian model for the dispersion of primary pollutants in a street canyon: sensitivity analysis and first validation trials, *Atmospheric Environment,* 23, pp. 3465–3475.

Lee, I.Y. and Park H.M. (1994), Parameterization of the pollutant transport and dispersion in urban street canyons, *Atmospheric Environment,* 28, No. 14, pp. 2343–2349.

Leitl, B. and Schatzmann, M. (2000), Physical modeling of emissions from naturally ventilated underground parking garages, *Environmental Monitoring and Assessment,* 65, pp. 221–229.

Lohmeyer, A., Mueller, W.J., and Baechlin, W. (2002), A comparison of street canyon concentration predictions by different modelers: final results now available from the Podbi-exercise, *Atmospheric Environment,* 36, pp. 157–158.

Lucas, D.H. (1958), The atmospheric pollution of cities, *International Journal of Air Pollution,* 1, pp. 71–86.

Luhar, A.K. and Hurley, P.J. (2003), Evaluation of TAPM, a prognostic meteorological and air pollution model, using urban and rural point-source data, *Atmospheric Environment,* 37, pp. 2795–2810.

Martin, D.O. (1971), An urban diffusion model for estimating long term average values of air quality, *Journal of the Air Pollution Control Association,* 21, pp. 16–19.

Martins, E.M. and Arbilla, G. (2003), Computer modeling study of ethanol and aldehyde reactivities in Rio de Janeiro urban air, *Atmospheric Environment,* 37, pp. 1715–1722.

Meade, P.J. and Pasquill, F. (1958), A study of the average distribution of pollution around Staythorpe, *International Journal of Air Pollution,* 1, pp. 60–70.

Mensink, C. (2000), Validation of urban emission inventories, *Environmental Monitoring and Assessment,* 65, pp. 31–39.

Mensink, C., De Vlieger, I., and Nys, J. (2000), An urban transport emission model for the Antwerp area, *Atmospheric Environment,* 34, pp. 4595–4602.

Mensink, C., Lewyckyj, N., and Janssen, L. (2002), A new concept for air quality modelling in street canyons, *Water, Air, and Soil Pollution: Focus,* 2, pp. 339–349.

Miller, M.E. and Holzsworth, G.C. (1967), An atmospheric model for metropolitan areas, *Journal of the Air Pollution Control Association,* 17, No. 1, pp. 46–50.

Morandi, M.T., Daisey, J.M., and Lioy, P.J. (1987), Development of a modified factor analysis/multiple regression model to apportion suspended particulate matter in a complex urban airshed, *Atmospheric Environment,* 21, No. 8, pp. 1821–1831.

Moses, H. and Carson, J.E. (1968), Stack design parameters influencing plume rise, *Journal of the Air Pollution Control Association,* 18, No. 7, p.484.

Munn, R.E., Thomas, D.A., and Cole, A.R.W. (1969), A study of suspended particulate and iron concentration in Windsor, Canada, *Atmospheric Environment,* 3, No. 1, pp. 1–10.

Nava, S., Prati, P., Lucarelli, F., Mando, P.A., and Zucchiati, A. (2002), Source apportionment in the town of la Spezia (Italy) by continuous aerosol sampling and PIXE analysis, *Water, Air, and Soil Pollution: Focus,* 2, pp. 247–260.

Nicholson, J.P., Weston, K.J., and Fowler, D. (2001), Modeling horizontal and vertical concentration profiles of ozone and oxides of nitrogen within high-latitude urban areas, *Atmospheric Environment,* 35, pp. 2009–2022.

Okamoto, S., Hayashi, M., Nakajima, M., Kainuma, Y., and Shiozawa, K. (1990), A factor analysis-multiple regression model for source apportionment of suspended particulate matter, *Atmospheric Environment,* 24A, No. 8, pp. 2089–2097.

Ott, W., Thomas, J., Mage, D., and Wallace, L. (1988), Validation of the simulation of human activity and pollutant exposure (SHAPE) model using paired days from the Denver, CO, carbon monoxide field study, *Atmospheric Environment,* 22, No. 10, pp. 2101–2113.

Palacios, M., Kirchner, F., Martilli, A., Clappier, A., Martin, F., and Rodríguez, M.E. (2002), Summer ozone episodes in the Greater Madrid area: analyzing the ozone response to abatement strategies by modeling, *Atmospheric Environment,* 36, pp. 5323–5333.

Panofsky, H.A. and Brier, G.W. (1958), *Some Applications of Statistics to Meteorology,* Pennsylvania State University Press, University Park, PA.

Panofsky, H.A. and Brier, G.W. (1984), *"Air Pollution Meteorology" Encyclopedia of Environmental Science and Engineering,* Vol. 1, Gordon and Breach, New York.

Panofsky, H.A. and Prasad, B. (1967), The effect of meteorological factors on air pollution in a narrow valley, *Journal of Applied Meteorology,* 6, No. 3, pp. 494–499.

Pelliccioni, A. and Poli, U. (2000), Use of neural net models to forecast atmospheric pollution, *Environmental Monitoring and Assessment,* 65, pp. 297–304.

Petersen, R.L. (1996), A wind tunnel evaluation of methods for estimating surface roughness length at industrial facilities, *Atmospheric Environment,* 31, pp. 45–57.

Pitts, R.O. and Lyons, T.J. (1992), A coupled mesoscale/particle model applied to an urban area, *Atmospheric Environment,* 26B, No. 3, pp. 279–289.

Pokrovskya, O.M., Kwokb, R.H.F., and Ng, C.N. (2002), Fuzzy logic approach for description of meteorological impacts on urban air pollution species: a Hong Kong case study, *Computers and Geosciences,* 28, pp. 119–127.

Pooler, F., Jr. (1961), A prediction model of mean urban pollution for use with standard wind roses, *International Journal of Air and Water Pollution,* 4, Nos. 3/4, pp. 199–211.

Pooler, F., Jr. (1966), A tracer study of dispersion over a city, *Journal of the Air Pollution Control Association,* 16, No. 12, pp. 677–681.

Qin, Y. and Oduyemi, K. (2003), Atmospheric aerosol source identification and estimates of source contributions to air pollution in Dundee, UK, *Atmospheric Environment,* 37, pp. 1799–1809.

Raghavan, N., Goyal, P., and Basu, S. (1983), A Gaussian model for predicting SO_2 concentration in the city of Agra, *Atmospheric Environment,* 17, No. 11, pp. 2199–2203.

Rao, K.S. (2002), ROADWAY-2: a model for pollutant dispersion near highways, *Water, Air, and Soil Pollution: Focus,* 2, pp. 261–277.

Rao, K.S., Ku, J., and Rao, S.T. (1989), A comparison study of three urban air pollution models, *Atmospheric Environment,* 23, No. 4, pp. 793–801.

Reddy, P.J., Barbarick, D.E., and Osterburg, R.D. (1995), Development of a statistical model for forecasting episodes of visibility degradation in the Denver metropolitan area, *Journal of Applied Meteorology,* 34, No. 3, pp. 616–625.

Riccio, A. and Barone, G. (2001), Real-time forecasting of photosmog episodes: the Naples case study, *Journal of Chemometrics,* 15, pp. 571–589.

Roberts, J.J., Croke, E.S., and Kennedy, A.S. (1970), An Urban Atmospheric Dispersion Model, Proceedings of the Symposium on Multiple Source Urban Diffusion Models, pp. 6-1–6-70.

Roberts, J.J., Croke, E.S., and Kennedy, A.S. (1971), Chicago Air Pollution Systems Analysis Program, Final Report, Argonne National Laboratory, ANLIES-CC-009, February, pp. 198–209.

Sahm, P., Louka, P., Ketzel, M., Guilloteau, E., and Sini, J.F. (2002), Intercomparison of numerical urban dispersion models—part I: street canyon and single building configurations, *Water, Air, and Soil Pollution: Focus,* 2, pp. 587–601.

San José, R., Rodríguez, M. A., Salas, I., and González, R.M. (2000), On the use of MRF/AVN global information to improve the operational air quality model Opana, *Environmental Monitoring and Assessment,* 65, p. 477–484.

Schatzmann, M. and Leitl, B. (2002), Validation and application of obstacle-resolving urban dispersion models, *Atmospheric Environment,* 36, pp. 4811–4821.

Schatzmann, M., Leitl, B., and Liedtke, J. (2000), Dispersion in urban environments: comparison of field measurements with wind tunnel results, *Environmental Monitoring and Assessment,* 65, pp. 249–257.

Scheffe, R.D. and Morris, R.E. (1993), A review of the development and application of the urban airshed model, *Atmospheric Environment,* 27B, No. 1, pp. 23–39.

Schlink, U. and Volta, M. (2000), Grey box and component models to forecast ozone episodes: a comparison study, *Environmental Monitoring and Assessment,* 65, pp. 313–321.

Schroeder, W.H. and Lane, D.A. (1988), The fate of toxic airborne pollutants, *Environmental Science and Technology,* 22, pp. 240–246.

Scott, H.J. (1954), Cleveland cooperative meteorological and air pollution program, paper no. 54-A-166, presented at Annual Meeting of the American Society for Mechanical Engineers, New York, November 28–December 3.

Silva, C., Pérez, P., and Trier, A. (2001), Statistical modeling and prediction of atmospheric pollution by particulate material: two nonparametric approaches, *Environmetrics,* 12, pp. 147–159.

Simpson, R.W., Daly, N.J., and Jakeman, A.J. (1983), The prediction of maximum air pollution concentrations for TSP and CO using Larsen's model and the ATDL model, *Atmospheric Environment,* 17, No. 12, pp. 2497–2503.

Slade, D.H. (1967), Modeling air pollution in the Washington, DC to Boston megalopolis, *Science,* 157, No. 3794, pp. 1794–1307.

Sloane, C.S. and Wolff, G.T. (1985), Prediction of ambient light scattering using a physical model responsive to relative humidity: validation with measurements from Detroit, *Atmospheric Environment,* 19, No. 4, pp. 669–680.

Smith, F.B. (1982), In tropospheric chemistry and air pollution, WMO Technical Note No. 176, World Meteorological Organization, Geneva, pp. 11–32.

Smith, M.E. (1961), International Symposium for Chemical Reaction of the Lower and Upper Atmosphere, San Francisco, Advance Papers, Stanford Research Institute, pp. 273–286.

Soulhac, L., Puel, C., Duclaux, O., and Perkins, R.J. (2003), Simulations of atmospheric pollution in Greater Lyon: an example of the use of nested models, *Atmospheric Environment,* 37, pp. 5137–5156.

Steven, R., Hanna, S.R., Britter, R., and Franzese, P. (2003), A baseline urban dispersion model evaluated with Salt Lake City and Los Angeles tracer data, *Atmospheric Environment,* 37, pp. 5069–5082.

Theodoridis, G., Karagiannis, V., and Valougeorgis, D. (2002), Numerical prediction of dispersion characteristics in an urban area based on grid refinement and various turbulence models, *Water, Air, and Soil Pollution: Focus,* 2, pp. 525–539.

Turner, D.B. (1964), A diffusion model for an urban area, *Journal of Applied Meteorology,* No. 1, pp. 85–91.

Turner, D.B. (1967), Workbook of atmospheric dispersion estimates, Public Health Service Publication, No. 999-AP-26, Environmental Health Series, Department of Health, Education, and Welfare.

Turner, D.B. (1968), The diurnal and day-to-day variations of fuel usage of space heating in St. Louis, Missouri, *Atmospheric Environment,* 2, pp. 339–351.

Ulke, A.G. and Andrade, M.F. (2001), Modeling urban air pollution in Sao Paulo, Brazil: sensitivity of model predicted concentrations to different turbulence parameterizations, *Atmospheric Environment,* 35, pp. 1747–1763.

Vardoulakis, S., Fisher, B.E.A., Gonzalez-Flesca, N., and Pericleous, K. (2002), Model sensitivity and uncertainty analysis using roadside air quality measurements, *Atmospheric Environment,* 36, pp. 2121–2134.

Vardoulakis, S., Fisher, B.E.A., Pericleous, K., and Gonzalez-Flesca, N. (2003), Modeling air quality in street canyons: a review, *Atmospheric Environment,* 37, pp. 155–182.

Vardoulakis, S., Gonzalez-Flesca, N., Fisher, and B.E.A. (2002), Assessment of traffic-related air pollution in two street canyons in Paris: implications for exposure studies, *Atmospheric Environment,* 36, pp. 1025–1039.

Vautard, R., Beekmann, M., Roux, J., and Gombert, D. (2001), Validation of a hybrid forecasting system for the ozone concentrations over the Paris area, *Atmospheric Environment,* 35, pp. 2449–2461

Viotti, P., Liuti, G., and Di Genova, P. (2002), Atmospheric urban pollution: applications of an artificial neural network (ANN) to the city of Perugia, *Ecological Modeling,* 148, pp. 27–46.

Vlachogiannis, D., Andronopoulos, S., Passamichali, A., Gounaris, N., and Bartzis, J. (2000), A three-dimensional model study of the impact of AVOC and BVOC emissions on ozone in an urban area of the eastern Spain, *Environmental Monitoring and Assessment,* 65, pp. 41–48.

Vlachogiannis, D., Rafailidis, S., Bartzis, J.G., Andronopoulos, S., and Venetsanos, A.G. (2002), Modelling of flow and pollution dispersion in a two dimensional urban street canyon, *Water, Air, and Soil Pollution: Focus,* 2, pp. 405–417.

Wanta, R.C. (1967), Meteorology and air pollution, *Air Pollution,* Vol. 1, edited by A.C. Stem, Academic Press, New York.

Wexler, A.S., Lurmann, F.W., and Seinfeld, J.H. (1994), Modeling urban and regional aerosols—I: model development, *Atmospheric Environment,* 28, No. 3, pp. 531–546.

Worley, F.L., Jr. (1971), Report on Mathematical Modeling of Photochemical Smog, prepared for presentation to the Panel on Modeling, NATO/CCMS Pilot Project on Air Pollution, Paris, France, July 26–27.

Yocum, J.E., Chisholm, D.A., Collins, G.F., Farrow, J.A., Gagosz, R.M., and Magyar, J.C. (1967), Air Pollution Study of the Capitol Region, December 1967, Travelers Research Center Service Corporation, Hartford, CT, prepared for the capitol.

Zanis, P., Zerefos, C.S., Gilge, S., Melas, D., Balis, D., Ziomas, I., Gerasopoulos, E., Tzoumaka, P., Kaminski, U., and Fricke, W. (2001), Comparison of measured and modeled surface ozone concentrations at two different sites in Europe during the solar eclipse on August 11, 1999, *Atmospheric Environment,* 35, pp. 4663–4673.

Ziv, A., Berkowicz, R., Genikhovich, E., Palmgren, F., and Yakovleva, E. (2002), Analysis of the St. Petersburg traffic data using the OSPM model, *Water, Air, and Soil Pollution: Focus,* 2, pp. 297–310.

ALESSANDRO ANZALONE
Polytechnic University of Puerto Rico

URBAN RUNOFF

INTRODUCTION

This article is a survey of control and treatment of combined sewer overflows (CSOs). The discussions of control/treatment technologies that consist mostly of downstream treatment have been divided into seven sections:

1) Source Control: Street Sweeping
2) Collection System Control
3) Storage
4) Physical (with/without) Chemical Treatment
5) Biological Treatment
6) Advanced Treatment
7) Disinfection

Storage is the oldest documented abatement measure currently practiced, and it should be considered at all times in system planning because it allows for maximum use of existing dry-weather facilities. Physical (with/without) chemical treatment will generally be the minimum required to meet discharge or receiving-water-quality goals. If a higher degree of organics removal is needed, biological treatment should be examined. If maintaining a viable microorganism population is not feasible, but removal of dissolved and colloidal organics is desired, advanced treatment may be attractive. If disinfection is required, it would follow some level of physical treatment.

General discussions of CSO control/treatment can be found in several documents, including the following:

EPA-600/2-76-286, "Cost Estimating Manual-Combined Sewer Overflow Storage and Treatment." (NTIS PB 266 359).

EPA-600/8-77-014, "Urban Stormwater Management and Technology: Update and User's Guide." (NTIS PB 275 654).

EPA-600/8-80-035, "Urban Stormwater Management and Technology: Case Histories." (NTIS PB 81 107153).

EPA-670/2-74-040, "Urban Stormwater Management and Technology: An Assessment." (NTIS PB 240 687).

Field, R. and Lager, J.A. "Urban Runoff Pollution Control State-of-the-Art. Journal of Environmental Engineering Division, ASCE, Vol. 101, No. EE1, February 1975.

Field, R. and Struzewski, Jr., E.J. "Management and Control of Combined Sewer Overflows," Journal of Water Pollution Control Federation, Vol. 44, No. 7, July 1972.

1 SOURCE CONTROL: STREET SWEEPING

Street sweeping to remove accumulated dust, dirt, and litter, has been shown to be an effective but limited method of attacking the source of stormwater-related pollution problems. Street-cleaning effectiveness is a function of (1) pavement type and condition, (2) cleaning frequency, (3) number of passes, (4) equipment speed, (5) sweeper efficiency, and (6) equipment type. Pavement type and condition affect performance more than do differences in equipment: In general, smooth asphalt streets are easier to keep clean than those consisting of loosely bound aggregate in a thick, oily matrix; and of course, the poorer a pavement's condition, the more difficult to keep it clean. The most important measure of street-cleaning effectiveness is "pounds per curb-mile removed" for a specific program condition. This removal value, in conjunction with the curb-mile costs, allows the cost for removing a pound of pollutant for a specific street-cleaning program to be calculated.

In the San Jose, California, street-sweeping project (EPA-600/2-79-161), experimental design and sampling procedures were developed that can be used in different cities to obtain specific information about street-dirt characteristics and their effects on air and water quality. At the test site in San Jose, it was determined that frequent street cleaning on smooth asphalt streets (once or twice per day) can remove up to 50% of the total solids and heavy-metal mass yields of urban runoff, whereas typical street-cleaning programs (once or twice per month) remove less than 5% of the total solids and heavy metals in the runoff. It was also determined that removal-per-unit effort decreased with increasing numbers of passes per year. This is shown in Figure 1, which relates the annual total solids removed to the street-cleaning frequency, for different street-surface conditions in San Jose.

Street-sweeping results are highly variable. Therefore, a street-sweeping program for one city cannot be applied to other cities, unless the program is shown to be applicable through experimental testing. This may be seen when comparing street-sweeping test results from San Jose with those of Bellevue, Washington. In Bellevue, it was demonstrated that additional cleaning, after a certain level of effort, is not productive and that the additional street-cleaning effort would be better applied to other areas. For the study area in Bellevue, it is estimated that street cleaning operations of about two or three passes per week would remove up to

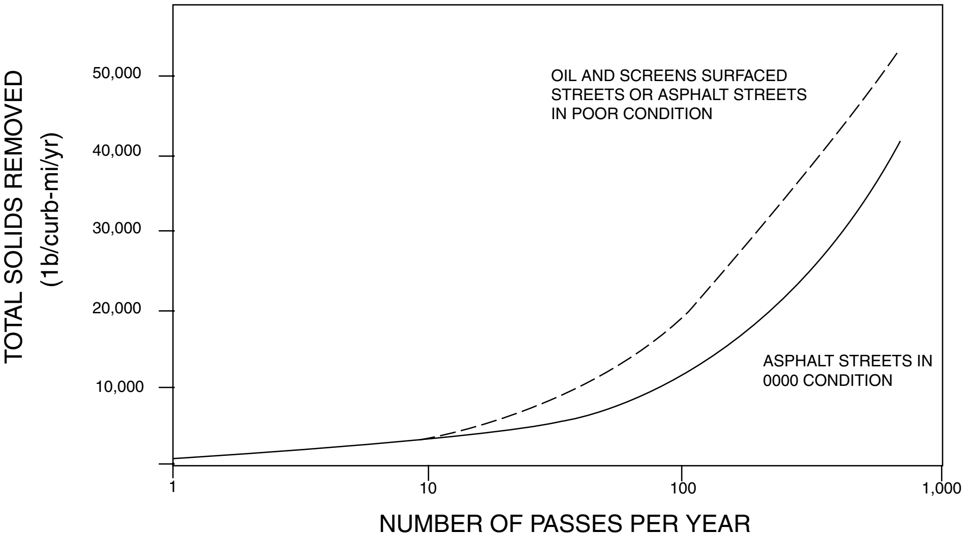


FIGURE 1 Street sweeping: annual amount removed as a function of the number of passes per year at San Jose test site.

68 kg (150 lb/curb-mi), or up to 25% of the initial street-surface load. Increased utilization of street-cleaning equipment would result in very little additional benefit. This is illustrated, for total solids and chemical oxygen demand (COD) removals, in Figures 2 and 3, respectively. Increased street-cleaning operations beyond two or three times per week are likely to increase the street-surface loadings due to erosion of the street surface. Increasing the cleaning frequency from once per week to two or more times per week will have only a very small additional benefit. Cleaning very infrequently (once every two months) may not be beneficial at all, except in cities where it may be possible to schedule street cleaning so that it is coordinated with rainfall events.

Street cleaning not only affects water quality but has multiple benefits, including the improvement of air quality, aesthetic conditions, and public health. Since street cleaning alone will probably not ensure that water-quality objectives are met, a street-cleaning program would have to be incorporated into a larger program of “best management practices” and/or downstream treatment. Costs of street cleaning have been reported to range from $4.92 to $19.03/curb-km ($7.92 to $30.61/curb-mi). The wide variation in these costs was attributed to differences in labor rates and equipment costs.

2 COLLECTION SYSTEM CONTROL

Catch Basins

A catch basin is defined as a chamber or well, usually built at the curbline or a street, for the admission of surface water to a sewer or subdrain, having at its base a sedimentation sump to retain grit and detritus below the point of overflow. It should be noted that a catch basin is designed to trap sediment, while an inlet is not. Historically, the role of catch basins has been to minimize sewer clogging by trapping coarse debris (from unpaved streets) and to reduce odor emanations from low-velocity sewers by providing a water seal.

In a project conducted in the West Roxbury section of Boston, three catch basins were cleaned, and subsequently four runoff events were monitored at each catch basin. Average pollutant removals per storm are shown in Table 1.

Catch basins must be cleaned often enough to prevent sediment and debris from accumulating to such a depth that the outlet to the sewer might become blocked. The sump must be kept clean to provide storage capacity for sediment and to prevent resuspension of sediment. Since the volume of stormwater detained in a catch basin will reduce the amount of overflow by that amount (it eventually leaks out or evaporates), it is also important to clean catch basins to provide liquid storage capacity. To maintain the effectiveness of catch basins for pollutant removal will require a cleaning frequency of at least twice per year, depending upon conditions. The increased cost of cleaning must be considered in assessing the practicality of catch basins for pollution control.

Typical cost data for catch basins are presented in Table 2. The reported costs will vary, depending on the size of the catch basin used by a particular city. Catch basin cost multiplication factors, as a function of sump storage capacity, are shown in Figure 4.

Estimated national average costs for three catch-basin cleaning methods are presented in Table 3.

Sewer Flushing

The deposition of sewage solids in combined sewer systems during dry weather has long been recognized as a major contributor to “first-flush” phenomena occurring during wet-weather runoff periods. The magnitude of these loadings during runoff periods has been estimated to range

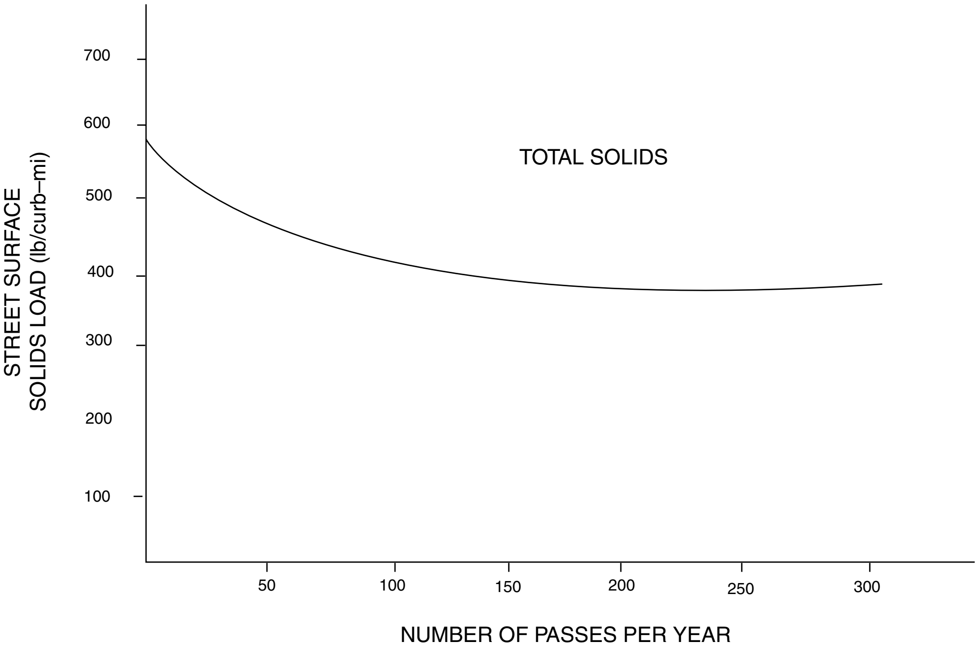


FIGURE 2 Street cleaner productivity in Bellevue, Washington.

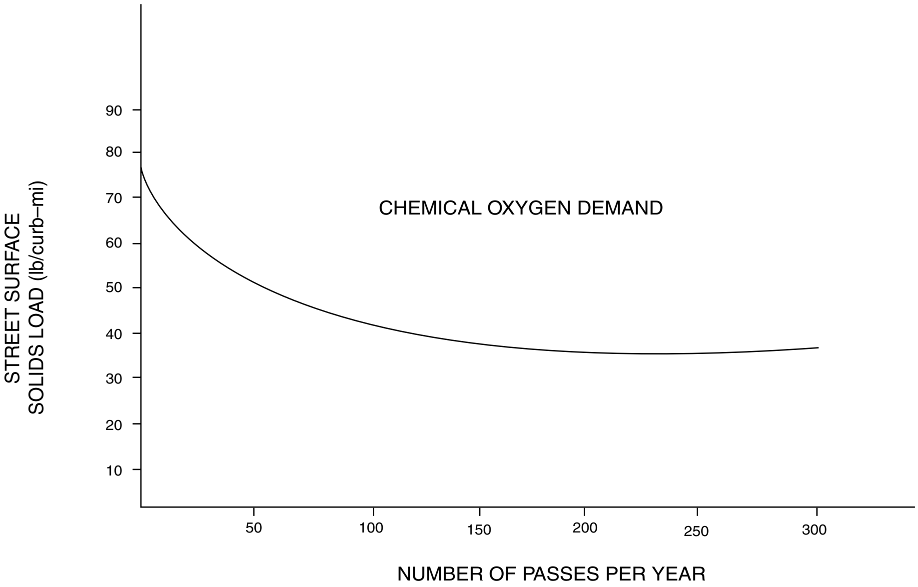


FIGURE 3 Street cleaner productivity in Bellevue, Washington.

TABLE 1
Pollutants retained in catch basins

Constituent	% Retained
SS	60–97
Volatile SS	48–97
COD	10–56
BOD_5	54–88

TABLE 2
Catch basin costs (ENR = 5,000)

	Range	Average
Total Installed Cost, $	1,014–2,453	2,019

(EPA-600/2-77-051).

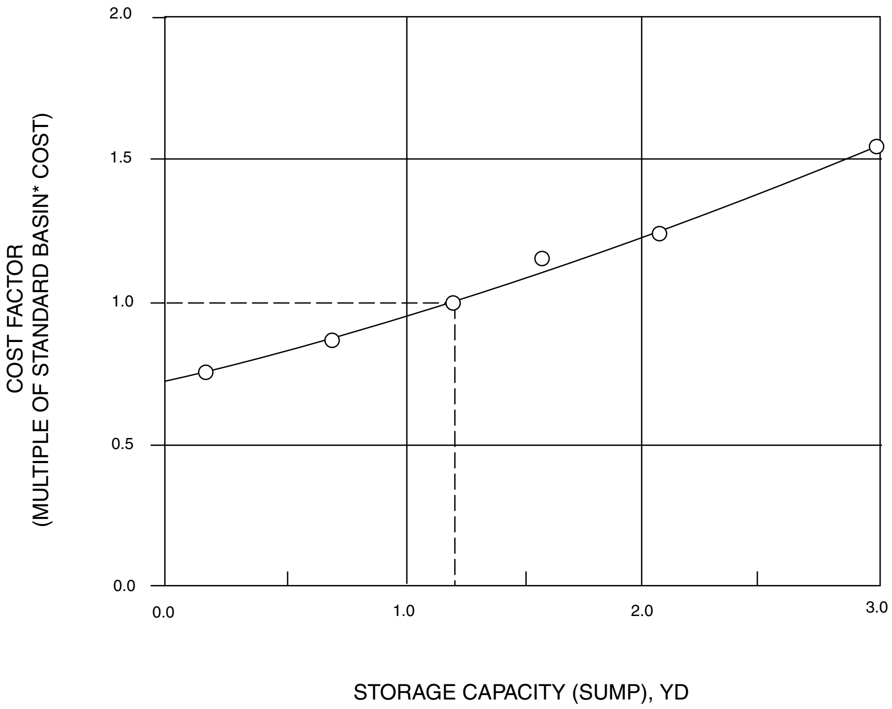


FIGURE 4 Catch basin cost factors versus storage capacity (EPA-600/2-77-051).

TABLE 3
Catch basin cleaning costs (ENR = 5,000)

Manual			Eductor			Vacuum		
$/catch basin	$m^3	$/yd^3	$/catch basin	$/m^3	$/yd^3	$/catch basin	$/m^3	$/yd^3
19.20	47.07	36.21	14.77	13.40	10.14	20.00	28.10	21.44

(EPA-600/2-77-051).

up to 30% of the total annual dry-weather sewage loadings. Sewer flushing during dry weather is designed to periodically remove the material, as it accumulates, and hydraulically convey it to the treatment facilities, thus preventing resuspension and overflow of a portion of the solids during storm events and lessening the need for CSO treatment. Flushing is particularly beneficial for sewers with grades too flat to be self-cleansing and also helps ensure that sewers can carry their design flow capacities. Sewer flushing requires cooperation between the authorities with jurisdiction over collection system maintenance and wastewater treatment.

For developing sewer-flushing programs, it is necessary to be able to estimate deposition buildup. Predictive equations have been developed, based on field studies in Boston, to relate the total daily mass of pollutant deposition in a collection system to the system characteristics, such as per-capita waste-production rate, service area, total pipe length, average pipe slope, and average pipe diameter. A simple model is given by the equation

$$TS = 0.0076(L')^{1.063}(\bar{S})^{-0.4375}(q)^{-0.51} \quad (R^2 = 0.845)$$

where

TS = deposited solids loading, lb/d
$\bar{S}$ = mean pipe slope, ft/ft
L' = total length of sewer system, ft
q = per-capita waste rate (plus allowance for infiltration), gpcd

(EPA-600/2-79-133)

The total pipe length (L') of the system is generally assumed to be known. In cases where this information is not known, and where crude estimates will suffice, the total pipe length can be estimated from the total basin area, A (acres), using the expressions that follow.

For low population density (10–20 people/acre):

$$L' = 168.95(A)^{0.928} \quad (R^2 = 0.821).$$

For moderate-high population density (30–60 people/acre):

$$L' = 239.41(A)^{0.928} \quad (R^2 = 0.821).$$

If data on pipe slope are not available, the mean pipe slope can be estimated using the following equation:

$$\bar{S} = 0.348(\bar{S}_g) \quad (R^2 = 0.96)$$

where $\bar{S}_g$ = mean ground slope, ft/ft.

It has been found that cleansing efficiency of periodic flush waves is dependent upon flush volume, flush discharge rate, sewer slope, sewer length, sewer flowrate, sewer diameter, and population density. Maximum flushing rates at the downstream point are limited to the regulator/interceptor capacities prior to overflow. Internal automatic flushing devices have been developed for sewer systems. An inflatable bag is used to stop flow in upstream reaches until a volume capable of generating a flushing wave is accumulated. When the appropriate volume is reached, the bag is deflated with the assistance of a vacuum pump, releasing impounded sewage and resulting in the cleaning of the sewer segment. Field experience has indicated that sewer flushing by manual means (water-tank truck) is a simple, reliable method for CSO solids removal in smaller-diameter laterals and trunk sewers.

Pollutant removals as a function of length of pipe flushed (Dorchester, Massachusetts, EPA-600/2-79-133) are presented in Table 4. The relationship between cleaning efficiency and pipe length is important, since an aim of flushing is to wash the resuspended sediment to strategic locations, such as a point where sewage is flowing, to another point where flushing will be initiated, or to the sewage treatment plant.

Flushing is also an effective means for suspending and transporting heavy metals associated with light colloidal solids particles. Approximately 20–40% of heavy metals contained within sewage sediment—including cadmium, chromium, copper, lead, nickel, and zinc—have been found to be transported at least 305m (1,000 ft) by flush waves. Estimated costs of sewer-flushing methods are shown in Table 5.

TABLE 4

Pollutant removals by sewer flushing as a function of length of segment flushed (254–381 mm [10–15 in.] pipe)

	% Removals: organics and nutrients	% Removals: dry-weather grit/ inorganic material
Manhole-to-Manhole Segments	75–95	75
Serial Segments up to 213 m (700 ft)	65–75	55–65
305 m (1,000 ft)	35–45	18–25

TABLE 5

Estimated costs of sewer-flushing methods based on daily flushing program (ENR = 5,000)

Number of segments: 46(254–457mm [10–18 in.] pipe) Automatic-flushing module operation (one module/segment)	
Capital cost	$21,640
Annual O&M cost	$199,084
Manual flushing mode	
Capital cost	$137,050
Annual O&M cost	$236,738

Regulator/Concentrators

The dual-functioning swirl regulator/concentrator can achieve both flow control and good removals (90–100%, laboratory determined) of inert settleable solids (effective diameter 0.3 mm, specific gravity 2.65) and organics (effective diameter 1.0 mm, s.g. 1.2). It should be noted that the laboratory test solids represent only the heavier fraction of solids found in CSO. Actual CSO contains a wider range of solids, so removals in field operations are closer to 40–50%.

Swirls have no moving parts. Flow is regulated by a central circular weir spillway, while simultaneously, solid–liquid separation occurs by way of flowpath-induced inertial separation and gravity settling. Dry-weather flows are diverted through the foul sewer outlet to the intercepting sewer for subsequent treatment at the municipal plant. During higher-flow storm conditions, 3–10% of the total flow—which includes sanitary sewage, storm runoff, and solids concentrated by swirl action—is diverted by way of the foul sewer outlet to the interceptor. The relatively clear, high-volume supernatant overflows the central circular weir and can be stored, further treated, or discharged to a stream. The swirl is capable of functioning efficiently over a wide range of CSO rates and has the ability to separate settleable solids and floatable solids at a small fraction of the detention time normally required for sedimentation. Suspended solids (SS) removals for the Syracuse, New York, prototype unit, as compared to hypothetical removals in a conventional regulator, are shown in Table 6. The BOD_5 removals for the Syracuse unit are shown in Table 7 (see EPA-600/2-79-134. *Disinfection/ Treatment of Combined Sewer Outflows*).

The *helical bend regulator/concentrator* is based on the concept of using the helical motion imparted to fluids at bends when a total angle of approximately 60 degrees and a radius of curvature equal to 16 times the inlet pipe diameter (*D*) are employed. Figure 5 illustrates the device. The basic structural features of the helical bend are: (1) the transition section from the inlet to the expanded straight section before the bend, (2) the overflow side weir and scum baffle, and (3) the foul outlet for removing concentrated solids and controlling the amount of underflow going to the treatment works. Dry-weather flow goes through the lower portion of the device and to the intercepting sewer. As the liquid level increases during wet weather, helical motion beings and the solids are drawn to the inner wall and drop to the lower level of the channel leading to the treatment plant. When the

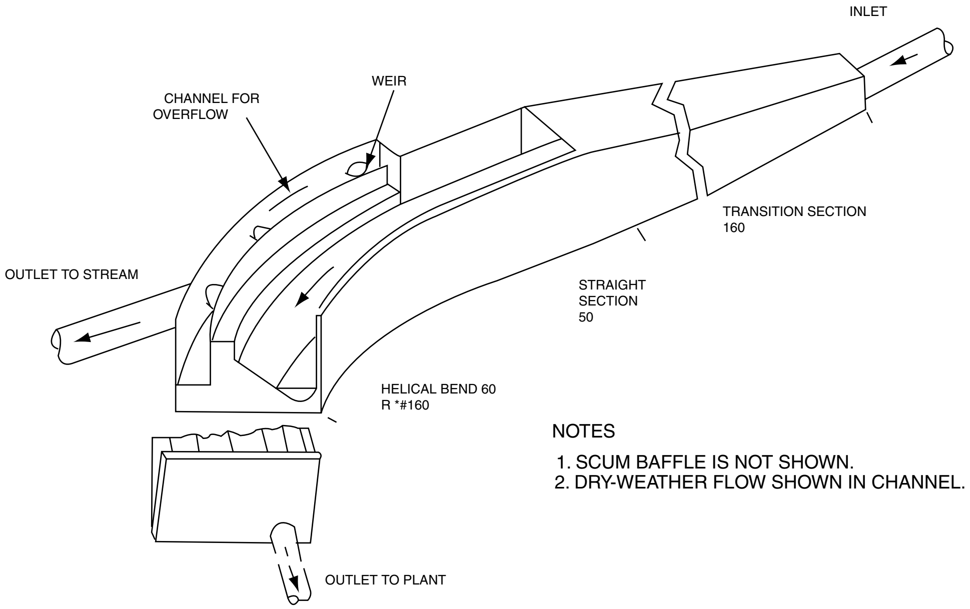


FIGURE 5 Helical bend.

TABLE 6
Suspended solids removal

	Swirl regulator/concentrator						Conventional regulator (hypothetical)			
	Average SS per storm mg/l			Mass Loading, kg			Mass Loading, kg			
Storm #	Inf.	Eff.	(%) Rem.†	Inf.	Eff.	(%) Rem.†	Inf.	Under-flow	(%) Rem.‡	Swirl Net Removal Benefit, %*
2–1974	535	345	36	374	179	52	374	101	27	25
3–1974	182	141	23	69	34	51	69	33	48	3
7–1974	110	90	18	93	61	34	93	20	22	12
10–1974	230	164	29	255	134	48	256	49	19	29
14–1974	159	123	23	99	57	42	99	26	26	16
1–1975	374	167	55	103	24	77	103	66	64	13
2–1975	342	202	41	463	167	64	463	170	37	27
6–1975	342	259	24	112	62	45	112	31	28	17
12–1975	291	232	20	250	168	33	250	48	19	14
14–1975	121	81	33	83	48	42	83	14	17	25
15–1975	115	55	52	117	21	82	117	72	62	20

* Calculated by subtracting the hypothetical percent SS removals in a conventional regulator, from the percent SS removals in a swirl regulator/concentrator.

† Data reflecting negative SS removals at tail end of storms not included.

‡ For the conventional regulator removal calculation, it is assumed that the SS concentration of the foul underflow equals the SS concentration of the inflow

TABLE 7
Swirl regulator/concentrator BOD_5 removal

	Mass loading, kg			Average BOD_5 per storm, mg/l		
Storm#	Inf.	Eff.	(%) Rem.	Inf.	Eff.	(%) Rem.
7–1974	26,545	4,644	82	314	65	79
1–1975	3,565	1,040	71	165	112	32
2–1975	12,329	6,164	50	99	70	20

(EPA-600/2-79-134; EPA-625/2-77-012).

storm subsides, the velocity of flow increases, due to the constricted channel. This helps prevent the settling of solids. As with the swirl, the proportion of concentrated discharge will depend on the particular design. The relatively clean CSO passes over a side weir and is discharged to the receiving water or to storage and/or treatment facilities. Floatables are prevented from overflowing by a scum baffle along the side weir and collect at the end of the chamber. They are conveyed to the treatment plant when the storm flow and liquid level subside.

Based on laboratory tests, pollutant removals in a helical bend unit are comparable to those in a swirl (a helical bend was demonstrated in Boston). Helicals and swirls are, in effect, upstream treatment devices for the removal of relatively heavy, coarse material, but they cannot be used to substitute for primary clarification. A comparison of construction costs for helical bend and swirl regulator/concentrators is presented in Table 8. It should be noted that these costs do not reflect the real cost-effectiveness of swirls and helicals, since these units actually serve dual functions (i.e., flow control and wastewater treatment). Even though the construction cost for the helical bend is higher than for the swirl, the helical may be more appropriate for a particular site, based on space availability and elevation difference between the interceptor and the incoming combined sewer (the helical requires a smaller elevation difference than the swirl). If there is not sufficient hydraulic head to allow dry-weather flow to pass through the facility, an economic evaluation would be necessary to determine the value of one of three alternatives: (1) pumping the foul sewer flow continuously, (2) pumping the foul flow during storm conditions, or (3) bypassing the facility during dry-weather conditions.

3 STORAGE

Because of the high volume and variability associated with CSO, storage is considered a necessary control alternative. Storage is also the best documented abatement measure currently practiced. Storage facilities are frequently used to attenuate peak flows associated with CSO. Storage must be considered at all times in system planning because it allows for maximum use of existing dry-weather treatment plant facilities and results in the lowest-cost system in terms of treatment. The CSO is stored until the treatment plant can accept the extra volume. At that time, the CSO is discharged. Storage facilities can provide the following advantages: (1) They respond without difficulty to intermittent and random storm behavior, and (2) they are not upset by water-quality changes.

TABLE 8
Comparison of constructor costs for helical bend and regulator/concentrators (ENR = 5,000)

Capacity	Swirl	Helical
1.42 m^3/s (50 cfs)	$459,360	$1,121,500
2.83 m^3/s (100 cfs)	744,910	2,102,300
4.67 m^3/s (165 cfs)	1,034,595	2,963,090

Note: Land costs not included.

Figure 6 shows that there is an increase in BOD_5 and SS percent removals, with an increase in tank volume per drainage area. Figure 7, however, demonstrates decreasing removal efficiencies per unit volume as tank size increases. Also, beyond an optimum tank volume, the rate of cost increase for retaining the extra flow increases; therefore, it is not economical to design storage facilities for the infrequent storm. During periods when the tank is filled to capacity, the excess that overflows to the receiving water will have had a degree of primary treatment by way of sedimentation.

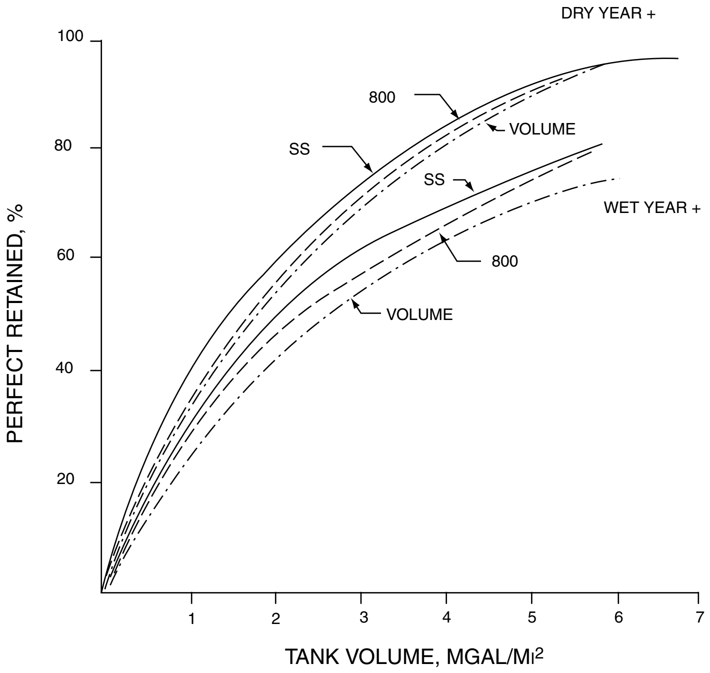
Storage facilities can be classified as either in-line or off-line. The basic difference between the two is that in-line storage has no pumping requirements. In-line storage can consist of either storage within the sewer pipes ("in-pipe") or storage in in-line basins. Off-line storage requires detention facilities (basins or tunnels) and facilities for pumping CSO to either storage or sewer system. Examination of storage options should begin with in-pipe storage. If this is not suitable, the use of in-line storage tanks should be considered; however, head allowances must be sufficient since no pumps will be used. Off-line storage should be considered last, since this will require power for pumping. Since the idea of storage is to lower the cost of the total treatment system, the storage capacity must be evaluated simultaneously with downstream treatment capacity so that the least cost combination for meeting water/CSO quality goals can be implemented.

If additional treatment capacity is needed, a parallel facility can be built at the existing plant, or a satellite facility can be built at the point of storage.

In-Pipe Storage

Because combined sewers are designed to carry maximum flows occurring, say, once in 5 years (50–100 times the average dry-weather flow), during most storms there will be considerable unused volume within the conduits. In-pipe storage is provided by damming, gating, or otherwise restricting flow passage causing sewage to back up in the upstream lines. The usual location to create the backup is at the regulator, or overflow point, but the restrictions can also be located upstream.

For utilization of this concept, some of all of the following may be desirable: sewers with flat grades in the vicinity of the interceptor, high interceptor capacity, and extensive control and monitoring networks. This includes installation of effective regulators, level sensors, tide gates, rain gauge networks, sewage and receiving water quality monitors, overflow detectors and flow-meters. Most of the systems are computerized, and to be safe, the restrictions must be easily and automatically removed from the flow stream when critical flow levels are approached or exceeded. Such systems have been successfully implemented in Seattle and Detroit. In-pipe storage was also demonstrated in Minneapolis–St. Paul. Costs associated with in-pipe storage systems are summarized in



*48.5 cm (19.1 in.) rainfall
+103.4 cm (40.7in.) rainfall
runoff coefficient (C) = 0.5

FIGURE 6 Percent retained versus tank volume (EPA-600/8-77-014).

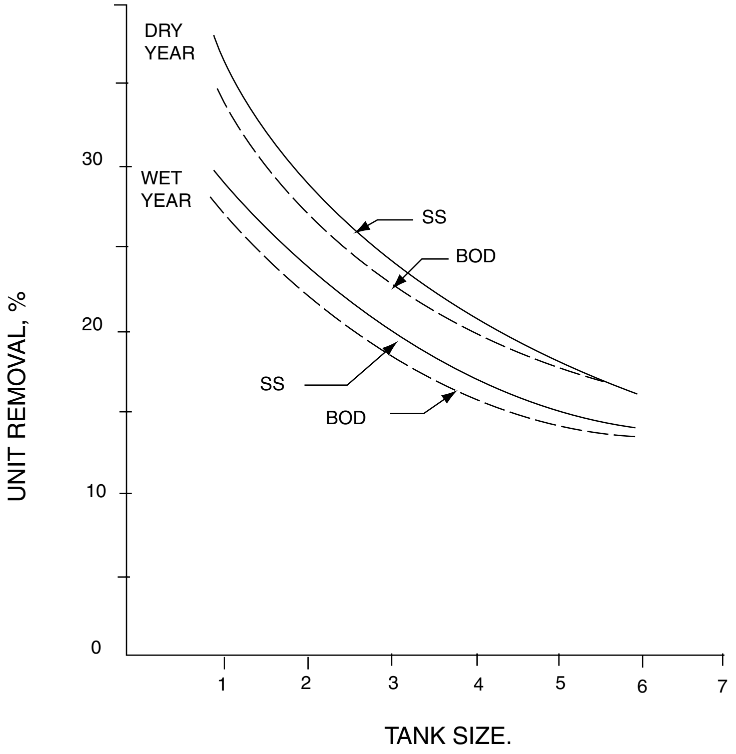


FIGURE 7 Unit removal efficiencies for CSO detention tanks (EPA-600/7-77-014).

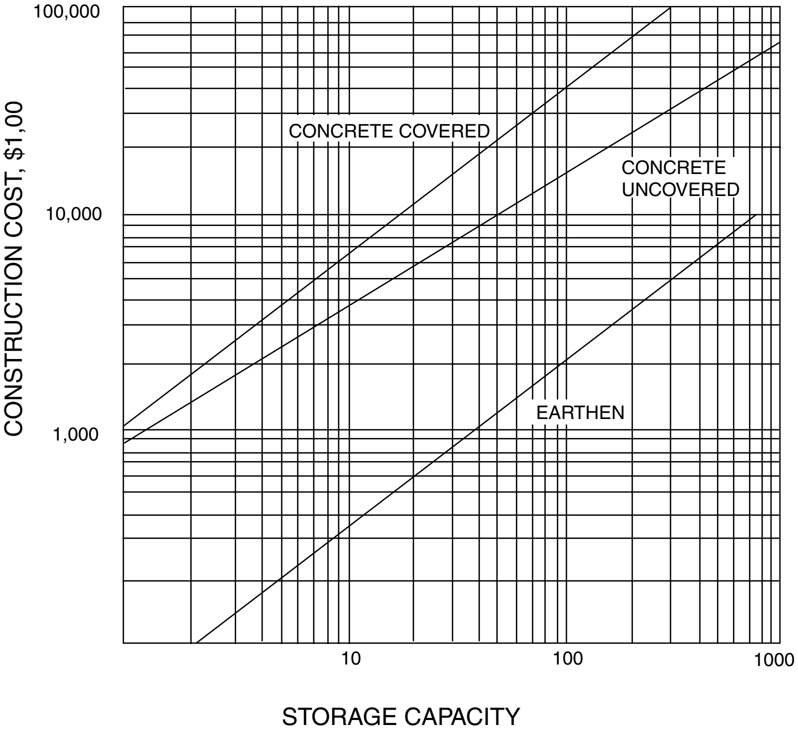


FIGURE 8 Storage basin construction costs: ENR = 5,000 (EPA-600/8-77-014).

Table 9. Costs include regulator stations, central monitoring and control systems, and miscellaneous hardware.

Off-Line Storage

Off-line storage facilities can be located at overflow points or near dry-weather treatment plants. Typical storage facilities include lagoons and covered or uncovered concrete tanks. Tunnels are also used where land is not available. Costs for basin storage facilities are presented in Table 10, and construction cost curves are shown in Figure 8. Note that these curves do not include pumping facilities, so these curves are applicable to in-line basins; the costs for earthen basins include liners.

Innovative Storage Technology

In-receiving Water Flow Balance System Karl Dunkers, an independent research engineer from Sweden, has developed, under the auspices of the Swedish EPA counterparts, an approach to lake protection against pollution from stormwater runoff. Instead of using conventional systems for equalization (i.e. concrete tanks or lined ponds), which are relatively expensive and require a lot of land area, the flow balanced method uses a wooden pontoon tank system in the lake, which performs in accordance with the plug-flow principle. The tank bottom is the lake bottom itself. The tank volume is always filled up, either with polluted stormwater runoff or with lake water. When it is raining, the stormwater runoff will "push" the lake water from one compartment to another. The compartment walls are of flexible PVC (polyvinyl chloride) fiberglass cloth. When not receiving runoff, the system reverses by automatically flowing back and the lake water fills up the system. Thus, the lake water is utilized as a flow balance medium. These units are sized to yield an effective in-water volume equal to the storage required for the storm size selected for design. Added benefits are the ease of construction and the flexibility to expand the volume if deemed necessary after initial installation; initial storage-volume estimates need not be as exact. Costs have been estimated to be one-fifth to one-tenth the cost of conventional land-side storage without real estate costs.

Sweden has invested in three of the installations so far. Two have been in operation for eight to nine years, and a third for seven years. The systems seem to withstand wave action up to 0.9 m (3 ft) as well as severe icing conditions. If a wall is punctured, patching is easily accomplished. Maintenance has been found to be inexpensive. The in-receiving water-flow balance system has been successfully demonstrated with urban runoff in freshwater lakes only. If used with CSO, consideration would have to be given to sludge handling and disposal. EPA's Storm and Combined Sewer Pollution Control Program is demonstrating this unique system with CSO in a much harsher marine/estuarine environment in New York City. Testing for seven storms indicates effectiveness as both a floatables trap and a temporary storage of CSO volumes. The estimated cost of this system is $1,641/lin m ($500/lin ft).

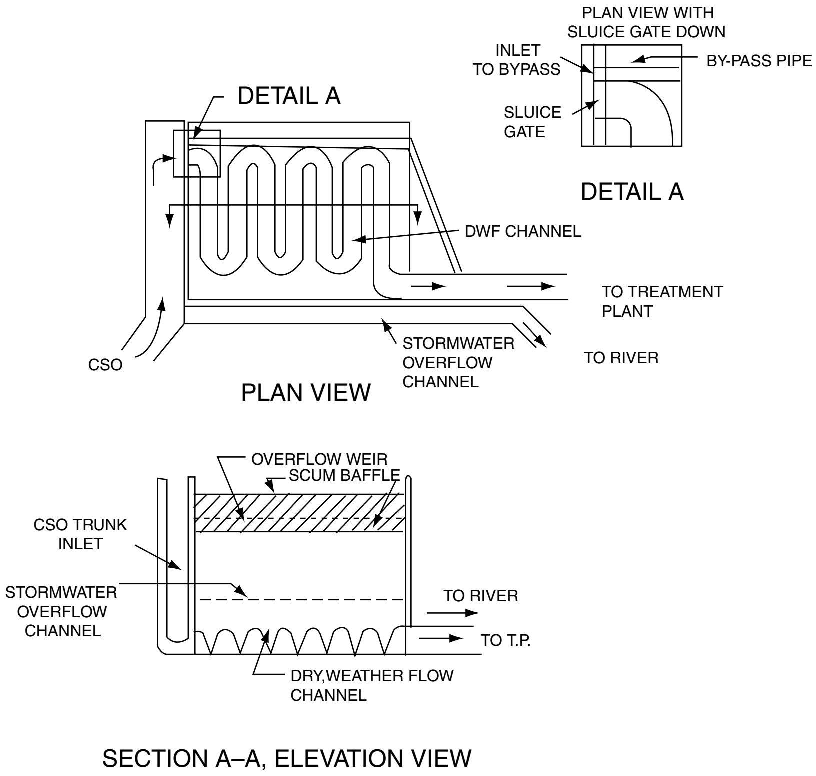


FIGURE 9 Self-cleaning storage-sedimentation basin.

TABLE 9
Summary of in-pipe storage costs (ENR = 5,000)

Location	Storage Capacity (Mgal)	Drainage Area (acre)	Capital Cost ($)	Storage Cost ($/gal)	Cost per Acre ($/acre)	Annual O&M Cost ($/yr)
Seattle, WA Control & Monitoring System	—	—	8,748,550	—	—	182,500
Automated Regulator Stations	—	—	9,747,970	—	—	550,410
Total	17.8	13,120	18,496,520	1.06	1,411	732,910
Minneapolis—St. Paul, MN	NA	64,000	7,531,870	—	117	—
Detroit, MI	140	89,000	7,024,910	0.6	78	—

NA = Not available.

Self-cleaning Storage/Sedimentation Basin In Zurich, Switzerland, an in-line sedimentation storage tank was designed to prevent solids from shoaling after a storm and to provide for solids transport to the interceptor. The floor of the tank contains a continuous dry-weather channel (which is an extension of the tank's combined sewer inlet) that meanders from side to side (see Figure 9) through the tank. This channelized floor arrangement allows for complete sediment transport to the interceptor both during dry weather and upon drawdown after a storm event. The dry-weather flow comes through the meandering bottom channel. During wet-weather flows, the water level in the tank rises above the channel. If the storm intensity is low enough, there is complete capture, and if the storm intensity continues to rise, an overflow occurs through a weir at the tail end of the tank. A scum baffle prevents entrained solids from overflowing. This arrangement allows for sedimentation to take place during a tank overflow condition and, at the same time, for transport of solids that settle by way of the bottom channel.

TABLE 10
Summary of basin storage costs (ENR = 5,000)

Location	Storage capacity (Mgal)	Drainage area (acres)	Capital cost ($)	Storage cost ($/gal)	Cost per acre	Annual O&M cost ($/year)
Akron, OH (1)	1.1	188.5	1,139,000	1.04	6,100	7,200
Milwaukee, WI Humboldt Ave.	3.9	570	4,435,000	1.14	7,800	127,000
Boston, MA Cottage Farm Detection and Chlorination Station	1.3	15,600	16,237,000	12.49	1,000	203,000
Charles River Marginal Conduit Project (2)	1.2	3,000	23,725,000	19.77	8,000	246,000
New York, NY Spring Creek Auxiliary Plant Pollution Control Plant						
Storage	12.39	3,620	29,838,000	2.40	9,100	251,000
Sewer	13.00	—	—	—	—	—
Total	25.39	3,620.00	29,838,000	1.18	9,100	251,000
Chippewa Falls, WI	2.82	90	1,868,000	0.65	21,000	7,000
Storage Treatment	—	—	427,000	—	5,400	20,000
Total	2.82	90	2,295,000	—	—	—
Chicago, IL (3)	—	—	—	—	—	—
Phase I Tunnels under Construction or Completed, and Pumping Stations	1,016	—	1,676,040,000	1.65	—	—
Phase I Tunnels Remaining	1,033	—	1,556,973,000	1.54	—	—
Phase I Subtotal	2,049	240,000	1,958,460,000	0.05	—	14,973,000
Phases I and II Total	44,374	240,000	5,191,473,000	1.17	21,800	26,770,000
Sandusky, OH (4)	0.36	14.86	1,303,590	3.61	87,500	15,500
Washington, D.C. (5)	0.20	30.0	2,172,650	11.02	73,600	8,400
Columbus, OH (6) Whittier Street	3.75	29,250	15,353,420	4.10	500	—
Cambridge, MD (7)	0.25	20	769,640	3.20	46,600	36,000

† Estimated values.

‡ Estimated area.

(1) EPA-600/2-76-272 (media-void space storage).

(2) Environmental assessment statement for Charles River Marginal Conduit Project in the cities of Boston and Cambridge: Commonwealth of Massachusetts, Metropolitan District Commissions, September 1974.

(3) The Metropolitan Sanitary District of Greater Chicago: personal communication from Mr. Forrest Neil, chief engineer, September 1981.

(3a) Actual award TARP Status Report: July 1, 1981.

(3b) Estimate as for January 1, 1981, TARP Status Report July 1, 1981.

(3c) TARP Phase II estimate as of July 31, 1981.

(3d) Estimate.

(4) EPA 11022ECV09/71 (underwater).

(5) EPA 11020DWF12/69 (underwater).

(6) EPA-600/2-77-064; EPA 11020FAL03/71.

(7) EPA 11022DPP10/70 (underwater).

(EPA-600/8-77-014).

4 PHYSICAL (WITH/WITHOUT) CHEMICAL TREATMENT

Physical-chemical processes are of particular importance in CSO treatment because of their adaptability to automatic operation (including almost instantaneous startup and shutdown), excellent resistance to shockloads, and ability to consistently produce a low SS effluent. In this discussion, physical-chemical systems will be limited to screening, filtration, chemical clarification, and dissolved air flotation.

TABLE 11
Description of screening devices used in CSO treatment

Type of screen	General description	Process application	Comments
Drum Screen	Horizontally mounted cylinder with screen fabric aperture in the range of 100–841 microns. Operates at 2–7 r/min.	Pretreatment	Solids are trapped on inside of drum and are back-washed to a collection trough.
Microstrainers	Horizontally mounted cylinder with screen fabric aperture in the range of 23–100 microns. Operates at 2–7 r/min.	Main treatment	Solids are trapped on inside of drum and are back-washed to a collection trough.
Rotostrainer	Horizontally mounted cylinder made of parallel bars perpendicular to axis of drum. Slot spacing in the range of 250–2,500 microns. Operates at 1–10 r/min.	Pretreatment	Solids are retained on surface of drum and are removed by a scraper blade.
Disk Strainer	Series of horizontally mounted woven wire discs mounted on a center shaft. Screen aperture in the range of 45–500 microns. Operates at 5–15 r/min.	Pretreatment, or posttreatment of concentrated effluents	Unit achieves a 12–15% solids cake
Rotary Screen	Vertically aligned drum with screen fabric aperture in the range of 74–167 microns. Operates at 30–65 r/min.	Main treatment	Splits flow into two distinct streams: unit effluent and concentrate flow, in the proportion of approximately 85:15.
Static Screen	Stationary inclined screening surface with slot spacing in the range of 250–1,600 microns.	Pretreatment	No moving parts. Used for removal of large suspended and settleable solids.

* A vertically mounted microstrainer is available, which operates totally submerged at approximately 65 r/min. Aperture range is 10–70 microns. Solids are moved from the screen by a sonic cleaning device.
(EPA-600/8–77/014).

Screening

Screens have been used to achieve various levels of SS removal contingent with three modes of screening process applications:

1) *Main treatment.* Screening is used as the primary treatment process.
2) *Pretreatment.* Screening is used to remove suspended and coarse solids prior to further treatment to enhance the treatment process or to protect downstream equipment.
3) *Dual use.* Screening provides either main treatment or pre-treatment of stormwater and is used as an effluent polisher during periods of dry weather.

Screens can be divided into four categories:

Screen Type	*Opening Size*
bar screen	>25.4 mm (>1 in.)
coarse screen	4.8–25.4 mm (3/16–1 in.)
fine screen	0.1–4.8 mm (1/250–3/16 in.)
microscreen	<0.1 mm (<1/250 in.)

TABLE 12
Design parameters for static screens

Hydraulic loading, gal/min/ft of width	100–180
Incline of screens, degrees from vertical	35*
Slot space, microns	250–1,600
Automatic controls	None

* Bauer Hydrasieves™ have 3-stage slopes on each screen: 25°, 35°, and 45° .gal/min/ft × 0.207 = 1/m/s.
(EPA-600/8-77-014).

No special studies have been made to evaluate bar and coarse screens in relation to CSO, so the basis for design should be the same as for their uses in dry-weather treatment facilities. Because CSO contains a significant amount of coarse debris, which is aesthetically undesirable, providing coarse screening as the minimum CSO treatment may be useful. Fine screens and microscreens are discussed together because, in most cases, they operate in a similar manner.

Several distinct types of screening devices have been developed and used for SS removal from CSO, and are described in Table 11. Design parameters for static screens, microstrainers, drum screens, disc screens, and rotary screens are presented in Tables 12, 13, and 14. Removal efficiency of screening devices is adjustable by changing the aperture size of the screen placed on the unit, making these devices very versatile. Solids removal efficiencies are affected by two mechanisms: (1) straining by the screen, and (2) filtering of

TABLE 13
Design parameters for microstrainers, drum screens, and disc screens

Parameter	Micro-strainers	Drum screen	Disc screen
Screen aperture, microns	23–100	100–420	45–500
Screen material	Stainless steel or plastic	Stainless steel or plastic	Wire cloth
Drum speed, r/min			
Speed range	2–7	2–7	5–16
Recommended speed	5	5	—
Submergence of drum, %	60–80	60–70	50
Flux rate, gal/ft²/min of submerged screen	10–45	20–50	20–25
Headloss, in.	10–24	6–24	18–24
Backwash			
Volume, % of inflow	0.5–3	0.5–3	—*
Pressure, lb/in.²	30–50	30–50	—

* Unit's waste product is a solids cake of 12–15% solids content.
gal/min/ft² × 2.445 = m³/h/m².
in. × 2.54 = cm.
ft × 0.305 = cm.
lb/in.² × 0.0703 = kg/cm².
(EPA-600/8-77-014).

TABLE 14
Design parameters for rotary screens

Screen aperture, microns	
Range	74–167
Recommended aperture	105
Screen material	stainless steel or plastic
Peripheral speed of screen, ft/s	14–16
Drum speed, r/min	
Range	30–65
Recommended speed	55
Flux rate, gal/ft²/min	70–150
Hydraulic efficiency, % of inflow	75–90
Backwash	
Volume, % of inflow	0.02–2.5
Pressure, lb/in.²	50

ft/s × 0.305 = m/s.
gal/ft²/min × 2.445 = m³/m²/h.
lb/in.² × 0.0703 = kg/cm².
(EPA-600/8-77-014).

smaller particles by the mat (schmutzdecke) deposited by the initial straining.

The efficiencies of screens treating a waste with a typical distribution of particle sizes will increase as the size of screen opening decreases.

The second-most important condition affecting removal efficiencies, especially for microstrainers, is the thickness of filtered material on the screen. Whenever the thickness of this filter mat is increased, the suspended matter removal will also increase because of the decrease in effective pore size and the filtering action of the filtered mat. This will also increase headloss across the screen. It was found, during experimental microstrainer operation in Philadelphia, that because of extreme variation in the influent SS concentration of CSO, removal efficiency would also vary, while effluent concentration remained relatively constant. For example, an effluent concentration of 10 mg/l SS would yield a reduction of 99% for an influent concentration of 1,000 mg/l (representative of "first-flush"), whereas the SS reduction would be only 50% if the influent concentration were 20 mg/l (representative of tail end of storm). This concentration-dependent phenomenon is apt to recur in other physical-chemical stormwater treatment operations (R. Field and E. Struzeski, JWPCF, Vol. 44, No. 7, July 1972).

Microstrainers and fine screens remove 25–90% of the SS, and 10–70% of the BOD_5, depending on the size of the screens used and the type of wastewater being treated. At Philadelphia, polyelectrolytes addition (0.25–1.5 mg/l) improved the operating efficiency of the microstrainer. Suspended solids removal increased from 70% to 78%, and the average effluent SS was reduced from 40 to 29 mg/l. The flux rate also increased from an average of 56.2 m³/m²/h (23 gal/ft²/min) to 95.4 m³/m²/h (39 gal/ft²/min). After an extensive laboratory coagulation study, moderately charged, high-molecular-weight cationic polyelectrolytes were found to be the most suitable for this application. Microstrainer, drum screen, static screen, and rotary screen performances as a function of influent SS concentration, for several experimental projects, are shown in Figures 10, 11, 12 and 13, respectively.

Costs of screening facilities are presented in Table 15.

Screening/Dual Media High-Rate Filtration

Dual media high-rate filtration (DMHRF) (>29 m³/m²/h [>8 gal/ft²/min]) removes small-sized particulates that remain after screening, and floc remaining after polyelectrolyte and/or coagulant addition. Principal advantages of the proposed system are: (1) high treatment efficiencies, (2) automated operation, and (3) limited space requirements. To be most effective, filtration through media that are graded from coarse to fine in the direction of filtration is desirable. A uniform sized, single specific-gravity medium filter cannot conform to this principle since backwashing of the bed automatically grades the bed from coarse to fine in the direction of washing; however, the concept can be approached by using a two-layer bed. A typical case is the use of coarse anthracite particles on top of less coarse sand. Since anthracite is less dense than sand, it can be coarse and still remain on top of the bed after the backwash operation. Another alternative would be an upflow filter, but these units have limitations in that they cannot accept high filtration rates.

The principal parameters to be evaluated in selecting a DMHRF system are media size, media depth, and filtration rate. Since much of the removal of solids from the water takes

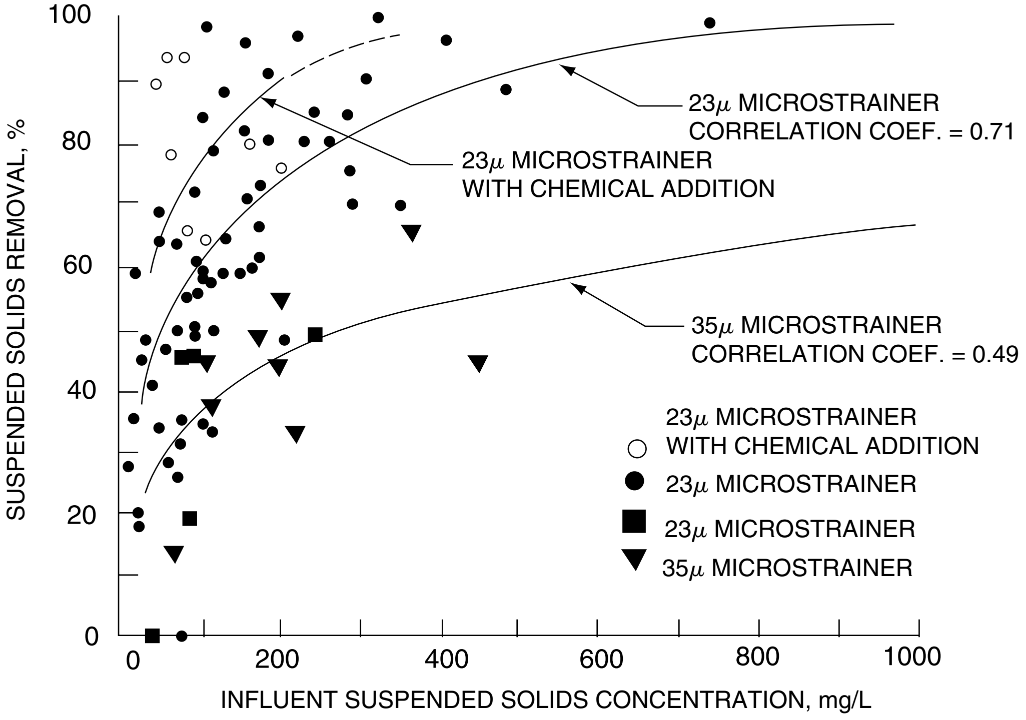


FIGURE 10 Microstrainer performance as a function of SS concentration (EPA-600/8-77-014).

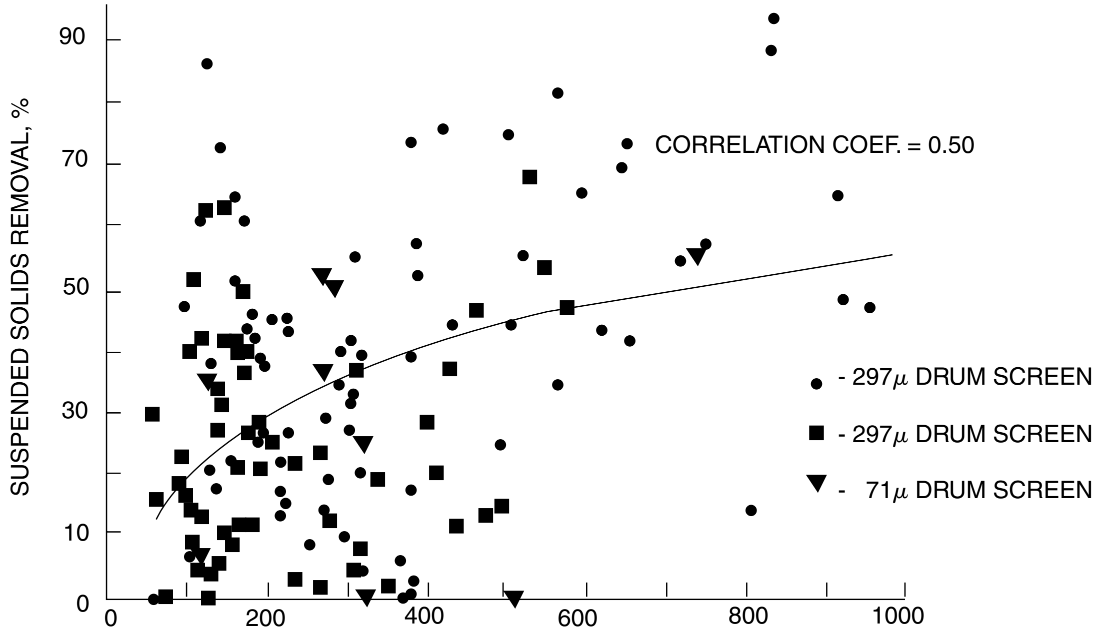


FIGURE 11 Drum Screen performance as a function of SS concentration (EPA-600/8-7-014).

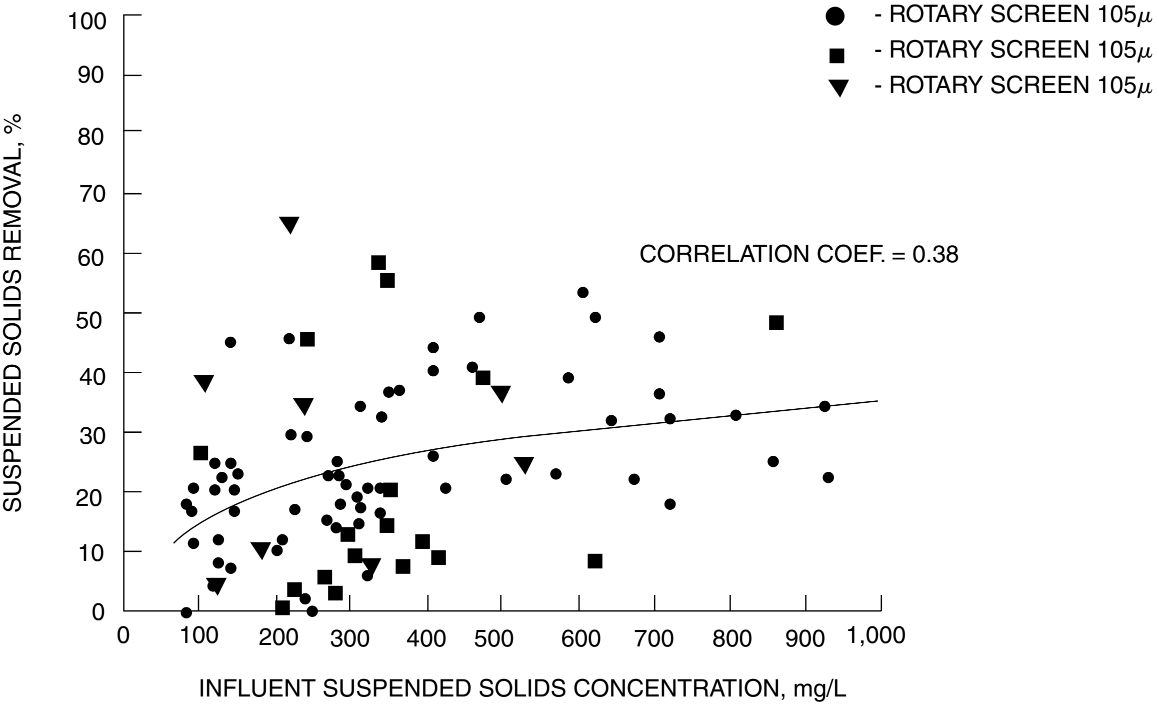


FIGURE 12 Static screen performance as a function of SS concentration (EPA-600/8-77-014).

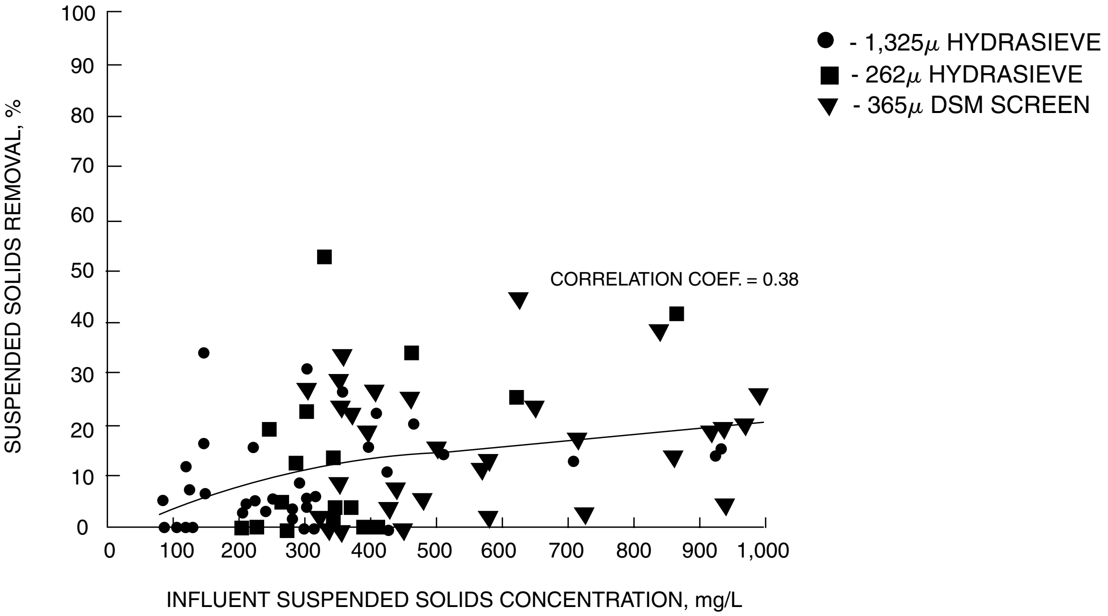


FIGURE 13 Rotary screen performance as a function of SS concentration (EPA-600/8-77-014).

TABLE 15
Cost summary of selected screening alternatives (ENR = 5,000)

Project location	Type of screen	Screening capacity (Mgal/d)	Capital cost ($)	Annual cost ($/Mgal/d)	O&M cost ($ 1,000 gal)
Belleville, Ont. (1)	Rotary screen	1.8	83,700	43,600	0.207
		5.4	244,200	44,800	0.207
		7.2	321,000	44,500	0.207
	Static screen	0.75	37,200	49,700	0.106
		5.3	239,000	45,500	0.106
		7.5	326,800	43,500	0.106
Cleveland OH(2)*†	Drum screen	25	1,521,300	60,800	—
		50	2,219,400	44,300	—
		100	4,363,00	43,600	—
		200	8,350,800	41,700	—
Fort Wayne, IN	Static screen	18	681,000	37,800	0.051
	Drum screen	18	635,700	35,200	0.097
	Rotary screen	18	1,461,800	38,500	0.114
Mount Clemens,	Microstrainer	1.0	65,500	65,500	—
MI(3)					—
Philadelphia, PA	Microstrainer with chemical addition	7.4	227,000	30,700	0.120
	Microstrainer without chemical addition	7.4	370,000	50,000	0.120
Racine, WI	Drum screen	3.9	56,500	14,500	—
Seattle, WA (4)*	Rotary screen	25	1,500,000	60,000	0.245
Syracuse, NY (5)†	Rotary screen	5	323,700	64,700	—
	Drum screen	10	642,500	64,300	—

* Estimated costs for several sizes of facilities.

† Estimates include supplemental pumping stations and appurtenances.

(1) Operational data for the Belleville Screening Project, Ontario Ministry of the Environment, August 6, 1976.

(2) EPA 11023EY104/72.

(3) EPA-670/2-75-010.

(4) EPA 11023 FDD03/70.

(5) EPA-600/2-76-286.

Mgal/d × 0.0438 = m^3/s.

$/1,000 gal × 0.264 = $/$m^3$.

(EPA-600/8-77-014).

place within filter media, their structure and composition is of major importance. Too fine a medium may produce a high-quality effluent, but also may cause excessive headlosses and extremely short filter runs. On the other hand, media that is too coarse may fail to produce the desired clarity of the effluent. Therefore, the selection of media for DMHRF should be made by pilot testing using various materials in different proportions and at different flow rates. Depth of media is limited by headloss and backwash considerations. The deeper the bed, the greater the headloss and the harder it is to clean. On the other hand, the media should be of sufficient depth so as to be able to retain the removed solids within the depth of the media for the duration of the filter run at the design flux rate without permitting breakthrough. The design filtration flux must be such that the effluent will be of a desired quality without causing excessive headloss through the filter, which in turn requires frequent backwashing. At high flux, shear forces seem to have significant effect on solids retention and removal.

Several DMHRF pilot-study installations have been demonstrated for control of CSO pollution. These facilities have used 15.2-, 30.5-, and 76.2-cm (6-, 12-, and 30-in.)-diameter filter columns with anthracite and sand media, together with various dosages of coagulants and/or polyelectrolytes. Descriptions of the DMHRF facilities are summarized in Table 16. Suspended solids removal by DMHRF was found to vary directly with influent SS concentrations and inversely with flux or hydraulic loading rate. Experimental results (from New York and Cleveland) have shown that SS removals from CSO increase appreciably with appropriate

TABLE 16
Description of CSO–DMHRF pilot-plant demonstration facilities

Project location	Process description	No. of filter columns	Diameter of columns (in.)	Pretreatment facilities	Filter media[†]	Period of operation
Cleveland, OH	Pilot deep bed, dual media high-rate filtration, with chemical addition. Facilities include pretreatment, storage and filtration.	3 1	6 12	420-micron drum screen	5 ft of No. 3 anthracite over 3 ft of No. 612 sand	1970–71
New York, NY (Newtown Creek)	Pilot deep bed, dual media high-rate filtration, with polyelectrolyte addition. Facilities include pretreatment, storage, and filtration. Dry-weather and combined sewer flow is pumped from grit chamber of Newton Creek plant.	1 2	30 6	420-micron rotostrainer, later replaced with a 420-micron disc strainer	5 ft of No. 3 anthracite over 2 ft of No. 612 sand	1975–78
Rochester, NY	Pilot deep bed, dual media high-rate filtration with chemical addition.	3	6	swirl separator	5 ft of No. 1–1/2 or No. 2 anthracite over 3 ft of No. 1220 sand	1955–76

[*] Systems operate at flux rates ranging from 20 to 73 $m^3/m^2/h$ (8–30 $gal/ft^2/min$).

in. × 2.54 = cm.
ft × 0.305 = m.
(EPA-600/8-77-014).

[†] Media	Effective size (m)	Uniformity coefficient
No. 3 anthracite	4.0	1.5
No. 2 anthracite	1.78	1.63
No. 1–1/2 anthracite	0.98	1.73
No. 612 sand	2.0	1.32
No. 1220 sand	0.95	1.41

chemical additions. The DMHRF treatment of CSO at New York's dual-use facility (40 $m^3/m^2/h$ [16 $gal/ft^2/min$] constant flux) provided overall average SS removals of 61% across the filter and 66% across the system, with an average influent SS concentration of 182 mg/l. Average SS removals for the three testing models(l) no chemicals, (2) polymer only, (3) polymer and alum—and test ranges are shown in Table 17.

A measure of the capability of a filter to remove SS, which is useful for predicting removals and filter-run cycle, is the *specific capture* or *mass capture.* This can be expressed as pounds of solids removed per filter surface, or pounds of solids removed per media volume. Table 18 presents average SS mass captures obtained across the filter (New York) during CSO tests of at least 3 hours duration, and the average for tests S-13, -14, and –16, which used more optimal chemical feeds and occurred during the storms of greatest intensity. It should be noted that these mass capture values are specific to the Newtown Creek filter and the test conditions.

BOD removals (New York) from CSO averaged 32% across the filter and 41% across the system with an average influent BOD_5 of 136 mg/l. The removals improved with chemical additions. Average BOD removals for the three testing modes and test ranges are shown in Table 19. It should be noted that the nature of the CSO tested (e.g., the presence of dissolved industrial organic contaminants) may account for variable BOD removals.

Limited tests were also run (New York) to determine heavy-metals reduction. These results, shown in Table 20, represent composite samples. Design parameters for DMHRF are shown in Table 21. Costs of CMHRF facilities are summarized in Table 22. These costs are based on facilities similarly designed to that of the Cleveland demonstration project. Comparison with alternate treatment systems show that DMHRF is cost-competitive with conventional sedimentation facilities for dual process (sanitary and CSO) or CSO treatment, yet DMHRF has only 5–7% the area requirements. For strict CSO treatment, DMHRF is competitive with dissolved air flotation and micro-straining processes.

Screening/Dissolved Air Flotation

Dissolved air flotation (DAF) is a unit operation used to separate solid particles or liquid droplets from a liquid phase. Separation is brought about by introducing fine air bubbles into the liquid phase. As the bubbles attach to the solid particles or liquid droplets, the buoyant force of the combined particle and air bubble is great enough to cause the particle to rise. Once the particles have floated to the surface, they are removed by skimming. The most common

TABLE 17
CSO–CMHRF average SS removals (New York, NY)

	Plant influent (mg/l)	Filter influent (mg/l)	Filter effluent (mg/l)	Filter removals (%)	System removals (%)
No Chemicals	175	150	67	55	62
Poly Only	209	183	68	63	67
Poly and Alum	152	143	47	67	69

(EPA-600/2–79/015).

TABLE 18
DMHRF average mass capture of CSO (New York, NY)

	Capture per filter surface		Capture per media volume	
CSO Test Nos.	lb/ft^2/run*	lb/ft^2/hr	lb/ft^3/run[†]	lb/ft^3/hr
S4B, S9-16	3.7	0.76	0.54	0.11
S-13,-14,-16	5.2	1.2	0.76	0.17

* 1 lb/ft^2 = 4.88 kg/m^2.

[†] 1 lb/ft^3/ = 16.02 kg/m^3.

(EPA-600/2–79–015).

TABLE 19
CSO–DMHRF average BOD removals (New York, NY)

	Plant influent (mg/l)	Filter influent (mg/l)	Filter effluent (mg/l)	Filter removals (%)	System removals (%)
No Chemicals	164	131	96	27	41
Poly Only	143	129	84	35	41
Poly and Alum	92	85	53	38	43

(EPA-600/2–79–015).

process for forming the air bubbles is to dissolve air into the waste stream under pressure and then release the pressure to allow the air to come out of solution. The pressurized flow carrying the dissolved air to the flotation tank is either (1) the entire stormwater flow, (2) a portion of the stormwater flow (split flow pressurization), or (3) recycled DAF effluent.

High overflow rates 3.2–25 m^3/m^2/h (1.3–10.0 gal/ft^2/min) and shorter detention times (0.2–1.0 h) can be used for DAF than for conventional settling (0.5 m^3/m^2/h [0.2–0.7 gal/ft^2/min]; 1.0–3.0 h). This process has a definite advantage over gravity sedimentation when used on CSO, since particles with densities both higher and lower than the liquid can be removed on one skimming operation. Dissolved air flotation also aids in the removal of oil and grease, which are not as readily removed during sedimentation. The principal parameters that affect removal efficiencies are: (1) overflow rate, (2) amount of air dissolved in the flows, and (3) chemical addition. Chemical addition has been used to improve removals, and ferric chloride has been the chemical most commonly added. A treatment system consisting of screening followed by DAF has been found to be an effective method of reducing pollutants in CSO.

The basis of this system is that screening will remove particles that are too heavy for the air bubbles to carry. Average reported percent removals (Milwaukee pilot: 18,925 m^3/d [5 Mgal/d]), with and without chemical addition, are listed in Table 23. The chemical flocculant addition required to achieve the removals therein was 20 mg/l ferric chloride, and 48 mg/l cationic polyelectrolytes.

At Racine, Wisconsin, with full-scale prototype plants (54,126 m^3/d and 168,054 m^3/d [14.3 and 44.4 Mgal/d]), 40 mg/l ferric chloride and 2 mg/l cationic polyelectrolytes were used. The percent removals (concentration basis) are presented in Table 24. The results from Site II are better than from Site I because the hydraulic loading was usually lower at Site II than at Site I, resulting in lower overflow rates and longer tank detention times at Site II. Typical design parameters for DAF facilities are presented in Table 25.

TABLE 20
Removal of heavy metals by DMHRF

	Heavy metal constituent						
	Cadmium	Chromium	Copper	Mercury	Nickel	Lead	Zinc
Average removal, %*	56	50	39	0	13	65	48

* concentration basis.
(EPA-600/2–79–015).

TABLE 21
Design parameters for DMHRF

Filter Media Depth (ft)	
No. 3 anthracite	4–5
No. 612 sand	2–3
Effective Size (mm)	
Anthracite	2
Sand	2
Flux Rate (gal/ft2/min)	
Range	8–40
Design	24
Headloss (ft)	5–30
Backwash	
Volume (% of inflow)	4–10
Air	
Rate (standard $ft^3/min/ft^2$)	10
Time (min)	10
Water	

TABLE 22
Summary of costs for DMHRF facilities (ENR = 5,000)

	Construction costs ($)		Construction costs ($/Mgal/d)		O&M costs ($/yr)	
Capacity (Mgal/d)	24 gal/ ft^2/min	16 gal/ ft^2/min	24 gal/ ft^2/min	16 gal/ ft^2/min	24 gal/ ft^2/min	16 gal/ ft^2/min
25	3,600,000	4,200,000	144,000	168,000	110,100	113,100
50	5,424,000	6,550,000	108,500	130,900	137,600	142,000
100	9,951,000	12,150,000	99,500	121,500	244,800	254,900
200	16,900,000	20,100,000	84,400	100,200	323,000	334,600

* Includes low-lift pumping station, prescreening, and chemical addition facilities; excludes engineering and administration.
Mgal/d × 0.0438 = m^3/s.
gal/ft^2/min × 2.445 = m^3m^2/min.
(EPA-600/8-77-014).

For the two full-scale CSO test sites in Racine, capital costs (including land) were $1,064,000 and $2,132,000 for 54,100- and 168,000-m^3/d (14.3- and 44.4-Mgal/d) facilities, respectively (ENR = 5,000). Construction cost curves (ENR = 2,000) for DAF facilities, based on the experienced cost of the demonstration facilities, are presented in Figure 14. The operation and maintenance cost (ENR = 5,000) of the systems was $0.16/$m^3$ ($0.58/1,000 gal). It was thought that these costs could be reduced to $0.09/$m^3$ ($0.30/1,000 gal) by process and procedural modifications as described in EPA-600/2-79-106a. The major reason for the high O&M cost is the cost of labor for maintenance of the sites and cleanup of the sites after a system operation. These costs were $0.10/$m^3$ ($0.38/1,000 gal), or 65% of the total. Therefore, maintenance becomes the major cost item in

TABLE 23
Percent removals achieved with screening/DAF (Milwaukee, WI)

	Without chemical flocculant addition	With chemical flocculant addition
BOD_5	35 ± 8	60 ± 11
COD	41 ± 8	57 ± 11
SS	43 ± 7	71 ± 9
Volatile SS	48 ± 11	71 ± 9
Nitrogen	29 ± 14	24 ± 9

(EPA-600-77-069a).

TABLE 24
Percent removals achieved with screening/DAF (Racine)

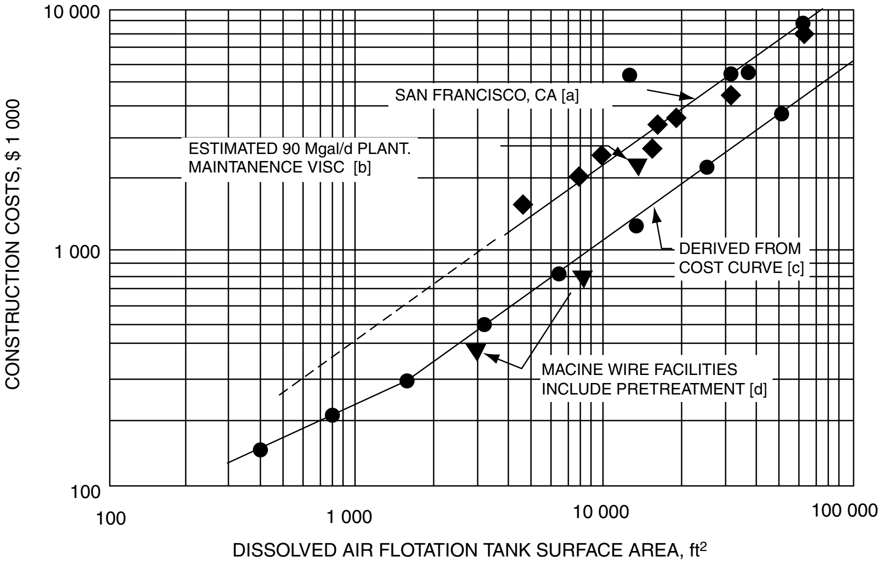
	Percent Removal	
	Site I	Site II
BOD_5	57.5	65.4
TOC	51.2	64.7
SS	62.2	73.3
Volatile SS	66.8	70.9
Total Phosphorus	49.3	70.0

(EPA-600/2-79-106a).

TABLE 25
DAF design parameters

Overflow Rate (gal/ft²/min)	
Low rate	1.3–4.0
High rate	4.0–10.0
Horizontal Velocity (ft/min)	1.3–3.8
Detention Time (min)	
Flotation cell range	10–60
Floating cell average	25
Saturation tank	1–3
Mixing chamber	1
Pressurized Flow (% of total flow)	
Split-flow pressurization	20–30
Effluent recycle pressurization	25–45
Air to Pressurized Flow Ratio (standard ft³/100 gal)	1.0
Air to Solids Ratio	0.05–0.35
Pressure in Saturation Tank (lb/in.²)	40–70
Float	
Volume (% of total flow)	0.75–1.4
Solids concentration (% dry weight basis)	1–2

lb/in.² × 0.0703 = kg/m².
(EPA-600/8-77-014).



NOTE: $11^2 = 0.0626\text{-m}^3$
$M_{gal/s} = 0.0438\text{-m}^3/s$

FIGURE 14 DAF construction costs: ENR = 2,000 (EPA/600/8-77-014).

TABLE 26
Comparison of physical treatment systems

Physical Unit Process	Suspended Solids	BOD_5	COD	Settleable Solids	Total Phosphorus	Total Kjeldahl Nitrogen
Sedimentation						
Without chemicals	20–60	30	34	30–90	20	38
Chemical assisted	68	68	45	—	—	—
Swirl Regulator/ Concentrator	40–60	25–60	—	50–90	—	—
Screening						
Microstrainers	50–95	10–50	35	—	20	30
Drum screens	30–55	10–40	25	60	10	17
Rotary screens	20–35	1–30	15	70–95	12	10
Disc screens	10–45	5–20	15	—	—	—
Static screens	5–25	0–20	13	10–60	10	8
Dissolved Air Flotation*	45–85	30–80	55	93[†]	55	35
High Rate Filtration[‡]	50–80	20–55	40	55–95	50	21
High Gradient Magnetic Separation[§]	92–98	90–98	75	99	—	—

* Process efficiencies include both prescreening and dissolved all flotation with chemical addition.

[†] From pilot-plant analysis EPA-600/8-7-014.

‡ Includes prescreening and chemical addition.

[§] From bench scale pilot-plant operation, 1–4 L/min (0.26–1.06 gal/min).

the full-scale application of screening/DAF for the treatment of CSO. Table 26 represents a tabulated summary of various physical treatment systems, including the swirl regulator/concentrator previously described.

5 BIOLOGICAL TREATMENT

Biological treatment is a means for stabilizing dissolved organic matter and removing nonsettleable colloidal solids. This can be accomplished either aerobically or anaerobically. Several biological processes have been applied to CSO treatment. These include contact stabilization, trickling filters, rotating biological contractors (RBC), and treatment lagoons. Descriptions of these processes and typical combined sewage treatment installations are provided in Tables 27 and 28. Biological treatment processes are generally classified as secondary treatment processes, capable of removing 70–95% of the BOD_5 and SS from waste flows at dry-weather design flowrates and loadings. When biological treatment processes are used for combined sewage treatment, removal efficiencies are lower (the percent organic matter is smaller for CSO solids than for dry-weather solids) and are controlled to a large degree by hydraulic and organic loading rates. Most biological systems are susceptible to overloading conditions and shock loads as compared to physical treatment processes. However, RBC have achieved high removals at flows 8–10 times dry-weather design flows.

Typical pollutant removals for contact stabilization, trickling filters, and RBCs (wet-weather loading conditions) are presented in Table 29. These processes include primary (except contact stabilization) and final clarification. Final clarification greatly influences the overall performance of the system by preventing the carry-over of biological solids produced by the processes. Pollutant removal efficiencies by treatment lagoons have varied from highs of 85–95% to negative values due to excessive algae production and carry-over. In addition to the type of lagoon and the number of cells in series (stages), several major factors that influence removal efficiencies include: (1) detention time, (2) source of oxygen supply, (3) mixing, (4) organic and hydraulic loading rates, and (5) algae removal mechanisms. A single-cell storage/oxidation lagoon in Springfield, Illinois averaged 27% BOD_5 removal and 20% SS removal; however, fish kills in the receiving water were greatly reduced as compared to that prior to the construction of the facility. Multiple-cell facilities with algae control systems constructed at Mount Cleimens, Michigan and Shelbyville, Illinois provide 75–90% SS and BOD_5 removal efficiencies during wet-weather conditions.

An operational problem common to all stormwater biological systems is that of maintaining a viable biomass to treat

TABLE 27
Description of biological processes used in CSO treatment

Biological process	Process description	Source of biomass	Requires additional treatment	Type of additional treatment
Contact Stabilization	Process is a modified activated sludge process in which the absorption phase, or contact, and the oxidation phase (stabilization) takes place in two separate tanks. Sludge is wasted from the stabilization tank to maintain constant biomass concentrations.	From conventional activated sludge treatment facility	Yes	Secondary clarification
Trickling Filters	Standard trickling filter process in which a biological growth is supported on a stationary medium and the stormwater distributed over the surface and allowed to flow through the media. Process can include standard rate or deep bed plastic media designs.	Must be continuously maintained with a source of food	Yes	Secondary qualification
Rotating Biological Contactors	Process operates on the same principle as trickling filters; however, the biological growth is supported on large diameter, closely-spaced disks that are partially submerged and rotate at slow speeds.	Must be continuously maintained with a source of food	Yes	Secondary clarification
Oxidation Ponds	Shallow aerobic ponds that rely on surface reaeration for oxygen supply to maintain biological uptake of organics. Sedimentation also occurs in oxidation ponds.	Allowed to generate for each storm	Optional	Final clarification, screening, or sand filtration
Aerated Lagoons	Similar to oxidation ponds except they are deeper and rely on artificial means of oxygen supply such as surface aerators or diffused air systems. System operates under aerobic conditions.	Allowed to generate for each storm	Optional	Final clarification, screening, or sand filtration
Facultative Lagoons	Facultative lagoons are the deepest of the lagoons and rely on surface reaeration. The lagoons have three distinct layers: (1) aerobic near the surface due to algae and aeration, (2) a transitions zone, and (3) an anaerobic zone near the bottom sludge deposits. The biological oxidation and anaerobic stabilization occur simultaneously.	Allowed to generate for each storm	Optional	Final clarification, screening, or sand filtration.

(EPA-600/8-77-014).

flows during wet-weather conditions. At New Providence, New Jersey trickling filters are operated in series during dry weather and in parallel during wet weather. This type of operation maintains a viable microorganism population during dry weather and also provides greater capacity for the wet-weather flows. For processes that borrow biomass from dry-facilities or allow the biomass to develop, a lag in process efficiency may be experienced as the biomass becomes acclimated to the changing waste strength and flowrate. Also, because of the limited ability of biological systems to handle fluctuating and high hydraulic shock loads, storage/detention facilities preceding the biological processes may be required. General maintenance problems experienced by wet-weather biological facilities are similar to those experienced at conventional biological installations. Winter operations of mechanical surface aerators have had some serious drawbacks, including icing, tipping, or sinking. Other methods of providing the required oxygen that show promise and have been demonstrated at many dry-weather facilities include diffused air systems and submerged tube aerators.

A comparison of construction and O&M costs for biological treatment systems and treatment lagoons is presented in Table 30. Costs of final clarification are included where control of solids and sludge produced by the biological treatment system are required. Costs also include pumping, disinfection, and algae control systems, where applicable. Engineering, administration, and land costs are not included in the estimates; however, land costs may be the controlling economic factor in the evaluation of lagoon treatment systems, and therefore must be evaluated for each specific location. Biological CSO treatment systems generally are integrated with or are a part of dry-weather treatment facilities. Cost estimates of the wet-weather portion of these facilities were separated from total costs of the total treatment systems. The cost of the in-line RBC at Milwaukee was used together with an estimated cost for a final clarifier to develop an estimated cost of a complete RBC treatment system. The final clarifier cost was based on one 19.8-m (65-ft) diameter clarifier with a surface loading rate of 2.04 $m^3/m^2/h$ (1,200 $gal/ft^2/d$). Costs of lagoon treatment systems vary widely and are a function of the type of lagoon (oxidation, aerated, or facultative), the number of cells and the miscellaneous equipment requirements (e.g., aeration equipment, disinfection equipment, instrumentation, pumping, and algae control provisions). Costs for many of these CSO facilities are based on only one installation of each biological treatment process. Therefore, these costs should be considered only coarse estimates and may be greatly influenced by the degree of integration with dry-weather treatment required to produce a viable system.

TABLE 28
Summary of typical biological CSO treatment installations

Project location	Type of biological treatment	Tributary area	Design capacity (Mgal/d)	Major process components	No. of units	Total size	Period of operation
Kenosha, WI	Contact stabilization	1,200	20	Contact tank	2	32,700 ft^3	1972–75
				Stabilization tank	2	97,900 ft^3	
Milwaukee, WI	Rotating biological contractors	35	0.05*	3-ft diameter RBC units	24	28,300 ft^3	1969–70
Mount Clemens, MI Demonstration system	Treatment lagoons in series with recirculation between storms	212	1.0†		1	750,000 ft^3	1972–75
					1	1,100,000 ft^3	
					1	930,000 ft^3	
Citywide full-scale system	Storage/treatment lagoons in series with recirculation between storms	1,471	4.0†		1	4,440,000 ft^3	Under construction
					1	508,000 ft^3	
				Storage/aerated lagoon	1	1,100,000 ft^3	
					1	922,000 ft^3	
				Oxidation lagoon			
				Aerated lagoon	1	36-ft diameter	
				Aerated lagoon			
				Aerated storage basin	1	65-ft diameter	
				Aerated lagoon			
				Oxidation lagoon			
				Aerated/oxidation lagoon			
New Providence, NJ	Trickling filters		6.0	High-rate plastic media	1	36-ft diameter	1970–
				High-rate rock media	1	65-ft diameter	
Shelbyville, IL	Treatment lagoons						
	Southeast site	44	28	Oxidation lagoon	1	255,600 ft^3	1969–
	Southwest site	450	110	Detention lagoon plus 2-cell facultative lagoon	1	2,782,700 ft^3	1969–
Springfield, IL	Treatment lagoon	2,208	67	Storage/oxidation lagoon	1	5,330,000 ft^3	1969–

TABLE 29
Typical wet-weather BOD and SS removals for biological treatment processes

	Expected range of pollutant removal, %	
Biological treatment process	BOD	COD
Contact Stabilization	70–90	75–95
Trickling Filters	65–85	65–85
RBC*	40–80	40–80

* Removal reflects low ranges from 30 to 10 times dry weather flow (EPA-600/8-77-014).

These costs can be used as a preliminary guide, but detailed analyses should be performed to compare and evaluate biological treatment alternatives with other methods of treatment and control. Initial capital investments of integrated dual-use facilities can be reduced by apportioning part of the costs to the dry-weather facility. The cost reduction is in proportion to the net benefit that the wet-weather facility provides to the overall treatment efficiency during dry-weather periods.

6 ADVANCED TREATMENT

High Gradient Magnetic Separation

High gradient magnetic separation (HGMS) is a new treatment technology applied to CSO management. In its simplest form, the high gradient magnetic separator consists of a canister packed with a fibrous ferromagnetic material that is magnetized by a strong external magnetic field (coils surround the canister). An iron frame increases the efficiency of the electromagnetic coils. The device operates in a sequence of feed and flush modes.

The magnetic particles are trapped on the edges of the magnetized fibers while the nonmagnetic particles and slurry pass through the canister. The matrix offers only a small hydraulic resistance to the feed flow, occupying less than 5% of the canister volume (95% void volume). When the matrix has become loaded with magnetic particles, the particles are easily washed from the matrix by reducing the magnetic field to zero and opening valves and backflushing. High gradient magnetic separation may also be used to remove nonmagnetic contaminants from water. This is accomplished by binding finely divided magnetic seed particles, such as magnetic iron oxide (magnetite), to the nonmagnetic contaminants, thus creating a "magnetic handle" ("indirect filtration" or "seeded water treatment"). Binding of the magnetic seed is accomplished in two general ways: (1) adsorption of the contaminant to magnetic seed and (2) chemical coagulation (alum). Particles ranging in size from soluble through settleable (0.001 μ) may be removed with this process. Design parameters for HGMS are presented in Table 31.

Magnetic separation can provide the rapid filtration of many pollutants from water, with a small expenditure of energy. Removal is much more efficient than with sedimentation because the magnetic forces on fine particles may be many times greater than gravitational forces. To date, only bench scale tests and a pilot-plant scale system of 1–4/min (0.26–1.06 gal/min) have been operated. Typical pollutant removals are shown in Tables 32, 33 and 34.

Costs of HGMS have been evaluated for a 94,625-m^3/d (25-Mgal/d) facility and are summarized in Table 35. Capital costs include pre-treatment, chemical addition, thickening and dewatering equipment, pumps, backflush system, instrumentation, and disinfection system. Operation and maintenance costs include chemicals, labor, electrical utilities, and maintenance.

Powdered Activated Carbon-Alum Coagulation

Several combined sewage treatment demonstration projects have evaluated the benefits of chemical aids to process operations, but only one pilot operation representing a complete physical–chemical treatment system has been implemented. It was demonstrated at a 379-m^3/d (100,000-gal/d) pilot unit in Albany, New York. In this project, raw municipal sewage and CSO were mixed with powdered activated carbon to remove dissolved organics. Alum was then added to aid in subsequent clarification. Addition of polyelectrolytes was followed by a short flocculation period. Solids were separated from the liquid stream by gravity settling, and the effluent

TABLE 30

Summary of capital, operation and maintenance costs for biological treatment alternatives (ENR = 5,000)

Project location	Type of biological treatment	Peak plant capacity (Mgal/d)	Construction cost ($1000)	Cost/ capacity ($/Mgal/d)	Cost/ tributary area ($/acre)	Annual O&M ($/1,000 gal. except as noted)
Kenosha, WI	Contact stabilization	20	3410	170,000	2900	34.5
Milwaukee, WI*	Rotating biological contactor	4.3	747	173,000	21,400	11.0
Mount Clemens, MI						
Demonstration system	Aerated treatment lagoons	64	1,607	25,000	7,600	50.0
Citywide system	Storage/aerated treatment lagoons	260	14,343	55,000	9,800	47.5
New Providence, NJ†	High-rate trickling filter	6	1,188	198,000	—	30.7
Shelbyville, IL						
Southeast site	Oxidation lagoon	28	108	3,900	2,790	3,820/yr‡
Southwest site	Storage and facultative lagoons	110	844	7,700	1,900	14,460/yr‡
Springfield, IL	Oxidation lagoon	67	440	6,500	200	5,260/yr

* Includes estimate of final clarifier.

† Includes plastic media trickling filter, final clarifier, plus one-half of other costs.

‡ Based on estimated man-day labor requirements.

TABLE 31
Preliminary design parameters for high gradient magnetic separators

Magnetic field strength, kG*	0.5–1.5
Maximum flux rate, gal/ft^2/min	100
Minimum detention time, min	3
Matrix loading, g solids/g of matrix fiber	0.1–0.5
Magnetic addition, mg/l	100–500
Magnetic to suspended solids ratio	0.4–3.0
Alum addition, mg/l	
Range	90–120
Average	100
Polyelectrolyte addition, mg/l	0.5–1.0

* Kilogauss.

TABLE 32
Removal of solids by HGMS for CSO and Raw Sewage Samples

	Removal %	
Solids parameter	CSO	Raw sewage
SS	95	91
Settleable Solids	95+	99+
Apparent Color, PCU	87	82
Turbidity, NTU	93	88

TABLE 33
Removal of biological and chemical constituents by HGMS

Pollutant parameter	Average removal %
BOD_5	92
COD	74
Total Coliforms on EMB Agar at 37°C	99.3
Fecal Coliforms on EMB Agar at 37°C	99.2
Algae	99.9
Virus, Bacteriphage T_7	100
Virus, Polio	99–100

(EPA-600/8-77-014).

was then disinfected and discharged, or filtered (tri-media), then disinfected prior to discharge. Carbon regeneration in a fluidized bed furnace and alum recovery from the calcined sludge were also demonstrated, as was reuse of the reclaimed chemicals. Average carbon losses per regeneration cycle were 9.7%. Average removals in excess of 94% COD, 94% BOD, and 99% SS were consistently achieved (without filtration) in treating combined sewage. Representative capital and O&M costs for a physical-chemical treatment plant designed for raw stormwater treatment, projected from data developed during the Albany project, are summarized in Table 36.

DISINFECTION

Conventional municipal sewage disinfection generally involves the use of chlorine gas or sodium hypochlorite as the disinfectant. To be effective for disinfection purposes, a contact time of not less than 15 min at peak flowrate and a chlorine residual of 0.2–2.0 mg/l are commonly recommended. Disinfection of CSO is generally practiced at treatment facilities to control the discharge of pathogens and other microorganisms in receiving waters. However, an approach other than that used for the conventional municipal sewage is required, mainly because such flows have characteristics of intermittency, high flowrate, high SS content, wide temperature variation, and variable bacterial quality. Several other aspects of disinfection practices require consideration for CSO treatment applications:

1) A residual disinfecting capability may not be feasible for CSO (and all wastewater) discharges. Chlorine residuals and compounds discharged to natural waters may be harmful to aquatic life.
2) The coliform count is increased by surface runoff in quantities unrelated to pathogenic organism concentration. Total coliform or fecal coliform levels may not be the most useful indication of disinfection requirements and efficiencies.
3) Discharge points requiring disinfection are often at outlying points on the sewer system and require unmanned, automated installations.

The disinfectant used at a facility for treatment of CSO should be adaptable to intermittent use. Other considerations include the disinfection effectiveness and the safety and ease

TABLE 34
Removal of heavy metals by HGMS

	Heavy metal constituent						
	Cadmium	Chromium	Copper	Mercury	Nickel	Lead	Zinc
Average removal, %	43	41	53	71	0–67	0–67	84

(EPA-600/8–77–014).

of feeding. Table 37 shows disinfectants that might be used for storm disinfection. Chlorine and hypochlorite will react with ammonia to form chloramines and with phenols to form chlorophenols.

These are toxic to aquatic life and the latter also produce taste and odor in the water. Chlorine dioxide does not react with ammonia and completely oxidizes phenols. Ozone also is effective in oxidizing phenols. Other disinfectants that could be considered for future applications are bromine and ultraviolet light. However, certain factors limit their application at this time. While bromine can be shown to be a more effective disinfectant (Alleman, 1988), it may pose a greater toxicity concern to the aquatic system (White, 1986). Ultraviolet disinfection represents an appealing simplicity for remote locations (no disinfectant storage requirements and on-off control), but the high degree of solids removal and capital costs remain as impediments to its application to CSOs.

High-rate disinfection refers to achieving either a given percent or a given bacterial count reduction through the use of (1) decreased disinfectant contact time, (2) increased mixing intensity, (3) increased disinfectant concentration, (4) chemicals having higher oxidizing rates, or (5) various combinations of these. Where contact times are less than 10 min, (usually in the range of 1–5 min), adequate mixing is a critical parameter, providing complete dispersion of the disinfectant and forcing disinfectant contact with the maximum number of micro-organisms. The more physical collisions high-intensity mixing causes, the lower the contact time requirements.

Mixing can be accomplished by mechanical flash mixers at the point of disinfectant addition and at intermittent points,

TABLE 35

Construction and O&M costs for a 25-Mgal/d HGMS facility (ENR = 5,000)

Construction Cost	
Total, $	5,282,000
$/Mgal/d	211,000
O&M Cost	
$/yr	1,360,000
$1,000 gal treated	0.30

TABLE 36

Estimated capital and O&M costs for a physical-chemical treatment plant (ENR = 5,000)

Capital costs*				O&M costs[†] (cents/1,000 gal)			
1 Mgal/d	10 Mgal/d	25 Mgal/d	100 Mgal/d	1 Mgal/d	10 Mgal/d	25 Mgal/d	100 Mgal/d
448,000	447,800	9,110,000	26,675,000	48.0	47.1	39.0	29.3

* Capital costs include screens, grit chambers, overflow facilities, pipe reactor vessels, pumps, chemical storage, carbon slurry tanks, sludge storage, agitators, flocculators, tube settlers, filtration, chlorination, carbon regeneration/sludge incineration, fluidized bed furnace, chemical make-up system, 10% contingencies, and land.

[†] O&M costs include all materials, power, and labor. (Plant is designed for raw stormwater treatment.)

TABLE 37

Characteristics of principal stormwater disinfection agents.

Characteristics	Chlorine	Hypochlorite	Dioxide	Ozone
Stability	Stable	6-month half-life	Unstable	Unstable
Reacts with ammonia to form chloramines	Yes	Yes	No	No
Destroys phenols	At high concentrations	At high concentrations	Yes	Yes
Produces a residual	Yes	Yes	Short-lived*	No
Affected by PH	More effective	More effective	Slightly	No
Hazards	Toxic	Slight	Toxic; explosive	Toxic

* Chlorine dioxide dissociates rapidly.

(EPA-600/8-77-014)

or by specially designed plug flow contact chambers containing closely spaced, corrugated parallel baffles that create a meandering path for the wastewater.

High-rate disinfection was shown to be enhanced beyond the expected additive effect by sequential addition of Cl_2, followed by ClO_2 at intervals of 15–30s (EPA-670–2-75–021; EPA-600/2-76-244). A minimum effective combination of 8 mg/l of Cl_2 followed by 2 mg/l of ClO_2 was found as effective in reducing total and fecal coliforms, fecal streptococci, and viruses to acceptable target levels as adding 25 mg/l Cl_2 or 12 mg/l ClO_2 individually. It was surmised that the presence of free Cl_2 in solution with chlorite ions (ClO_2^- [the reduced state of ClO_2]) may cause the oxidation of ClO_2^- back to its original state. This process would prolong the existence of ClO_2, the more potent disinfectant.

Ozone has a more rapid disinfecting rate than chlorine and also has the further advantage of supplying additional oxygen to the wastewater. The increased disinfecting rate of ozone requires shorter contact times and results in a lower capital cost for a contactor, as compared to that for a chlorine contact tank. Ozone does not produce chlorinated hydrocarbons or a long-lasting residual as chlorine does, but it is unstable and must be generated on-site just prior to application. Thus, unlike chlorine, no storage is required. In tests on CSO in Philadelphia (see "microstraining and disinfection" reports listed in the references at the end of this section),

TABLE 38

Cost data on chlorine gas and hypochlorite disinfection (ENR = 5,000)

Location, agent, source	Capital cost ($)	Operating cost ($/yr)	Cost/lb available chlorine ($)
Akron, OH*			
Sodium hypochlorite Purchased	1,104,000	58,200	0.37–0.67
Cambridge and Somerville, MA			
Sodium hypochlorite Purchased	—	—	0.97
On-site generation	—	—	0.51
New Orleans, LA[†]			
Sodium hypochlorite			
On-site generation	1,448,000	724,000	0.30
Saginaw, MI[‡]			
Chlorine gas	406,000	5,800	0.87
Sodium hypochlorite			
Purchased	49,000	16,000–30,000	0.45–0.78
On-site generation	239,000–403,000	11,700–13,000	0.69–1.0
South Essex Sewage District, MA[§]			
Chlorine gas	1,667,000	582,000	0.09
Sodium hypochlorite			
Purchased	1,054,000	913,000	0.12
On site generation Seawater	4,200,000	406,000	0.09
Brine	4,200,000	758,000	0.13

* CSO disinfection.

[†] Storm sewer discharge disinfection.

[‡] CSO disinfection at use rate of 42,000 lb/yr of chlorine.

[§] Sewage treatment plant effluent disinfection at use rate of 24,000 lb/day of chlorine.

TABLE 39

Comparison of estimated capital costs for three different disinfection methods (ENR = 5,000)

Disinfection method	Capital cost ($/mgd)
2-min ozone contact (chamber with once-through oxygen-fed ozone generator)	32,530
2-min chlorine contact (chamber with hypochlorite feeder)	3,800
5–10-min conventional chlorine contact	4,230

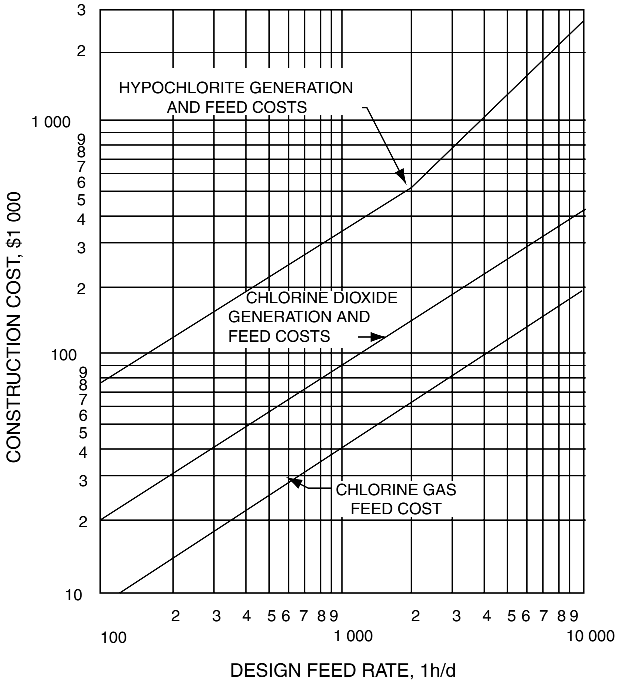


FIGURE 15 Chlorine disinfection feed facilities cost curves: ENR = 2,000.

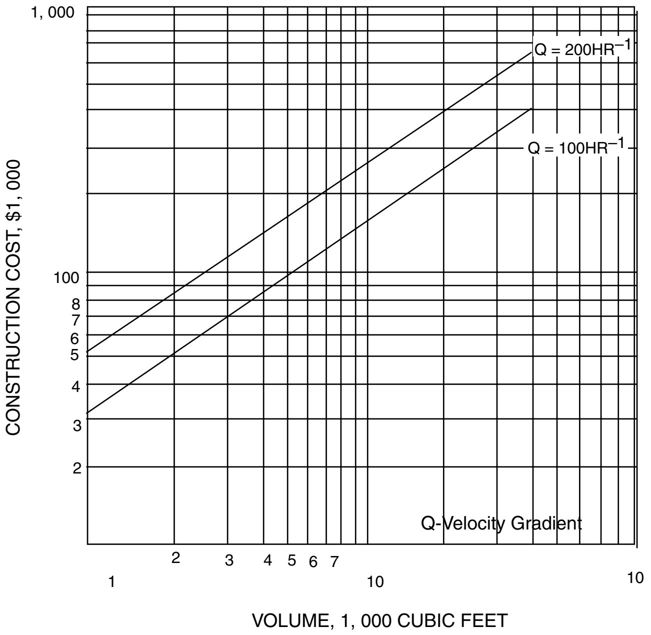


FIGURE 16 High-intensity mixing/chlorine contact basin cost: ENR = 5,000 (EPA-600/2–76–286).

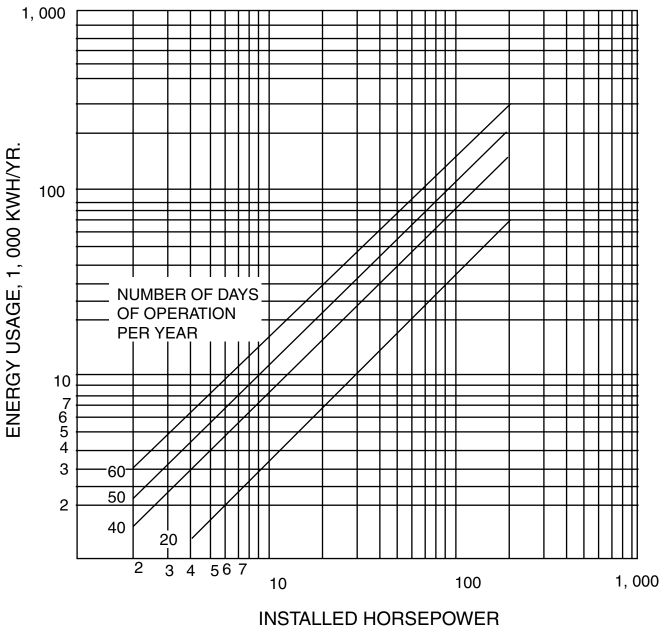


FIGURE 17 High-intensity mixing (chlorine contact) power requirements.

equivalent disinfection was obtained using either 3.8 mg/l of ozone or 5 mg/l of chlorine.

Because of the characteristic intermittent operation associated with treatment of CSO, reduction of construction cost with a potential increase in operating costs often results in overall minimum costs. In the case of chlorination facilities, as applied to treatment of CSO, the construction costs associated with contact basins having conventional contact time of 15–30 min are high and difficult to justify. Therefore, consideration should be given to higher mixing intensities, to make better usage of the chlorine and/or higher chlorine dosages and smaller, shorter detention-time contact basins to effect the same end results. Disinfectant costs for CSO treatment are higher than those for sewage treatment. This is the result of smaller total annual disinfectant volume requirements, increased disinfectant concentration requirements, and higher unit O&M costs for CSO treatment facilities. These costs could be reduced by using the facilities in conjunction with dry-weather flow treatment plants, whenever possible. Curves comparing generation and feed costs for chlorine gas, chlorine dioxide, and hypochlorite generation disinfection systems for CSO have been developed and are presented in Figure 15. These costs include manufactured equipment, labor, piping, housing, electrical and instrumentation, and miscellaneous items; no allowance for land was included. Capital and operating costs for several CSO and stormwater disinfection facilities are presented in Table 38.

As previously mentioned, conventionally long contact times may not be economical. Short-term contact times with more intense mixing, using a basin and mixer similar to those used in coagulant mixing, can effect the same disinfection results. Construction cost curves for high-intensity mixing/chlorine contact basins are presented in Figure 16. Power requirement curves, for high-intensity mixing, are presented in Figure 17.

The capital costs for different disinfection agents and methods resulting from the Philadelphia study are shown in Table 39. The capital costs for ozone generation are usually the highest of the most commonly used processes. Ozone operation costs are very dependent on the cost of electricity and the source of the ozone (air or pure oxygen).

REFERENCES

SECTION 1

EPA-600/2-75-004, "Contributions of Urban Roadway Usage to Water Pollution," D.G. Shaheen, Biospherics Inc., Rockville, MD, March 1975 (NTIS PB 245 854).

EPA-600/2-79-161, "Demonstration of Nonpoint Pollution Abatement Through Improved Street Cleaning Practices," R.E. Pitt, Wood-Ward-Clyde Consultants, San Francisco, CA, August 1979 (NTIS PB 80–108988).

EPA-600/8-80-035, "Urban Stormwater Management and Technology: Case Histories," W.G. Lynbard, *et al.,* Metcalf & Eddy, Inc., Palo Alto, CA, August 1980 (NTIS PB 81 107153).

EPA-R2-72-81, "Water Pollution Aspects of Street Surface Contaminants," J.D. Sartor and G.B. Boyd, URS Research Co., San Mateo, CA, November 1972 (NTIS PB 214 408).

EPA-R2-73283, "Toxic Materials Analysis of Street Surface Contaminants," R.E. Pitt and G. Amy, URS Research Co., San Mateo, CA, August 1973 (NTIS PB 224 677).

SECTION 2

11010DNO03/72, “A Flushing System for Combined Sewer Cleansing,” Central Engineering Laboratories, FMC Corp., Santa Clara, CA, March 1972 (NTIS PB 210 858).

11020DNO06/67, “Feasibility of a Periodic Flushing System for Combined Sewer Cleansing,” FMC Corp., Santa Clara, CA, August 1967 (NTIS PB 195 223).

EPA-600/2-83-043, “Evaluation of Catchbasin Performance for Urban Stormwater Pollution Control,” G.L. Aronson, *et al.,* Environmental Design and Planning, Inc. Boston, MA, 1983 (NTIS PB 83-217745).

EPA-600/S2-83-043, project summary of above report.

EPA-600/2-75-062, “The Helical Bend Combined Sewer Overflow Regulator,” R.H. Sullivan, *et al.,* American Public Works Assoc., Chicago, IL, December 1975 (NTIS PB 250 619).

EPA-600/2-77-120, “Procedures for Estimating Dry-Weather Pollutant Deposition in Sewerage Systems,” W. Pisano and C.S. Queriroz, Energy and Energy Environmental Analysis, Inc., Boston, MA, July 1977 (NTIS PB 270 695).

EPA-600/2-79-133, “Dry-Weather Deposition and Flushing for Combined Sewer Overflow Pollution Control,” W. Pisano, *et al.,* Boston, MA, August 1979 (NTIS PB 80-118524).

EPA-600/2-79-134, “Disinfection/Treatment of Combined Sewer Overflows, Syracuse, New York,” F. Drehwing, *et al.,* O’Brien & Gere Engineers, Inc., Syracuse, NY, August 1979 (NTIS PB 80-113459).

EPA-600/2-80-118, “Review of Alternatives for Evaluation of Sewer Flushing-Dorchester Area-Boston,” H.L. Kaufman and F. Lai, Clinton Bogert Associates, Ft. Lee, NJ, August 1980 (NTIS PS 81-111850).

EPA-625/2-77-012, “Swirl Device for Regulating and Treating Combined Sewer Overflows,” EPA Technology Transfer Capsule Report, prepared by R. Field and H. Masters, USEPA, Edison, NJ, 1977 (ERIC 2012, Cincinnati).

EPA-670/2-73-059, “The Dual-Functioning Swirl Combined Sewer Overflow Regulator/Concentrator,” R. Field, USEPA, Edison, NJ, September 1973 (NTIS PB 227 182/3).

EPA-670/2-74-039, “Relationship between Diameter and Height for Design of a Swirl Concentrator as a Combined Sewer Overflow Regulator,” R.H. Sullivan, *et al.,* American Public Works Assoc., Chicago, IL, July 1974 (NTIS PB 234-646).

EPA-R2-72-008, “The Swirl Concentrator as a Combined Sewer Overflow Regulator Facility,” R. Sullivan, American Public Works Asso., Chicago, IL, September 1972 (NTIS PB 214 687).

SECTION 3

11020FAQ03/71, “Dispatching Systems for Control of Combined Sewer Losses,” Metropolitan Sewer Board, St. Paul, MN, March 1971 (NTIS PB 203 678).

11022ELK12/71, “Maximizing Storage in Combined Sewer Systems, Metropolitan Seattle, WA, December 1971.

11023-80/70, “Retention Basin Control of Combined Sewer Overflows,” Springfield Sanitary District, Springfield, IL, August 1970 (NTIS PB 200 828).

EPA-670/2-74-075, “Surge Facility for Wet- and Dry-Weather Flow Control,” H.L. Welborn, Rohnert Park, CA, November 1974 (NTIS PB 238 905).

EPA-600/2-75-071, “Detention Tank for Combined Sewer Overflow—Milwaukee, WI, Demonstration Project,” Consoer, Townsend and Assoc., Chicago, IL, December 1975 (NTIS PB 250 427).

EPA-600/2-76-222a, “Wastewater Management Program, Jamaica Bay—Vol. I, Summary Report,” D.L. Feurstein and W.O. Maddaus, September 1976 (NTIS PB 260 887).

EPA-600/2-76-222b, “Wastewater Management Program, Jamaica Bay—Vol. II, Supplemental Data, NYC Spring Creek,” D.L. Feurstein and W.O. Maddaus, 1976 (NTIS PB 258 308).

EPA-600/2-77-046, “Cottage Farm Combined Sewer Detention and Chlorinaton Station,” Cambridge, MA, Metropolitan District Commission, November 1977 (NTIS PB 263 292).

EPA-670/2-74-022, “Computer Management of a Combined Sewer System” C.P. Leiser Metropolitan Seattle, WA, July 1974 (NTIS PB 235 717).

EPA-670/2-75-010, “Multi-Purpose Combined Sewer Overflow Treatment Facility, Mount Clemens, Michigan,” V.U. Mahida, *et al.,* Spalding, DeDecker & Assoc., Inc., Madison Heights, MI, May 1975 (NTIS PB 242 914).

EPA-670/2-75-020, “Sewerage System Monitoring and Remote Control,” T.R. Watt, *et al.,* Detroit Metro Water Dept, Detroit, MI, May 1975 (NTIS PB 242 107).

SECTION 4

11020EXV07/69, “Strainer/Filter Treatment of Combined Sewer Overflows,” Fram Corp., East Providence, RI, July 1969 (NTIS PB 185 949).

11020FDC01/72, “Screening/Flotation Treatment of Combined Sewer Overflows,” Ecology Division, Rex Chainbelt, Inc., January 1972 (No NTIS number).

11023EVO06/70, “Microstraining and Disinfection of Combined Sewer Overflows,” Cochrane Div., Crane Co., King of Prussia, PA, June 1970 (NTIS PB 195 674/BA).

110EY104/72, “High-Rate Filtration of Combined Sewer Overflows” (Cleveland), R. Nebolsine, *et al.,* Hydrotechnic Corp., New York, April 1972 (NTIS PB 211 144).

11023FDD03/70, “Rotary Vibratory Fine Screening of Combined Sewer Overflows,” Cornell, Howland, Hayes, and Merryfield, Corvallis, OR, March 1970 (NTIS PB 195 168).

11023FDD07/71, “Demonstration of Rotary Screening for Combined Sewer Overflows,” Dept of Public Works, Portland, OR, July 1971 (NTIS PB 206 814).

EPA-600/2-75-033, “Treatment of Combined Sewer Overflows by Dissolved Air Flotation,” T.A. Bursztynsky, *et al.,* Engineering Science, Inc., Berkeley, CA, September 1975 (NTIS PB 248 186).

EPA-600/2-77-069a, “Screening/Flotation Treatment of Combined Sewer Overflows—Vol. I; Bench-Scale and Pilot Plant Investigations,” M.K. Gupta, *et al.,* Envirex Environmental Science Div., Milwaukee, WI, August 1977 (NTIS PB 272 834).

EPA-600/2-79-015, “Dual Process High-Rate Filtration of Raw Sanitary Sewage and Combined Sewer Overflows,” H. Innerfield, *et al.,* NYC Dept of Water Resources, New York, April 1979 (NTIS PB 296 626/AS).

EPA-600/2-79-031b, “Combined Sewer Overflow Abatement Program, Rochester, NY—Vol. II, Pilot Plant Evaluations,” F.J. Drehwing, *et al.,* O’Brien & Gere Engineers, Inc., Syracuse, NY, July 1979 (NYTIS PB 80-159262).

EPA-600/2-79-085, “Combined Sewer Overflow Treatment by Screening and Terminal Ponding, Fort Wayne, IN,” D.H. Prah and P.T. Brunner, Fort Wayne, IN, August 1979 (NTIS PB 80-119399).

EPA-600/2-79-106a, “Screening/Flotation Treatment of Combined Sewer Overflows—Vol. II; Full-Scale Operation, Racine, WI,” T.L. Meinholz, Envirex, Inc., Milwaukee, WI, August 1979 (NTIS PB 80-130693).

EPA-600/8-77-014, “Urban Stormwater Management and Technology, Update and User’s Guide.”

EPA-670/2-74-049, “Microstraining and Disinfection of Combined Sewer Overflows—Phase III, M.B. Maher, Crane Co., King of Prussia, PA, August 1974 (NTIS PB 235 771).

EPA-R2-73-124, “Microstraining and Disinfection of Combined Sewer Overflows—Phase II,” G.E. Glover and G.R. Herbert, Crane Co., King of Prussia, PA, January 1973 (NTIS PB 219 879).

EPA-R2-73-222, “Ultra High-Rate Filtration of Activated Sludge Plant Effluent,” R. Nebolsine, *et al.,* Hydrotechnic Corp., (NTIS ONLY).

SECTION 5

11023-08/70, “Retention Basin Control of Combined Sewer Overflows,” Springfield Sanitary District, Springfield, IL, August 1970 (NTIS PB 200 828).

EPA-600/2-81-196 “Evaluatin of a Treatment Lagoon for Combined Sewer Overflow,” D.J. Connick, *et al.,* Environmental Design and Planning, Inc., 1981 (NTIS PB 82-105214), EPA-600/52-81-196.

EPA-670/2-73-071, "Utilization of Trickling Filters for Dual-Treatment of Dry- and Wet-Weather Flows," P. Homack, *et al.,* E.T. Killam Assoc., Inc., Milburn, NJ, September 1973 (NTIS PB 231 251).

EPA-670/2-74-050, "Combined Sewer Overflow Treatment by the Rotating Biological Contactor Process," F. L. Welsh and D.J. Stucky, Autotrol Corp., Milwaukee, WI, June 1974 (NTIS PB 231 892).

EPA-670/2-74-079, "An Evaluation of Three Combined Sewer Overflow Treatment Alternatives (Shelbyville, IL)," J.W. Parks, *et al.,* Urbana, IL, August 1974 (NTIS PB 239 115).

EPA-670/2-75-010 "Multi-Purpose Combined Sewer Overflow Treatment Facility, Mount Clemens, Michigan," V. U. Mahidal, *et al.,* Spalding, DeDecker & Assoc., Inc., Madison Heights, MI, May 1975 (NTIS PB 242 914).

EPA-670/2-75-019, "Biological Treatment of Combined Sewer Outflow at Kenosha, WI," R.W. Agnew, *et al.,* Envirex, Milwaukee, WI, April 1975 (NTIS PB 242 126).

EPA-R2-73-149, "Physical-Chemical Treatment of Combined and Municipal Sewage," A.J. Shuckrow, *et al.,* Pacific NW Lab, Battelle Memorial Institute, Richland, WA, February 1973 (NTIS PB 219 668).

SECTION 6

11023EVO06/70, "Microstraining and Disinfection of Combined Sewer Overflows," Cochrane Div., Crane Co., King of Prussia, PA, June 1970 (NTIS PB 195 67).

Alleman, J.E., Etzel, J.E., Gendron, D.E., Conley, J.C., MaCoy, W.F., and Hein A.J. Bromine-Based Disinfection Performance, presented at the 61st Annual Conference of the Water Pollution Control Federation, Dallas, TX, October 1986.

EPA-600/2-76-244, "Proceedings of Workshop on Microorganisms in Urban Stormwater," R. Field, *et al.,* Edison, NJ, (March 24, 1975), November 1976 (NTIS PB 263 030).

EPA-600/2-79-031b, "Combined Sewer Overflow Abatement Program, Rochester, NY—Vol. II, Pilot Plant Evaluations," F., J. Drehwing, *et al.,* O'Brien & Gere Engineers, Inc., Syracuse, NY, July 1979 (NTIS PB 80-159262).

EPA-600/2-79-134, "Disinfection/Treatment of Combined Sewer Overflows, Syracuse, NY," F.J. Drehwing *et al.,* O'Brien & Gere Engineers, Inc., Syracuse, NY, August 1979 (NTIS PB 80-113459).

EPA-670/2-73-077, "Combined Sewer Overflow Seminar Papers," November 1973, USEPA and NYS-DEC (NTIS PB 231 836).

EPA-670/2-74-049, "Microstraining and Disinfection of Combined Sewer Overflows—Phase III," M.B. Maher, Crane Co., King of Prussia, PA, August 1974 (NTIS PB 235 771).

EPA-670/2-75-021, "Bench-Scale High-Rate Disinfection of Combined Sewer Overflows with Chlorine and Chlorine Dioxide," P.E. Moffa, *et al.,* O'Brien & Gere Engineers, Inc., Syracuse, NY, April 1985 (NTIS PB 242 296).

EPA R2-73-124, "Microstraining and Disinfection of Combined Sewer Overflows—Phase II," G.E. Glover and G.R. Herbert, Crane Co., King of Prussia, PA, January 1973 (NTIS PB 219 879).

White, G.C. "Handbook of Chlorination," 2nd ed., 1986, Van Nostrand Reinhold, NY.

EPA-600/2-77-120, "Microorganisms in Urban Stormwater," V.P. Olivieri, *et al.,* The Johns Hopkins University, Baltimore, MD, 1977 (NTIS PHB 272).

Wet-Weather Flow in the Urban Watershed: Technology and Management. Richard Field and Daniel A. Sullivan, Eds. Lewis Publishers. Boca Raton, Fl. 2003.

RICHARD FIELD
Environmental Protection Agency

V

VAPOR AND GASEOUS POLLUTANT FUNDAMENTALS

The toxic gases produced during combustion and other chemical processes may be removed by destructive disposal, dispersive dilution or as recoverable side products. The removal path chosen is at the present time motivated primarily by economics, but public pressure and awareness of environmental problems also influence the choice.

This section will concern itself with destructive disposal and/or various recovery processes, the subject of dispersion being ably handled in the sections on *Air Pollution Meteorology and Urban Air Pollution Modeling.* The main emphasis will be on principles of gaseous reaction and removal with the description of equipment for air pollution abatement covered by pollutant. Problems specifically concerned with the automobile can be found under *Mobile Source Pollution.* Although the control principles to be described below are general, it is usually necessary to design equipment for each installation because of variations in physical and chemical properties of effluents; also, in general, the cost of adding pollution devices to an existing unit (retrofit) will be higher than if they were placed in the original design, because of construction difficulty and downtime.

Although the majority of effluent material from combustion occurs in the gaseous state, it is important to characterize the total effluent stream for control purposes. For example, the effluent may be condensible at operating temperature (a vapour) or noncondensible (a gas), but it usually is a mixture of the two. Particulate matter (solids) and mists (liquids) are often suspended in the gaseous stream; if the particles do not separate upon settling they are called aerosols. The considerations in this section deal with gas or vapor removal only and not with liquid or solid particle removal.

SULFUR DIOXIDE, SO_2, AND TRIOXIDE, SO_3

Sulfur dioxide is generated during combustion of any sulfur-containing fuel and is emitted by industrial processes that use sulfuric acid or consume sulfur-containing raw material. The major industrial sources of SO_2 are sulfuric acid plants, smelting of metallic ores, paper mills, and refining of oil. Fuel combustion accounts for roughly 75% of the total SO_2 emitted. Associated with utility growth is the continued long term increase in utility coal consumption from some 650 million tons/year in 1975 to between 1400 and 1800 million tons/year in 1990. Also the utility industry is increasingly converting to coal. Under the current performance standards for power plants, national SO_2 emissions are projected to increase approximately 15 to 16% between 1975 and 1990 (Anon. 1978). The SO_2 emitted from power plants is usually at low concentration ($<0.5\%$ by volume). However, a 900 MW unit will emit over 13,000 pounds of SO_2 per hour for a 1% sulfur coal. The SO_2 emitted from industrial processes is at higher concentrations and lower flow rates. The emitted SO_2 combines readily with mists and aerosols, thus compounding the removal problem.

Information concerning emissions standards is essential to pollution control engineering design. The current US federal SO_2 emissions limits for a stack are 1.2 lb/10^6 BTU for new oil and gas fired plants. Also, uncontrolled SO_2 emissions from new plants firing solid, liquid, and gaseous fuels are required to be reduced by 85%. The percent reduction requirement does not apply if SO_2 emissions into the atmosphere are less than 0.2 lb/10^6 BTU.

Flue gas desulfurization (FGD) methods are categorized as nonregenerable and regenerable. Nonregenerable processes produce a sludge that consists of fly ash, water, calcium sulfate and calcium sulfite. In regenerable processes, SO_2 is recovered and converted into marketable by-products such as elemental sulfur, sulfuric acid or concentrated SO_2. The sorbent is regenerated and recycled. The US Environmental Protection Agency believes the following types of FGD systems are capable of achieving the emissions limit standards: lime, limestone, Wellman-Lord, magnesium oxide and double alkali. Due to the process economics, utility industry prefer the lime/limestone systems. Limestone processes constitute about 58% of the current calcium-based capacity in service and under construction, and 69% of that planned, which amounts to 63% of the total (De Vitt *et al.*, 1980). The following reactions take place in the limestone systems:

$$\begin{aligned}
&\mathrm{CaCO_3(s)} \rightleftharpoons \mathrm{CaCO_3(aq)}\\
&2\,\mathrm{CaCO_3(aq)} + 2\mathrm{H^+} \rightleftharpoons \mathrm{CaCHCO_3} + \mathrm{Ca^{+2}}\\
&\mathrm{SO_2(g)} \rightleftharpoons \mathrm{SO_2(aq)}\\
&\mathrm{SO_2(aq)} + \mathrm{H_2O} \rightarrow \mathrm{H_2SO_3} \rightleftharpoons \mathrm{HSO_3^-} + \mathrm{H^+}\\
&\mathrm{HSO_3^-} + \tfrac{1}{2}\mathrm{O_2(aq)} \rightarrow \mathrm{SO_4^{-2}} + \mathrm{H^+}\\
&\mathrm{Ca^{+2}} + \mathrm{SO_4^{-2}} \rightleftharpoons \mathrm{CaSO_4(aq)} \rightarrow \mathrm{CaSO_4(s)} \downarrow
\end{aligned}$$

Or

$$\begin{aligned}
&\mathrm{Ca^{+2}} + 2\mathrm{Cl^-} \rightleftharpoons \mathrm{CaCl_2(aq)} \rightarrow \mathrm{CaCl_2(s)} \downarrow\\
&2\mathrm{H^+} + \mathrm{SO_4^{-2}} \rightleftharpoons \mathrm{H_2SO_4}\\
&\mathrm{H^+} + \mathrm{Cl^-} \rightleftharpoons \mathrm{HCl}\\
&\mathrm{Ca(HCO_3)_2} + 2\mathrm{H^+} \rightleftharpoons \mathrm{Ca^{+2}} + 2\mathrm{H_2CO_3}\\
&\mathrm{H_2CO_3} \rightleftharpoons \mathrm{CO_2} + \mathrm{H_2O}
\end{aligned}$$

The dissolution rate of limestone depends on the pH values. The pH values encountered in practical operations of limestone systems is in most cases between 5.5 and 6.5. If the limestone systems operate at too low pH, SO_2 removal efficiency will decrease. At too high pH, the scale formation will be promoted. Other factors affecting the performance of limestone systems include solids content, liquid-gas ratio, and corrosion. A discussion can be found elsewhere (De Vitt *et al.*, 1980).

In selecting the FGD processes, the following should be considered:

1) process type: wet or dry, regenerable or non-regenerable.
2) chemical reagent used.
3) end-product produced: saleable product or disposable waste.

The reagent, end product, principle of operation, and SO_2 removal efficiency of major FGD processes are shown in Table 1 (Princiotta, 1978). For details of SO_2 removal refer to the *Stack gas cleaning* sections.

Sulfur dioxide reacts slowly with a large excess of oxygen in the presence of sunlight to form trioxide. Gerhard (1956) showed that the process occurs with O_2 at a rate of 0.1–0.2% per hour and Cadle (1956) with ozone, O_3, at <0.1% per day. Niepenburg (1966) illustrates the effects of oxygen in the waste gas during combustion of oil.

The conversion of SO_2 to SO_3 is believed to be possible at realistic rates because of the presence, on diverse surfaces, of Fe_2O_3 which acts as a catalyst. The SO_3 has a short lifetime since it readily combines with water vapor in the atmosphere to form sulfuric acid.

Oxides of Nitrogen, NO_x

NO_x is produced in all combustions which take place using air as an oxygen supply and in those chemical industries employing nitric acid. More than 55% of the total NO_x emissions of 20 million tons originate from stationary sources as shown in Figure 1, and 93% of all stationary source NO_x emissions are from combustion of fossil fuels for utilities. Direct industry-related emissions account for only 5% of the stationary source total. Approximately 30% of all stationary source NO_x is emitted by coal-fired utility boilers. Uncontrolled NO_x emissions from coal-fired sources have been measured in the range 0.53 to 2.04 lb/10^6 BTU at full load (Ziegler and Meyer, 1979). The NO_x formed in combustion is from fixation of atmospheric nitrogen and/or fuel nitrogen. Ermenc (1956) found that at high temperature nitrogen and oxygen combine to form both NO and NO_2. The yield of NO increases from 0.26% at 2800°F to 1.75% at 3800°F. If the temperature is reduced slowly the reverse reaction will take place, but if the products are quenched by rapid heat exchange, the reverse reaction rate becomes small and the oxides remain in the exhaust stream. The oxide NO can usually be oxidized to form NO_2 according to:

$$2\mathrm{NO} + \mathrm{O_2} \rightarrow 2\mathrm{NO_2}.$$

Because this is a tri-molecular gas phase reaction, the concentration of NO and NO_2 tremendously affects the rate at which the oxidation takes place. At low concentration, for example 1–5 ppm in air, the reaction is so slow that it would be negligible except for the photochemical reactions which take place in the presence of sunlight. The dioxide also reacts with oxygen to form ozone. The existence of nitrogen trioxide at low concentration in polluted atmospheres is postulated (Hanst, 1971) to form by the reaction with ozone.

$$\mathrm{NO_2} + \mathrm{O_3} \rightleftharpoons \mathrm{NO_3} + \mathrm{O_2}$$

and remain in equilibrium with N_2O_5

$$\mathrm{N_2O_5} \rightleftharpoons \mathrm{NO_2} + \mathrm{NO_3}.$$

The NO_3 formation reaction only takes place after NO is substantially depleted in the atmosphere and O_3 begins to appear. Without more stringent control of new sources, the NO_x emissions by 1995 are projected to be 66% higher than than in 1985. Even with application of the best control method to all new sources there is still a projected 24% increase over 10 year emissions (McCutchen, 1977). The typical NO_x emission from nitric acid plants is 1000–3000 ppm. The federal standard for new nitric acid plants is 3 lb NO_x/ton 100% HNO_3—that is, about 200 ppm (Ricci, 1977). The most widely used process for nitric acid plant tailgas cleanup is catalytic decomposition of NO_x to nitrogen and oxygen.

The current and projected values of the New Source Performance Standards (NSPS) for NO_x are discussed later in this article. During recent years N_2O formation rates have been the subject of controversy, especially in fuel NO_x mechanisms.

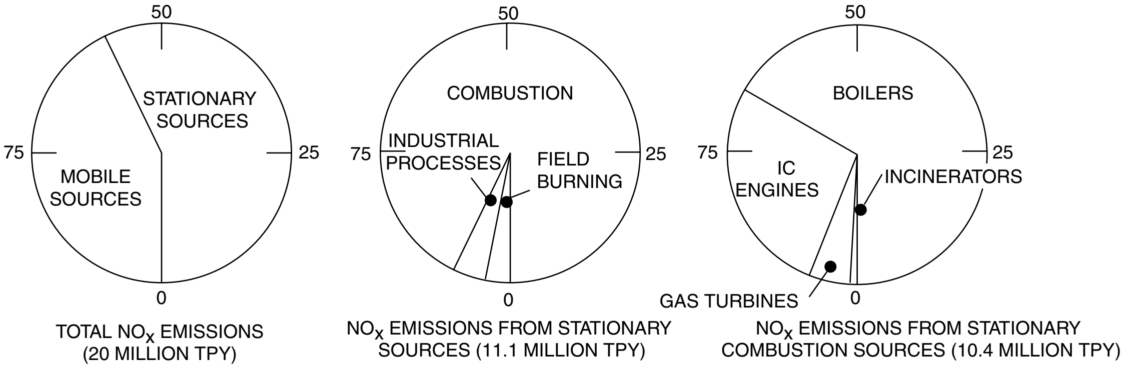


FIGURE 1 Sources of NO_x emissions.

TABLE 1
Status of Commercial FGD Processes (Adapted from Princiotta, 1978)

FGD Process	Reagent	End Product	Principle of Operation	SO_2 Removal Efficiency(%)
Limestone Scrubbing	Limestone (LS)	$CaSO_3/CaSO_4$ Sludge	LS slurry reacts in scrubber absorbing SO_2 and producing insoluble sludge.	80–90
Lime Scrubbing	Lime	$CaSO_3/CaSO_4$ Sludge	Lime slurry reacts in scrubber absorbing SO_2 and producing insoluble sludge.	85–95
Wellman Lord	Sodium carbonate (regenerated)	Sulfuric Acid Sulfur	Soluble sodium sulfite absorbs SO_2 in scrubber, the sodium bisulfite produced is thermally regenerated, yielding sodium sulfite and SO_2 for either acid or S production.	85–95
Double Alkali	Sodium carbonate (regenerated)	$CaSO_3/CaSO_4$ sludge	Soluble sodium sulfite absorb SO_2 in scrubber; the sodium bisulfite produced is reacted with lime precipitating $CaSO_3/CaSO_4$.	85–95
Magnesium Oxide	Magnesium oxide (regenerated)	Sulfuric Acid	Magnesium oxide slurry absorbs SO_2 in scrubber; the magnesium sulfite produced is thermally treated, yielding MgO and SO_2 for acid production.	85–95

Oxides of Carbon

The carbon dioxide, CO_2, concentration threshold for humans is 5% (5000 ppm) for an 8-hour exposure. This compares with a normal atmospheric CO_2 concentration of 0.03% (300 ppm). With a perfect stoichiometric combination of pure carbon in air, a CO_2 concentration of about 21% could be attained. Considering the usual dispersion of combustion gases it would take an unusual isolation to produce a CO_2 health hazard. A more detailed description of CO_2 consequences may be found in the Appendix.

Incomplete combustion of fuels is the more serious problem, since carbon monoxide, CO, will form. This rarely happens in stationary furnaces for which efficiencies of combustion are high and oxygen is available in excess of the theoretical requirements.

It has been estimated (Anon., 1970) that slightly more than 100 million tons of CO are emitted annually in the USA, of which the major sources are automobiles (59%), various open burnings (16%), chemical industry (10%) and other transportation means (5%). New York City, with its acute urban traffic problem, has established a first alert at 15 ppm of CO over

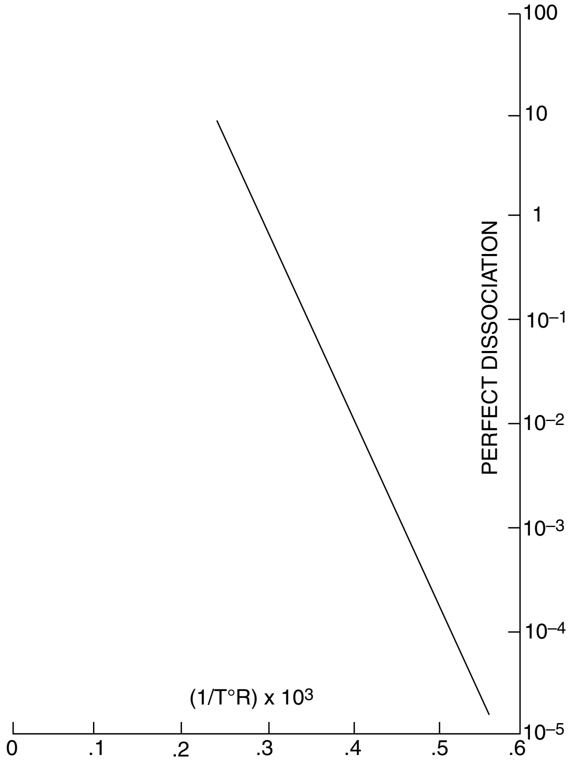


FIGURE 2

at 8-hour period. High temperatures favor the equilibrium dissociation of CO_2 to CO, with the latter being very stable at high temperatures. Thus is a CO_2–CO mixture is quenched from its high temperature zone the percentage CO may remain high, since at lower temperatures longer times are required to reach equilibrium. Rich fuel-air mixtures favor the formation of CO over CO_2. A complete description of CO control methods may be found in the section *Mobile Source Pollution.*

Miscellaneous Gases

Compounds of *fluorine* are known to have negative effects (Fluorisis) at concentrations as low as 5×10^{-3} ppm. They are generated as waste gases of fertilizer aluminum and ceramic processes, but are present to a lesser extent in most flue gases. A concentration of 0.1 ppm (vol.) of fluorine has been set as a maximum permissible value by the American Conference of Governmental Industrial Hygienists; USSR standards are roughly one tenth as stringent.

Ozone, O_3, is one of the strongest gaseous oxidants and is formed naturally from oxygen during electrical discharges in the atmosphere and at the high temperatures of combustion.

$$O_2 \rightarrow 2O$$
$$O_2 + O\cdot \rightarrow O_3$$

Taken as oxidants, New York City classified an ozone level above 0.03 ppm as unsatisfactory and above 0.07 ppm as unhealthy over a 6 hour period. Eye irritation commences at concentrations of about 0.1 ppm. Interestingly enough ozone in the lower stratosphere affords part of the protective shield against ultraviolet radiation from the sun, which could destroy land vegetation.

$$O_3 \rightarrow O_2 + O\cdot$$

Some scientists are concerned that nitric oxide formed by supersonic jets may deplete the ozone supply in the lower stratosphere, eroding the barrier to the destructive rays.

High temperature processes involving metal recovery from ores emit *mercury* vapor in addition to sulfur dioxide. Mercury is available at concentrations up to a few hundred ppm (Kangas *et al.,* 1971) during zinc sulfide ore processing for example. *Hydrochloric* and hydrofluoric acids also appear in the roaster gases of such processes. No danger levels for mercury vapor have been officially established in ambient air quality standards.

A few limits have been established for less common pollutants of the process industries in USSR standards given below.

The US ambient air quality standards call for *hydrocarbon* concentrations below 160 mg/m³ (0.24 ppm) between 6–9 am.

Aldehydes and other *oxygenated hydrocarbons* are formed by the action of ozone on unburned hydrocarbons in the presence of sunlight. For example,

$$O_3 + \text{1-3 Butadiene} \rightarrow \text{A crolein} + \text{Formaldehyde.}$$

In the above reaction both products have been linked to the severe eye irritation encountered in urban environments.

TRANSPORT OF POLLUTANTS

The feed and waste materials of any combustion or chemical process travel through ducts or pipes. Control devices may be placed at various stages of the process, depending on the separation technique to be employed. It will be valuable to review the flow and transport behavior for fluids and then the separation methods. The important process variables to be considered are the mass flow rate of the waste gas, its temperature, pressure and composition. The raw material feed rate variables may also be of significance, as in the desulfurization of fuel oil. Control devices may broadly be classified according to the physical separation process being used, adsorption: absorption: extraction: distillation: or to the chemical process, homogenous or heterogeneous catalytic reaction. In each instance, equations which account for the transport of material and energy must be developed.

In a sense almost any process may be considered as taking place in a pipeline. The simplest model of flow is called plug flow and assumes that no mixing takes place along the axis of the pipeline, but that lateral mixing is complete. Also, this assumes a flat velocity profile exists at each

TABLE 2
Properties of Selected Gaseous Pollutants

					Solubility, CC per 100 GMS		
Name	Formula	Molec WT	Sense Properties	Boiling Pt, °C	Cold H_2O	Warm H_2O	Other
Ammonia	NH_3	17.03	Colorless Pungent	−33.4	Very soluble	> 1000 (99°)	—
Carbon monoxide	CO	28.01	Colorless Odorless	−192	3.5 (0°)	2.32 (20°)	Alcohol, Cu_2Cl_2
Chlorine	Cl_2	70.91	Gn-yellow Pungent	−34.6	310 (10°)	177 (30°)	Aq. NaOH or KOH
Fluorine	F_2	38.00	Gn-yellow	−187	decomposes	—	—
Hydrochloric acid	HCl	36.47	Colorless	−85	very soluble	> 1000 (99°)	Alcohol, Ethylether
Hydrafluoric acid	HF	20.01	Colorless	19.4	very soluble	> 1000 (99°)	—
Hydrogen sulfide	H_2S	34.08	Colorless Decay odor	−59.6	437 (0°)	186 (40°)	Alcohol, CS_2
Mercury	Hg	200.61	—	356.9	—	—	—
Nitric oxide	NO	30.01	Colorless	−151	7.34 (0°)	0 (100°)	Alcohol, H_2SO_4
Nitrogen dioxide	NO_2	46.01	Red-brown	21.3	decomposes	—	HNO_3, H_2SO_4, CS_2
Ozone	O_3	48.00	Faint blue	−112	0.494 (0°)	0 (50°)	Oil turp., oil cinn.
Sulfur dioxide	SO_2	64.07	Colorless Choking	−10	8000 (9°)	1600 (50°)	H_2SO_4, alcohol, acetic acid
Sulfur trioxide	SO_3	80.66	Colorless	44.6	decomposes	—	H_2SO_4

longitudinal position, or that the average velocity is the same at each lateral position.

Continuity

If the density, κ, of the fluid at a distance along the pipe, *Z*, of cross section, *S*, changes, the velocity must also change as seen by an elemental mass balance across *d Z* distance, i.e. setting the mass accumulation rate equal to the sum of net input and generation rates (see Figure 3).

$$\frac{\partial \varrho}{\partial t} = \frac{1}{S}\frac{\partial(\varrho v S)}{\partial Z}. \tag{1}$$

For steady state results, $\varrho v S$ = const. = W_o and the mass flow rate becomes the same at all axial positions. If the fluid is incompressible, ϱ= const., as for most liquids, *vS*, the volumetric flow rate, does not vary with position even during transient conditions.

Motion

In a comparable manner an elemental momentum (force) balance may be made over length *dZ*, which for incompressible flow reduces to

$$\frac{\partial v}{\partial t} = \frac{1}{2}\frac{\partial v^2}{\partial Z} - \frac{g_c}{\varrho}\frac{\partial p}{\partial Z} - \frac{g_c(F+F_w)}{\varrho S_o} + g_z. \tag{2}$$

TABLE 3

	Daily		Instantaneous	
	mg/m³	ppm/wt	mg/m³	ppm/wt
Cl_2	0.03	(0.024)	0.10	(0.081)
H_2S	0.01	(0.0081)	0.03	(0.024)
CS_2	0.15	(0.122)	0.50	(0.406)
P_2O_5	0.05	(0.049)	0.15	(0.122)
Phenol	0.10	(0.081)	0.30	(0.24)

For both steady and incompressible flow

$$\frac{dp}{dZ} = \left[\frac{F}{S_o} - \frac{\varrho g_z}{g_c}\right] - const. \tag{3}$$

The equation describes the relation between velocity and pressure along the pipe. The quantities *F* and F_w are the magnitudes of skin frictional force and force doing work on external surfaces, respectively, both per unit length of pipe.

ENERGY

The First Law of Thermodynamics may be written for the differential element of length, *dz*, at steady state

$$\frac{dH}{dz} + \frac{g}{g_c} + \frac{v}{g_c}\frac{dv}{dz} = \delta Q - \delta W_s. \tag{4}$$

For unsteady behaviour where temperature gradients are desired the equation of thermal energy may be applied assuming a uniform temperature at any cross-section and no axial conduction.

$$\varrho c_v \frac{\partial T}{\partial t} = -T\left(\frac{\partial p}{\partial T}\right)_{\&}\left(\frac{\partial v}{\partial z}\right) + \delta q - \delta w_s - \frac{\partial(\varrho_v v T)}{\partial z} \tag{5}$$

in which *q* and w_s are the volumetric thermal energy input rate (produced for example by an electrical or chemical phenomenon) and the work output rate, respectively. For a constant density fluid equation (5), the left hand side represents the accumulation of internal energy, and the right hand terms represent the influence of pressure on the energy transport rate, the combined energy input rate per unit volume by generation and forces and the net energy input rate by flow (force convection), respectively.

Component Balance

The equations of continuity, motion and energy often may be applied to describe the situation in stacks of power plants, in the flow of fuels and effluents, and in the analysis of material, momentum and energy requirements of a

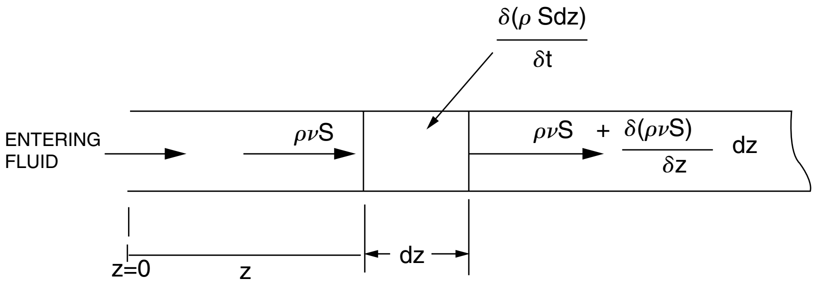


FIGURE 3

pollution producing process. To analyze the concentrations of pollutant it still remains to make component material balances of $n - 1$ species within the system (for which n components exist). For separation processes an additional phase equation is usually required for transfer of pollutants between the rich and lean phases.

The mass balance on a particular species may be found for component A by examining the imaginary stationary differential element of thickness dz. Assuming plug flow we may derive an expression for c_A:

$$\begin{array}{ccccc} \text{Accum.} & = & \text{Net Input} & + & \text{Generation} \\ \text{rate of } A & & \text{rate of } A & & \text{rate of } A \end{array} \tag{6}$$

c_A—*molar concentration*
F_a—*flux of* A at position Z, moles A flowing/(time) (cs. area) $= vc_A$
R_A—production rate of component A, by chemical or nuclear reaction, moles A formed/(time) (vol.)
v—average fluid velocity.

Also, r_A is usually some empirical function of c_A such as kc_A^n for an irreversible decomposition reaction of nth order.

Gas Adsorption

Adsorption is the process by which a solid surface attracts fluid phase molecules and forms a chemical or physical bond with them. The mechanism of adsorption includes:

1) diffusion of the pollutant from bulk gas to the external surface of the particles,
2) migration of the adsorbate molecules from the external surface of the absorbent to the surface of the pores within each particle,
3) adsorption of the pollutant to active sites on the pore.

The attraction for a specific gas phase component will depend on properties such as the concentration of the gas phase component, the total surface area of absorbent, the temperature, polarity of the component and adsorbent, and similar properties of competing gas molecules. Adsorption is used to concentrate (30–50 fold) or store pollutants until they can be recovered or destroyed in the most economical manner. Adsorption is an exothermic process. The heat of adsorption for chemical adsorption is higher than that for physical adsorption. In the former, if the amount of pollutants to be removed large, it is necessary to remove the heat of adsorption, since the concentration of adsorbed gas decreases with increasing temperature at a given equilibrium pressure. For chemical adsorption, properties which affect reaction kinetics will also come into play (see section on Gas Reaction).

Activated carbon, silicon, aluminum oxides, and molecular sieves make up the majority of commercially significant adsorbents. Activated carbon is the least affected by humidity and physically adsorbs nonpolar compounds since it has no great electrical charge itself. The adsorption rate of activated carbon can be increased with chemical impregnation. For instance, activated carbon impregnated with oxides of copper and chromium are found very useful to remove the hydrogen sulfide in gas streams where oxygen is not present (Lovett and Cunniff, 1974). Alumina and silica materials preferentially adsorb polar compounds. Molecular sieves have greater capture efficiencies than activated carbons but they often have a lower retention efficiency and are considerably more expensive.

The ease of adsorbent regeneration depends on the magnitude of the force holding the pollutants on the surface of adsorbent. The usual methods for the adsorbent regeneration include stripping (steam or hot inert gas), thermal desorption, vacuum desorption, thermal swing cycle, pressure swing cycle, purge gas stripping, and in situ oxidation.

In many respects the equilibrium adsorption characteristics of a gas or vapor upon a solid resemble the equilibrium solubility of a gas in a liquid. For simple systems, a single curve can be drawn of the solute concentration in the solid phase as a function of solute concentration or partial pressure in the fluid phase. Each such curve usually holds at only one specific temperature, and hence is known as an isotherm. Five types of commonly recognized isotherms are shown by the curves in Figure 4. There are three commonly used mathematical expressions to describe vapor or gas adsorption equilibrium: the Langmuir, the Brunauer-Emmett-Teller (BET), and the Freundlich isotherm. The Langmuir isotherm applies to adsorption on completely homogeneous surfaces, with negligible interaction between adsorbed molecules. It might be surmised that these limitations correspond to a constant heat of adsorption. The Freundlich isotherm can be derived by assuming a logarithmic decrease in heat of adsorption with fraction of coverage. Gas adsorption is an unsteady state process. The curve of effluent concentration as a function of time is commonly referred to as the break-through curve. It usually has an S shape. The break-through curve may be steep or relatively flat, depending on the rate of adsorption, the adsorption isotherm, the fluid velocity, the inlet concentration, and the column length. The time at which the break-through curve first begins to rise appreciably is called breakpoint.

The design of an adsorption column requires prediction of the breakthrough curve, and thus the length of the adsorption cycle between elutions of the beds, given a bed of certain length and equilibrium data. Because of the different forms of equilibrium relationship encountered, and the unsteady nature of the process, prediction of the solute break-through curve can be quite difficult. At present, detailed design of adsorption columns is still highly dependent on pilot scale evaluations of simulated or real systems.

Before discussing the method of predicting the breakthrough curves, one should consider the isotherm. For Langmuir isotherm (Langmuir, 1917), if it is assumed that A_I reacts with an unoccupied site X_0 to form adsorbed component X_1,

$$A_I + X_0 \underset{k-1}{\overset{k_1}{\rightleftharpoons}} X_1 \tag{7}$$

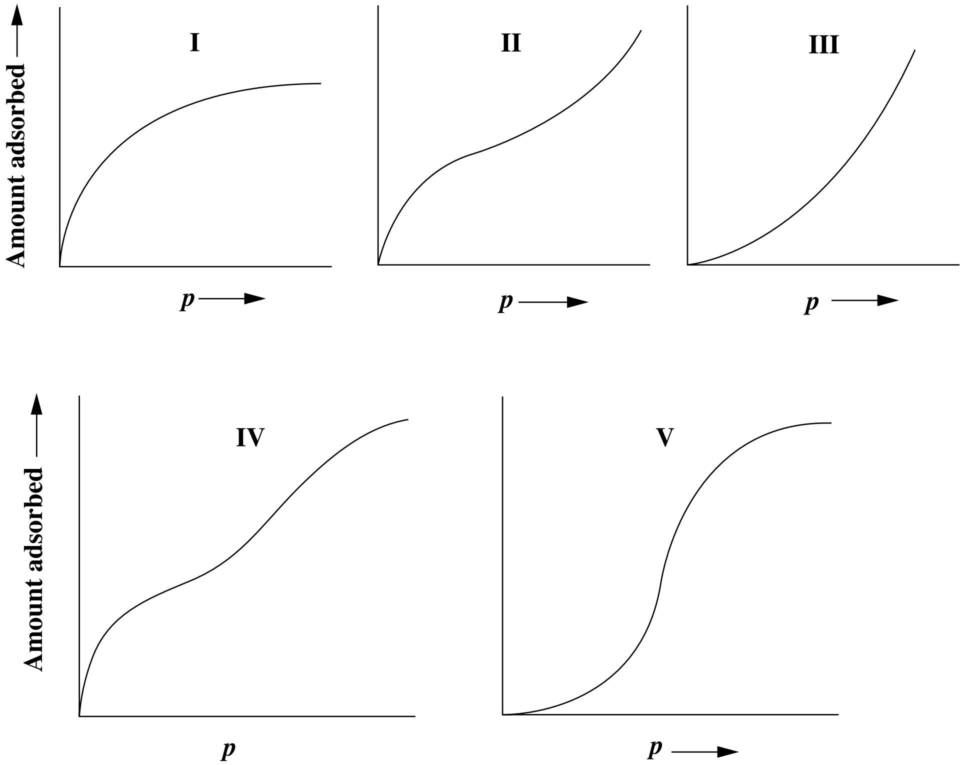


FIGURE 4 Types of adsorption isotherms.

the equilibrium adsorption concentration, C, is obtained in terms of the gas phase concentration C_1 and the total adsorption site concentration, C_0

$$\frac{C}{C_0} = \frac{K_1C_1}{1 + K_1C_1}. \tag{8}$$

Here, K_1 is the adsorption equilibrium constant which varies only if temperature varies.

Diatomic molecules such as chlorine might be expected to simultaneously adsorb and dissociate on adjacent sites. Such an adsorption might be described symbolically by

$$A_1 + 2X_0 \overset{k_1}{\underset{k-1}{\square}} 2X_1$$

in which case the equilibrium isotherm expression can readily be shown to be

$$\frac{C}{C_0} = \frac{(K_1C_1)^{1/2}}{1 + (K_1C_1)^{1/2}}. \tag{9}$$

If more than one pollutant is being adsorbed, each component, A_j, undergoes

$$A_j + X_0 \square \; X_j.$$

The equilibrium for single site adsorption of component A_j becomes

$$\frac{C}{C_0} = \frac{K_1C_1}{1 + \sum_{j=1}^{n} K_jC_j}. \tag{10}$$

The latter may be referred to as competitive adsorption. Figure 5 depicts the different dependence on gas phase concentration for the adsorption types described thus far.

Another isotherm finding wide use, particularly in multilayer adsorption, is that of Brunauer, Emmett and Teller (1938), the BET equation

$$\frac{C_s}{C_{\max}} = \frac{Q_2X}{(1 - X)[1 + (Q_2 - 1)X]} \tag{11}$$

in which:

C_s, $C_{\max}$—amount of gas absorbed per gm. of solid, and maximum amount, respectively

Q_2—term exponentially dependent on heat of adsorption

X—ratio of equilibrium gas phase concentration of saturation Value.

If neither the Langmuir or BET equations are satisfactory, a plynomial fit to adsorption data may be required.

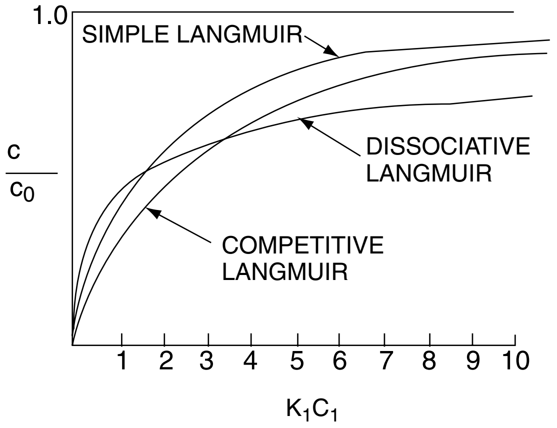


FIGURE 5

For adsorption the particles are usually dumped into a column as a packed bed (Figure 6). In commercial adsorption columns, equilibrium concentrations are not attained uniformly, but for convenience the rate of adsorption is assumed to be proportional to *C*. See below under Reaction.

The time t_3 in which breakthrough (C_{ig} at the exit equals the permissible set amount) occurs may be established by analyzing the differential component mass balance assuming that the transport of material again is governed by diffusion through a film

$$N_i = k_g(C_{ig} - C^*_{ig}) \tag{12}$$

in which C^*_{ig} is the gas phase concentration in equilibrium with the absorbed concentration C_{is} at the same elevation.

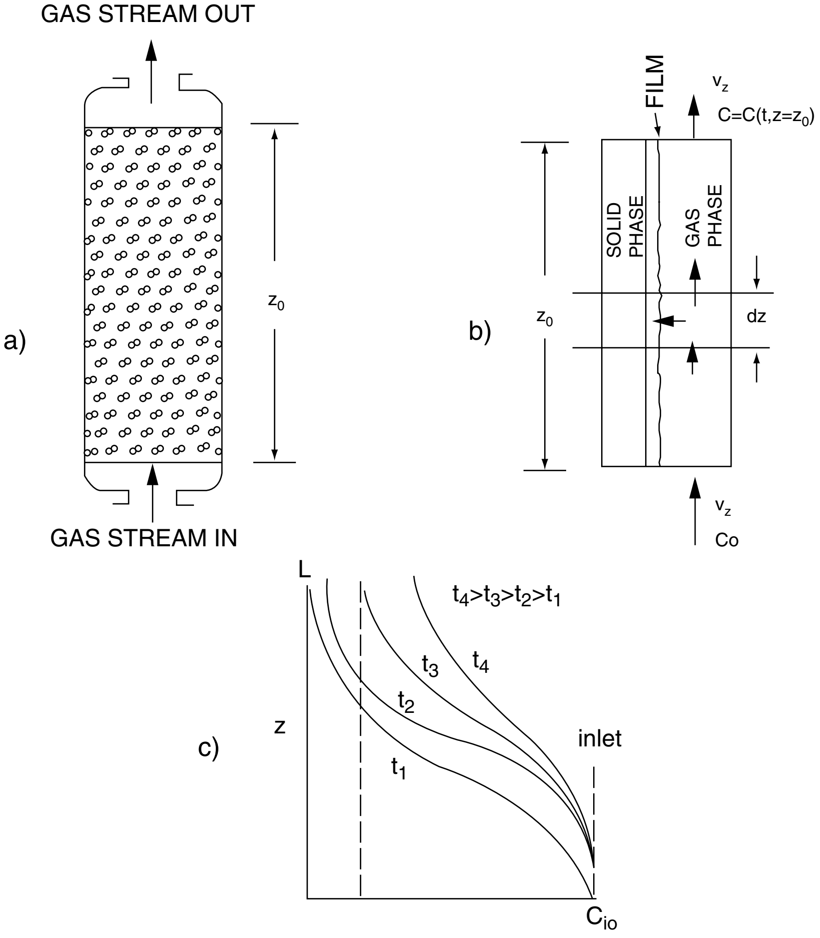


FIGURE 6 (a) Pictorial representation. (b) Schematic model showing a differential element over which a mass balance is made. (c) Pollutant concentration as a function of time at various heights.

The equilibrium concentrations have been determined for most commercially available adsorbents and typical pollutants and are presented as either Langmuir or BET isotherms. In general such equations take the form

$$C_{is} = f(C_{ig}^*) \tag{13}$$

The generation term is excluded, as it is assumed that chemical reaction is not taking place in the system. To develop an expression for the net rate of accumulation of the mass of A_i in the column, that is, the rate expression for the adsorption process, one also needs the function of the concentration of A_i in the fluid phase.

Considering a differential column segment and writing the continuity relationship for pollutant A_i in each phase for this differential section, one gets for the solid phase in this section.

$$\frac{1}{M_a} P_s(1-\varepsilon)\left(\frac{\partial C_{is}}{\partial t}\right) = k_g(C_i - C^*) \tag{14}$$

where:

k_g: is the gas phase mass transfer coefficient in units of (time – 1).

M_a: is the molecular weight of A_1.

P_s: is the averaged density of the solids.

ε: is the fractional voidage in the bed.

For the gas phase in this differential segment,

$$\varepsilon\frac{\partial C_{if}}{\partial t} = -V\frac{\partial C_{if}}{\partial z} - k_g(C_i - C_g^*) \tag{15}$$

where:

V_z: is the superficial velocity of the fluid.

These partial differential equations (13)–(15) may be solved simultaneously by numerical analysis using difference formulas to approximate the partial derivatives. In such a way the breakthrough curves of hazardous organic vapors may be predicted for a given adsorbent.

Smoothed computerized results were plotted on Figure 7 for five different compounds having Langmuir type behavior on activated carbon under the same hypothetical operating conditions. If one wishes to attain a 90% removal of certain organic vapor, one could easily see from Figure 7 that diethyl ether requires the shortest re-cycle time and methyl isobutyl ketone the longest among the five materials on the graph.

Properties of Adsorbents

Figures 8 and 9 are adsorption isotherms for activated carbon with nitrous oxide and carbon dioxide respectively. A more sophisticated correlation of adsorption data is presented in Figures 10–12 for pure CO, C_2H_4 and CO_2 gases. Here $(RT/V_s)\ln f_s/f_g$ is plotted versus N_s (in which:—gas constant, T—temperature, °K, V_s—molar volume of adsorbate, cc/mole, f_s and f_g—fugacites of adsorbate and gas and N—amount of gas adsorbed, g—moles/gm. adsorbent. Hydrocarbons and SO_3 adsorb readily on activated carbon. SO_2 has a maximum retention of 10 wt.% on carbon at 20°C, 760 torr. Ozone decomposes to oxygen on carbon (Ray and Box, 1950).

Figure 13 has comparable results plotted for CO_2 adsorption on silica gel. Activated carbon has significantly better equilibrium properties than does silica gel (vis Figure 9 vs. Figure 13).

Other results for activated carbon and zeolites may be found in the book by Strauss (1968). Basic facts about adsorption properties of activated charcoal, system types and components and applications are discussed by Lee (1970). He tabulated data on the air purification applications for inexpensive, nonregenerative, thin bed adsorbers and for regenerative systems, and discusses the design of a solvent vapor recovery system.

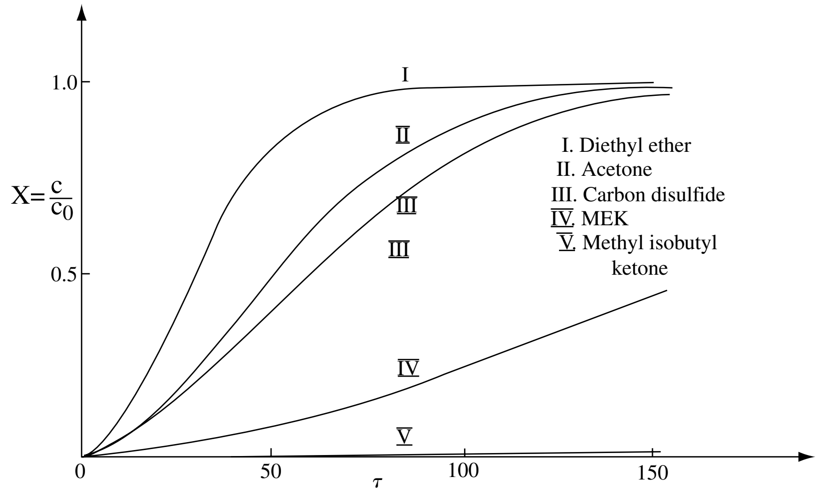


FIGURE 7 Break through curves for various compounds at 20°C and 1 atm with C_0 = 0.00548 mole/liter.

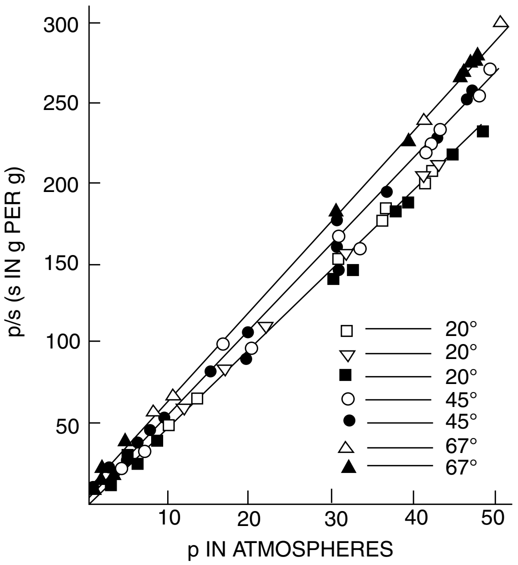


FIGURE 8 The system nitrous oxide-carbon. McBain and Britton: *J. Am. Chem. Soc.* **52,** 2217 (1930) (Fig. 14).

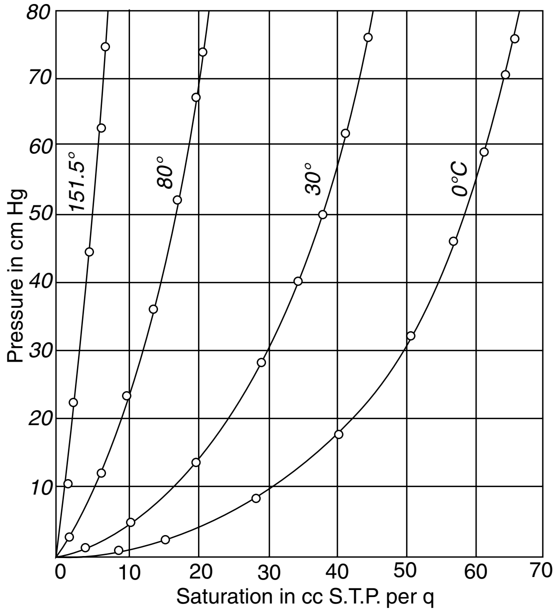


FIGURE 9 Absorption isotherms for the system CO_2-carbon (note that $t_c = 31°C$).

"Activated carbon filters were used to concentrate atmospheric mixtures of acrolein, methyl sulfide, and *n*-propyl mercaptan. Removal efficiency and carbon capacity for each of the odor compounds were investigated using two different carbones, Cliffchar (4–10 mesh) and Barnebey–Cheney (C-4). A closed system was devised to establish a known atmospheric odor concentration for each filter run. Solvent extraction techniques were employed to desorb and recover the odor compounds from the carbon filters. All quantitative analyses were conducted with gas liquid chromatography utilizing the hydrogen flame ionization detector. The removal studies conducted indicate that the efficiency of removal of a carbon filter is essentially 100% up to the point of filter breakthrough. This breakthrough point is governed by the filter's capacity for a particular compound. This study indicated that the filter capacity is dependent both on the type of carbon employed and the particular odor compound adsorbed. Solvent recovery of the odor compounds from the carbons varied from 0 to 4.5% for the mercaptan up to 96 to 98% for acrolein. Per cent recovery was found to vary for a given odor compound with different carbons and for a given carbon with different odor pollutants." (Brooman and Edgerley, 1966.)

Gas Absorption

Adsorption is a diffusional process that involves the transfer of molecules from the gas phase to the liquid phase because of the contaminant concentration gradient between the two phases. Adsorption of any species occurs either at the surface of the liquid film surrounding the packing or at the bubble surface when the gas is the dispersed phase. When a gas containing soluble components is brought into contact with a liquid phase an exchange of the soluble components will occur until equilibrium in a batch system or steady state in a flow system is attained. Adsorption may involve only a simple physical solubility step or may be followed by chemical reaction for more effective performance. The latter is usually used for flue gas desulfurization and denitrification. Rates of adsorption depend on the solubility of the gas. At equilibrium, for gaseous species of low or moderate solubility, the partial pressure of the component is related to its liquid mole fraction according to Henry's law,

$$p_i = HX_i$$

where H is Henry's constant. Both partial pressures and mole fraction may be related to concentration

$$p_i = C_{ig}RT \text{ and } X_i = C_{il}/C_l.$$

At constant temperature, $C_{ig} = H'C_{il}$, where $H' = H/C_tRT$. If the gas is highly soluble in the liquid, H will be small. The solubility of the gas is affected by the concentration of ions in the solution at the interface. Van Krevelen and Hoftijzer (1948) proposed an empirical equation to correct the effect of concentration of ions on Henry's constant.

The rate of mass transfer is proportional to both the interfacial area and the concentration driving force. The proportional constant is known as the mass transfer coefficient. Because material does not accumulate at the interface (Figure 14) the flux in each phase must be the same. Thus the rate of transfer per unit area is

$$N_i = j_g(C_{ig} - C_{igI}) = h_L(C_{iLI} - C_{iL}).$$

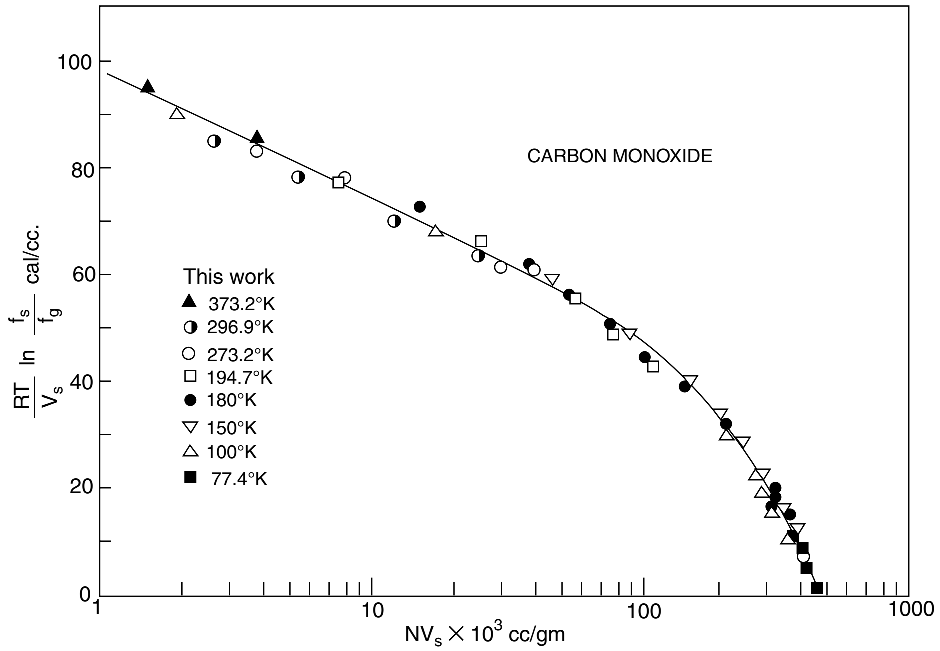


FIGURE 10

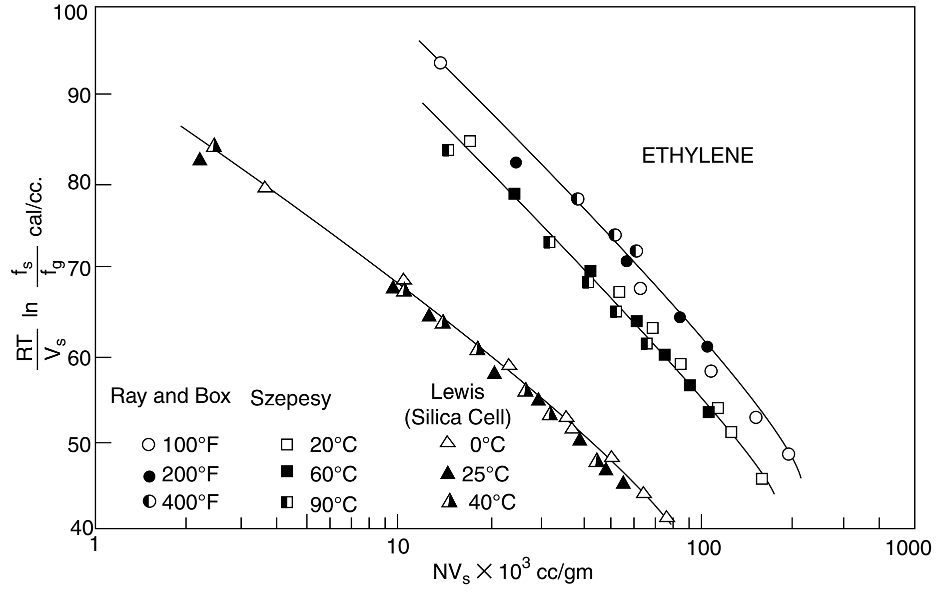


FIGURE 11

In most industrial adsorption processes, the gas is reacted with some substance to form a semistable compound in the liquid phase. This technique permits a great deal more gas to be adsorbed per gallon of liquid circulated, and, in most instances, will increase the mass transfer coefficient. In this situation,

$$\begin{aligned} N_i &= k_g(C_{ig} - C_{igI}) = k_1(C_{iLI} - C_{iL}) \\ &= k_{LE}(C_{iLI} - C_{iL}) \end{aligned}$$

where E = enhancement factor, k_L = mass transfer coefficient in absorption $E = 1$ for physical adsorption. The enhancement factor can be found elsewhere (Astarita, 1967; Danckwert, 1970; Sherwood *et al.*, 1975).

Consider a system whose equilibrium line is straight over the range of compositions which need be considered. The mass transfer rate may be described in forms of pseudo concentration values, thus,

$$N_i = K_g(C_{ig} - C_{ig}^*) = K_L(C_{iL}^* - C_{iL}),$$

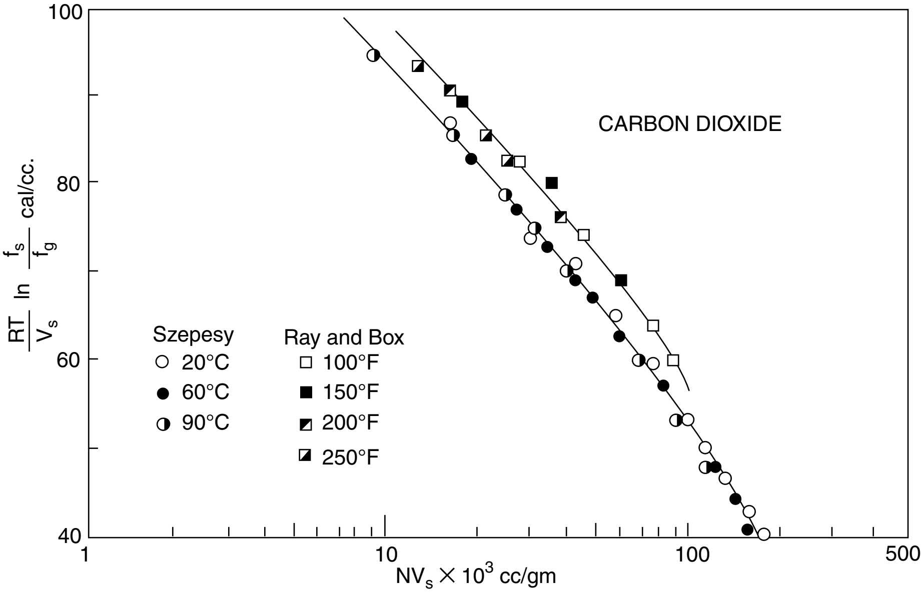


FIGURE 12

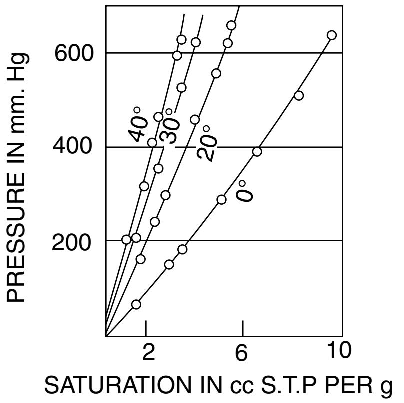


FIGURE 13 Adsorption isotherms for the system CO_2-silica gel (note that $t_c = 31$°C). Patrick, Preston and Owens: *J. Phys. Chem.* **29,** 421 (1925) (Fig. 1).

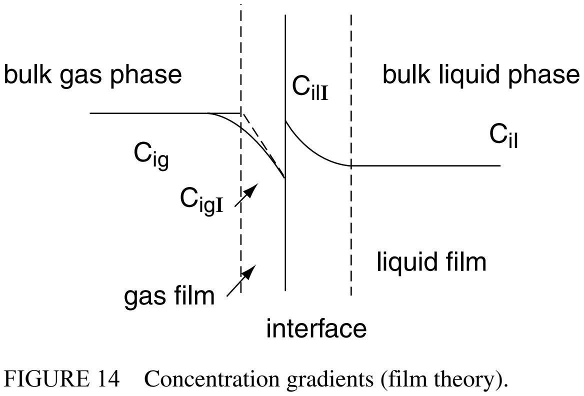


FIGURE 14 Concentration gradients (film theory).

where K_g and K_L are overall mass transfer coefficient in gas and liquid phase, respectively. If concentrations at equilibrium can be represented by Henry's law

$$C_{ig}^* = H'C_{iL},$$

then

$$\begin{aligned} N_i &= k_g(C_{ig} - C_{igI}) = k_L E(C_{iLI} - C_{iL}) \\ &= k_g(C_{ig} - C_{ig}^*) = K_L(C_{ig}^* - C_{iL}), \end{aligned}$$

whence

$$\frac{1}{K_g} = \frac{1}{k_g} + \frac{H'}{Ek_L} = \frac{H'}{K_L}.$$

If the gas is highly soluble in the liquid, H' will be small and $K_g = k_g$. Hence the mass transfer rate is

$$N_i = K_g(C_{ig} - C_{ig}^*)$$

and absorption is said to be gas phase controlled.

In cleaning an effluent stream of low pollutant species concentration physical absorption alone is often insufficient to produce the required removal and a reactant may be added to the absorbing solution to enhance the rate. For example potassium permanganate has been found to be an excellent absorbent for NO (see excellent review by C. Strombald (1988)).

The design of absorbers involves the estimation of column diameter, height, and pressure drop. The column diameter is fixed by the contaminated gas flow rate. The determination of the height of the two-phase contacting zone involves an estimation of the mass transfer coefficients, the alternating use of equilibrium concentration relationship, and the law of mass conservation. For nonisothermal or adiabatic operating conditions, the law of energy conservation needs to be considered. There are many types of equipment and configurations for absorbers or scrubbers. For example, McCarthy (1980) discussed the scrubber types and selection criteria.

For packed towers the interfacial area, *a,* differs from the packing surface area, a_T, because the packing is not always completely wetted. A fraction of the surface may not be active in mass transfer. Also, stagnant pockets will be less effective than flowing streams. The correlation of Onda *et al.* (1968) may be used to estimate the value of interfacial area.

$$\frac{a}{a_1} = 1 - \exp \left\{ -1.45 \left(\frac{\sigma_c}{\sigma} \right)^{.75} \left(\frac{L}{a_1 \mu_L} \right)^{.1} \left(\frac{L^2 a_t}{\rho_L^2 g} \right)^{-.5} \left(\frac{L^2}{\rho_L \sigma a_t} \right)^{-.2} \right\}$$

Where σ_c represents the critical surface tension above which the packing can't be wetted. The values of σ_c for various packing materials are shown in Table 4 (Onda *et al.*, 1967)

L : the superficial liquid mass flow rate

P_L, μ_L : density and viscosity of the liquid, respectively

σ : surface tension of the liquid, dynes/cm.

TABLE 4
Critical surface tension of packing materials

Material	σc dynes/cm
Carbon	56
Ceramic	61
Glass	73
Paraffin	20
Polyethylene	33
Polyvinylchloride	40
Steel	75

This equation correlates results with a maximum error of 20% except in the case of Pall rings where it is conservative. The reason is probably that the interfacial and wetted area are different for Pall rings whose shape forces a fraction of the liquid phase to be dispersed in small droplets that are not accounted for in the values of *a.* Values of a Pall rings are underestimated about 50% according to Charpentier (1976).

The liquid phase mass transfer coefficient can be estimated using Mohunta's equation (1969).

$$k_L a = 25 \times 10^{-4} \left(\frac{g \rho_L}{a_t \mu_L} \right)^{.66} \left(\frac{g^2 \rho_L}{\mu_L} \right)^{.111} \left(\frac{\mu_L \mathrm{L}^3 \mathrm{a}_t^3}{g^2 \rho_L^4} \right)^{.25} \times \left(\frac{\mu_L}{\rho_L D_L} \right)^{-.5}$$

within a range of ±20%.

The range of variables and physical properties of Mohunta's equation are:

Variables	*Range*
L	0.1–42 k_g/m^2 sec
G	0.015–1.22 k_g/m^2 sec
μ_L	0.7–1.5 CP
$\mu_L/\rho_L \mathrm{D}_L$	140–1030
d	0.6–5 cm
D (column)	6–50 cm

where G : superficial gas mass flow rate
D_L : liquid diffusivity
d : packing diameter.

For the gas phase mass transfer coefficient, Laurent and Charpentier (1974) derived the correlation

$$\frac{k_g P}{G} = \frac{C}{M} (a_t d)^{-1.7} \left(\frac{\mathrm{Gd}}{\mu_\mathrm{G}} \right)^{-.3} \left(\frac{\mu_G}{\rho_G D_G} \right)^{-.5}$$

where P : total pressure, atm
M : gas molecular weight
D_G : solute gas diffusively
C = 2.3 for d , < 1.5 cm
C = 5.23 for d > 1.5 cm.

Tray type towers have also been used successfully. Bubble cap plates correlations have been proposed by Andrew (1961)

$$k_g = 7u^{1/4} S^{-1/2} D_g^{1/2} \text{ cm/sec}$$

$$k_l = 11u^{1/4} S^{-1/2} D_l^{1/2} \text{ cm/sec}$$

$$a = 0.7u^{1/2} S^{5/6}$$

in which:

S = effective liquid area on plate, cm^2
u = superficial gas velocity, cm/sec
a = interfacial area/unit area of plate.

A conservative estimate for the liquid phase coefficient in a sieve plate may be obtained from the Equation of Claderbank and Moo–Young (1961)

$$k_l = 0.31(g v_1)^{1/3}(d_l/v_1)^{2/3} \text{ cm/sec}$$

where v_1 is the kinematic liquid viscosity. For CO_2 and water at ordinary temperature, $k_l \approx 0.01$ cm/sec. Typical sieve plate N_i values are about 10^{-4} g mole/cm^2-sec-atm.

"Each tray of Figure 15 has a series of drawn orifices fitted with a cage and cap. The orifice has a flared entrance. This reduces the dry pressure drop allowing a greater percentage of the work expended to be utilized for scrubbing. The floating cap maintains the scrubbing efficiency even with variations in gas flow as wide as 40–110% of capacity.

Gas enters the vessel flowing upward through the valve trays. Liquid is introduced on the top tray and flows across each tray over a weir and then to a sealed downcomer to the next lower tray. A level is maintained over each tray by the weir. The upward flowing gas is given a horizontal component by the cap causing atomization of the liquid on the tray. The froth formed consisting of a myriad of small droplets traps the particles and absorbs or reacts with acid or alkaline vapors. The liquid agitation on the tray surface prevents buildup. This has been demonstrated by the many successful applications in the Petrochemical Industry involving tarry solids and liquids.

Each tray has a 1½ W.G. pressure drop. A typical installation with four (4) trays would require less than 8″ W.G. pressure loss."

An excellent review of gas phase absorption may be found in the work of Danckwerts (1970). The molecular diffusivities in the vapor phase D_r, and in the liquid D_1, may be found from existing correlations, for example see Bird *et al.* (1960). Unlike the solid in adsorption the liquid solvent in absorption usually leaves the system where it can be regenerated. Hence a steady state plug flow analysis in either phase in terms of overall coefficients is possible

$$\begin{aligned} LdC_{il} = GdC_{ig} &= K_g a(C_{ig} - C^*_{ig})dz \\ &= K_1 a(C^*_{il} - C_{il})dz. \end{aligned}$$

The required tower height is then given by either of the following integral relationships:

$$Z = \frac{G}{K_g a}\int_{C_{go}}^{C_{ge}} \frac{dC_g}{C_g - C^*_g} = \frac{L}{K_1 a}\int_0^{C_{le}} \frac{dc_{il}}{C^*_{il} - C_{il}}$$

in which: a is the surface area of contact per unit volume of bed; L and G are per superficial mass velocities. Further discussion on the subject may be found in the work of Cooper and Alley (1994).

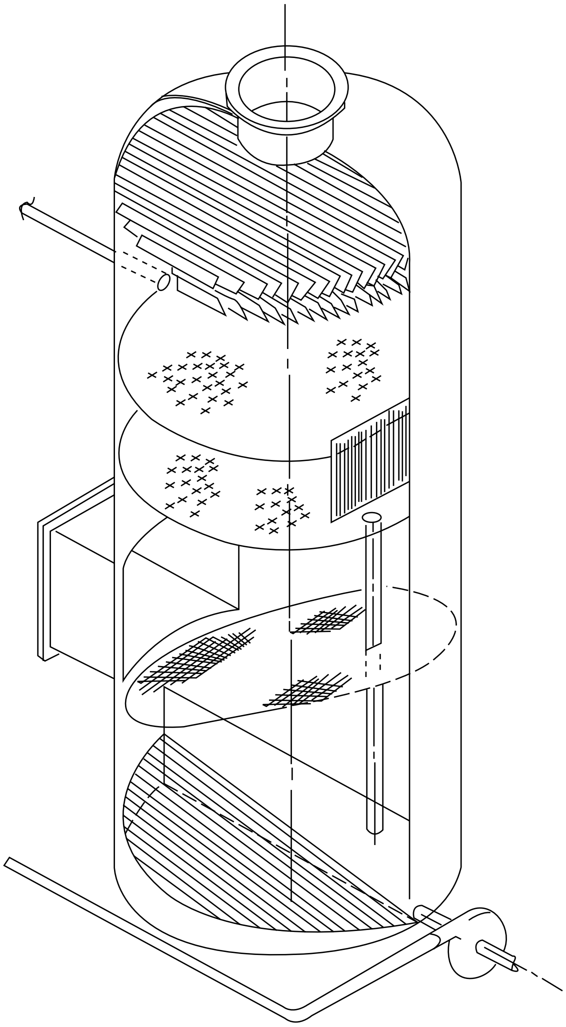


FIGURE 15 Flexitrary Scrubber (Courtesy of Koch Engineering).

PROPERTIES OF ABSORBENTS

Henry's law constants for CO, CO_2, NO and H_2S, are presented in Table 5. Lower values of H such as those for H_2S correspond to higher solubility values. Table 6 contains specific wt. fraction absorbed at equilibrium vs. gas partial pressure for both ammonia and SO_2 in water; Table 7 has similar material for HCl.

Highly soluble materials have absorption rates which are controlled by diffusion through the gas phase (see Table 8).

REACTION

Processes exist for catalytically removing gaseous pollutants by either forming harmless products or products more amenable to recovery. The behavior of most catalytic reactions requires a more substantial analysis than homogeneous systems because of the presence of at least two phases. One of the

TABLE 5

Henry's Law constant for slightly soluble pollutants, H $\times 10^{-4}$ atm/mole-fraction

T,°C	CO	CO_2	NO	H_2S	SO_2
0	3.52	0.0728	1.69	0.0268	0.0016
20	5.36	0.142	2.64	0.0483	0.0033
40	6.96	0.233	3.52	0.0745	0.0062
60	8.21	0.341	4.18	0.103	—
80	8.45	—	4.48	0.135	—
100	8.46	—	4.54	0.148	—

primary differences is that heterogeneous reactions require the adsorption of a fluid phase component on the catalyst surface. In addition to adsorption, diffusion rates may be significant for both main stream to surface gas transfer and for transport inside catalyst pores. The analysis is often simplified by lumping the parameters of the system and assuming plug flow.

Thus for a packed bed of catalyst as in Figure 16 the catalyst mass is lumped as a single substance and the gas phase as another, the latter in plug flow.

The mass of catalyst required to attain a given conversion may be calculated if the dependence of u and P_i on c_i is known.

$$M = \&_B V = S\int_{c_{io}}^{c_i} \frac{d(u\bar{c}_i)}{P_i}$$

where:

u = superficial velocity based on total cross-sectional area S

K_B = bulk density, mass of solids/unit volume

$P_i(\bar{C_i})$ = production rate, moles of A_l formed/(mass solids) (time).

The dependence of the production rate on concentration is usually determined assuming either a Langmuir-Hinshelwood or a Langmuir Rideal mechanism, if diffusion rate constants are large.

If we let X_i represent an adsorbed component A_i on an active site X_0, the *Langmuir-Rideal* sequence, which assumes that a fluid phase reactant combines with a surface adsorbed molecule, the sequence may be expressed as

$$A_1 + X_0 \overset{k_1}{\underset{k-1}{\&}} X_1 \qquad \text{Langmuir adsorption}$$

$$A_2 + X_1 \overset{k_2}{\underset{k-2}{\&}} X_3 + A_4 \qquad \text{Precursor — surface reaction}$$

$$X_3 \overset{k_3}{\underset{k-3}{\&}} A_3 + X_0 \qquad \text{Langmuir desorption.}$$

Usually, to simplify the analysis all the equations save one are assumed to be at equilibrium and the unsteady equation is said to control the rate. Thus for the Langmuir-Hinshelwood equation, the rate of product formation is if

Surface Reaction Controls

$$P = \frac{(k_s K_1 K_2 C_1 C_2 - k_{-s} K_3 K_4 C_3 C_4) C_0^2}{\left(1 + \sum_{j=1}^{4} k_j C_j\right)^2}.$$

The denominators indicate that each component competes for adsorption sites on the catalyst surface. The rate constants and equilibrium constants are denoted by k and K, respectively, with subscripts denoting surface reaction and j referring to adsorption of component A_j. The concentration of each component can be put in terms of C_3 by stoichiometry considerations.

CATALYST PROPERTIES

A typical conversion described by Langmuir adsorption followed by chemical reaction is that for SO_2 removal over vanadium oxide catalyst (Mars and Maessen, 1961)

$$SO_2 + \partial V^{5+} + O^{2-} \square \; SO_3 + \partial V^{4+}$$

95% Yield at 450°C

$$r = kp_{O_2} \frac{Kp_{SO_2}/P_{SO_3}}{[1 + (Kp_{SO_2}/P_{SO_3})^{1/2}]^2}$$

k = rate constant,

K = equilib. constant for above reaction,

p_i = partial pressure of ith species.

Similarly for the oxidation of ethylene

$$r = \frac{kp_{c_2H_2}^2 p_{O_3}}{[1 + K_1 C_{C_2H_4} + K_2 C_{CO_2}^2]^2}.$$

TABLE 6

Solubility data for ammonia and sulfur dioxide in water ammonia

Mass NH_3 per 100 Masses H_2O	Partial pressure of NH_3, mm Hg						
	0°C	10°C	20°C	30°C	40°C	50°C	60°C
100	947	—	—	—	—	—	—
90	785	—	—	—	—	—	—
80	636	987	—	—	—	—	—
70	500	780	—	—	—	—	—
60	380	600	945	—	—	—	—
50	275	439	686	—	—	—	—
40	190	301	470	719	—	—	—
30	119	190	298	454	692	—	—
25	89.5	144	227	352	534	825	—
20	64	103.5	166	260	395	596	834
15	42.7	70.1	114	179	273	405	583
10	25.1	41.8	69.6	110	167	247	361
7.5	17.7	29.9	50.0	79.7	120	179	261
5	11.2	19.1	31.7	51.0	76.5	115	165
4	—	16.1	24.9	40.1	60.8	91.1	129.2
3	—	11.3	18.2	29.6	45.0	67.1	94.3
2	—	—	12.0	19.3	30.0	44.5	61.0
1	—	—	—	—	15.4	22.2	30.2

Sulfur Dioxide

Mass SO_2 per 100 Masses H_2O	Partial pressure of SO_2, mm Hg							
	0°C	7°C	10°C	15°C	20°C	30°C	40°C	50°C
20	646	657	—	—	—	—	—	—
15	474	637	726	—	—	—	—	—
10	308	417	474	567	698	—	—	—
7.5	228	307	349	419	517	688	—	—
5.0	148	198	226	270	336	452	665	—
2.5	69	92	105	127	161	216	322	458
1.5	38	51	59	71	92	125	186	266
1.0	23.3	31	37	44	59	79	121	172
0.7	15.2	20.6	23.6	28.0	39.0	52	87	116
0.5	9.9	13.5	15.6	19.3	26.0	36	57	82
0.3	5.1	6.9	7.9	10.0	14.1	19.7		
0.1	1.2	1.5	1.75	2.2	3.2	4.7	7.5	12.0
0.05	0.6	0.7	0.75	0.8	1.2	1.7	2.8	4.7
0.02	0.25	0.3	0.3	0.3	0.5	0.6	0.5	1.3

From Sherwood, T.H., Ind. Eng. Chem., 17 (1925), 745.

The simplifications in the above expressions come about because of differences in the order of magnitudes of various rate and equilibrium constants.

Dependence of gaseous diffusion coefficient on temperature: $D \propto T^{3/2}_{\text{absolute}}$. Values for many gases can be estimated from Reid, Sherwood and Prausnitz (1977).

One other place where reaction occurs frequently is in the *combustion* process. Oxygen in the air usually combines with a fuel containing carbon, hydrogen, sulfur, hydrogen sulfide, and hydrocarbons. In addition, the nitrogen in the air will also react with oxygen at the elevated temperatures of a combustion furnace. In Table 9 many

TABLE 7
Equilibrium data for hydrogen chloride gas and water (Washburn, 1926)

Mass of HCl per 100 Masses H_2O	Partial pressure of HCl, mm Hg (Torr)				
	10°C	30°C	50°C	80°C	110°C
78.6	840	—	—	—	—
66.7	233	627	—	—	—
56.3	56.4	188	535	—	—
47.0	11.8	44.5	141	623	—
38.9	2.27	9.90	35.7	188	760
31.6	0.43	2.17	8.9	54.5	253
25.0	0.084	0.48	2.21	15.6	83
19.05	0.016	0.106	0.55	4.66	28
13.64	0.00305	0.0234	0.136	1.34	9.3
8.70	0.000583	0.00515	0.0344	0.39	3.10
4.17	0.000069	0.00077	0.0064	0.095	0.93
2.04	0.0000117	0.000151	0.00140	0.0245	0.280

TABLE 8
Molecular diffusivities and Schmidt numbers for gaseous pollutants in air at 0°C and 1 atm

			D_v	$\mathrm{Sc} = \frac{\mu}{\& D_v}$
Typical Sc	SO_2	0.40 ft²/hr	1.11×10^{-3} lb mole/ft-hr	1.28
High Scn-octane	—	0.196	0.55	2.62
Low Sc	NH_3	0.836	2.33	0.61
	CO_2	0.535	1.49	0.96
	CO	0.713	$(\mu/k)_{\mathrm{air}} = 0.512$ ft²/hr	—

Prandtl numbers for gases	$C_p\mu/k$
Air	0.69
CO_2	0.75
CO	0.72
NO_2, NO	0.72

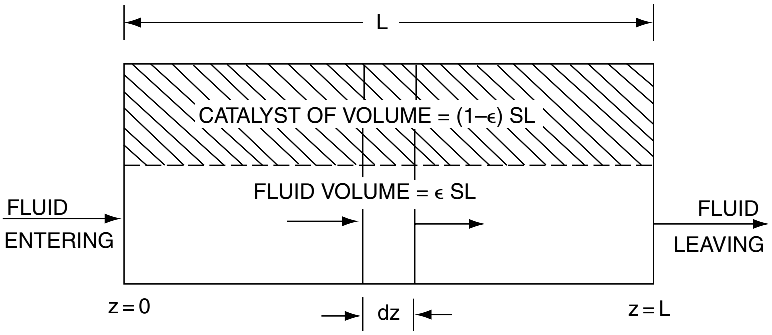


FIGURE 16

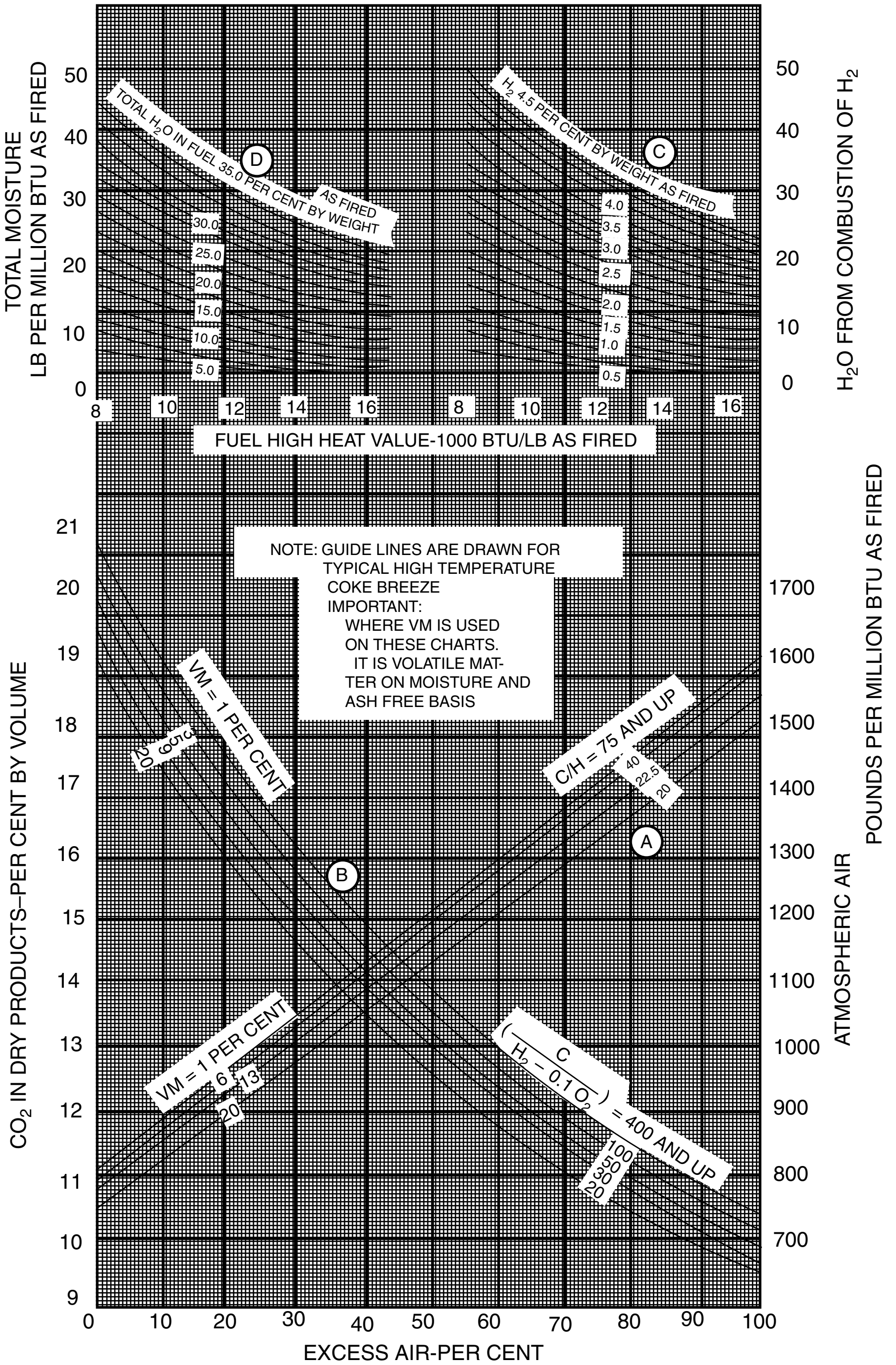


FIGURE 17 Graphical calculation curves for coal—million Btu basis (Fryling, 1966).

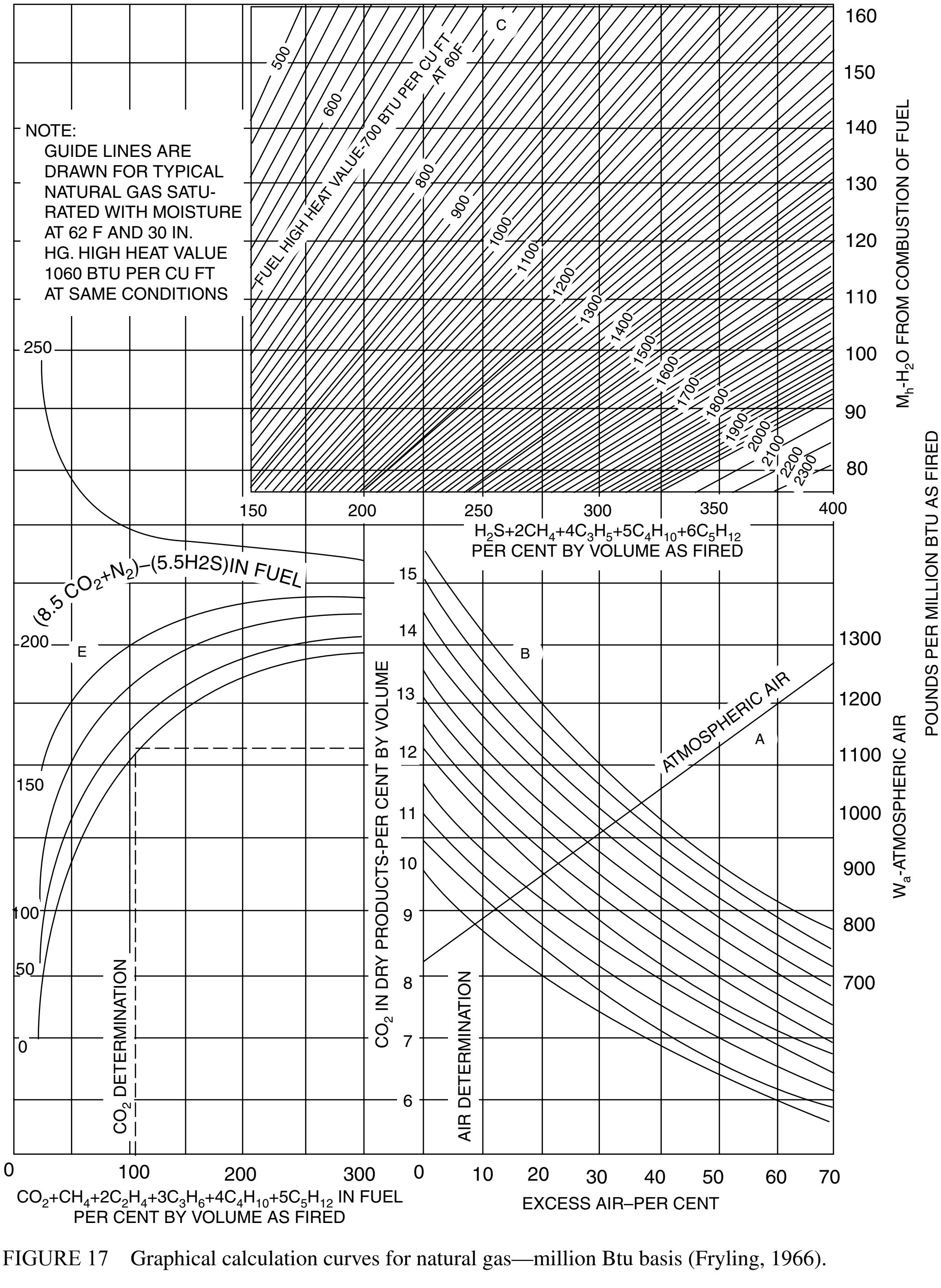


FIGURE 17 Graphical calculation curves for natural gas—million Btu basis (Fryling, 1966).

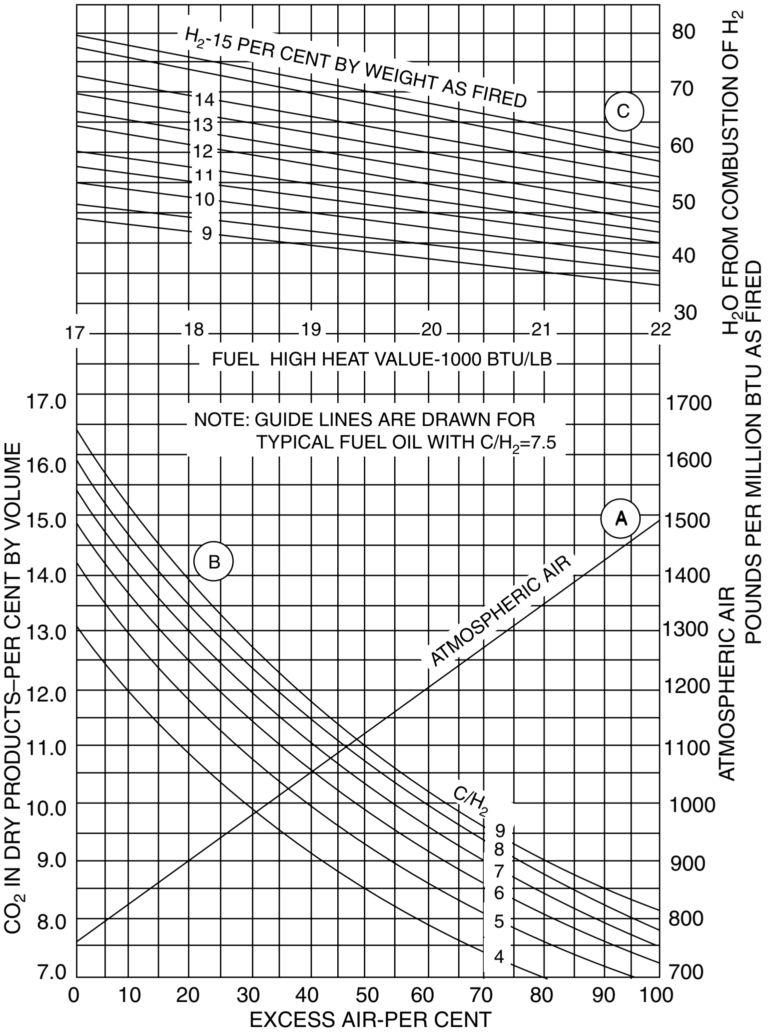


FIGURE 17 Chart for fuel oil (Fryling, 1966).

gaseous properties including the heat liberated during combustion are presented.

An excess of air is typically used over the theoretical amount of oxygen actually required for combustion; see Table 10.

In both instances safety considerations are important for preventing explosive mixtures. The detonation limits for various pure gases with air are presented in Table 12 below.

Talmage (1971) describes a flammability envelope diagram which must be considered for handling flammable vapors, Figure 18(a),(b). By appropriately adding inerts or other fuel, it is possible to operate outside of such an envelope. Another review focuses on the addition of nitrogen to combustion mixtures (Subramaniam, 1990).

REACTION RATE PARAMETERS

Table 13 is a compilation of bimolecular reaction rate constants involving typical pollutants.

The decomposition of ozone takes place with a rate constant (Laider, 1965)

$$k = 4.6 \times 10^{15}\, e^{-2400/RT} \text{ cc mole}^{-1} \text{ sec}^{-1}.$$

POLLUTANT CONTROL METHODS

Gases containing compounds of sulfur such as SO_2, SI_3, H_2S and mercaptans have received the widest attention for the purpose of control of all noxious gases. For this reason the

TABLE 9

Combustion constants of pure gases (Schmidt and List, 1962)

							Heat of Combustion$_c$			
							Btu/ft^3		Btu/lb	
No.	Substance	Formula	Mol wt[a]	Lb/ft^3[b]	ft^3/lb[b]	Sp. Gr. Air = 1.000[b]	Gross	Net[d]	Gross	Net[d]
1	Carbon	C	12.01	—	—	—	—	—	14,093[g]	14,093[g]
2	Hydrogen	H_2	2.016	0.005327	187.723	0.06959	325.0	275.0	61,100	51,623
3	Oxygen	O_2	32.000	0.08461	11.819	1.1053	—	—	—	—
4	Nitrogen (atmos.)	N_2	28.016	0.07439[e]	13.443[e]	0.9718[e]	—	—	—	—
5	Carbon monoxide	CO	28.01	0.07404	13.506	0.0672	321.8	321.8	4,347	4,347
6	Carbon dioxide	CO_2	44.01	0.1170	8.548	1.5282	—	—	—	—
7	Methane	CH_4	16.041	0.04243	23.565	0.5543	1013.2	913.1	23,879	21,520
8	Ethane	C_2H_6	30.067	0.08029[e]	12.455[e]	1.04882	1792	1041	22,320	20,432
9	Propane	C_3H_8	44.092	0.1196[e]	8.365[e]	1.5617[e]	2590	2385	21,661	19,944
10	*n*-butane	C_4H_{10}	58.118	0.1582[e]	6.321	2.06654	3370	3113	21,308	19,680
11	Isobutane	C_4H_{10}	58.118	0.1582[e]	6.321	2.06654[e]	3363	3105	21,257	19,629
12	*n*-pentane	C_5H_{12}	72.144	0.1904[e]	5.252[e]	2.4872	4016	3709	21,091	19,517
13	Isopentane	C_5H_{12}	72.144	0.1904[e]	5.252[e]	2.4872[e]	4008	3716	21,052	19,478
14	Neopentane	C_5H_{12}	72.144	0.1907[e]	5.252[e]	2.4872[e]	3993	3693	20,970	19,396
15	n-hexane	C_6H_{14}	86.169	0.2274[e]	4.398[e]	2.9704[e]	4762	4412	20,940	19,403
16	Ethylene	C_2H_4	28.051	0.07456	13.412	0.9740	1613.8	1513.2	21,644	20,295
17	Propylene	C_3H_6	42.077	0.1110[e]	9.007[e]	1.4504[e]	2336	2186	21,041	19,691
18	*n*-butene (Butylene)	C_4H_8	56.102	0.1480[e]	6.756[e]	1.9336[e]	3084	2885	20,840	19,496
19	Isobutene	C_4H_8	56.102	0.1480[e]	6.756[e]	1.9336[e]	3068	2869	20,730	19,382
20	*n*-pentene	C_5H_{10}	70.128	0.1852[e]	5.400[e]	2.4190[e]	3836	3586	20,712	19,363
21	Benzene	C_6H_6	78.107	0.2060[e]	4.852	2.6920[e]	3751	3601	18,210	17,480
22	Toluene	C_7H_8	92.132	0.2431	4.113[e]	3.1760[e]	4484	4284	18,440	17,620
23	Xylene	C_8H_{10}	106.158	0.2803[e]	3.567[e]	3.6618	5230	4980	18,650	17,760
24	Acetylene	C_2H_2	26.036	0.06971	14.344	0.9107	1499	1448	21,500	20,776
25	Naphthalene	$C_{10}H_8$	128.162	0.3384	2.955[e]	4.4208[e]	5854[f]	5654[f]	17,298[f]	16,708[f]
26	Methyl alcohol	CH_3OH	32.041	0.0846[e]	11.820[e]	1.1052	867.9	768.0	10,259	9,078
27	Ethyl alcohol	C_2H_5OH	46.067	0.1216[e]	8.221[e]	1.5890[e]	1600.3	1450.5	13,161	11,929
28	Ammonia	NH_3	17.031	0.0456[e]	21.914[e]	0.5961[e]	441.1	365.1	9,668	8,001
29	Sulphur	S	32.06	—	—	—	—	—	3,983	3,983
30	Hydrogen sulphide	H_2S	34.076	0.09109[e]	10.979[e]	1.1898[e]	647	596	7,100	6,545
31	Sulphur dioxide	SO_2	64.06	0.1733	5.770	2.264	—	—	—	—
32	Water vapor	H_2O	18.016	0.04758[e]	21.017[e]	0.6215[e]	—	—	—	—
33	Air		28.9	0.07655	13.063	1.0000	—	—	—	—

[†] From *Gaseous Fuels*, Ed. by L. Shnidman (New York: American Gas Association, 1954).

All gas volumes corrected to 60°F and 30 in. Hg dry. For gases saturated with water at 60°F, 1.73% of the Btu value must be deducted.

[a] Calculated from atomic weights, Am. Chem. Soc., February 1937.

[b] Densities calculated from values given in grams per liter at 0°C and 760 mm in the International Critical Tables allowing the known deviations from the gas laws. Where the coefficient of expansion was not available, the assumed value was taken as 0.0037 per°C. Compare the with 0.003662 which is the coefficient for a perfect gas. Where no densities were available the volume of the mol was taken as 22.4115 liters.

[c] Converted to mean Bru/lb (1/180 of the heat per pound of water from 32°F to 212°F) from data by Frederick D. Rossini, National Bureau of Standards, letter of April 10, 1937, except as noted.

TABLE 9 (*continued*)

	ft^3/ft^3 of combustible						*lb/lb of combustible*						*Experimental error in heat of combustion*
	Required for combustion			*Flue products*			*Required for combustion*			*Flue products*			
No.	*O^2*	*N_2*	*Air*	*CO_2*	*H_2O*	*N_2*	*O_2*	*N_2*	*Air*	*CO_2*	*H_2O*	*N_2*	*% ±*
1	—	—	—	—	—	—	2.664	8.863	11.527	3.664	—	8.863	0.012
2	0.5	1.882	2.382	—	1.0	1.882	7.937	26.407	34.344	—	8.937	26.407	0.015
5	0.5	1.882	2.382	1.0	—	1.882	0.571	1.900	2.471	1.571	—	1.900	0.045
7	2.0	7.528	9.528	1.0	2.0	7.528	3.990	13.275	17.265	2.744	2.246	13.275	0.033
8	3.5	13.175	16.675	2.0	3.0	13.175	3.725	12.394	16.119	2.927	1.798	12.394	0.030
9	5.0	18.821	23.821	3.0	4.0	18.821	3.629	12.074	15.703	2.994	1.634	12.074	0.023
10	6.5	24.467	30.967	4.0	5.0	24.467	3.579	11.908	15.487	3.029	1.550	11.908	0.022
11	6.5	24.467	30.967	4.0	5.0	24.467	3.579	11.908	15.487	3.029	1.550	11.908	0.019
12	8.0	30.114	38.114	5.0	6.0	30.114	3.548	11.805	15.353	3.050	1.498	11.805	0.025
13	8.0	30.114	38.114	5.0	6.0	30.114	3.548	11.805	15.353	3.050	1.498	11.805	0.071
14	8.0	30.114	38.114	5.0	6.0	30.114	3.548	11.805	15.353	3.050	1.498	11.805	0.11
15	9.5	35.760	45.260	6.0	7.0	35.760	3.528	11.738	15.266	3.064	1.464	11.738	0.05
16	3.0	11.293	14.293	2.0	2.0	11.293	3.422	11.385	14.807	3.138	1.285	11.385	0.021
17	4.5	16.939	21.439	3.0	3.0	16.939	3.422	11.385	14.807	3.138	1.285	11.385	0.031
18	6.0	22.585	28.585	4.0	4.0	22.585	3.422	11.385	14.807	3.138	1.285	11.385	0.031
19	6.0	22.585	28.585	4.0	4.0	22.585	3.422	11.385	14.807	3.138	1.285	11.385	0.031
20	7.5	28.232	35.732	5.0	5.0	28.232	3.422	11.385	14.807	3.138	1.285	11.385	0.037
21	7.5	28.232	35.732	6.0	3.0	28.232	3.073	10.224	13.297	3.381	0.692	10.224	0.12
22	9.0	33.878	42.878	7.0	4.0	33.878	3.126	10.401	13.527	3.344	0.782	10.401	0.21
23	10.5	39.524	50.024	8.0	5.0	39.524	3.165	10.530	13.695	3.317	0.849	10.530	0.36
24	2.5	9.411	11.911	2.0	1.0	9.411	3.073	10.224	13.297	3.381	0.692	10.224	0.16
25	12.0	45.170	57.170	10.0	4.0	45.170	2.996	9.968	12.964	3.434	0.562	9.968	—
26	1.5	5.646	7.146	1.0	2.0	5.646	1.498	4.984	6.482	1.374	1.125	4.984	0.027
27	3.0	11.293	14.293	2.0	3.0	11.293	2.084	6.934	9.018	1.922	1.170	6.934	0.030
28	0.75	2.823	3.573	—	1.5	3.323	1.409	4.688	6.097	—	1.587	5.511	0.088
										SO_2			—
29							0.998	3.287	4.285	1.998		3.287	0.071
				SO_2						SO_2			—
30	1.5	5.646	7.146	1.0	1.0	5.646	1.409	4.688	6.097	1.880	0.529	4.688	0.30

---

[d] Deduction from gross to net heating value determined by deducting 18,919 Btu per pound mol of water in the products of combustion. Osborne, Stimson, and Ginnings, *Mechanical Engineering*, p. 163, March 1935, and Osborne, Stimson, and Fiock, National Bureau of Standards Research Paper 209.

[e] Denotes that either the density or the coefficient of expansion has been assumed. Some of the materials cannot exist as gases at 60°F and 30 in. Hg pressure, in which case the values are theoretical ones given for ease of calculation of gas problems. Under the actual concentrations in which these materials are present their partial pressure is low enough to keep them as gases.

[f] From Third Edition of "combustion."

[g] National Bureau of Standards, RP 1141.

subject has been treated separately in the overview article entitled *Sulfur Removal.* In this section we shall consider the removal of NO_x, ozone, HF, HCl, and mercury vapor. Additional information on NO_x, HC, and CO removal may be found in the section entitled *Automotive Pollution.*

NITROGEN OXIDE

Before discussing the NO_x removal techniques, one should consider the differences between the combustion flue gas and nitric acid plant tailgas. The differences are shown in Table 14. Also in coal fired units particulate matter is present in gas phase making elimination by certain techniques impractical.

NITRIC ACID PLANT TAILGAS

The amount of NO_x emitted from nitric acid plants in the US is approximately 27,000 metric tons per year. The systems for nitric acid plants tailgas cleanup involve

TABLE 10
Excess air at furnace outlet (Frying, 1996). At inlet (Stein, 1968)

	Fuel	Excess air, %	Gas fuels	air/gas (vol)
Solid fuels	Coal	10–40	Natural	9.8
	Coke	20–40	L.P.	29.5
	Wood	25–50	Manufacturing gas	4.9
	Bagasse	25–45	Coke oven	4.9
Liquid fuels	Oil	3–15	Blast furnaces	0.7
			Producer gas	1.0
Gaseous fuels	Natural gas	5–10	Sewage gas	6.3
	Refinery gas	8–15		
	Blast furnace gas	15–25		
	Coke over gas	5–10		

TABLE 11
Typical complete combustion of bituminous coal

Gas	Weight/100 lb coal	Molar Basis	Volume fraction
SO_2	3	3/64 = 0.047	0.015
CO_2	207	207/44 = 4.70	0.145
O_2	40	40/32 = 1.25	0.040
N_2	750	750/28 = 24.8	0.800
	1000 lb	= 30.8	1,000

TABLE 12
Detonation limits

	g/m^3 with air
C_2H_2	26–844
NH_3	106–198
CoH_6	45–217
CO	145–863
C_2H_4	36–373
H_2	3–62
CH_4	35–93
H_2S	61–645

catalytic reduction, molecular sieve adsorption, extended water-absorption, tailgas scrubbing with nitric acid, and a there-stage absorber which combines gas chilling and urea scrubbing (Ricci, 1977).

There are two types of catalytic reduction: selective and nonselective catalytic reductions. Nonselective reduction uses natural gas (or H_2) as reductant to decompose the NO_x to nitrogen and oxygen. However, it has high capital and operating costs, and results in lower plant efficiency and higher CO emissions, nonselective catalytic reduction is not an attractive method. Selective catalytic reduction involves addition of a species, usually NH_3, which selectively reduces NO_x in an oxygen-containing environment to N_2 and N_2O. The selective catalytic reduction may be used to control the SO_2 and NO_x emissions simultaneously.

Removal of entrained water and acid mist by using a chilling method and mist eliminator is followed by gas drying with a desiccant. Then the gas stream passes through a molecular-sieve bed that catalytically converts NO to NO_2. Finally, the sieve selectively absorbs the NO_2. The absorbed NO_x is then released by heating and returned to the absorber. This method has high efficiency of NO_x removal. However, platinum catalyst may be lost and the method requires higher energy.

The extended water-absorption process involves routing tailgas from the existing absorber tower to a second unit for additional NO_x absorption. Counter-current washing with water produces a weak acid that is pumped to the first absorber. In order to increase the conversion of NO to NCO_2 which is relatively easy to scrub out, high pressure is required. For new plants, the pressure should surpass 150 psia.

There is no chemical consumption in the tailgas scrubbing with nitric acid process since nitric acid is recycled internally. Low energy consumption is the other advantage. However, its high liquid-gas ratio may lower tower capacity.

About 80% of the initial NO_x input is recovered as weak nitric acid in the three-stage absorber which combines gas chilling and the urea scrubbing process. The remainder is urea and ammonium nitrate. They can serve to prepare liquid fertilizers.

COMBUSTION FLUE GAS

There are two major categories of NO_x control for stationary combustion sources: combustion modification and flue gas denitrification. Combustion modification involves change of either operating or design conditions.

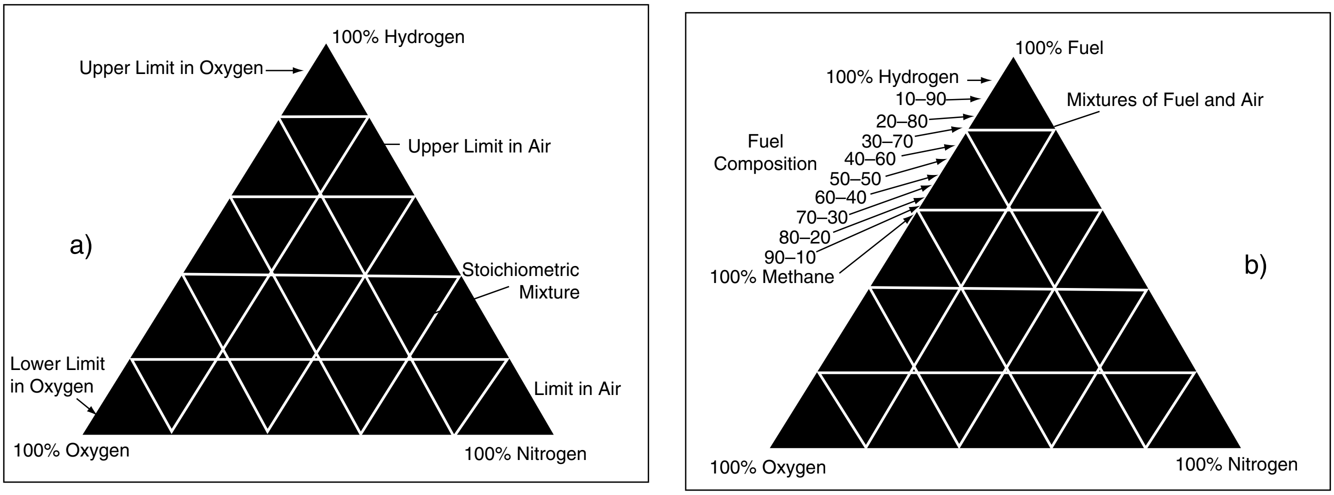


FIGURE 18 Flammability diagram. (a) Hydrogen–oxygen 1 atm 20°C. (b) Fuel–oxygen 1 atm 20°C (Talmage, 1971).

TABLE 13
Kinetic parameters for some bimolecular reactions (Laider, 1965)

Reaction	Activation energy, kcal per mole	Logarithm of frequency factor, cc mole^{-1} sec^{-1}			
		Observed	Calculated by absolute rate theory	Calculated by simple collision theory	Reference
$NO + O_3 \rightarrow NO_2 + O_3$	2.5	11.9	11.6	13.7	*a*
$NO + O_3 \rightarrow NO_3 + O$	7.0	12.8	11.1	13.8	*b*
$NO_2 + F_2 \rightarrow NO_2F + F$	10.4	12.2	11.1	13.8	*c*
$NO_2 + CO \rightarrow NO + CO_2$	31.6	13.1	12.8	13.6	*d*
$2NO_2 \rightarrow 2NO + O_2$	26.6	12.3	12.7	13.6	*e*
$NO + NO_2Cl \rightarrow NOCl + NO_2$	6.9	11.9	11.9	13.9	*f*
$2NOCl \rightarrow 2NO + Cl_2$	24.5	13.0	11.6	13.8	*g*
$NO + Cl_2 \rightarrow NOCl + Cl$	20.3	12.6	12.1	14.0	*h*
$F_2 + ClO_2 \rightarrow FClO_2 + F$	8.5	10.5	10.9	13.7	*i*
$2ClO \rightarrow Cl_2 + O_2$	0	10.8	10.0	13.4	*j*

TABLE 14
Differences between combustion flue gas and nitric acid plant tailgas

Source Item	Flue gas	Tailgas
NO_x concentration	low	high
Key component of NO_x	NO	NO_2
Flow rate	high	low
Gas pollutant	$NO_x + SO_2$	NO^x

COMBUSTION MODIFICATION

Modification of Operating Conditions The production of thermal NO_x is very temperature sensitive. Reducing the flame temperature is effective in reducing thermal NO_x production. This can be achieved by using flue gas recirculation by reduced air preheat, and by steam or water injection. In flue gas recirculation, the recirculated gas must be returned to the combustion zone. The greatest reduction in flame temperature is achieved by mixing the gas directly with the combustion air. The above methods are not as effective for coal fired boilers since coal contains high fuel nitrogen. Both thermal and fuel NO_x can be reduced by staged combustion, low excess air, reduced heat release rate, and a combination of these methods. In staged combustion, fuel is mixed with sub-stoichiometric amounts of air and burned in the first stage. In the second stage fuel burn-out is completed by injecting secondary air into the stage. Formation of NO is thereby limited in the first stage because of the low air level. By using interstage cooling, temperatures can be held down in the second stage where the excess air is injected. Low excess air decreases the NO_x emissions by reducing oxygen availability. The effectiveness of low excess air combustion has been demonstrated for gas and oil combustion by Sensenbaugh and Jonakin (1960). Low excess air combustion for coal may result in furnace slagging, smoke, flame instability, and efficiency loss due to carbon carry over. This technique is always implemented concurrently with other combustion modifications. Reduced heat release reduces the flame temperature and thereby NO_x emissions. The fuel NO_x can be reduced by changing to a fuel with lower %N. On an equivalent heat generation basis, fuels in order of decreasing NO_x formation are coal, oil, gas. Duhl (1976) pointed out that NO_x found in the methanol flue gas was lower than that produced by natural gas, and even lower than that derived from oil combustion. Southern California Edison successfully tested methanol in a 26-MW gas turbine and got NO_x emissions of 45–50 ppm against 260 ppm for distillate fuel in an identical unit (Parkinson, 1981). The effects of the modification of operating conditions to reduce NO_x emissions by fossil fuel combustion are shown in Table 15 (Ricci, 1977).

MODIFICATION OF DESIGN CONDITIONS

Low NO_x Burner Low NO*x* burners are either the delayed or minimum mixing of fuel and air method to increase the locally fuel-rich zones and thereby decrease volatile nitrogen conversion. Also secondary air or tertiary air is added to ensure the carbon burn-out. For instance, the split-flame burners, Figure 19, of Babcock and Wilcox, reduce NO_x emissions by minimizing mixing between the coal and primary air. The secondary air stream is introduced from the outer annulus to complete the carbon burn-out. According to the manufacturer, the NO_x emission is 0.45lb/10^6 BTU without overfire air or flue gas recirculation (Parkinson, 1981).

TABLE 15
How current NO_x control methods fare with fossil fuels

	NO_x reduction		
Method	Gas (%)	Oil (%)	Coal (%)
Flue-gas circulation	60	20	Not effective
Reduced combustion-air preheat	50	40	Not competitive
Steam or water injection	60	40	Not competitive
Staged combustion	55	40	40
Low excess air	20	20	20
Reduced heat-release rate	20	20	20
Combined staging, low excess air, and reduced heat release	50	35	40
Change to fuel with lower percent N	Not effective	40	20

Source: EPA report No. 650/2-74-066.

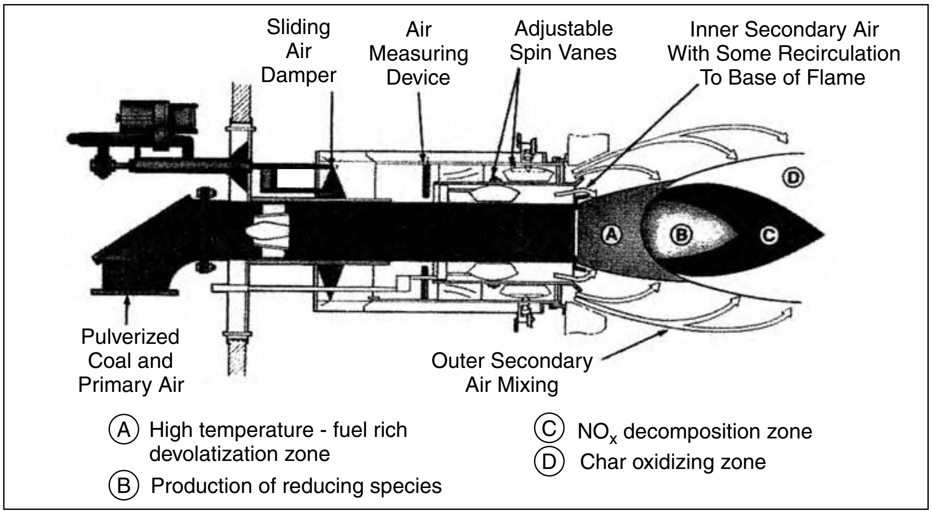


FIGURE 19 B & W dual register, low NO_x burner (1993). (See Color Plate III)).

Fluidized Bed Combustion (FBC) According to the operating pressure of the beds, there are two classifications of FBC—atmospheric pressure and pressurized. The former is appropriate for both utility and industrial heating applications. The latter is for electricity-generating plants. In FBC, crushed coal is fed to a bed of fine limestone or dolomite solvent particles which is fluidized by hot air. Water which is circulated through tubes immersed in the bed is converted into steam by the heat released during combustion. The bed temperature is in the range of 1500–1600°F. The raw limestone ($CaCO_3$) will calcine to lime (CaO) at the normal bed operating temperature. The lime will react with SO_2 released during coal combustion and form gypsum ($CaSO_4$) on the surface of the particles. The necessary sulfur capture is achieved by continuously feeding limestone and maintaining proper bed inventory.

A fluidized bed offers very high heat transfer rates with resulting low combustion temperatures. In addition, the limestone or dolomite absorbs much of the sulfur in the fuel. Thus, fluidized bed combustion promises lower SO_2 and NO_x emissions, coupled with high overall efficiency generating steam or electricity. Also it allows for the use of lower grade fuels. A comparison of EPA permissible limits with NO_x and SO_2 emissions from the 100,000 lb saturated steam/hr FBC unit at Georgetown University is shown in Table 16 (Gamble, 1980). FBC produces more particles in the offgas than conventional systems, since there is no slag to carry them off and an offgas dust collector of high efficiency is required.

FLUE GAS DENITRIFICATION

There are two major classifications of flue gas denitrification: dry and wet. (Dry techniques tests to date include selective catalytic reduction.)

Dry Processes (*SCR*) These include nonselective catalytic reduction, selective noncatalytic reduction, adsorption and electron-beam radiation. The selective catalytic reduction of NO_x to nitrogen and water by injection of ammonia in the presence of a catalyst has been developed and commercialized by a number of Japanese companies because of Japan's special need. In selective catalytic reduction of NO_x with ammonia as the reductant, the main reactions are

$$4NH_3 + 4NO + O_2 \rightarrow 4N_2 + 6H_2O$$
$$4NH_3 + 2NO_2 + O_2 \rightarrow 3N_2 + 6H_2O.$$

TABLE 16
Comparison of FBC stack emission with EPA limit (Gamble, 1980)

	EPA limit	FBC stack emission
SO_2lb/10^6 BTU	1.2	0.48
NO_xlb/10^6 BTU	0.7	0.3

The first reaction predominates when the NO concentration is about 90–95% of the NO_x in combustion flue gas. The application of SCR techniques to coal fired sources does not have long commercial experience, because of catalyst pore plugging problems in the presence of particulate fines. Two commercial units with a fixed bed "parallel passage reactor" to remove the NO_x with efficiency of 98–99% have been reported. Japan Gas Co. (JGC) claimed that this configuration allows NO_x reduction while minimizing dust plugging and pressure drop. Moving bed reactors have also been developed (Ricci, 1977).

Exxon's Thermal De NO_x system is a typical selective noncatelytic reduction process. The technique uses ammonia, but no catalyst and operates in a fairly narrow temperature range 1700–1800°F. The reaction temperature can be reduced to around 1300–1400°F by introducing hydrogen. Due to the narrow reaction temperature range, it is difficult to maintain optimum reaction temperature. However, Exxon's system claims low denitrification costs and no catalyst plugging problem. Mitsui Petrochemical Industries, Ltd. was the first one to commercially apply the process. The NO_x reduction at full load on a 120 tons/hr heavy fuel oil fired steam boiler is about 50%.

As previously described, the molecular sieves can be used to reduce the NO_x emissions from nitric acid plants. NO_x reduction using activated carbon adsorption is tested in pilot plant. The Activated carbon can (1) adsorb NO_2 to form nitric acid (2) promote the oxidation of NO to NO_2 (3) catalyze reduction NO to N_2. Chemical impregnations, such as those of copper and vanadium, enhance carbon's catalytic effect on NO_x/NH_3 reaction. NH_3 is injected into the column. The carbon in the columns adsorb SO_2, NO_x reacts with ammonia to form N_2 and water.

Wet Processes Wet processes fall into four categories, absorption-reduction, oxidation-absorption-reduction, absorption-oxidation, and oxidation-absorption. The classification system for wet NO_x removal processes is shown in Figure 20 (Fawcett *et al.*, 1977). Due to the high capital and operating costs, and the formation of NO_3 containing wastewater, the wet NO*x* removal processes haven't been able to match the commercial gains posted by selective catalytic reduction and other dry techniques. The details of dry and wet methods for NO*x* removal have been reviewed by Fawcett *et al.* (1977). Also, Siddiqi and Tenini (1981) reviewed the commercial flue gas treating applications. The flue gas treating applications in the US and Japan shown in Table 17.

FLUORIDES

Not all fluoride emissions are gases. The system employed must have the capability of handling particulates, dispersed in huge quantities of air. Phosphoric acid production is a typical fluoride producer.

The process begins with the grinding of rock phosphate (typically $Ca_{10}F_2(PO_4)_6$). This is naturally a source of particulates, as is the drying of the pulverized phosphate. Acidulation

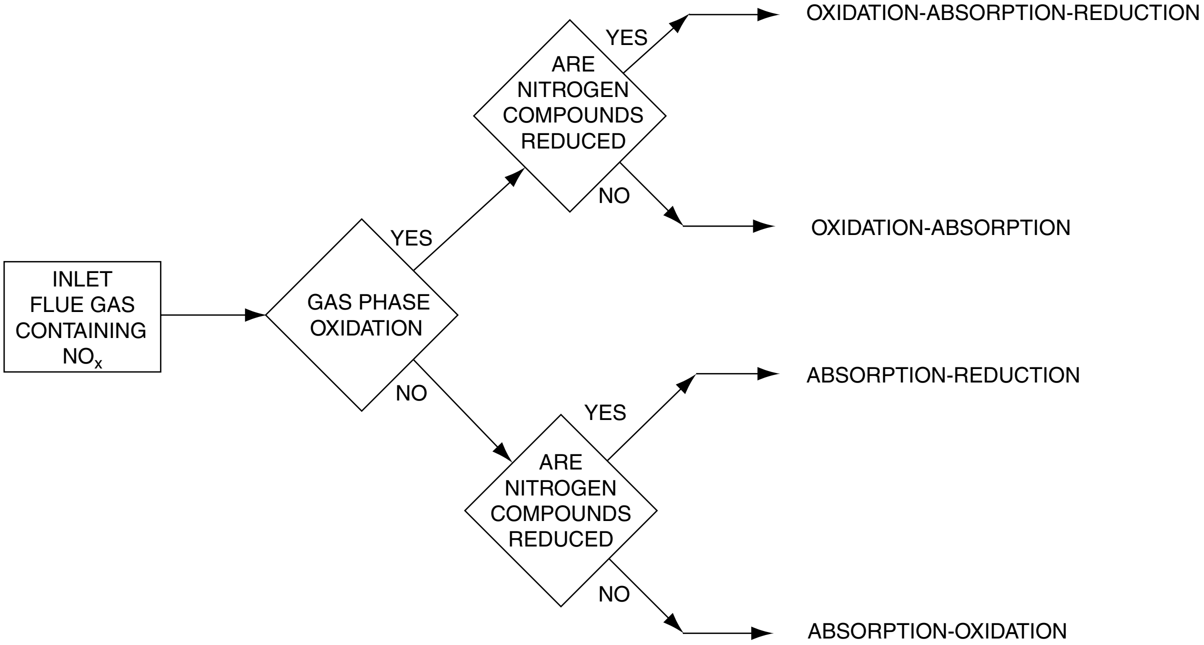


FIGURE 20 Classification system for wet NO_x removal processes (Fawcett *et al.*, 1977).

TABLE 17
Flue gas treating applications in the US and Japan

Process type	For	Units	Per cent of total
1 Selective catalytic reduction	NO_x	70	63.7
2 Selective reduction Exxon Thermal DeNOx	NO_x	25	22.7
3 Shell/UOP selective	NO_x and SO_2	2	1.8
4 Shell/UOP selective reduction-adsorption	NO_x and SO_2	2	1.8
5 Wet oxidation-adsorption reduction	NO_x and SO_2	9	8.2
6 Wet adsorption-reduction	NO_x and SO_2	4	3.6
	110	100	

with sulphuric acid produces the gaseous pollutants associated with this process, most notably hydrogen fluoride, and silicon tetrafluoride. Both of these gases can be scrubbed in aqueous medium, but each one causes a severe handling problem, after scrubbing. The HF is highly corrosive, and the SiF_4 combines with water to form a gelatinous mass which can clog an improperly designed scrubber.

Another constraint is the extremely low effluent concentrations allowed by law, about 0.02 mg/ft^3. Since each unit in the process stream is a source, there are at least 250,000 ft^3/min which much be scrubbed down from 200–500 mg/ft^3. This is a reasonable requirement, considering the toxicity of the pollutants, but it is still a difficult objective to attain.

Teller claims to solve all these problems at once with a cross flow scrubber and a high water flow rate. The device is a baffled tank with wetted baffles so that any solids which do not immediately dissolve will just drop to the bottom after colliding and not clog the air ducts. The large water flow rate keeps the fluorine concentration low, so that the same efficiency (4.8 transfer units) typical of a countercurrent flow scrubber can be maintained.

Cost factors cause the designer to look very hard at the size of this system and its pumping costs. It is obvious that a scrubber which has a greater than necessary number of transfer units will be un-economical. Evan with a tight design on both size and pressure drop, the enormous volumes of gas can still make this system undesirable. A way must be found to decrease the volume.

The best way to do this is by taking advantage of the fact that not all the units produce pollutants at the same concentration level. Since the total flow is large enough that multiple scrubbers would probably be necessary anyway, the first scrubbers in the line take the dirtiest gas and clean it to the level of the next stream. The two are mixed and sent to the next scrubber, and so on. Only the last scrubber need be large enough to accommodate the entire flow. This should solve the air pollution problems to a sufficient degree.

There are two problems which deserve mention: these are water pollution from the dissolved fluorides and solid waste sludges which slough off the baffles. Both of these are highly contaminated and must ultimately lead to a disposal problem.

Teller suggests the water problem be solved by the use of a closed cycle holding pond. Byproduct recovery would appear to be a possible way to solve this problem, but the situation requires further study. There is mention of neutralization of HF, but this will give a CaF_2 precipitate, which would eventually fill even a large holding pond. Regardless of the method chosen, the water must be cleaned of excess fluoride before it can be recycled.

Efficiency of fume control in the aluminium smelting industry has been the subject of a thorough study (Cook *et al.*, 1971). They note that the recovery process may be divided into two categories: "interception of total particulate which includes both fluorine compounds and relatively innocuous aluminium oxide.

Hydrogen fluoride is by far the dominant gaseous fluorine compound in pot gas. Recovery of HF is a function of surface area of alumina, the HF concentration in pot gas delivered to the reactors, quantity of alumina through the system, and contacting efficiency between pot gas and alumina. Reaction between HF and alumina is rapid and with proper reactor design, contacting is no problem. With a sufficient feed rate of alumina to the reactor, reaction efficiency remains high even at very low bed levels. On the other hand, insufficient feed rate or total cessation of feed for a period of time can impair the efficient recovery of HF as the bed becomes saturated. Similarly high concentration of HF in pot gas due to poor operating conditions also can overload the sorptive capacity of available alumina. Provided an adequate alumina supply, the recovery of HG gas is invariably above 99% and frequently above 99.9%.

The interception of particulate fluoride compounds takes place primarily in the fluid bed of the reactor. Fluoride concentration in the bed is found to be about 3½% nearly twice the concentration of hydrogen fluoride sorbed in alumina.

The final interception of particulate, both alumina and fluoride, is a mechanical filtration step which depends upon bag condition. Old or worn bags or poor connection to the tube sheet cause particulate loss. Bag fabric quality, a highly variable factor in its own right, also weighs heavily on particulate loss. Bag fabric quality, a highly variable factor in its own right, also weighs heavily on particulate recovery performance.

Solids efficiency tends to be less than gas efficiency, although this is not always the case. The two are not related because of factors previously discussed. As we are more vulnerable to particulate losses than to gas phase fluoride losses, so are we more vulnerable to fluoride loss in particulate than to alumina loss. About 10% of total particulate loss is fluoride compounds.

The Alcoa 398 system, or any other treatment system, can only capture fluorides from pot fume collected and delivered to the system. Hooding and fume duct, therefore, become important parts of the control system. In newer potlines, hooding is designed to achieve 95% interception.

"The recovery of fluorine in a directly reusable form in the process results in reduced fluoride additions to smelting installations. In addition to the performance measurements taken from operating reactors and pot-room monitors, patterns of fluorine consumption by individual pots, potlines, and entire plants were examined. Observation of the cells served by the prototype installation or Alcoa 398 Process revealed a reduction in aluminium fluoride consumption from 101 lb/ton aluminium produced to 32lb/ton aluminium. Since AlF3 is added to bath neutralize Na_2O incoming with alumina as well as to compensate for loss of fluorides to potlining and atmosphere, the fluorine cycle in smelting cells is complex, and part of this reduction in aluminium fluoride consumption was compensated by additional cryolite requirements. A more general measure of fluoride recovery was derived by evaluation of calcium fluoride sludge collected in ponds serving wet scrubber installations. All of the HF and solids which end in a sludge pond, and a little more besides, are returned directly to the pots with this process. On a fluorine basis, 30 lb of fluorine are retrieved per ton of aluminium produced."

Packed towers have also been used successfully (Specht and Calaceto, 1967) particularly in phosphate fertilizer manufacture.

Use of the packed tower principle has been twofold:

1) To handle high SiF_4 gas loadings in a single, long vessel. When high SiF_4 gas loading are scrubbed, the plugging effects can be disastrous unless there are provisions for a spray chamber to reduce the concentration of SiF_4 and for the subsequent removal of silica. A working design includes a cross-flow packed section for optimum removal of SiF_4 and then a counterflow packed section.
2) To handle very weak SiF_4 concentration (less than 0.5mg as F/std. cu. ft.) in a simplified counterflow tower or to serve as a positive entrainment eliminator.

"Generally, the cyclonic scrubber or packed tower requires 7 to 10 gal for every 30mg of F reduction. Depending on the severity of service, liquid pressure at the cyclonic and spray chamber may be 50 to 1001lb/sq in. gauge, while the liquid pressure at the packing area, depending on the method of distribution, may be 5 to 20lb/sq in. gauge."

Mercury Vapor

Activated carbon, impregnated with elemental sulfur, has been found effective for removing mercury vapor from air or other gas streams (Lovett and Cunniff, 1974). Impregnating the sulfur on activated carbon can increase the reaction rate of mercury vapor and sulfur due to the enormous internal surface area. When the carbon is spent, mercury can be recovered by thermal oxidation in a retort.

Measurements of losses through roof vents help to determine overall fume control performance as summarized in the *Indoor Air Pollution* article. The emission factors for various sources are available from the US EPA's Air Chief software. These factors, referred to as AP-42 factors, are presented for uncontrolled facilities as well as for facilities having air pollution control equipment in place. Factors for which control technology is indicated, do not necessarily reflect the best available or state-of-the-art controls, but rather reflect the level of (typical) control for which data were available at the time the information was published.

The factors are strongly influenced by the type of fuel being combusted and a separate chapter is devoted to EXTERNAL COMBUSTION SOURCES. These include steam/electric generating plants, industrial boilers, and commercial and domestic combustion units. Coal, fuel oil, and natural gas are the major fossil fuels used by these sources. Liquefied petroleum fuels are also used in relatively small quantities. A number of other fuel types are reviewed. Power generation, process heating, and space heating are some of the largest fuel combustion sources of sulfur oxides, nitrogen oxides, and particulate emissions. The commercially available control technology for removing gaseous pollutants is discussed for each fuel. In addition to the external combustion sources, other categories covered are as follows:

(1) Solid Waste Disposal, (2) Stationary Internal Combustion Sources, (3) Evaporation Loss Sources, (4) Petroleum Industry, (5) Organic Chemical Process Industry, (6) Liquid Storage Tanks, (7) Inorganic Chemical Industry, (8) Food And Agricultural Industries, (9) Wood Product Industries, (10) Mineral Products Industry, (11) Metallurgical Industry, (12) Miscellaneous Sources, and (13) Greenhouse Gas Biogenic Sources.

Because emission factors essentially represent an average of a range of emission rates, approximately half of the subject sources will have emission rates greater than the emission factor and the other half will have emission rates less than the factor. As such, a permit limit using an AP-42 emission factor would result in half of the sources being in noncompliance. Also, for some sources, sources often are tested more frequently when they are new and when they are believed to be operating properly, and either situation may bias the results.

REFERENCES

1. Andrew, S.P.S., *Alta Technologia Chimica*, Roma. P. 153, 1961.
2. Anonymous, Federal Register, **43,** No. 182, 42156, 1978.
3. Anonymous, National Air Pol. Contr. Admin., U.S. Depart. of H.E.W Report AP-66, March 1970.
4. Astarita, G., *Mass Transfer with Chemical Reaction,* Elsevier Publishing Co., Amsterdam, 1967.
5. Bird, R. *et al., Transport Phenomena,* Wiley, 1960.
6. Brooman, D.L. and E. Edgerley, Jr., *J.A.P.C.A.* **16,** 25, 1966.
7. Brunaner, S., P.H. Emmett and E. Teller, *J.A.C.S.* **60,** 309, 1938.
8. Cadle, R.D., in *Air Pollution Handbook,* Magill *et al.,* eds., McGraw-Hill, NY 1956.
9. Calderbank, P.H. and M.B. Moo-Young, *Chem. Eng. Sci.,* **16,** 39, 1961.
10. Charpentier, J.C., Recent progress in two phase gas-liquid mass transfer in packed beds, *The Chem. Eng. J.,* **11,** 161, 1976.
11. Cook, W.H. and D. Basmadjion, *Can. J. of Chem. Eng.,* **42,** 146, 1964.

11a. Cooper, C.D. and F.C. Alley, Air Pollution Control a Design Approach, Waveland Press; 2nd Ed. (1994).

12. Danckert, P.V., *Gas Liquid Reaction,* McGraw-Hill, NY 1970.
13. Devitt, T.W., B.A. Laseke and N. Kaplan, Utility flue gas desulfurization in the U.S., *Chem. Eng. Prog.,* **76**(5), p. 45, 1980.
14. Dulh, R.W., Methanol as boiler fuel, *Chem. Eng. Prog.,* **72**(7), p. 75, 1976.
15. Ermence, E.D., *Chem. Eng. Prog.,* **52,** p. 1956.
16. Faucett, H.L., J.D. Maxwell and F.A. Burnett, Technical assessment of NO_x removal process for utility application, *Joint Report, EPRIAF-568* and *EPA-60017-77-127,* 1977.
17. Fryling, G.R. ed., *Combustion Engineering* C.E. Inc., 1966.
18. Gamble, R.L., Fluidized bed combustion at Georgetown University, *Heat Engineering,* April–June, p. 81, 1980.
19. Gerhard, E.R., in *Air Pollution Handbook* Magill *et al.,* eds., McGraw-Hill, NY 1956.
20. Hanst, P.L., *J.A.P.C.A.* **21,** 269, 1971.
21. Hill, G.R., in *Energy Technology IV* (Proc. 4th Energy Technology Conference, Washington, D.C.) p. 154, R.F. Hill ed., Government Institutes, Inc., 1977.
22. Kangas, J. *et al., Chem. Eng* Sept. **6,** 55–57, 1971.
23. Krevelen, V. and P.J. Hoftijzer, *Chimie et Industrie: Numero Speiale du XXIe Congre International de Chimie Industrielle,* Bruxelles, p. 168, 1948 (cited by Dackwert, 1970).
24. Laider, K.J., *Chemical Kinetics,* McGraw-Hill, N.Y., 1965.
25. Langmuir, I., *J.A.C.S.* **40,** 1361, 1918.
26. Laurent, A. and J.C. Charpentier, *The Chem. Eng. J* **8,** 85, 1974.
27. Lee, D.R., *Heat Piping and Air Cond.,* **42**(4), p. 76, 1970.
28. Lovett, W.D. and F.T. Cunniff, Air pollution control by activated carbon, *Chem. Eng. Prog* **70**(s), 43, 1974.
29. Mars, P. and J.G.H. Maessen, *J. Catal.,* **10,** 1, 1968.
30. Mohunta, D.M., A.S. Vaidynanathan and G.S. Laddha, *Indian Chem. Eng.,* **11,** 39, 1969 (Cited by Charpentier, 1976).
31. McCarthy, J.E., Scrubber types and selection criteria, *Chem. Eng. Prog.,* **76**(5), p. 58, 1980.
32. McCutchen, G.D., NO_x Emissions trends and Federal Regulation, *Chem. Eng. Prog.,* **73**(8), p. 58, 1977.
33. Niepenburg, H., *J. Fuel Heat Technol.,* **14,** 34, 1966.
34. Onda, K., H. Takeuchi and Y. Okumato, *J. Chem. Eng.,* (Japan) **1,** 56, 1968.
35. Onda, K., H. Takeuchi and Koyama, Kagaku Kogaku, **31,** 126, 1967.
36. Parkinson, G., NO_x Controls: Many new systems undergo trials, *Chem. Eng.,* March 9, p. 39, 1981.
37. Princiotta, F.T., Advances in SO_2 Stack Gas Scrubbing, *Chem. Eng. Prog.,* **74**(2), p. 58, 1978.
38. Ricci, L.J., Nixing NO_x *Chem. Eng.* Feb. 14. p. 33, April 11, p. 48, and April 25, p. 70, 1977.
39. Ray, G.C. and E.O. Boxes, Jr., *Ind and Eng.,* **42,** 1315, 1950.
40. Reid, R., T.K. Sherwood and J.M. Prausnitz, *The Properties of Gases and Liquids,* McGraw-Hill, N.Y., 1977.
41. Sensenbauge, J.D., and J. Jonakin, Effect of combustion conditions on nitrogen oxide formation in boiler furnaces, *Am. Soc. Mech. Eng.*, Paper 60-WA-334, 1960.
42. Sherwood, T.K., R.L. Pigford and C.R. Wilke, *Mass Transfer,* McGraw-Hill, N.Y., 1975.
43. Schmidt, A.X. and H. List, *Material and Energy Balances,* Prentice Hall, 1962.
44. Specht, R.C. and R.R. Calaceto, *Chem. Eng. Prog.*, **63,** May, p. 78, 1967.

45. Stern, A.C., et., *Air Pollution,* **1–3,** Academic Press, 1968.
46. Strauss, W., *Industrial Gas Cleaning,* Pergaman Press Ltd., 1968.
47. Talmage, W.P., *Chem. Tech.*, p. 117, Feb. 1971.
48. Ziegler, E.N. and R.E. Meyers, Control technology for coal-fired combustion in Northeastern E.S., *Water, Air, and Soil Pollution,* **12,** 371, 1979.
49. Washburn, E.W., ed., *The Critical Tables,* McGraw-Hill, N.Y., 1926.
50. Siddiqi, A.A. and Tenini, J.W., Hydrocarbon Processing, **60,** No. 10, p. 115, 1981.
51. Stromblad, C., Absorption of Nitrogen Oxides-Licentiate Dissertation, University of Lund-Sweden, 1988.
52. Subramaniam, T.K. and J.V. Cangelos; *Chem. Eng.,* Dec. 1989, p. 108.
53. *Ibid,* Jan. 1990, p. 8.
54. Air Chief, Emission Factor and Inventory Group, EMAD/OAQPS, EPA 454/c-04–001 Version 11, Jan 2004.

TAI-MING CHIU
Institute of Nuclear Energy Research

EDWARD N. ZIEGLER
Polytechnic University

VAPOR AND GASEOUS POLLUTANT FUNDAMENTALS: see **Nitrogen Oxides Reduction**

W

WATER AND WASTE MANAGEMENT SYSTEMS IN SPACE

INTRODUCTION

Environmental Control and Life Support System (ECLSS) is the NASA terminology for the systems which allow people to exist and work in confined spaces and in uninhabited locations and hostile environments. These systems supply air, water, temperature and humidity controls to enable personnel to survive and work under hostile conditions. The purpose of this paper is to review the various devices that have been developed for the recovery and reuse of liquid wastes in spacecraft and space stations, and their possible terrestrial applications.

A SHORT HISTORY OF U.S. SPACECRAFT ECLS SYSTEMS

The U.S. spacecraft ECLS systems have grown in capabilities and complexity as the spacecraft have grown and the duration of their missions has been extended. Initial systems were almost completely open loop, with only CO_2 removal in a closed loop. There was no recycling of water or wastes. On the Orbiter, water is not recycled, but is vented overboard. Air is recycled. Solid wastes are returned to the earth. This system will be changed as the Orbiter is reconfigured for longer duration missions. Due to the high cost of transporting fresh water and oxygen into orbit, the space station is designed to operate in an almost completely closed loop mode. Only solid body waste and trash will be returned to the earth. Air and liquid wastes will be recycled.

Regenerative systems have been evaluated by NASA for the different functions of the ECLSS (Figure 1).[1] A partial list of the various technologies is given in Table 1. The emphasis is on purification and recycling the air and wastewater streams for atmosphere revitalization and water recovery and management. The high cost of transporting mass into orbit, approximately \$22,000/kg, causes NASA to place a premium on low weight, low volume, high efficiency, and low maintenance requirements.[2]

DRINKING WATER QUALITY STANDARDS

NASA has very strict standards for water consumed on the shuttle, and even stricter standards for the proposed space station. This is logical, as the water on the space station will be continuously recycled, and hence the astronauts will be exposed to this water for an extended duration. Table 2 lists the present potable water standards for the shuttle and Space Station Freedom (SSF), and compares these to the existing EPA potable water standards. The SSF hygiene water standards are also shown in the table. These standards were established early in the program.[3]

In order that hardware designers should have constant reference solutions to use when testing their equipment, NASA found it necessary to determine the compositions of urine and wastewater that would be produced on the space station. Hence NASA analyzed average urine composition, and determined a chemical model for urine, which could be made up and used to test recycling equipment. Similarly, a chemical model formula for hygiene water, i.e., water used for personal washing, laundry, and food preparation, was derived. These model compositions for urine and wastewater were then used to test individual items of hardware and complete processing systems. However, the final evaluation tests must include the processing of urine and wastewater obtained from volunteers living in a closed test environment.

TECHNOLOGIES TESTED FOR USE IN THE ECLSS

Various technologies were tested and/or evaluated for the ECLSS systems, with the emphasis on the water and air revitalization systems. Less attention has been paid to the treatment of solid waste, as this will only be required for long space voyages and for long term inhabitation of stations on the moon and planets. A number of water recovery and waste management systems have been tested for application in the space station. These systems have been evaluated by comparing them to the baseline technology that was originally

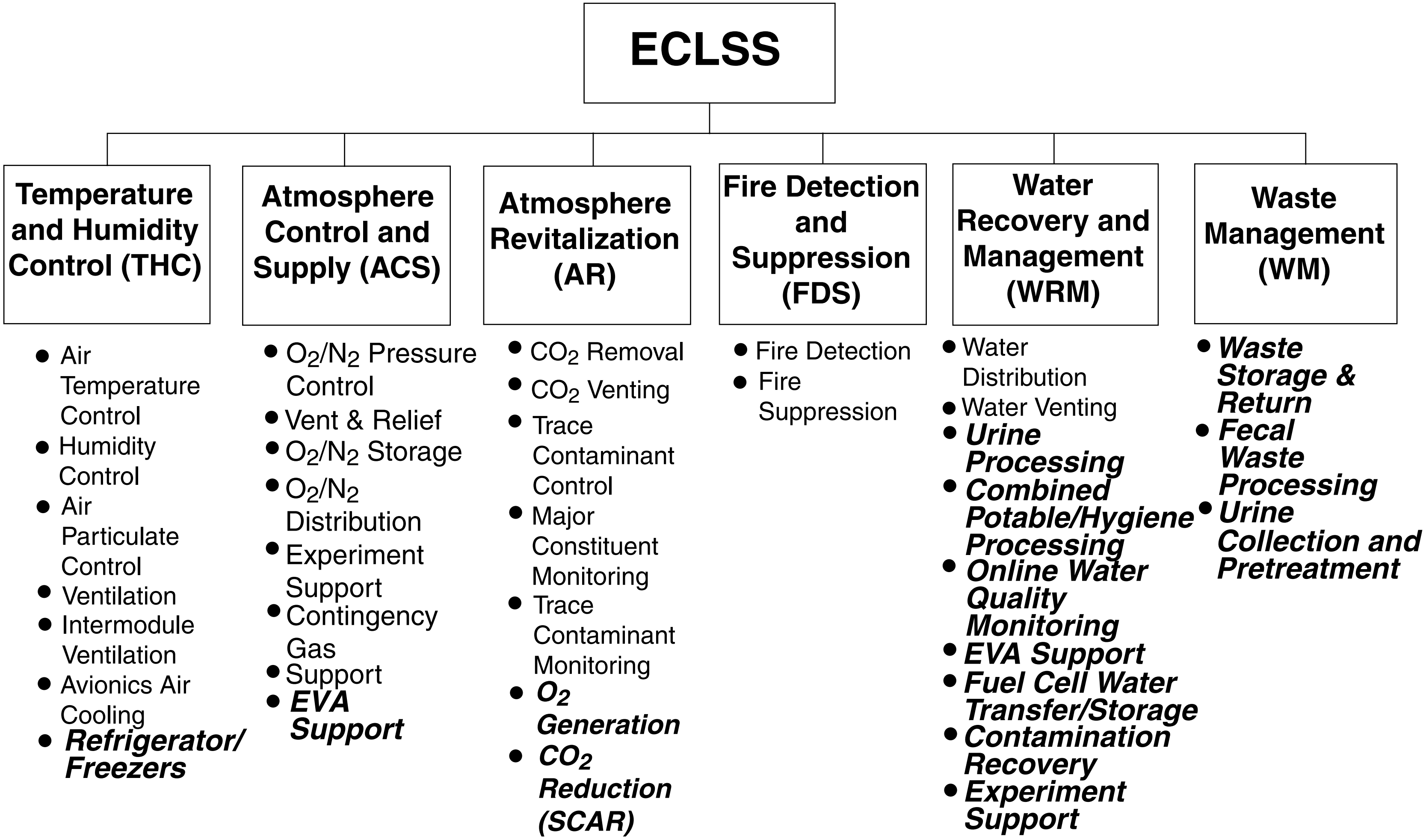


FIGURE 1 Space Station Alpha Environmental Control and Life Support System (ECLSS) Functions. *Source:* Ref. 4.

selected for the space station. The original baseline system was the Thermoelectric Integrated Membrane Evaporation Subsystem (TIMES). This was replaced by a new baseline subsystem which is a combination of multibed filtration (Unibed) and Vapor Compression Distillation (VCD). These baseline systems along with many other promising technologies are described below.

In all cases urine is pretreated immediately upon collection to minimize bacterial growth and possible damage to collection and handling equipment. In addition to inhibiting bacteria, the treatment is intended to control ammonia, stabilize dissolved solids to prevent precipitation, and allow the overall subsystem to function in a zero-gravity environment. Present pretreatment comprises the injection of either Oxone

TABLE 1
ECLSS Technologies Used or Evaluated

ECLSS Subsystem Category	Used/Evaluated	Technology
Atmosphere revitalisation	Used	LiOH
	Used	Molecular sieve
	Used	Sabatier reactor
	Used	Static feed water electrolysis
	Evaluated	Solid amine fixed bed
	Evaluated	Liquid sorbent closed loop
	Evaluated	Bosch system
	Evaluated	Algal bioreactor
	Evaluated	Growing green plants
Trace Contaminant Removal	Used	Activated charcoal
	Used	Catalytic oxidiser
	Used	Particulate filters
Water recovery and management	Used*	Vapor compression distillation
	Used	Chlorine
	Used	Sodium hypochlorite injection
	Used	Iodine injection
	Used	Heat sterilisation
	Used	Fuel cell by-product water
	Evaluated*	Unibed filter
	Evaluated*	TIMES membrane filter
	Evaluated*	Reverse osmosis
	Evaluated*	Electrodialysis
	Evaluated*	Electrooxidation
	Evaluation*	Supercritical water oxidation
	Evaluation*	Electrodeionisation
	Evaluation*	Air evaporation
	Evaluation*	Vapor phase catalytic ammonia removal
	Evaluation*	Immobilized cell or enzyme bioreactors
	Evaluation*	Plant transpiration and water recovery
Temperature and humidity control	Used	Condensing heat exchangers
	Used	Water cooled suits
Atmosphere control and supply	Used	Compressed gas storage
	Used	Cryogenic gas storage
Waste management	Used	Urine stored in bags
	Used	Feces stored in bags
	Used	Urine vented
	Used	Feces stored in bags and vacuum dried
	Used	Urine stored in tank and vented
	Used	Feces stored in bags and compacted

* Described in this paper.

TABLE 2
Water Quality Standards Comparison

QUALITY PARAMETERS	Shuttle Potable	EPA Current	SSF Potable	SSF Hygiene
PHYSICAL PARAMETERS	—	—	—	—
TOTAL SOLIDS (mg/L)	5 to 10	500	100	500
"COLOR, TRUE (Pt/Co units)"	15	15	15	15
CONDUCTIVITY	Reference only	—	—	—
TASTE (TTN)	3	—	3	N/A
ODOR (TON)	3	3	3	3
PARTICULATES (max size—microns)	—	—	40	40
pH	Reference only	6.5–8.5	6.0–8.5	5.0–8.5
TURBIDITY (NTU)	11	1	1	1
DISSOLVED GAS (free @ 37°C)	None at 1 AT	—	(see Note 1)	N/A
FREE GAS (@ STP)	—	—	(see Note 1)	Note 1
INORGANIC CONSTITUENTS (mg/L)(see Notes 2 and 5)	—	—	—	—
ALUMINUM	—	0.05	—	—
AMMONIA	—	—	0.5	05
ARSENIC	—	0.05	0.01	0.01
BARIUM	—	2	1	1
CADMIUM	0.01	0.01	0.005	0.005
CALCIUM	—	—	30	30
CHLORINE (Total—Includes chloride)	—	250	200	200
CHROMIUM	0.05	0.1	0.05	0.05
COPPER	1.00	1	1.00	1.00
IODINE (Total—Includes organic iodine)	Reference only	—	15	15
IRON	0.3	0.3	0.3	0.3
LEAD	0.05	0.05	0.05	0.05
MAGNESIUM	—	—	50	50
MANGANESE	0.05	0.05	0.05	0.05
MERCURY	0.005	0.002	0.002	0.002
NICKEL	0.05	0.1	0.05	0.05
NITRATE (NO_3-N)	—	10	10	10
POTASSIUM	1	—	340	340
SELENIUM	0.01	0.05	0.01	0.01
SILVER	0.1	0.1	0.05	0.05
SULFATE	—	250	250	250
SULFIDE	—	—	0.05	0.05
ZINC	5	5	5	5
BACTERICIDE (mg/L)	—	—	—	—
RESIDUAL IODINE (minimum)	Reference only	—	0.5	0.5
RESIDUAL IODINE (maximum)	Reference only	—	4	6
AESTHETICS (mg/L)	—	—	—	—
CATIONS	—	—	30	N/A
ANIONS	—	—	30	N/A
CO_2 (see Note 1)	—	—	15	N/A

(*continued*)

TABLE 2 (*continued*)

QUALITY PARAMETERS	Shuttle Potable	EPA Current	SSF Potable	SSF Hygiene
MICROBIAL	None Viable	—	—	—
BACTERIA (CFU/100 mL)	—	—	—	—
TOTAL COUNT	—	—	1	1
ANAEROBES	—	—	1	1
COLIFORM	—	1	1	1
ENTERIC	—	—	—	—
VIRUS (PFU/100 mL)	—	—	1	1
YEAST and MOLD (CFU/100 mL)	—	—	1	1
RADIOACTIVE CONSTITUENTS (pCi/L)	**NRC LIMITS**	**(see Note3)**	—	—
ORGANIC PARAMETERS (mg/L) (See Note 2)	—	—	—	—
TOTAL ACIDS	—	—	500	500
CYANIDE (total including organic cyanides)	—	—	200	200
HALOGENATED HYDROCARBONS	—	0.1 (THM)*	10	10
TOTAL PHENOLS	—	—	1	1
TOTAL ALCOHOLS	—	—	500	500
TOTAL ORGANIC CARBON (TOC)	Reference only	—	500	10000
UNCHARACTERIZED TOC (UTOC)	—	—	—	—
(see Note 4)	—	—	100	1000
ORGANIC CONSTITUENTS (mg/L)	—	—	—	—
(see Notes 2 and 5)	—	—	—	—

Note 1: No detectable gas using a volumetric gas vs. fluid measurement system. This excludes CO_2 used for aesthetics purposes.
Note 2: MCLs considered independently of others.
Note 3: The maximum contaminant levels for radioactive constituents in potable and personal hygiene water shall conform to Nuclear Regulatory Commission (NRC) regulations (10CFR20, *et al.*). These maximum contaminant levels are listed in the "Federal Register Vol 51, No: 6, 1986, Appendix B, Table 2.
Note 4: Total organic carbon minus identifiable organic contaminants.
Note 5: MCLs for others, if found, will be established as necessary.

* THM = Trihalomethanes
SSF = Space Station Freedom
Source: Ref 3.

(a potassium monopersulfate compound) combined with sulfuric acid, or hypochlorite (bleach) before distillation.

Thermoelectric Integrated Membrane Subsystem (TIMES)

This was the original baseline water recovery subsystem for the space station (Figure 2). Wastewater is heated to 66°C in a heat exchanger, and is then pumped through bundles of small diameter hollow Nafion fiber membranes in the evaporator module. The Nafion allows only water, gases, and small neutral molecules to pass through. The pressure on the exterior of the membrane is reduced to 17 kPa (2.5 psi) to assist in both transport through the membrane, and subsequent evaporation. The water vapor is then condensed, and the latent heat of condensation is conducted to the input heat exchanger. The TIMES system was found to produce poorer quality water than the VCD.[4,5]

Vapor Compression Distillation (VCD)

The VCD process involves spreading a thin film of the wastewater on the inside wall of a thin-walled rotating drum under low pressure, typically about 4.8 kPa (0.7 psi). The VCD is shown in Figure 3, and in-depth discussions may be found in Refs. 4, and 6–11. The system operates at 35°C which is slightly above ambient temperature. Heat is applied to the outside of the wall causing the thin film of water to boil. The vapor is extracted from the drum interior and compressed by a pump. The compressed vapor is then condensed on the exterior wall of the drum. The compressed vapor condenses at a higher temperature than that at which it had originally evaporated. The latent heat of condensation thus supplied the heat required to evaporate the original feed water in the inside of the drum. The unevaporated brine, heavily loaded with contaminants, is recycled back to the inflow stream, or passed to another subsystem.

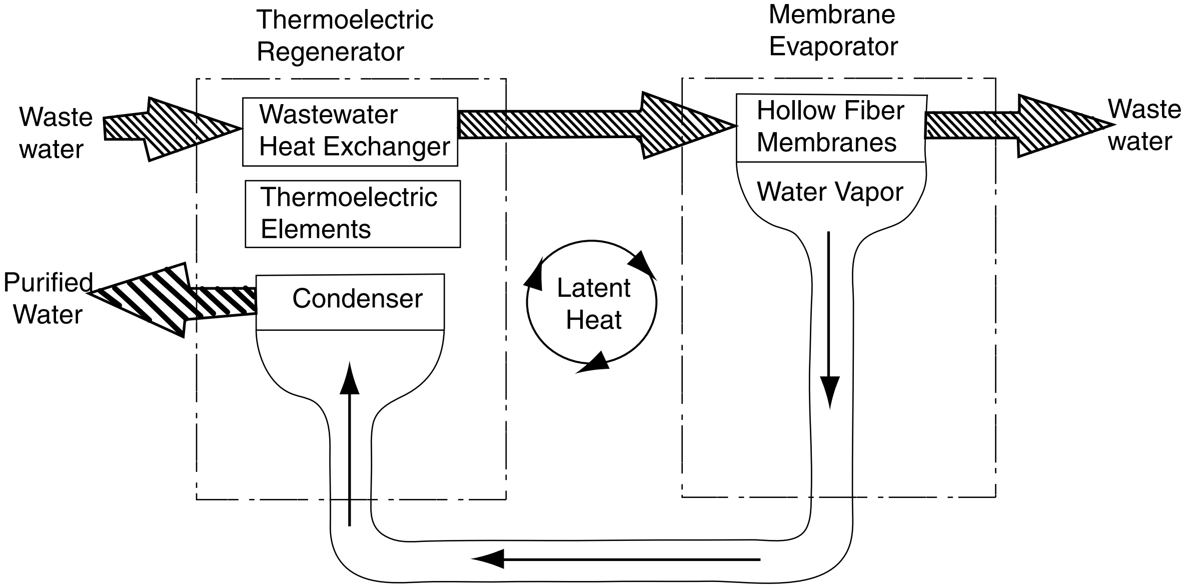


FIGURE 2 Schematic diagram of the TIMES (Thermoelectric Integrated Membrane Subsystem). *Source:* Ref. 4.

Reverse Osmosis (RO)

Reverse osmosis is a process in which pressure is applied to a concentrated solution which is on one side of a semipermeable membrane. The pressure forces the molecules of solvent through the membrane to the pure solvent or more dilute solution on the other side. Greater concentration differential between the two solutions on either side of the membrane require greater pressure. A test unit containing a bundle of tubes, capable of operating at up to 4140 kPa (600 psi), was evaluated. When wash water containing soap was processed a soap gel film formed on the membrane surfaces.[12] Reverse osmosis is a candidate for producing high quality water. A combination of several systems, including RO, produced water that met National Committee for Clinic Laboratory Standards.[13,14]

Multifiltration (Unibed)

The multifiltration system comprises a heat sterilization unit to kill microorganisms, followed by a series of progressively finer particulate filters down to 0.5 microns to avoid particulates clogging the sorbent beds.[15] The water then enters the Unibeds which remove the dissolved contaminants. The Unibed is a single replaceable unit which utilizes a set of beds of different sorptive materials arranged in a specified optimum sequence (Figure 4). The beds are designed to remove specified types and amounts of contaminants from a known waste stream in such a way that all the beds are exhausted at the same time. A sequence of Unibeds of identical design, each comprising three five-tube replaceable subunits, may be used. Unidentified compounds are generally removed by a mixture of several types of activated charcoal and nonionic sorbents near the outlet from the bed. Microbial control to avoid bacterial fouling of the various beds is achieved by iodinated resin beds at the inlet and outlet of each Unibed. Detailed discussions of the system may be found in Refs. 11, and 14–16.

Electrooxidation—Combined Electrolysis and Electrodialysis

Electrodialysis alone removes ions more efficiently than does reverse osmosis (RO); however, when combined with electrolysis the organic compounds are oxidized in the electrolysis and the inorganic salts are removed by the electrodialysis.[7] A combined electrolytic and electrodialytic cell is illustrated in Figure 5. No chemicals are used. Low voltages are used to avoid unwanted side effects such as the production of chlorine, or the formation of insoluble salts. Various types of electrode materials have been tested. Efficiency was increased considerably when the polarity across the electrodes was periodically reversed. This procedure is termed Periodic Reverse Pulsed Electrolysis (PRPE). The process has theoretical advantages over RO, as increasing the brine concentration improves the conductivity, and hence the efficiency of the process. Electrooxidation effectively kills bacteria in the feed.[17]

Supercritical Water Oxidation (SCWO)

Wastewater is heated to a temperature of 650°C under pressures of 250 atmospheres. When water is above its critical point, its properties as a solvent change. Organic compounds which are insoluble at normal temperature and pressure become soluble. The addition of sufficient oxygen then leads to the complete oxidation of these compounds. Most atmospheric gases and trace contaminant gases are also soluble in supercritical water, and will also be oxidized. The process also has the potential

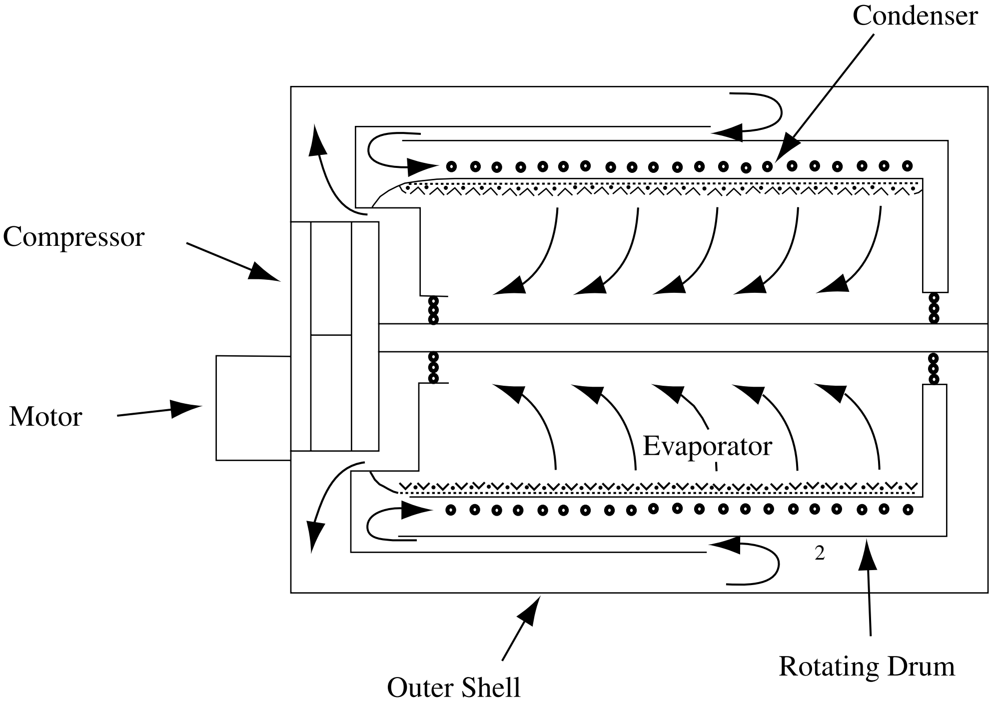


FIGURE 3 Cross section of a VCD (Vapor Compression Distillation) still. *Source:* Ref. 4.

for oxidizing much of the organic solid wastes produced on the space station.[18] Inorganic salts are produced when human metabolic wastes are subjected to supercritical water oxidation. These salts have very low solubility in supercritical water and precipitate out, and thus can be removed.[19] Metals are also precipitated out, with the exception of mercury, which carries over in the vapor phase, and has to be removed by ion adsorption.[20]

Electrodeionisation

Feed water moves through an ion exchange resin bed which selectively removes certain ions from the water. Simultaneously, the resin is regenerated by the action of an electric field imposed upon the resin bed.[21] A schematic process diagram may be found in Ref 21. Bacteria are not completely removed. It appears that this process works best to remove ionic contaminants. It is a candidate for the production of reagent grade water. Both benchtop and industrial capacity units are available.

Air Evaporation

A system with a heat pump and solar collectors for evaporative heat was tested. The system is capable of 100% water recovery from numerous types of contaminated sources. The wastewater is pretreated with a chemical solution to prevent decomposition and bacterial growth. It is then pumped through a wick filter to remove particles, in a series of pulses. The timing of the pulses is such that the liquid from one pulse is distributed along the wick by capillary action before the next pulse arrives. A heated air stream evaporates the water from the wick, leaving the solids behind. When the wick is full of solids, it is dried and replaced by a new wick. The air and vapor stream passes through a condensing heat exchanger and then through a water separator to extract the free water from the stream and test it for quality. It is then transferred to the post-treatment filter section and thence to the main water storage and distribution system. The wick system can be sterilized by heating it to 121°C while in a dry state.[22]

Vapor Phase Catalytic Ammonia Removal (VPCAR)

Neither pre-treatment nor post-treatment of the feed and product water are required. The high temperature employed also destroys microbes to a very great extent.[23] The present design utilizes the thin film evaporation technique of the VCD process in a rotating disk evaporator, combined with catalytic reactors for vapor phase chemical reactions. The ammonia and volatile hydrocarbons which are evaporated along with the water vapor are oxidized to innocuous gases by catalytic chemical reactions carried out in the vapor phase. The overall system schematic is shown in Figure 6.[10,11] The vapor from the VCD boiler passes over two catalytic beds. The first bed operates at a temperature of about 250°C, to oxidize organic volatiles to CO_2 and water, and ammonia to NO and water. The second bed operates at 450°C and reduces the NO to N_2 and O_2. The O_2 produced is more than sufficient for first bed usage. The vapor is recycled to maximize O_2 utilization. This high temperature vapor or steam supplies the

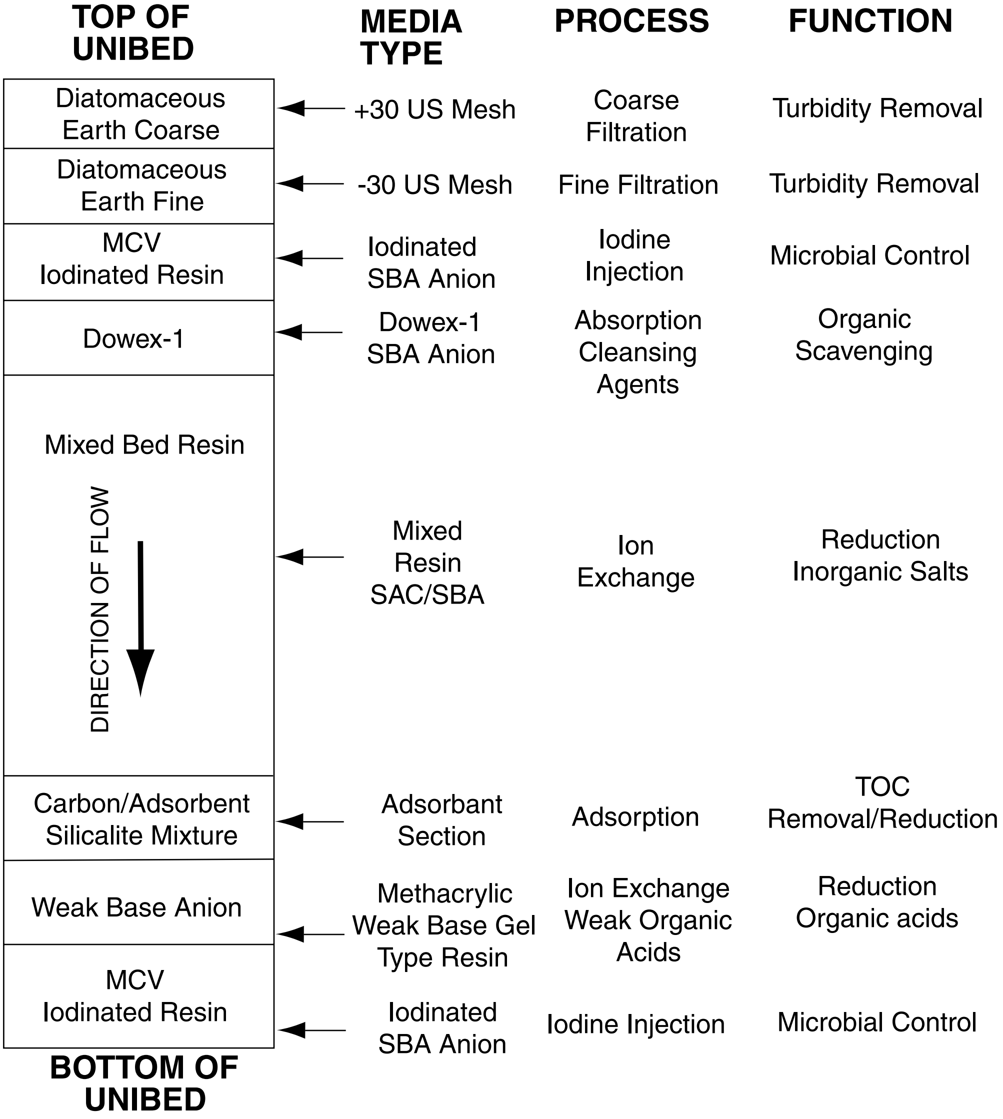


FIGURE 4 Multifiltration unibed schematic and process details. *Source:* Ref. 15.

heat to evaporate the feed, condensing to form the product water. The residual concentrated feed, or brine, is sent to a supercritical water oxidizer (SWCO) to remove the remaining contaminants. Bacterial growth is suppressed because bacteria do not thrive in the concentrated brine due to the buildup of both toxic compounds and of heavy metals in the brine. The vapor phase provides a barrier to non-airborne micro-organisms.[10]

Plant Transpiration and Water Recovery

The use of plants to produce clean water by transpiration is being examined by NASA. It is believed that plants can process wastewater, and transpire it with minimal contaminants. However, any volatile contaminants must be removed first, so that they cannot evaporate directly from the wastewater into the air. The type of pretreatment required will depend upon the waste steam being treated. Plant nutrients will have to be added to the water. Preliminary experiments using only a plant nutrient solution indicated that the amount of total organic carbon (TOC) in transpired water vapor from unstressed plants was very low, about 1.4 mg/L in the water that was collected in condensers outside the plant chamber. Water collected in the plant chamber had higher TOC values, averaging 2.5 mg/L. Stressed plants transpired water with TOC values of up to 7 mg/L. The average TOC level of tap water was 1.6 mg/L. Insect infestation stressed the plants, causing higher TOC levels in the transpired water. It was not clear if any of the increased TOC was derived from the insects themselves. There are many questions about the feasibility of using a compact growth facility to treat the wastewater. These questions need to be examined carefully.

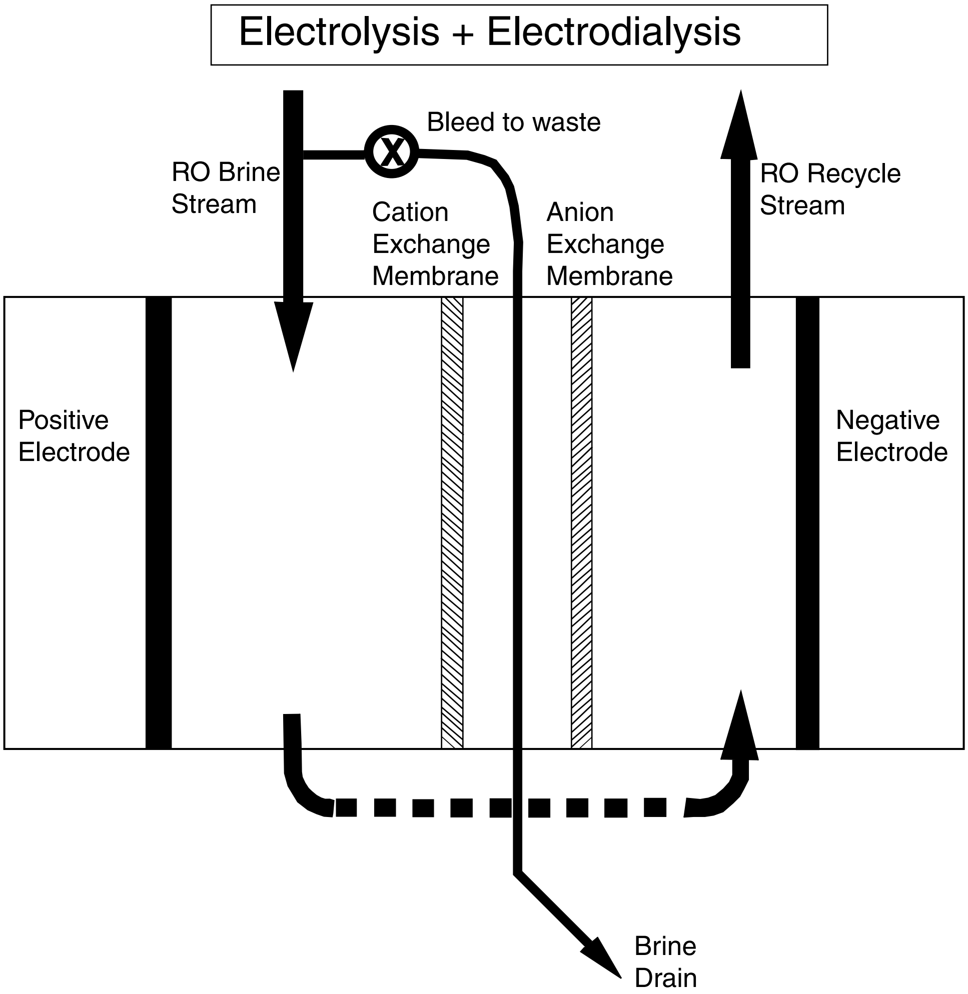


FIGURE 5 Electrooxidation—combined electrolytic and electrodialytic cell. *Source*: Ref. 7.

Discussions on these and related topics may be found in Refs. 24–27.

Immobilized Cell or Enzyme Bioreactors

Immobilized microorganisms convert contaminants into simpler molecules at near ambient temperatures. The immobilized cells can treat a wide variety of contaminants. This technology is under development, but it may have significant terrestrial uses.[28–30] These biological systems require little energy, and may leave little waste residue. Prototype reactors were tested using a packing of diatomaceous earth. When tested with feed containing 10 and 100 mg/L phenol in an oxygenated solution, they were able to produce an effluent with phenol content below the detection limit of 0.01 mg/L. The report did not say what organisms were used, nor how they were arranged for the test.[30] Immobilized microbial cell bioreactors have been used in industrial wastewater treatment applications. In a comparative test processing feed from a coal tar plant, a reactor using a porous polymeric biomass support achieved effluent phenol levels 100 times lower than those achieved by a commercial bioreactor utilizing a nonporous polyethylene support. Similar results were observed for minor contaminants, and the sludge output was about 20% of that from the nonporous system.

A bioreactor test using a consortium of enriched aerobic microorganisms immobilized in packed bed reactors was conducted, utilizing a simulated wastewater feed. A porous polymeric biomass support was used in the reactor. Influent concentrations were nominally 600 mg/L COD, and 1000 mg/L urea. COD reductions of 95% or more, and urea reductions of 95–99% were achieved when using hydraulic retention times of 24 or 48 hours. Effluent total suspended solids ranged from 1 to 3 mg/L.[28]

Tests have shown that low-molecular weight, polar, nonionic contaminants can be removed from solution by immobilized enzymes which catalyze the oxidation of various organic compounds, such as alcohols, aldehydes, and so on, to organic acids, for subsequent removal by ion exchange. Hence existing ion exchange technology (multifiltration Unibeds) can be

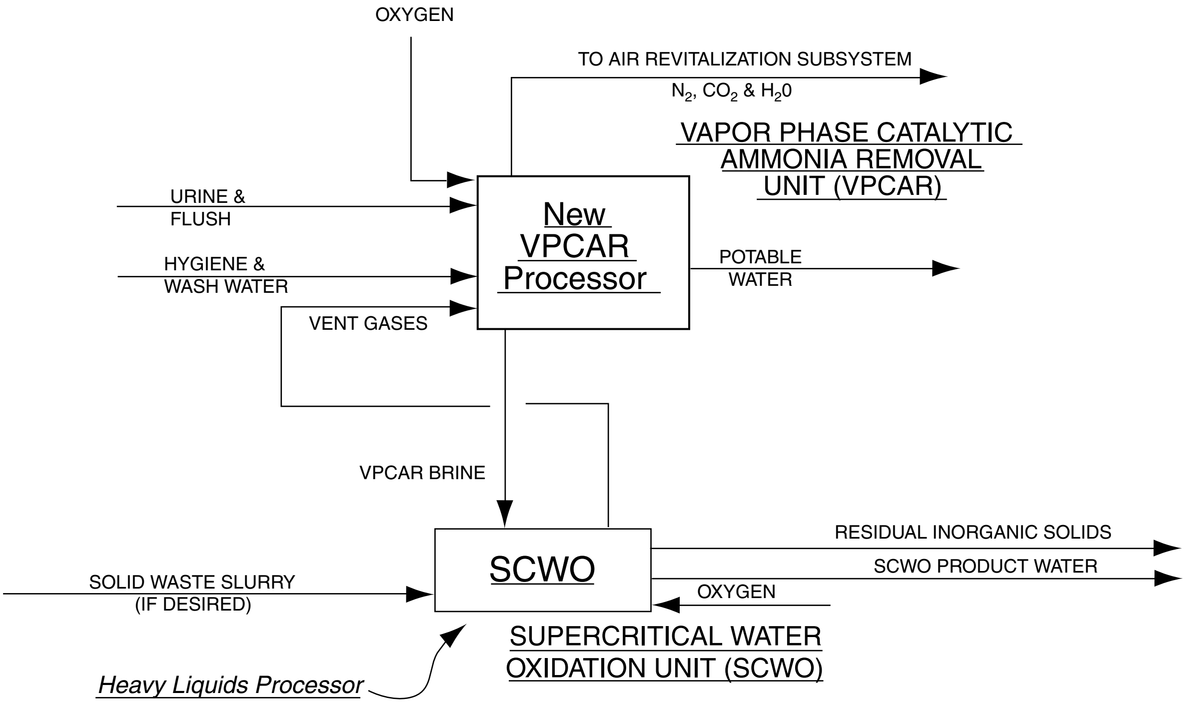


FIGURE 6 Schematic flow diagram of integrated water reclamation system utilizing a combination of VPCAR unit and SCWO unit. *Source:* Refs. 10 and 11.

made more effective. The enzymes are bound to support, usually, a silica-based material, which is in a form that makes it easy to utilize in a Unibed. Specific enzymes are used depending upon the compound or class of compound, to be removed. These Unibeds have to include iodine removers at the inlet, and iodinating resin beds at the outlets to maintain bacterial control in the water supply system of the spacecraft.[29]

Fuel Cell Product Water

When hydrogen and oxygen are combined in a fuel cell to produce electric power, pure water is also produced. This system was used in both the Apollo and Orbiter spacecraft.[5] Urine and wastewater can be electrolyzed to produce hydrogen and oxygen, thus providing a closed loop system (see Electrooxidation above).

COMPARATIVE PERFORMANCE OF TECHNOLOGIES

Many promising technologies for wastewater treatment and water recovery aboard the space station have been described briefly. Their performance levels vary depending upon the characteristics of the wastewater feed. The performance levels and the basic features of each system are compared in Table 3.

WATER QUALITY CONTROL

A major problem in these regenerative systems is bacterial contamination and biofilms forming on the internal surfaces of the water supply system. These biofilms are the result of colonies of bacteria forming on the surface. They occur even in nutrient-limited environments. Biofilms accelerate corrosion and are difficult to remove, or to prevent from forming.[31] Pathogenic bacteria have been shown to survive in sterile water and at elevated temperatures for many hours.[32] This bacterial contamination poses both health and long term mechanical problems in any recycling system. Adequate continuous sterilization procedures and equipment will be required in any recycling system. A high concentration of micro-organisms is found in shower water, reinforcing the necessity of ensuring the disinfection of the recycled potable water supply. A special soap is used for all washing on the Orbiter and is proposed for use on the space station, which reduces foaming and consequent problems in the recycling process.[33] Great care is taken to avoid the possibility of contamination due to cross connection of wastewater and processed water systems.

Pre- and Post-Treatment of Recycled Water

Pre- and post-treatment of water is generally necessary in order to obtain the desired water quality after the treatment of recycled water. NASA concentrated much of its research effort on distillation processes. The combination of distillation and ion exchange is quite successful in removing inorganic contaminants but is less successful in removing volatile organic contaminants. These volatiles must be removed to avoid their accumulation to toxic levels. Organic

TABLE 3
Summary of Described ECLSS Technologies for Water Recovery

Technology	Maximum Water Recovery Rate %	Comments	References
TIMES membrane filter	93	Max feed concentration 38%.	4,5
		Original baseline system. Volatile can be a problem.	
		Lower production rate than VCD.	
Vapor Compression Distillation (VCD)	95+	Max feed concentration 60%.	4,6–11
		Takes 50% power as TIMES, less susceptible to fouling, better quality of product water. Volatiles can be a problem.	
Reverse osmosis	95+	Poor removal of ionized gases.	12,13
		Poor rejection of urea, acids, and ammonia.	
Unibed filter	Nearly 100	Low power consumption, lightweight, simple.	11,14–16
		Filter bed units are expendable, must be replaced.	
		Wastestream must be of stable composition as filters are designed for a specific, stable, known wastestream composition.	
Electrooxidation—Combined Electrolysis and Electrodialysis	Nearly 100	Organic compounds are oxidised by electrolysis, and inorganic salts are removed by electrodialysis.	7,17
		High efficiency was achieved when polarity across the electrodes was periodically reversed.	
		Final TOC levels of 1 ppm is achieved when treating raw urine.	
		Solution has to be spiked with an expendable electrolyte.	
Supercritical Water Oxidation	Nearly 100	Organics, and most atmospheric and trace contaminant gases are completely oxidised.	18–20
		Process can also oxidise many solid wastes. Inorganic salts and metals produced have low solubility and precipitate out.	
		Mercury escapes into the vapor phase and must be removed by ion absorption.	
Electrodeionisation	Nearly 100	Bacteria not completely removed.	21
		Removes almost everything except silica.	
		Waste stream must be characterized for design of selective ion exchange resins.	
		Best for removal of known ionized contaminants.	
Air evaporation	Nearly 100	100% water recovery from urine and numerous other feeds.	22
		Consumables are required.	
		Use of solar collectors would reduce electrical power requirements significantly.	
Vapor Phase Catalytic Ammonia Removal	95	Energy efficient due to system integration.	10,23
		Bacterial growth in feed is suppressed, and vapor phase is a barrier to non-airborne organisms and contaminants. High temperatures destroy many micro-organisms.	
Plant transpiration and water recovery		Water vapor from unstressed plants had TOC of 1.4 ppm.	24–27
		No data on results using actual wastewater.	
Immobilised Cell or Enzyme Bioreactors	No water recovery	100 ppm phenol feeds were reduced to below detection limit of 10 ppb, i.e. 99 + % removal.	28–30
		COD reductions of 95 + %, and urea reductions of 95–99% are achieved.	
		Porous biomass support gives better results.	
Fuel cell product water	100	Pure water, byproduct of power generation.	5

contaminants may also encourage the growth of bacteria in the water storage and transport system. This is especially true when urine, with its heavy load of organic compounds, is treated and recycled. The specified maximum contaminant level for unidentified organic compounds is in the ppb range. Only about 70% of the organic content of urine distillate has been identified, hence it is difficult to design compound-specific removal techniques.

Oxidizers have been used to stabilize urine, but the use of such compounds produces many volatile organic compounds. Urine distillate obtained after such treatment was found to have a TOC content of 25–30 mg/L and a high ammonia concentration. When non-oxidizing compounds were used for pretreatment the TOC cotent was 8–10 mg/L and the ammonia concentration was lower. Certain of the processes described above are very effective in removing these organic compounds and ammonia, and hence require less pretreatment.

Post-treatment methods of polishing the product water are required after the main processes described above. Both multifiltration and UV assisted ozone oxidation reduced the TOC by over 90%. Reverse osmosis also achieved over 90% reduction of TOC and ammonia.[34] The post-treatment technique required depends upon the process used to perform the main water recovery.

Finished Water Quality Monitoring

In order to avoid contamination the processed water is continuously monitored for the levels of certain key parameters which are indicative of the water quality for the desired application. This is necessary to detect any loss of performance in the processing equipment, such as subsystem failure, or exhaustion of a consumable component in the system.

PLANNED TERRESTRIAL USE OF ECLSS

Attempts are being made to utilize ECLSS technology at the South Pole station. The proposed system is designed to recycle water and to grow plants, and later fish, for food at the South Pole. The Antarctic winter lasts some nine months, and the South Pole station is completely isolated during that time. Even emergency medical evacuations are very difficult and rare. A major supply problem is the need to transport some 240,000 gallons of diesel fuel to the station during the short 14 week summer flying season. The proposed ECLSS will reduce the amount of fuel needed. The system will be built in stages, and will ultimately include a waste processing facility. It will include some physical/chemical processing technology, particularly for the feed liquid to the plant growing facility. Experiments are being conducted to evaluate high yield varieties of edible crops. Experiments are planned to utilize wind power, solar obviously not being available during the Antarctic night.[35] This facility may prove more useful for evaluating possible terrestrial uses than many of the space studies, as the equipment is designed for normal gravity, with less emphasis on minimizing volume and weight, and maintenance requirements can be more relaxed than those for the space station.

SUMMARY

In summary, the systems described have high rates of water recovery, and remove or destroy a wide variety of contaminants. The technologies developed may have terrestrial applications. The emphasis on weight and volume saving that is such a major part of the NASA evaluation process may not apply to the same extent for terrestrial uses. Hence the production of reliable systems utilizing the principles tested by NASA should be relatively easy. It is not likely that the recycling of urine will be needed for many, if any, terrestrial uses, thus removing one of the most stringent NASA-imposed requirements for water recovery. Power efficiency, particularly in an integrated water recovery system, may be worth paying a high financial price for, in certain specialized needs. The present NASA system uses only 2 W per hour per kg of water recovered.[2] For some of the systems described it will be necessary to determine the composition of the contaminants in the waste stream as accurately as possible, in order to design and test the system. The requirements in possible terrestrial applications will vary tremendously according to the specific application. Recycling of urine is unlikely, but recycling of hygiene water is more likely. The make-up potable water may come from an original stored supply or low volume natural supply source. It may be possible to use metabolic wastes as input to a process for producing sterile fertilizer for plants, thus recovering water and reducing the volume and possibly the toxicity of the wastes generated. The water recovered from transpiration has potential for use as a potable water supply.

A major problem is that the conventional soap seriously affects some of the processes developed for space applications. They have all been designed to process wastewater which does not contain foaming cleansers. Hence any technology that is being considered must be tested with conventional cleaning products.

Other possible applications for ECLSS technologies may be to clean up well defined industrial waste streams. Several of the basic technologies can provide the capability to both concentrate and treat contaminants, and produce near-potable water as a bonus. As the basic technologies have been developed for space, it should be simple to produce mobile or package (skid-mounted) unites for industrial applications. It must be remembered that the hardware developed for the space program has been designed to handle the waste from a small number of people, and will probably have to be made much larger for most likely terrestrial applications.

REFERENCES

1. Mitchell, K.L., *et al., Technical Assessment of MIR-1 Life Support Hardware for the International Space Station,* Structures and Dynamics Laboratory, George C. Marshall Space Flight Center, NASA TM108441, March 1994.
2. Craig, C., R. Davenport, and M. Plam, Water reuse in space, *Worldwater,* February, 1996, pp. 11.

3. Sauer, R.L., R. Ramananathan, J.E. Straub, and J.R. Shultz, *Water Quality Program Elements for Space Station Freedom,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991. Society of Automotive Engineers, Warrendale, PA, pp. 1–21.
4. Wieland, P.O., *Designing for Human Presence in Space: An Introduction to Environmental Control and Life Support Systems,* George C. Marshall Space Flight Center, NASA RP-1324. 1994.
5. Miernik, J.H., B.H. Shah, and C.F. McGriff, *Waste Water Processing Technology for Space Station Freedom: Comparative Test Data Analysis,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 229–239.
6. Howard, S.G. and J.H. Miernik, *An Analysis of Urine Pretreatment Methods for Use on Space Station Freedom,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 157–166.
7. Herrman, C.C. and T. Wydeven, *Physical/Chemical Closed-loop Water Recycling for Long Duration Missions,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 233–245.
8. Zdankiewicz, E.M. and J. Chu, *Phase Change Water Recovery for Space Station—Parametric Testing and Analysis,* In Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 669–679.
9. Noble, L.D. Jr. *et al., Phase Change Water Recovery for the Space Station Freedom and Future Exploration Missions,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 145–158.
10. Noble, L.D. Jr. *et al., An Assessment of the Readiness of Vapor Compression Distillation for Spacecraft Wastewater Processing,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 271–282.
11. Flynn, M.T. *et al., Water Reclamation Technology Development for Future Long Range Missions,* Proceedings of the *22nd* International Conference on Environmental Systems, Seattle, WA, July 13–16, 1992, Society of Automotive Engineers, Warrendale, PA, pp. 1–9.
12. Ray, R.J. *et al., Membrane-Based Subsystem for Very High Recoveries of Spacecraft Waste Waters,* In Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 645–659.
13. Highsmith, A. *et al., Evaluation of Water Treatment Systems Producing Reagent Grade Water,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990 Society of Automotive Engineers, Warrendale, PA, pp. 113–117.
14. Carrasquillo, R.L. *et al., ECLSS Regenerative Systems Comparative Testing and Subsystem Selection,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 213–228.
15. Colley, C.D., *Functional Description of the Ion Exchange and Sorbent Media Used in the ECLSS Water Processor Unibeds,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 175–188.
16. Putnam, D.F., W.F. Michalek, and T. Van Pelt, *Space Station Hygiene Water Reclamation by Multifiltration,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 189–194.
17. Hitchins, G.D. *et al., Electrooxidation of Organics in Waste Water,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 179–188.
18. Hall, J.B. Jr., and D.A. Brewer, *Supercritical Water Oxidation: Concept Analysis for Evolutionary Space Station Application,* In Proceedings of the 16th Intersociety Conference on Environmental System, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 733–745.
19. Armellini, F.J. and J.W. Tester, *Salt Separation During Supercritical Water Oxidation of Human Metabolic Waste: Fundamental Studies of Salt Nucleation and Growth,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 189–203.
20. Swallow, K.C. *et al., Behavior of Metal Compounds in the Supercritical Water Oxidation Process,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 205–210.
21. Highsmith, A.K. *et al., Water Quality after Electrodeionisation.* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 90–96.
22. Morasko, G., D.F. Putnam, and R. Bagdigian, *Air Evaporation Closed Cycle Water Recovery Technology—Advanced Energy Saving Designs,* In Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 681–690.
23. Budininkas, P., F. Rasouli, and T. Wydeven, *Development of a Water Recovery Subsystem Based on Vapor Phase Catalytic Ammonia Removal,* In Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 661–667.
24. Macier, B.A., *Quality Assessment of Plant Transpiration Water,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 1–4.
25. Janik, D.S. and J.J. DeMarco, *Engineering Testbed for Biological Water/ Air Reclamation and Recycling,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 1–7.
26. Ray, R.J. *et al., Water Vapor Recovery from Plant Growth Chambers,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 89–97.
27. Blackwell, C.C. *et al., Options for Transpiration Water Removal in a Crop Growth System Under Zero Gravity Conditions,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 15–18.
28. Petrie, G.E. and M.S. Nacheff-Benedict, *Development of Immobilized Cell Bioreactor Technology for Water Reclamation in a Regenerative Life Support System,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 179–191.
29. Jolly, C.D., L.J. Schussel, and L. Carter, *Advanced Development of Immobilized Enzyme Rectors,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 109–117.
30. Miller, G.P. *et al., Using Biological Reactors to Remove Trace Hydrocarbon Contaminants from Recycled Water,* In Spacecraft Water Quality, Proceedings of the 21st International Conference on Environmental Systems, San Francisco, CA, July 15–18, 1991, Society of Automotive Engineers, Warrendale, PA, pp. 99–107.
31. Richardson, J.C. *et al., Bacterial Selectivity in the Colonization of Surface Materials from Groundwater and Purified Water Systems,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 101–111.
32. Kundsin, R. and R.E. Perkins, *Survival of Mycoplasmas and Ureaplasmas in Water and at Elevated Temperatures,* In Proceedings of the 20th Intersociety Conference on Environmental Systems, Williamsburg, VA, July 9–12, 1990, Society of Automotive Engineers, Warrendale, PA, pp. 97–100.

33. Verostko, C.E. *et al., Results on Reuse of Reclaimed Shower Water,* In Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 635–643.
34. Putnam, D.F., G.V. Colombo, and C. Chulien, *Pre- and Post-Treatment Techniques for Spacecraft Water Recovery,* In Proceedings of the 16th Intersociety Conference on Environmental Systems, San Diego, CA, July 14–16, 1986, Society of Automotive Engineers, Warrendale, PA, pp. 627–634.
35. Straight, C.L. *et al., The CELSS Antarctic Analog Project: A Validation of CELSS Methodologies at the South Pole Station,* In Proceedings of the 23rd Intersociety Conference on Environmental Systems, Colorado Springs, CO, July 12–15, 1993, Society of Automotive Engineers, Warrendale, PA, pp. 1–11.

ROBERT G. ZACHARIADIS

SYED R. QASIM
The University of Texas at Arlington

WATER CHEMISTRY

AQUATIC CHEMICAL EQUILIBRIA

In this section a few example will be given that demonstrate how elementary principles of physical chemistry can aid in the recognition of interrelated variables that establish the composition of natural waters. Natural water systems usually consist of numerous mineral assemblages and often of a gas phase in addition to the aqueous phase; they nearly always include a portion of the biosphere. Hence, natural aquatic habitats are characterized by a complexity seldom encountered in the laboratory. In order to distill the pertinent variables out of a bewildering number of possible ones, it is advantageous to compare the real systems with their idealized counterparts.

Thermodynamic equilibrium concepts represent the most expedient means of identifying the variables relevant in determining the mineral relationships and in establishing chemical boundaries of aquatic environments. Since minimum free energy describes the thermodynamically stable state of a system, a comparison with the actual free energy can characterize the direction and extent of processes that are approaching equilibrium. Discrepancies between equilibrium calculations and the available data of real systems give valuable insight into those cases where chemical reactions are not understood sufficiently, where non-equilibrium conditions prevail, or where the analytical data are not sufficiently accurate or specific.

Alkalinity and Acidity for Aqueous Carbonate Systems

Alkalinity and acidity are defined, respectively, as the equivalent sum of the bases that are titratable with strong acid and the equivalent sum of the acids that are titratable with strong base; they are therefore capacity factors which represent, respectively, the acid and base neutralizing capacities of an aqueous system. Operationally, alkalinity and acidity are determined by acidimetric and alkalimetric titrations to appropriate pH end points. These ends points (equivalence points) occur at the infection points of titration curves as shown in Figure 1 for the carbonate system.

The atmosphere contains CO_2 at a partial pressure of 3×10^{-4} atmosphere, while CO_2, H_2CO_3, HCO_3^- and CO_3^{2-} are important solutes in the hydrosphere. Indeed, the carbonate system is responsible for much of the pH regulation in natural waters.

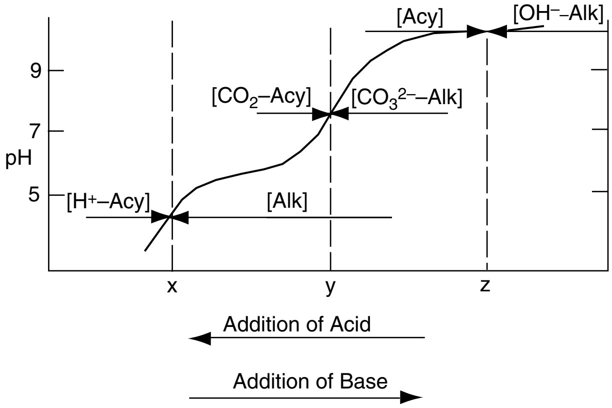


FIGURE 1 Alkalinity and acidity titration curve for the aqueous carbonate system. The conservative quantities alkalinity and acidity refer to the acid neutralizing and base neutralizing capacities of a given aqueous system. These parameters can be determined by titration to appropriate equivalence points with strong acid and strong base. The equations given below define the various capacity factors rigorously. Figure from Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 130.

The following equations define for aqueous carbonate systems the three relevant capacity factors: Alkalinity (Alk), Acidity (Acy), and total dissolved carbonate species (C_T):[†]

$$[\mathrm{Alk}] = [\mathrm{HCO_3^-}] + 2[\mathrm{CO_3^{2-}}] + [\mathrm{OH^-}] - [\mathrm{H^+}] \qquad (1)$$

$$[\mathrm{Acy}] = 2[\mathrm{H_2CO_3}\,*] + [\mathrm{HCO_3^-}] + [\mathrm{H^+}] - [\mathrm{OH^-}] \qquad (2)$$

$$C^b = [\mathrm{H_2CO_3}\,*] + [\mathrm{HCO_3^-}] + [\mathrm{CO_3^{2-}}] \qquad (3)$$

where $[H_2CO_3^*] = [CO_2(aq)] + [H_2CO_3]$.

These equations are of analytical value because they represent rigorous conceptual definitions of the acid neutralizing and the base neutralizing capacities of carbonate systems. The definitions of alkalinity and acidity algebraically

---

[†] Brackets of the form [ ] refer to concentration, e.g., in moles per liter.

express the net proton deficiency and net proton excess of the systems with repect to specific proton reference levels (equivalance points). The definitions can be readily amplified to account for the presence of buffering components other than carbonates. For example, in the presence of borate and ammonia the definition for alkalinity becomes

$$\begin{aligned}[\mathrm{Alk}] = [\mathrm{HCO_3^-}] + 2[\mathrm{CO_3^{2-}}] + [\mathrm{B(OH)_4^-}] \\ + [\mathrm{NH_3}] + [\mathrm{OH^-}] - [\mathrm{H^+}]\end{aligned} \tag{4}$$

Although individual concentrations or activities, such as $[H_2CO_3^*]$ and pH, are dependent on pressure and temperature, [Alk], [Acy], and C_T are conservative properties that are pressure and temperature independent. (Alkalinity, acidity, and C_T must be expressed in terms of concentration, e.g., as molarity, molality, equivalents per liter or parts per million as $CaCO_3$). Note that 1meq/l 5 50 ppm as $CaCO_3$.

The use of these conservative parameters facilitates the calculation of the effects of the addition or removal of acids, bases, carbon dioxide, bicarbonates, and carbonates to aqueous systems. As shown in Figure 2, each of these conservative quantities remains constant for particular changes in the chemical composition. The case of the addition or removal of dissolved carbon dioxide is of special interest. Respiratory activities of aquatic biota contribute carbon dioxide to the water whereas photosynthetic activities decrease the concentration of this weak acid. An increase in carbon dioxide increases both the acidity of the system and C_T, the total concentration of dissolved carbonic species, and it decreases the pH, but it does not affect the alkalinity. Alternatively, acidity remains unaffected by the addition or removal of $CaCO_3(s)$ or $Na_2CO_3(s)$. C_T, on the other hand, remains unchanged in a closed system upon addition of strong acid or strong base. For practical purposes, systems may be considered closed if they are shielded from the atmosphere and lithosphere or exposed to them only for short enough periods to preclude significant dissolution of CO_2 or solid carbonates.

Dissolution of Carbon Dioxide

Though much of the CO_2 which dissolves in solution may ionize to form HCO_3^- CO_3^{2-}, depending upon the pH, only a small fraction (0.3% at 25°C) is hydrated as H_2CO_3. Hence, the concentration of the unhydrated dissolved carbon dioxide, $CO_2(aq)$, is nearly identical to the analytically determinable concentration of $H_2CO_3^*$ (= $[CO_2(aq)]$ + $[(H_2CO_3])$.

The equilibrium of a constituent between a gas phase and a solution phase can be characterized by a mass law relationship. Table 1 gives the various expressions and their interrelations for the characterization of the CO_2 dissolution equilibrium. A water that is in equilibrium with the atmosphere ($P\mathrm{CO}_2 = 10^{-3.5}$ atm) contains at 25°C approximately 0.44 milligram per liter (10^{-5} M) of CO_2; K_H (Henry's Law constant) at 25°C is $10^{-1.5}$ mole per liter-atm.

Dissolved Carbonate Equilibria

Two systems may be considered: (1) a system closed to the atmosphere and (2) one that is in equilibrium with the atmosphere.

Closed Systems In this case $H_2CO_3^*$ is considered a non-volatile acid. The species $H_2CO_3^*$, HCO_3, CO_3^{2-} and are interrelated by the equilibria:[†]

$$[\mathrm{H^+}][\mathrm{HCO_3^-}]/[\mathrm{H_2CO_3}^*] = K_1$$

$$[\mathrm{H+}][\mathrm{CO_3^{2-}}]/[\mathrm{HCO_3^-}] = K_2$$

where K_1 and K_2 represent the equilibrium constants (acidity constants).

The ionization fractions, whose sum equals unity (see Eq. (3)), can be defined as follows:

$$\alpha_0 = [\mathrm{H_2CO_3}^*]/C_T \tag{7}$$

$$\alpha_1 = [\mathrm{HCO_3^-}]C_T \tag{8}$$

$$\alpha_2 = [\mathrm{CO_3^{2-}}]/C_T \tag{9}$$

From Eqs. (3) to (9) the ionization fractions can be expressed in terms of $[H^+]$ and the equilibrium constants:

$$\alpha_0 = (1 + K_1/[\mathrm{H^+}] + K_1K_2/[\mathrm{H^+}]^2)^{-1} \tag{10}$$

$$\alpha_1 = ([\mathrm{H^+}]/K_1 + 1 + K_2/[\mathrm{H^+}])^{-1} \tag{11}$$

$$\alpha_2 = ([\mathrm{H^+}]^2/K_1K_2 + [\mathrm{H^+}]/K_2 + 1]^{-1} \tag{12}$$

Values for K_1 and K_2 are given in Tables 2A and 2B. Because HCO_3^- and CO_3^{2-} may form complexes with other ions in the systems (e.g., in sea water, $MgCO_3$, $NaCO_3^-$, $CaCO_3$ $MgHCO_3^+$, it is operationally convenient to define a total concentration of the species to include an unknown number of these complexes. For example,

$$\begin{aligned}[\mathrm{CO_{3T}^{2-}}] = [\mathrm{CO_{3T}^{2-}}] + [\mathrm{MgCO_3}] \\ + [\mathrm{CaCO_3}] + [\mathrm{NaCO_3^-}] + \Lambda\end{aligned} \tag{13}$$

The distribution of carbonate species in sea water as a function of pH is given in Figure 3.

Systems Open to the Atmosphere A very elementary model showing some of the characteristics of the carbonate system in natural waters is provided by equilibrating pure water with a gas phase (e.g., the atmosphere) containing CO_2 at a constant partial pressure. Such a solution will remain in

---

[†] To facilitate calculations the equilibria are written here in terms of concentration quotients. The activity corrections can be considered incorporated into the equilibrium "constants" which therefore vary with the particular solution. Such constants for given media of constant ionic strength, as well as the true thermodynamic constants, are listed in Tables 2A and 2B.

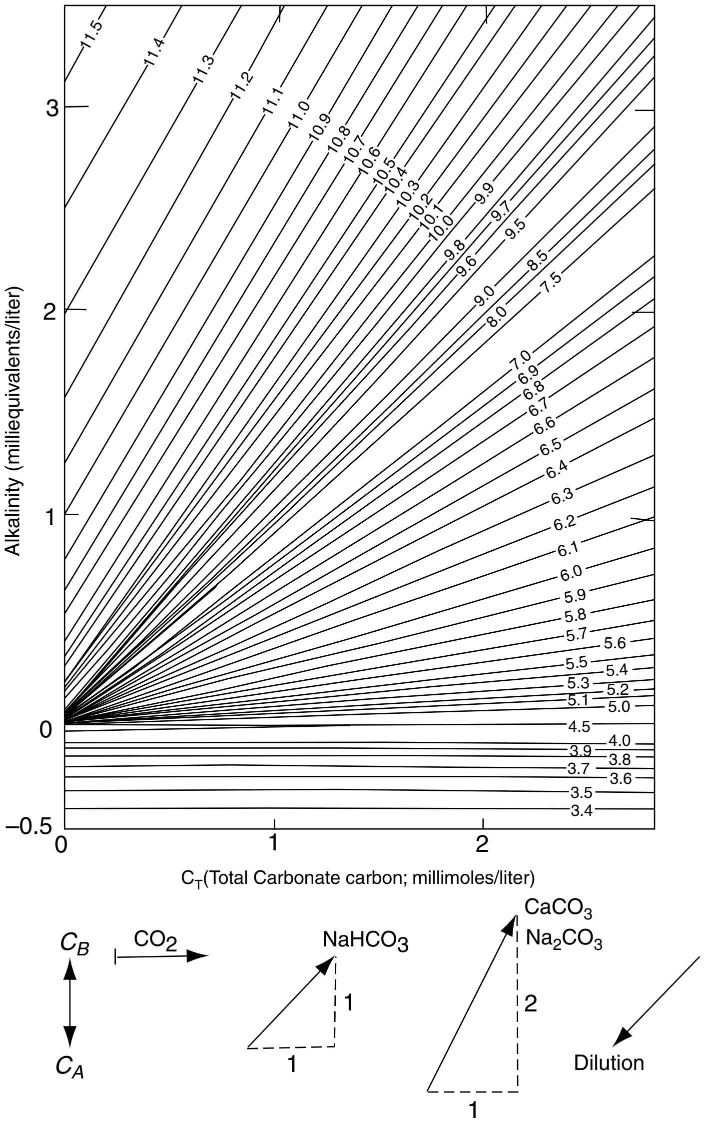


FIGURE 2 Closed system capacity diagram: pH contours for alkalinity versus C_T (total carbonate carbon). The point defining the solution composition moves as a vector in the diagram as a result of the addition (or removal) of CO_2, $NaHCO_3$, and $CaCO_3$ (Na_2CO_3) or C_B (strong base) and C (strong acid). (After K.S. Deffeyes, Limnol., *Oceanog.*, **10**, 412, 1965.) Figure from Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 133.

equilibrium with p_{co}, despite any variation of pH by the addition of strong base or strong acid. This simple model has its counterpart in nature when CO_2 reacts with bases of rocks, for example with clays and silicates.

Figure 4 shows the distribution of the solute species of such a model. A partial pressure of CO_2 equivalent to that in the atmosphere and equilibrium constants valid at 25°C have been assumed. The equilibrium concentrations of the individual carbonate species can be expressed as a function of and $[H^+]_2$. From Henry's Law,

$$[H_2CO_3{}^*] = K_H p_{CO_2}, \tag{14}$$

and Eqs. (5) to (9), one obtains

$$C_T = \frac{1}{\alpha_0} K_H p_{CO_2} \tag{15}$$

$$[HCO_3^-] = \frac{\alpha_1}{\alpha_0} K_H p_{CO_2} = \frac{K_1}{[H^+]} K_H p_{CO_2} \tag{16}$$

TABLE 1
Solubility of gases

Example[a]: $CO_2(g)$ $CO_2(aq)$

Assumptions: Gas behaves ideally; $[CO_2(aq)] = [H2CO_3^*]$

I. Expressions for Solubility Equilibrium[b]

(1) Distribution (mass law) constant, K_D:

$K_D = [CO_2(aq)]/[CO_2(g)]$ (dimensionless) (1)

(2) Henry's law constant, K_H:

In (1), $[CO_2(g)]$ can be expressed by Dalton's law of partial pressure:

$[CO_2(g)] = p_{CO_2}/RT$ (2)

Combination of (1) and (2) gives

$[CO_2\ (aq)] = (K_D/RT)p_{CO_2} = K_H\ p_{CO_2}$, (3)

where $K_H = K_D/RT$ (mole liter^1 atm $^{-1}$)

(3) Bunsen absorption coefficient, α_B:

$[CO_2(aq)] = (\alpha_B/22.414)p_{CO_2}$ (4)

where $22.414 = RT/p$ (liter mole^{-1}) and

$\alpha_B = K_H \times 22.414$ (atm^{-1}) (5)

Partial Pressure and Gas Composition

$p_{CO_2} = x_{CO_2}\ (P_T - w)$ (6)

where X_{CO_2} = mole fraction or volume fraction in dry gas, P_T = total pressure and w = water vapor pressure

Values of Henry's Law Constants at 25°C

Gas		K_H(mole liter^{-1} atm^{-1})
Carbon Dioxide	CO_2	33.8×10^{-3}
Methane	CH_4	1.34×10^{-3}
Nitrogen	N_{2-}	0.642×10^{-3}
Oxygen	O_2	1.27×10^{-3}

[a] Same types of expressions apply to other gases.

[b] The equilibrium constants defined by (1)–(4) are actually constants only if the equilibrium expressions are formulated in terms of activities and fugacities.

Table from Stumm, W. and J. Morgan, *Aquatic Chemistry*, Wiley-Interscience, New York, 1970, p. 125.

and

$$[CO_3^{2-}] = \frac{\alpha_2}{\alpha_0} K_H P_{CO_2} = \frac{K_1 K_2}{[H^+]^2} K_H p_{CO_2}. \qquad (17)$$

It follows from these equations that in a logarithmic concentration—pH diagram (Figure 4) the lines of $H_2CO_3^*$, HCO_3^2, CO_3^{2-} have slopes of 0, +1, and +2, respectively.

If we equilibrate pure water with CO_2, the system is defined by two independent variables, for example, temperature and Pco_2, In other words, the equilibrium concentrations of all solute components can be calculated by means of Henry's Law, the acidity constants and the proton condition or charge balance if, in addition to temperature, one variable, such as Pco_2, $[H_2CO_3^*]$ or $[H^+]$, is known or measured. Use of the proton condition instead of the charge balance generally facilitates calculations because species irrelevant to the calculation need not be considered. The proton condition merely expressed the equality between the proton excess and the proton deficiency of the various species with respect to a convenient proton reference level. Figure 4 furnishes a graphic illustration of its use.

Solubility Equilibria

Minerals dissolve in or react with water. Under different physico-chemical conditions minerals are precipitated and accumulate on the ocean floor and in the sediments of rivers and lakes. Dissolution and precipitation reactions impart to the water and remove from it constituents which modify its chemical properties.

TABLE 2A

Equilibrium constant for CO_2 solubility

Equilibrium: $CO_2(g) + aq = H_2CO_3$

Henry's law constant: $K = [H_2CO_3]/p_{CO_2}(M.atm^{-1})$

Temp., °C	→ 0	Medium, 1 M NaClO^4	Seawater, 19% C1$^-$
	–log K	–log^cK	–log ^{cK}
0	1.11[a]	—	1.19[a]
5	1.19[a]	—	1.27[a]
10	1.27[a]	—	1.34[a]
15	1.32[a]	—	1.41[a]
20	1.41[a]	—	1.47[a]
25	1.47[a]	1.51[c]	1.53[a]
30	1.53[a]	—	1.58[a]
35	—	1.59[c]	—
40	1.64[b]	—	—
50	1.72[b]	—	—

[a] Values based on data taken from Bohr and evaluated by K. Buch, Meeresforschung, 1951.

[b] A.J. Ellis, *Amer. J. Sci.*, **257,** 217 (1959).

[c] G. Nilsson, T. Rengemo, and L.G. Sillen, *Acta Chem. Sand.*, **12,** 878 (1958).

Ref.: Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 148.

It is difficult to generalize about rates of precipitation and dissolution other than to recognize that they are usually slower than reactions between dissolved species. Data concerning most geochemically important solid-solution reactions are lacking, so that kinetic factors cannot be assessed easily. Frequently the solid phase initially formed is metastable with respect to a thermodynamically more stable solid phase. Relevant examples of such metastability are the formation of aragonite under certain conditions instead of calcite, the more stable form of calcium carbonate, and the over-saturation of quartz in most natural waters. This over-saturation persists due to the extremely slow establishment of equilibrium between silicic acid and quartz.

The solubilities of most inorganic salts increase with increasing temperature. However, a number of compounds of interest in natural waters (e.g. $CaCO_3$, $CaSO_4$) decrease in solubility with increasing temperature. The dependence of solubility on pressure is very slight but must be considered for the extreme pressures encountered at ocean depths. For example, the solubility product of $CaCO_3$ will increase by approximately 0.2 logarithmic units for a pressure of 200 atmospheres (ca. 2000 meters).

TABLE 2B

First acidity constant: $H_2CO_3 = HCO_3^- + H^+$

$$K_1 = \frac{\{H^+\}\{HCO_3^-\}}{\{H_2CO_3^*\}} \qquad K_1' = \frac{\{H^+\}[HCO_{3T}^-]}{[H_2CO_3^*]} \qquad {}^cK_1 = \frac{[H^+][HCO_{3T}^-]}{[H_2CO_3^*]}$$

Temp., °C	Medium			
	→ 0	Seawater, 19% Cl$^-$	Seawater	1 M NaClO4
	$-\log K$	$-\log K_1'$	—	$-\log{}^cK_1$
0	6.579[a]	6.15[b]	—	—
5	6.517[a]	6.11[b]	6.01[e]	—
10	6.464[a]	6.08[b]	—	—
14	—	—	6.02[f]	—
15	6.419[a]	6.05[b]	—	—
20	6.381[a]	6.02[b]	—	—
22	—	6.00[c]	5.89[c]	—
25	6.352[a]	6.00[b], 6.09[d]	—	6.04[g]
30	6.327[a]	5.98[b]	—	—
35	6.309[a]	5.97[b]	—	—
40	6.298[a]	—	—	—
50	6.285[a]	—	—	—

[a] H. S. Harned and R. Davies, Jr., *J. Amer. Chem. Soc.*, **65,** 2030 (1943).

[b] After Lyman (1956), quoted in G. Skirrow, *Chemical Oceanography,* Vol. I, J.P. Riley and G. Skirrow, Eds., Academic Press, New York, 1965, p. 651.

[c] A Distèche and S. Distèche, *J. Electrochem. Soc.*, **114,** 330 (1967).

[d] Calculated as log ($K_1/f_{HCO_3^-}$) as determined by A. Berner, *Geochim. Cosmochim. Acta,* **29,** 947 (1964).

[e] D. Dyrssen, and L.G. Sillén, *Tellus,* **19,** 810 (1967).

[f] D. Dyrssen, *Acta Chem. Scand.*, **19,** 1265 (1965).

[g] M. Frydman, G.N. Nilsson, T. Rengemo, and L.G. Sillén, *Acta Chem. Scand.*, **12,** 878 (1958).

Ref.: Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 148.

TABLE 2B (*continued*)
Solid acidity constant: $HCO_3^- = H^+ + CO_3^{2-}$

$K_2 = \frac{\{H^+\}\{CO_3^{2-}\}}{\{HCO_3^-\}}$ $K_2' = \frac{\{H^+\}[CO_{3T}^{2-}]}{[HCO_{3T}]}$ ${}^cK_2 = \frac{[H^+][CO_{3T}^{2-}]}{[HCO_3^-]}$

	Medium			
Temp., °C	→ 0	Seawater	0.75 M NaCl	1 M $KclO_4$
—	$-\log K_2$	$-\log K_2'$	$-\log K_2'$	$-\log {}^cK_2$
5	10.625[a]	9.40[b]	—	—
5	10.557[a]	9.34[b]	—	—
10	10.490[a]	9.28[b]	—	—
15	10.430[a]	9.23[b]	—	—
20	10.377[a]	9.17[b]	—	—
22	—	9.12[c]	9.49[c]	—
25	10.329[a]	9.10[b]	—	9.57[d]
30	10.290[a]	9.02[b]	—	—
35	10.250[a]	8.95[b]	—	—
40	10.220[a]	—	—	—
50	10.172[a]	—	—	—

[a] H. S. Harned and S. R. Scholes, *J. Amer. Chem. Soc.*, **63,** 1706 (1941).
[b] After Lyman, quoted in G. Skirrow, *Chemical Oceanography,* Vol. I, J. P. Riley and G. Skirrow, Eds., Academic Press, New York, p. 651.
[c] A. Distèche and S. Distèche, *J. Electrochem. Soc.*, **114,** 330 (1967).
[d] M. Frydman, G. N. Nilsson, T. Rengemo, and L. G. Sillén, *Acta Chem. Scand.*, **12,** 878 (1958).
Ref.: Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, pp. 149 and 150.

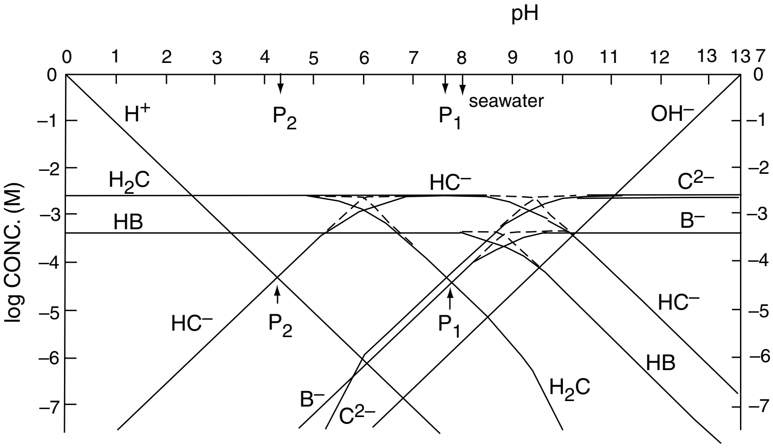


FIGURE 3 Logarithmic concentration—pH equilibrium diagram for seawater as a closed system. For seawater log $B_T = -3.37$, log $C_T = -2.62$ and the following pK values: 6.0 for $H_2CO_3^*$, 9.4 for and pK = 13.7. B_T = total borate boron and C_T = total carbonate carbon. Arrows gives [H^+] for seawater (pH = 8.0) and for two equivalence points (points of minimum buffer intensity): P_1, corresponding to a proton reference level of HB + HC^- + H_2O, and P_2, corresponding to a proton reference level of HB + H_2C + H_2O (From Dyrssen, D. and L.G. Sillén, *Tellus,* 19, 110, 1967).

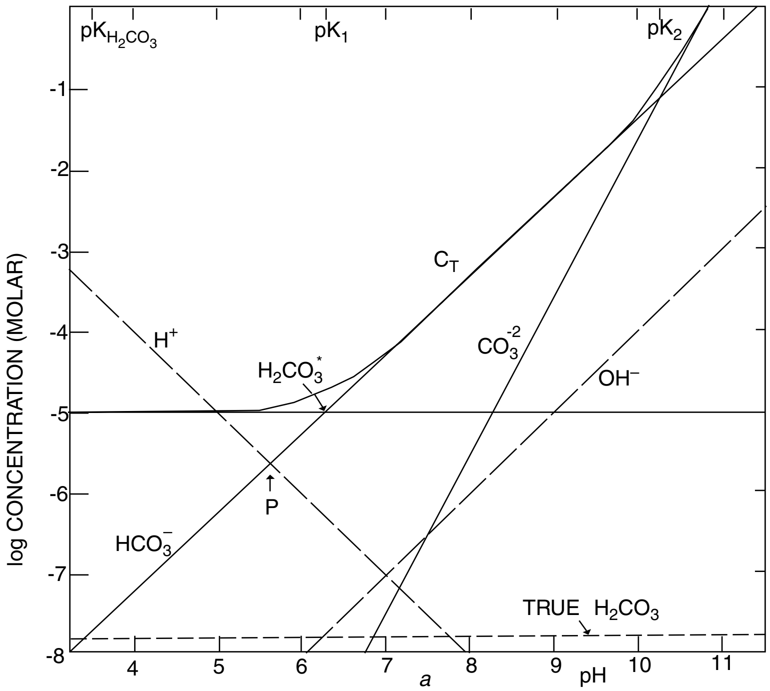


FIGURE 4 Logarithmic concentration—pH equilibrium diagram for the aqueous carbonate system open to the atmosphere. Water is equilibrated with the atmosphere ($pCO_2 = 10^{-3.5}$ atm) and the pH is adjusted with strong base or strong acid. Eqs. (14), (15), (16), (17) with the constants (25°C) $pK_H = 1.5$, $pK_1 = 6.3$, $pK_2 = 10.25$, pK(hydration of CO_2) $= -2.8$ have been used. The pure CO_2 solution is characterized by the proton condition $[H^+] = [HCO_3^-] + 2[CO_3^{2-}]+[OH^-]$ see point *P*) and the equilibrium concentrations $-\log[H^+] = -\log[HCO_3^-] = 5.65$; $-\log[CO_2aq] = -\log[H_2CO_3] = 5.0$; $-\log[H_2CO_3] \approx 7.8$; $-\log[CO_3^{2-}] = 8.5$. *Ref.:* Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 127.

Solubility of Oxides and Hydroxides

If pure solid oxide or hydroxide is in equilibrium with free ions in solution, for example,

$$Me(OH)_2(s) = Me^{2+} + 2OH^- \qquad (18)$$

$$MeO(s) + H_2O = Me^{2+} + 2OH^- \qquad (19)$$

the conventional (concentration) solubility product is given by

$$c^*_{K_{s0}} = [Me^{2+}][OH^-]^2(mole^3 liter^{-3}), \qquad (20)$$

where the subscript "0" refers to solution of the simple, uncomplexed forms of the metal ion.

Sometimes it is more appropriate to express the solubility in terms of reaction with protons, for example,

$$Me(OH)_2(s) + 2H^+ = Me^{2+} + 2H_2O \qquad (21)$$

$$MeO(s) + 2H^+ = Me^{2+} + H_2O. \qquad (22)$$

In the general case for a cation of charge z, the solubility equilibrium for Eqs. (21) and (22) is characterized by

$$c^*_{K_{s0}} = \frac{[Me^{z+}]}{[H^+]}[mole^{(1-z)}liter^{(z-1)} = \frac{c_{K_{s0}}}{K_w^z}, \qquad (23)$$

where K_w is the ion product of water. This constant and also a number of solubility equilibrium constants relevant to natural waters are given in Table 3.

Equation (23) can be written in logarithmic form to express the equilibrium concentration of a cation Me^{z+} as a function of pH:

$$\log[Me^{z+}] = \log^{c^*} K_{s0} - pH. \qquad (24)$$

Equation (24) is plotted for a few oxides and hydroxides in Figure 5.

TABLE 3
Equilibrium constants for oxides, hydroxides, carbonates, hydroxide carnonates, sulfates, silicates, and acids

Reaction	Symbol for equilibrium constants	log *K* (25°C)	I
I. OXIDES AND HYDROXIDES			
$H_2O(1) = H^+ + OH^-$	K_w	−14.00	0
		−13.77	1M $NaClO_4$
$(am)Fe(OH)_3(s) = Fe^{3+} + 3OH^-$	K_{s0}	−38.7	3M $NaClO_4$
$(am)Fe(OH)_3(s) = FeOH^{2+} + 2OH^-$	K_{s1}	−27.5	3M $NaClO_4$
$(am)Fe(OH)_3(s) = Fe(OH)_2^+ + OH^-$	K_{s2}	−16.6	3M $NaClO_4$
$(am)Fe(OH)_3(s) + OH^- = FE(OH)_4$	K_{s4}	−4.5	3M $NaClO_4$
$2(am)Fe(OH)_3(s) = Fe_2(OH)_2^{4+} + 4OH^-$	K_{s22}	−51.9	3M $NaClO_4$
$(am)Fe(OOH)(s) + 3H^+ = Fe^{3+} + 2H_2O$	$^*K_{s0}$	3.55	3M $NaClO_4$
α—$FeOOH(s) + 3H^+ = Fe^{3+} + 2H_2O$	$^*K_{s0}$	1.6	3M $NaClO_4$
α—$Al(OH)_3$(gibbsite) $+ 3H^+ = Al^{3+} + 3H_2$	$^*K_{s0}$	8.2	0
γ—$Al(OH)_3$(bayerite) $+ 3H^+ = Al^{3+} + 3H_2O$	$^*K_{s0}$	9.0	0
$(am)Al(OH)_3(s) + 3H^+ = Al^{3+} + 3H_2O$	$^*K_{s0}$	10.8	0
$Al^{3+} + 4OH^- = Al(OH)_4^-$	K_4	32.5	0
$CuO(s) + 2H^+ = Cu^{2+} + H_2O$	$^*K_{s0}$	7.65	0
$Cu^{2+} + OH^- = CuOH^+$	K_1	6.0 (18°C)	0
$2Cu^{2+} + 2OH^- = Cu_2(OH)_2^{2+}$	K_{22}	17.0 (18°C)	0
$Cu^{2+} + 3OH^- = Cu(OH)_3^-$	K_3	15.2	0
$Cu^{2+} + 4OH^- = Cu(OH)_4^{2-}$	K_4	16.1	0
$ZnO(s) + 2H^+ = Zn^{2+} + H_2O$	$^*K_{s0}$	11.18	0
$Zn^{2+} + OH^- = ZnOH^+$	K_1	5.04	0
$Zn^{2+} + 3OH^- = Zn(OH)_3^-$	K_3	13.9	0
$Zn^{2+} + 4OH^- = Zn(OH)_4^{2-}$	K_4	15.1	0
$Cd(OH)_2(s) + 2H^{2+} = Cd^{2+} + 2H_2O$	$^*K_{s0}$	13.61	0
$Cd^{2+} + OH^- = CdOH^+$	K_1	3.8	1M $LiClO_4$
$Mn(OH)_2(s) = Mn^{2+} + 2OH^-$	K_{s0}	−12.8	0
$Mn(OH)_2(s) + OH^- = Mn(OH)_3^-$	K_{s3}	−5.0	0
$Fe(OH)_2$(active) $= Fe^{2+} + 2OH^-$	K_{s0}	−14.0	0
$Fe(OH)_2$(inactive) $= Fe^{2+} + 2OH^-$	K_{s0}	−14.5 (−15.1)	0
$Fe(OH)_2$(inactive) $+ OH^- = Fe(OH)_3$	K_{s3}	−5.5	0
$Mg(OH)_2 = Mg^{2+} + 2OH^-$	K_{s0}	−9.2	0
$Mg(OH)_2$(brucite) $= Mg^{2+} + 2OH^-$	K_{s0}	−11.6	0
$Mg^{2+} + OH^- = MgOH^+$	K_1	2.6	0
$Ca(OH)_2(s) = Ca^{2+} + 2OH^-$	K_{s0}	−5.43	0
$Ca(OH)_2(s) = CaOH^+ + OH^-$	K_{s1}	−4.03	0
$Sr(OH)_2(s) = Sr^{2+} + 2OH^-$	K_{s0}	−3.51	0
$Sr(OH)_2(s) = SrOH^+ + OH^-$	K_{s1}	0.82	0
$AgOH(s) = Ag^+ + OH^-$	K_{s0}	−7.5	0
II. CARBONATES AND HYDROXIDE CARBNONATES			
$CO_2(g) + H_2O = H^+ + HCO_3^-$	K_{p1}	−7.82	0
		−7.5	Seawater
			5°C, 200 atm seawater
$HCO_3^- = H^+ + CO_3^{2-}$	(K_2)	−10.33	0
		−9.0	Seawater

(Continued)

TABLE 3 (*continued*)

Reaction	Symbol for equilibrium constants	log *K* (25°C)	I
		−9.0	5°C, 200 atm seawater
$CaCO_3$(calcite) = $Ca^{2+} + CO_3^{2-}$	K_{s0}	−8.35	0
		−6.2	Seawater
$CaCO_3$(aragonite) = $Ca^{2+} + CO_3^{2-}$	K_{s0}	−8.22	0
$SrCO_3(s) = Sr^{2+}CO_2^{2-}$	K_{s0}	−9.03	0
		−6.8	Seawater
$ZnCO_3(s) + 2H^+ = Zn^{2+} + H_2O + CO_2(g)$	$^*K_{ps0}$	7.95	0
$Zn(OH)_{1.2}(CO_3)_{0.4}(s) + 2H^+ = Zn^{2+} + H_2O + CO_2(g)$	$^*K_0$	9.8	0
$Cu(OH)(CO_3)_{0.5}(s) + 2H^+ = Cu^{2+} + 3/2H_2O + 1/2CO_2(g)$	$^*K_{ps0}$	7.08	0
$Cu(OH)_{0.67}(CO_3)_{0.67}(s) + 2H^+ = Cu^{2+} + 4/3H_2O + 2/3CO_2(g)$	$^*K_{ps}$	7.08	0
$MgCO_3$(magnesite) = $Mg^{2+} + CO_3^{2-}$	K_{s0}	−4.9	0
$MgCO_3$(nesquehonite) = $Mg^{2+} + CO_3^{2-}$	K_{s0}	−5.4	0
$Mg_4(CO_3)_3(OH)_2.3H_2O$(hydromagnesite) = $4Mg^{2+} + 3CO_3^{2-}\ 2OH^-$	K_{s0}	29.5	0
$CaMg(CO_3)_2$(dolomite) = $Ca^{3+} + Mg^{2+} + 2CO_3^{2-}$	K_{s0}	−16.7	0
$FeCO_3$(siderite) = $Fe^{2+} + 2CO_3^{2-}$	K_{s0}	−10.4 (−10.24)	0
$CdCO_3(s) + 2H^+ = CD^{2+} + H_2O + CO_2(g)$	K_{ps0}	6.44	1M $NaCl_4$
$MnCO_3(s) = Mn^{2+}2CO_3^{2-}$	K_{s0}	−10.41	0
III. SULFATES, SULFIDES, AND SILICATES			
$CaSO_4(s) = Ca^{2+} + SO_4^{2-}$	K_{s0}	−4.6	0
$H_2S = H + HS^-$	K_1	−7.0	0
$HS^- = H^+ + S^{2-}$	K_2	−12.96	0
MnS(green) = $Mn^{2+} + S^{2-}$	K_{s0}	−12.6	0
MnS(pink) = $Mn^{2+} + S^{2-}$	K_{s0}	−9.6	0
$FeS(s) = Fe^{2+} + S^{2-}$	K_{s0}	−17.3	0
SiO_2(quartz) + $2H_2 = H_4SiO_4$	K_{s0}	−3.7	
(am)$SiO_2(s) + 2H_2O = H_4SiO_4$	K_{s0}	−2.7	0
$H_4SiO_4 = H^+ + H_3SiO_4$	K_{s0}	−9.46	0
IV. ACIDS			
$NH_4^+ = N^+ + NH_3(aq)$	K_a	−9.3	0
$HOCl = H^+ + OCl^-$	K_a	−7.53	0

The constants given here are taken from quotations or selections in (a) L. G. Sillén and A. E. Martell, *Stability Constants of Metal Ion Complexes,* Special Publ., No. 17, the Chemical Society, London, 1964: (b) W. Feitknecht and P. Schindler, *Solubility Constants of Metal Oxides, Metal Hydroxides and Metal Hydroxide Salts in Aqueous Solutions,* Butterworths, London, 1963; (c) P. Schindler, “Heterogeneous Equilibria Involving Oxides, Hydroxides, Carbonates and Hydroxide Carbonates”, in *Equilibrium Concepts in Natural Water Systems, Advance in Chemistry Series,* No. 67, American Chemical Society, Washington, DC, 1967, p. 196; and (d) J. N. Butler, *Ionic Equilibrium, A Mathematical Approach,* Addison-Wesley Publishing, Reading, Mass., 1964. Unless otherwise specified a pressure of 1 atm is assumed.

[a] Most of the symbols used for the equilibrium constants are those given in *Stability Constants of Metal-Ion Complexes* Table from Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, pp. 168 and 169.

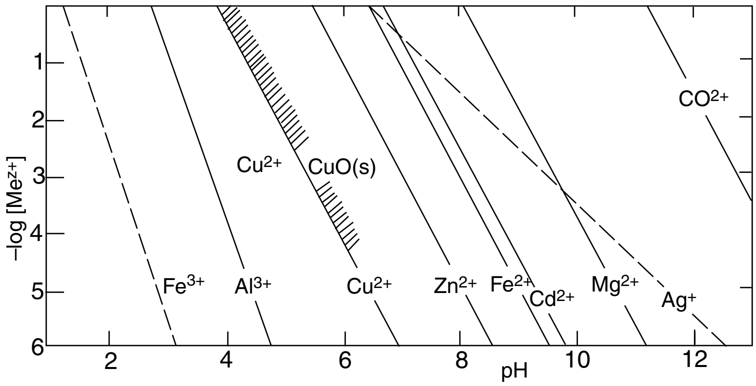


FIGURE 5 Solubility of oxides and hydroxides: free metal ion concentration in equilibrium with solid oxides ore hydroxides. As shown explicitly by the equilibrium curve for copper, free metal ions are constrained to concentrations to the left of (below) the respective curves. Precipitation of the solid hydroxides and oxides commences at the saturation concentrations represented by the curves. The formation of hydroxo metal complexes must be considered for the evaluation of complete solubility of the oxides or hydroxides. Ref.: Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 171.

The relations in Figure 5 do not fully describe the solubility of the corresponding oxides and hydroxides, since in addition to free metal ions, the solution may contain hydrolyzed species (hydroxo complexes) of the form. The solubility of the metal oxide or hydroxide is therefore expressed more rigorously as

$$\mathrm{Me}_T = [\mathrm{Me}^{z+}] + \sum_{1}^{n} [\mathrm{Me(OH)}_n^{z-n}]. \qquad (25)$$

Plots of this equation as a function of pH are given in Figure 6 for ferric hydroxide, zinc oxide, and cupric oxide.

Solubility of Carbonates

The maximum soluble metal ion concentration is a function of pH and concentration of total dissolved carbonate species. Calculation of the equilibrium solubility of the metal ion for a given carbonate for a water of a specific analytic composition discloses whether the water is over-saturated or undersaturated with respect to the solid metal carbonate. In the case of calcite

$$[\mathrm{Ca}^{2+}] = \frac{K_{s0}}{[\mathrm{CO}_3^{2-}]} = \frac{K_{s0}}{C_T \alpha_2} \cdot \qquad (26)$$

Since α_2 is known as a function of pH, Eq. (26) gives the equilibrium saturation value of Ca^{2+} as a function of C_T and pH. An analogous equation can be written for any metallic cation in equilibrium with its solid metallic carbonate. These equations are amenable to simple graphical representation in a log concentration versus pH diagram as illustrated in Figure 7.

Control of Solubility

Solubility calculations, such as those exemplified above, give thermodynamically meaningful conclusions, under the specified conditions (e.g., concentrations, pH, temperature and pressure), only if the solutes are in equilibrium with that solid phase for which the equilibrium relationship has been formulated. For a given set of conditions the solubility is controlled by the solid giving the smallest concentration of solute. For example, within the pH range of carbonate bearing natural waters, the stable solid phases regulating the solubility of Fe(II), Cu(II), and Zn(II) are, respectively, $FeCO_3$ (siderite), CuO (tenorite) and Zn (OH) (CO_3) (hydrozincite).

Unfortunately, it has not yet been possible to determine precise solubility data for some solids important in the regulation of natural waters. Among these are many clays and dolomite ($CaMg(CO_3)_2$), a mixed carbonate which constitutes a large fraction of the total quantity of carbonate rocks. The conditions under which dolomite is formed in nature are not well understood and attempts to precipitate it in the laboratory from solutions under atmospheric conditions have been unsuccessful. These difficulties in ascertaining equilibrium have resulted in a diversity of published figures for its solubility product, $(\{Ca^{2+}\}\{Mg^{2+}\}\{Co_3^{2-}\}2\}^{\dagger}$, ranging from $10^{-16.5}$ to $10^{-19.5}$ (25°C).

The Activity of the Solid Phase

In a solid-solution equilibrium, the pure solid phase is defined as a reference state and its activity is, because of its constancy,

---

† Note that this solubility product is expressed for activities, as represented by { }.

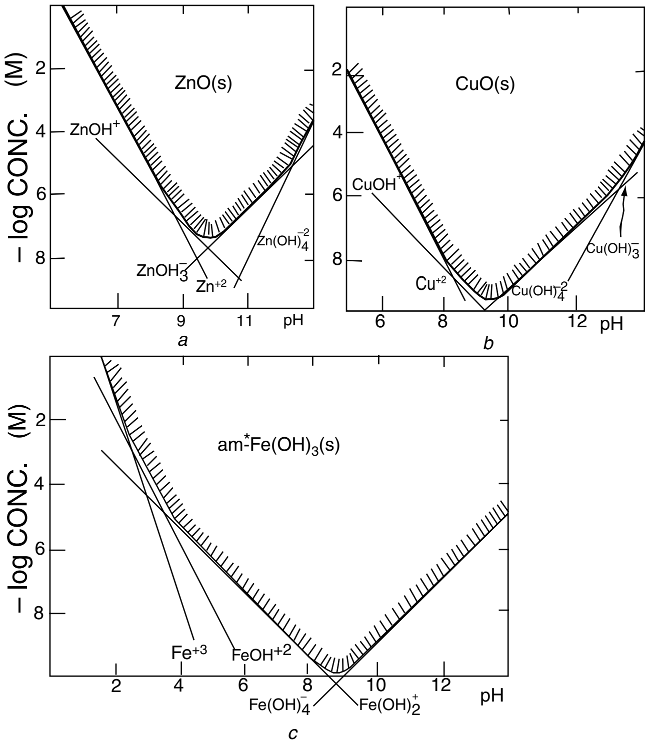


FIGURE 6 Solubility of amorphous $Fe(OH)_3$, ZnO, and CuO. The equilibrium solubility curves for the hydroxo metal complexes and for the free metal ion have been combined to yield the composite curve bordered by the cross hatching. The constituent curves were constructed from the data in Table 3. The possible occurrence of polynuclear complexes, for example $Fe_2(OH)_2^{2-}$, $CU_2(OH)_2^{2+}$, has been ignored. Such complexes do not change the solubility characteristics markedly for the solids considered. Also ignored is complexing with other ligands such as NH_3. Ref.: Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 173.

set equal to unity. This means that the activity of the solid phase is implicitly contained in the solubility equilibrium constant. Therefore, experimental differences from precisely known equilibrium constants can be used to deduce information about the constitution of the solid phase.

There are various factors that affect the activity of the solid phase: (1) the lattice energy, (2) the degree of hydration, (3) solid solution formation, (4) the free energy of the surface and (5) the presence of constituents affecting the purity of the solid.

The solids occurring in nature are seldom pure substances. For example, isomorphous replacement by a foreign constituent in the crystalline lattice is an important factor by which the activity of the solid phase may be decreased.

Redox Equilibria and Electron Activity

There is a conceptual analogy between acid-base and oxidation-reduction reactions. In a similar way that acids and bases have been interpreted as proton donors and proton acceptors, reductants and oxidants are defined as electron donors and electron acceptors. Since free electrons do not exist in solution, every oxidation is accompanied by a reduction and *vice versa.* An oxidant is thus a substance which causes oxidation to occur, while itself becoming reduced.

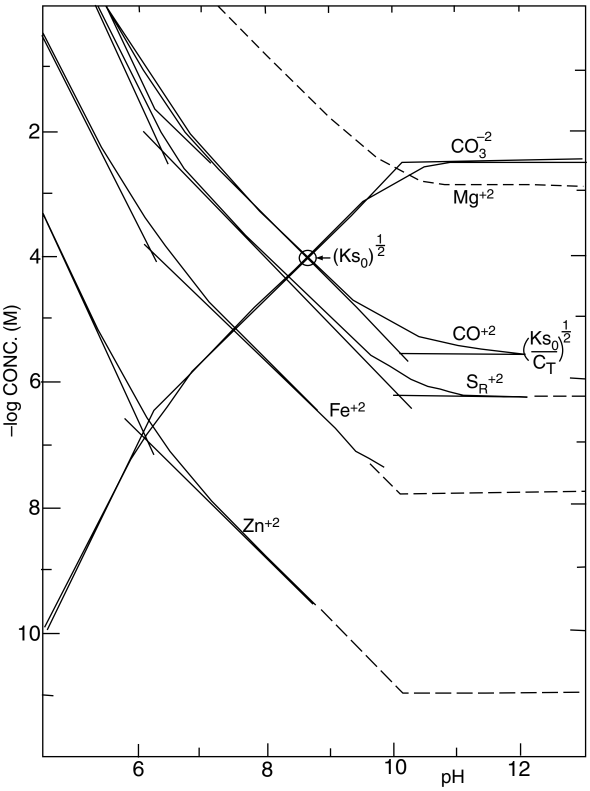


FIGURE 7 Solubility of carbonates in solutions of constant total dissolved carbonate carbon. The maximum soluble metal ion concentration ($[Me^{2+}] = k_{s0}/[(CO_3^{2+}] = K_{s0}/\alpha_2 C_T$) is shown as a function of pH for $C_T = 10^{-2.5}$ M. The acidity constants for $H_2CO_3{}^*$ are indicated on the horizontal axis. Dashed portions of the curves indicated conditions under which $MeCO_3(s)$ is not thermodynamically stable due to the formation of the more stable solid hydroxide or oxide. Ref.: Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 179.

The oxidation states of the reactants and products change as a result of the electron transfer which mechanistically may occur as a transfer of a group that carries one or more electrons. Since electrons are transferred in every redox reaction they can be treated conceptually like any other discrete reacting species, namely, as free electrons. The following redox reaction is illustrative:

$$\begin{array}{ll} O_2 + 4H^+ + 4e = 2H_2O & \text{reduction} \\ 4Fe^{2+} = 4Fe^{3+}\ 4e & \text{oxidation} \\ O_2 + 4Fe^{2+} + 4H^+ = 4Fe^{3+} + 2H_2O & \text{redox} \\ & \text{reaction} \quad (27) \end{array}$$

This treatment of electrons is no different from that of many other species such as hydrogen ions and silver ions, which do not actually exist in aqueous solution as free and unhydrated species but which are expressed as H^+ and Ag^+ in reactions. For example, the H^+ really takes the form of hydrated protons ($H_9O_4^+$) or hydrogen complexes (acids such as HCl).

Redox Intensity

By introducing the definition $pH = -\log[H^+]$, which under idealized conditions is formulated as

$$pH = -\log\{H^+\}, \tag{28}$$

Sørenson (1909) established a convenient intensity parameter that measures the relative tendency of a solution to donate or transfer protons. In an acid solution this tendency is high and in an alkaline solution it is low.

Similarly, Jørgensen (1945) has established an equally convenient redox intensity parameter,

$$p\varepsilon = -\log\{e\}, \tag{29}$$

that measures the relative tendency of a solution to donate or transfer electrons. In a highly reducing solution the tendency to donate electrons, that is, the hypothetical "electron pressure" or electron activity, is relatively large. Just as the activity of hypothetical hydrogen ions is very low at high pH, the activity of hypothetical electron is very low at high $p\varepsilon$. Thus a high $p\varepsilon\delta$ indicates a relatively high tendency for oxidation. In equilibrium equations H^+ and e are treated in an analogous way. Thus oxidation or reduction equilibrium constants can be defined and treated similarly to acidity constants as shown by the following equations:

For protolysis:

$$HA = H^+ + A^- \tag{30}$$

$$\{H^+\}\{A^-\}/\{HA\} = K_{HA}, \text{ or}$$

$$pH = -\log K_{HA} + \log(\{A^-\}/\{HA\}). \tag{31}$$

For the oxidation of Fe^{2+} to Fe^{3+}:

$$Fe^{2+} = Fe^{3+} + e \tag{32}$$

$$\{e\}\{Fe^{3+}\}/\{Fe^{2+}\} = K_{ox}, \text{ or}$$

$$p\varepsilon = -\log K_{ox} + \log(\{Fe^{3+}/Fe^{2+}\}). \tag{33}$$

As seen from Eq. (33) $p\varepsilon$ increases with the ratio of the activities (or concentrations) of oxidized to reduced species.[†]

From the following general reduction reaction:

$$aA + bB + ne = cC + dD$$

the generalized expression for any redox couple is given by

$$\begin{aligned} p\varepsilon &= p\varepsilon^0 - \frac{a}{n^{p^A}} - \frac{b}{n^{p^B}} + \frac{c}{n^{p^C}} + \frac{d}{n^{p^D}} \\ &= \frac{1}{n}\frac{\Delta G}{2.3RT'} \end{aligned} \tag{34}$$

where $pX = -\log[X]$ and

$$p\varepsilon = -\frac{1}{n}\log K_{ox} = \frac{1}{n}\log K_{red} = \frac{1}{n}\frac{\Delta G}{n2.3RT}$$

K_{ox} and K_{red} are the equilibrium constants for the oxidation and reduction reactions; n is the number of electrons transferred in the reaction. ΔG represents the free energy change for the reduction. Thus it is seen that $p\varepsilon$ is a measure of the electron free energy level per mole of electrons.

Equation (34) permits the expression of the redox intensity by $p\varepsilon$ for any redox couple for which the equilibrium constant is known. Numerical illustrations of the calculation of $p\varepsilon$ values (25°C) are given for the following equilibrium systems in which the ionic strength, I,[‡] is assumed to approach O:

a) An acid solution 10^{-5} M in Fe^{3+} and 10^{-3} M in Fe^{2+}.
b) A natural water at pH = 7.5 in equilibrium with the atmosphere ($Po_2 = 0.21$ atm.).
c) A natural water at pH = 8 containing 10^{-5} M Mn^{2+} in equilibrium with $\gamma - MnO_2(s)$.

"Stability Constant of Metal-Ion Complexes" gives the following equilibrium constants (K_{red}):

a) $Fe^{3+} + e = Fe^{2+}$;

$$K = \frac{\{Fe^{2+}\}}{\{Fe^{3+}\}\{e\}}; \log K = 12.53$$

b) $1/2\ O_2\,(g) + 2H^+ + 2e$;

$$H_2O(1)$$

[†] The sign convention adopted here is that recommended by IUPAC (International Union of Pure and Applied Chemistry).

[‡] Ionic strength, I, is a measure of the interionic effect resulting primarily from electrical attraction and repulsions between the various ions; it is defined by the equation $\tau = 1/2\ \Sigma_i C_i Z_i^2$. The summation is carried out for all types of ions, cations and anions, in the solution.

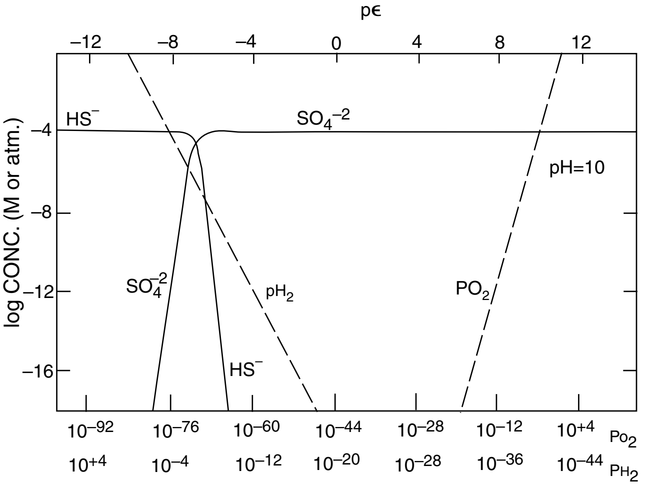


FIGURE 8 Equilibrium distribution of sulfur compounds as a function of pε at pH = 10 and 25°C. Total concentration is 10^{-4} M. The dotted curve shows that solid sulphur cannot exist thermodynamically at pH = 10, since its activity never becomes unity. Figure from Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 311.

$$K = \frac{1}{Po_2^{1/2}\{H^+\}^2\{e\}^2}; \log K = 41.55$$

c) $\gamma-MnO_2(s) + 4H^+ + 2e$

$= Mn^2 + 2H_2O(1);$

$$K = \frac{\{Mn^{2+}\}}{\{H^+\}^4\{e\}^2}; \log K = 40.84.$$

For the conditions stipulated the following pε values are obtained:

a) $p\varepsilon = 12.53 + \log\frac{\{Fe^{3+}\}}{\{Fe^{2+}\}} = 10.53$

b) $p\varepsilon = 20.78 + 1/2\log(Po_2^{1/2}\{H^+\}^2) = 13.11$

c) $p\varepsilon = 20.42 + 1/2\log\frac{\{H^+\}^4}{\{Mn^{2+}\}} = 6.92$

Equilibrium Distribution in the Sulphur System

Figure 8 shows the pε dependence of a 10^{-4}M SO_4^{2-} $-HS^-$system at pH = 10 and 25°C. The reaction is

$$SO_4^{2-} + 9H^+ + 8e = HS^- + 4H_2O(1) \qquad (35)$$

and the redox equilibrium equation is

$$p\varepsilon = 1/8\log K + 1/8\log\frac{[SO_4^{2-}][H^+]}{[HS^-]} \qquad (36)$$

where log *K* (for the reduction reaction) is 10^{34}. Hence,

$$p\varepsilon = 4.25 - 11.25pH + 1/8\log[SO_4^{2-}] - 1/\log[HS^-]$$

or, for pH = 10,

$$p\varepsilon = -7 + 1/8\log[SO_4^{2-}] - 1/8\log[HS^-].$$

HS^- is the predominant S(−II) species at pH = 10. Figure 8 shows that the lines for $[SO_4^{2-}]$ and $[HS^-]$ intersect at pε = −7. The asymptotes for $[SO_4^{2-}]$ have slopes of +8 and 0, whereas those for $[HS^-]$ have slopes of 0 and −8.

Lines for the equilibrium partial pressure of O_2 and H_2 are also given in the diagram. As the diagram shows, rather high relative electron activities are necessary to reduce SO_4^{2-}. At the pH value selected, the reduction takes place at pε values slightly less negative than for the reduction of water. Thus in the presence of oxygen and at pH = 10, only sulfate can exist; its reduction is possible only at pε values less than −6.

Equilibrium Constants for Redox Reactions

Equilibrium constants for some redox processes pertinent in aquatic conditions are listed in Table 4. A quite comprehensive reference source for such constants is *Stability Constants for Metal-Ion Complexes,* L. G. Sillén and A. E. Martell, The Chemical Society, London (1964) and its Supplement (1971). Significantly, the first section of this reference deals with the electron as a ligand, similarly to its treatment above. Another compilation, somewhat outdated though still very useful, is *Oxidation Potentials,* 2nd ed., W. M. Latimer, Prentice-Hall, (1952). This treatise lists redox potentials rather than equilibrium constants, but, as shown in the next section, the latter are readily obtained from the former.

The Determination of pε and Redox Potential

As with pH, pε can be measured with a potentiometer using an indicator electrode (e.g., a platinum or gold electrode) and a reference electrode. The result is read as a potential difference in volts. When a reversible hydrogen electrode, at which the electrode reaction is $H_2 = 2H^+ + 2e$, is used as the reference, the resulting potential difference is termed the redox potential, E_H, where the suffix *H* refers to the hydrogen electrode as the reference. Usually another reference electrode is used, e.g., a calomel electrode, but the addition of a constant factor (i.e., the potential difference between the calomel electrode and the hydrogen electrode) to the observed voltage gives E_H. pε is then readily computed by dividing E_H by a constant factor according to Eq. (27):

$$p\varepsilon\frac{E_H}{(2.3RT/F)},$$

where

F = the Faraday constant (96,500 coulombs per equivalent),

R = the gas constant (8.314 volt-coulombs per degree—equivalent),

and

T = the absolute temperature in degrees Kelvin.

At 25°C, $(2.3\ \omega T/F) = 0.059$ volt.

Both conceptually and computationally, it is more convenient to use the dimensionless pε rather than the more directly measured E_H, as a measure of the redox intensity. Every tenfold change in the activity ratio of Eq. (34) causes a change in pε of one unit divided by the number of electrons transferred in the redox reaction. The fact that one electron can reduce one proton is another reason for expressing the intensity parameter for oxidation in a form equivalent to that used for acidity.

Incidentally, pH is also not measured directly. It is determined by measuring the potential (in volts) of an indicator electrode (e.g., a glass electrode) with respect to a reference electrode. If the hydrogen electrode is used or referred to as the reference electrode, the resulting potential difference is called the acidity potential. The pH is calculated by dividing the acidity potential by 2.3 RT/F.

The Peters-Nernst Equation

Since much literature still describes the electron activity in volts rather than on a scale[†] similar to that for other reagents, the Peter-Nernst equation is still of utility. It is easily expressed by substituting Eq. (37) into Eq. (34):

$$\begin{aligned} E_H &= E_H^0 + \frac{2.3RT}{nF}\log\frac{\{\text{oxidized}\}}{\{reduced\}} \\ &= \frac{-\Delta G}{nF}, \end{aligned} \tag{38}$$

where

$$E_H^0 = (2.3RT/F)p\varepsilon; = \frac{-\Delta G^0}{nF}. \tag{39}$$

Measurement of E_H

It is essential to distinguish between the concept of a potential and the measurement of a potential. Redox or electrode potentials (as listed in "Stability Constants of Metal-Ion Complexes" and other references) have been derived from equilibrium data, thermodynamic data, the chemical behavior of a redox couple with respect to known oxidizing and reducing agents, and from direct measurements of electrochemical cells.

Direct measurement of E_H for natural water environments involves complex theoretical and practical problems in spite of the apparent simplicity of the electrochemical technique. For example, the E_H of aerobic (dissolved oxygen containing) waters, measured with a platinum or gold electrode, does not agree with that predicted by Eq. (38). Even when reproducible results are obtained, they often do not represent reversible Nernst potentials. Among the considerations hindering the direct measurement of E_H are the rates of electron exchange at certain electrodes and the occurrence of mixed potentials. A mixed potential results when the rate of oxidation of one redox couple is compensated by the rate of reduction of a different couple during the measurement. Although aqueous systems containing oxygen or similar oxidizing agents will usually give positive E_H values and anaerobic systems will usually give negative ones, detailed quantitative interpretation with respect to concentrations of redox species is generally unwarranted.

Since natural waters are normally in a dynamic rather than an equilibrium condition, even the concept of a single oxidation–reduction potential characteristic of the aqueous system as a whole cannot be justified. At best, measurement can give rise to an E_H value applicable to a particular redox reaction or to redox species in partial chemical equilibrium and even then only if these redox agents are electrochemically reversible at the electrode surface at a rate that is rapid compared with the electron drain or supply by the measuring electrode system.

pε–pH *Diagrams*

A pε–pH stability field diagram shows in a panoramic way which species predominate at equilibrium under any condition of pε (or E_H) and pH. The primary value of a pε–pH diagram is its simultaneous representation of the consequences of the equilibrium constants of many reactions for any combination of pε and pH. Figure 9 shows stability fields for the various species pertinent to the chlorine system. These diagrams are readily constructed from thermodynamic data such as those listed in Table 4.

Log Concentration—pε Diagrams

The predominant redox species are depicted as a function of pε or redox potential in pε-log concentration diagrams. Upper or lower bounds of pε values for the occurrence of specific redox reactions are immediately evident from these double log-arithmetic diagrams. Such diagrams, constructed from the data in Table 4 with pH = 7, are shown in Figure 10 for a few elements in the biochemical cycle.

[†] With dilute solutions it is convenient to express concentrations on a minus log concentration scale, e.g., $pNa^+ = -\log[Na^+]$.

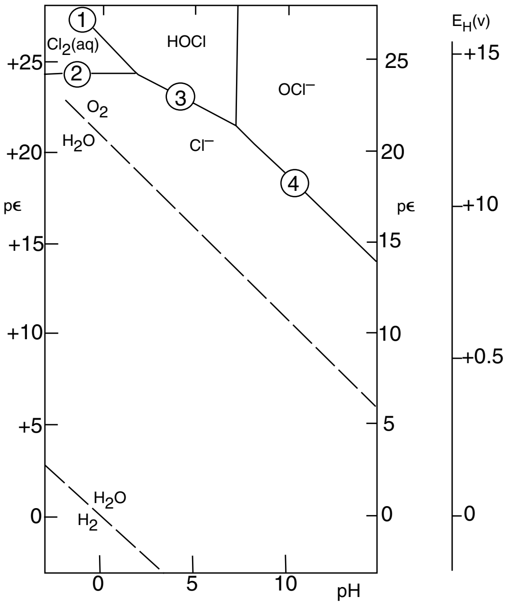


FIGURE 9 Stability field diagram for the chlorine system: pε versus pH. The curves labelled 1,2,3 and 4 represent equilibria derived from Eqs. (25) and (26) of Table 4 and pK_a for HOCl: This latter value is 7.53 at 25°C and is represented by the vertical line between HOCl and OCl$^-$ stability fields. In dilute solutions Cl$_2$ (aq) exists only at low pH. Cl$_2$, OCl$^-$, and HOCl are all unstable or metastable in water since they are all slowly reduced by water as shown by the position of their stability fields, with respect to the H$_2$O–O$_2$ equilibrium curve. Figure from Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 320.

The boundary conditions for the stability of water (Figure 10(d)) are given at high pε values by the oxidation of water to oxygen:

$$1/2\ H_2O = 1/4\ O_2 + H^+ + e \qquad (40)$$

and at low *pε* values by the reduction of water to hydrogen

$$H_2O + e = 1/2\ H_2 + OH^-. \qquad (41)$$

Water in equilibrium with the atmosphere ($P\ o_2 = 0.21$ atm.) at pH = 7.0 (25°C) has a pε = 13.6 (E_H = 0.8 volt.)

Figure 10(a) gives the relationships among several oxidation states of nitrogen as a function of pε. For most of the aqueous range of pε, N$_2$ gas is the most stable species. However, at large negative pε values ammonia becomes predominant and for pε greater than +12 nitrate dominates pH = 7. The fact that the nitrogen gas of the atmosphere has not been converted largely into nitrate under the prevailing aerobic conditions at the land and water surfaces indicates a lack of efficient biological mediation.

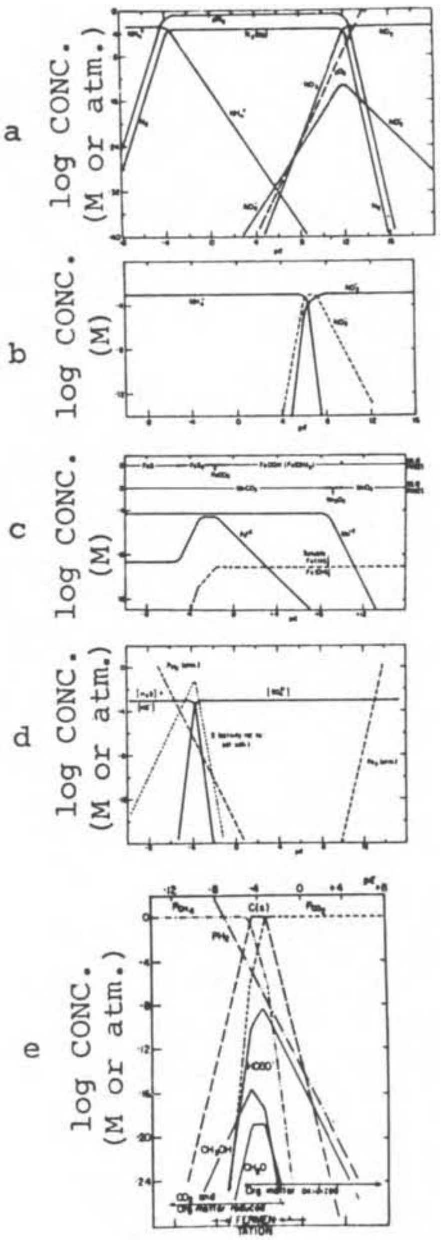


FIGURE 10 Equilibrium concentrations of biochemically important redox components as a function of pε at a pH of 7.0. These equilibrium diagrams have been constructed from equilibrium constants listed in Table 3 and 4 for the following concentrations representative of natural water systems: pH = 7.0; C_T (total carbonate carbon) = 10^{-3} M; [H$_2$(aq)] = [H$_2$S(aq)] + [HS$^-$] + [SO$_4^{2-}$] = 10^{-5}M; [NO^{3-}] + [NO^{2-}] + NO^{4+}] = 10^{-3}M; $p_{N2}{}^2$ = 0.78 atm. and thus [N$_2$(aq)] = 0.5 × 10^{-3}M. Figure from Stumm, W. and J. Morgan, *Aquatic Chemistry,* Wiley-Interscience, New York, 1970, p. 331.

TABLE 4

Equilibrium constants of redox processes pertinent in aquatic conditions (25°C)

	Reaction	$p\varepsilon°(= -\log K)$	$p\varepsilon°(W)$[a]
(1)	1/4 O_2(g) + H^+ (W) + e = 1/2 H_2O	+20.75	+13.75
(2)	1/5 NO_3^- + 6/5H^+ (W) + e = 1/10 N_2(g) + 3/5 H_2O	+21.05	+12.65
(3)	1/2 MnO_2 (S) + 1/2HCO_3^- + 3/2 H^+ (W) + e = 1/2$MnCO_3$(s) + 3/8 H_2O	+20.46	+9.96 +8.46
(3a)	1/2 MnO_2(s) + 2H^+ (W) + e = 1/2 Mn^{2+} + H_2O	+20.42	+6.42
(4)	1/2 NO_3^- + H^+ (W) + e = 1/2 NO_2^- + 1/2 H_2O	+14.15	+7.15
(5)	1/8 NO_2 + 5/4 H^+ (W) + e = 1/8 NH_4^+ + 3/8 H_2O	+14.90	+6.15
(6)	1/6 NO_2^- + 4/3 H^+ (W) + e = 1/6 NH_4^+ + 1/3 H_2O	+15.14	+5.82
(7)	1/2 CH_3OH + H^+ (W) + e = 1/2 CH_4(g) + 1/2 H_2O	+9.88	+2.88
(8)	1/4 CH_2O + H^+ (W) + e = 1/4 CH_4(g) + 1/4 H_2O	+6.94	−0.06
(9)	FeOOH(s)+ HCO_3^- + 2H^+ (W) + e = $FeCO_3$(s) + 2H_2O	+15.33	+1.33 −1.67[b]
(9a)	Fe^{3+} + e = Fe^{2+}	+12.53	+ 12.53
(10)	1/2 CH_2O + H^+ (W) + 3 = 1/2 CH_3OH	+3.99	−3.01
(11)	1/6 SO_4^{2-} + 4/3 H^+ (W) + e = 1/6 S(s) + 2/3 H_2O	+6.03	−3.30
(12)	1/8 SO_4^{2-} + 5/4H+ (W) + 3 = 1/8 H_2S(g) + 1/2 H_2O	+5.75	−3.50
(13)	1/8 SO_4^{2-} + 9/8 H^+ (W) + e = 1/8 HS^- + 1/2 H_2O	+4.13	−3.75
(14)	1/2 S(s) + H^+ (W) + e = 1/2 H_2S(g)	+2.89	−4.11
(15)	1/8 CO_2(g) + H^+ (W) + e = 1/8 CH_4(g) + 1/4 H_2O	+2.87	−4.13
(16)	1/6 N_2(g) + 4/3 H^+ (W) + e = 1/3 NH_4^+	+4.68	−4.68
(17)	1/2 ($NADP^+$) 1/2 H^+ (W) + e =1/2 (NADPH)	−2.0	−5.5[c]
(18)	H^+ (W) + e = 1/2 H_2(g)	0.0	−7.00
(19)	Oxidized ferredoxin + e = reduced ferredoxin	−7.1	−7.1[d]
(20)	1/4 CO_2(g) + H^+ (W) + e = 1/24 (glucose) + 1/4 H_2O	−0.20	−7.20[e]
(21)	1/2 $HCOO^-$ + 3/2H^+ (W) + e = 1/2 CH_2O + 1/2 H_2O	+2.82	−7.68
(22)	1/4 CO_2(g) + H^+ (W) + e = 1/4 CH_2O + 1/4 HO_2	−1.20	−8.20
(22a)	1/4 HCO_3^- + 5/4 H^+ (W) + e = 1/4 CH_2O + 1/2 H_2O	+0.76	−7.99 −8.74[b]
(23)	1/2 CO_2(g) + 1/2 H^+ (W) + e = 1/2 $HCOO^-$	−4.83	−8.73
(24)	1/2 $C1_2$(aq) + e = $C1^-$	+23.6	+23.6
(25)	HOCl + H^+ (W) + e = 1/2 $C1_2$ (aq) + H_2O	+26.9	+19.9

[a] (*W*) signifies that the pH = 7.

[b] These data correspond to (HCO_3^-) = 10^{-3}M rather than unity and so are not exactly $p\varepsilon(W)$; they represent typical aquatic data conditions more nearly than $p\varepsilon(W)$ values do.

[c] M. Calvin and J.A. Bassham, *The Photosynthesis of Carbon Compounds*, Benjamin, New York, 1962.

[d] D. I. Arnon, *Science*, **149,** 1460 (1965).

[e] A. L. Lehninger, *Bioenergetics*, Benjamin, New York, 1965:

Table from Stumm, W. and J. Morgan, *Aquatic Chemistry*, Wiley-Interscience, New York, 1970, p. 318.

Because of these peculiarities in equilibria with N_2, Figure 10(b) has been constructed for the metastable equilibria between NO_3^-, and NO_2^- by assuming the transition between bound nitrogen and N_2 to be hindered. This diagram shows that the shifts in relative predominance of the three species, NH_4^+, NO_2^- and NH_3^- occur within a rather narrow $p\varepsilon$ range. That each of these species has a dominant zone within this range would seem to be a contributing factor to the highly mobile characteristics of the fixed species of the nitrogen cycle. Note that the relevant species of N (−III) at pH = 7 is NH_4^+ rather than NH_3 since the ammonia system $pK_a = 9.3$.

The reduction of SO_4^{2-} to H_2S or HS^- provides a good example of the application of equilibrium concepts to aquatic relationships. Figure 10(d) shows that significant reduction of SO_4^{2-} to H_2S at pH = 7 requires $p\varepsilon < -3$. The biological enzymes that mediate this reduction must therefore operate at or below this $p\varepsilon$. Because the system is dynamic rather than static only an upper bound can be set in this way, for the excess driving force in terms of $p\varepsilon$ at the mediation site cannot be determined by equilibrium computations. Since, however, many biologically mediated reactions seem to operate with relatively high efficiency in utilizing free energy, it

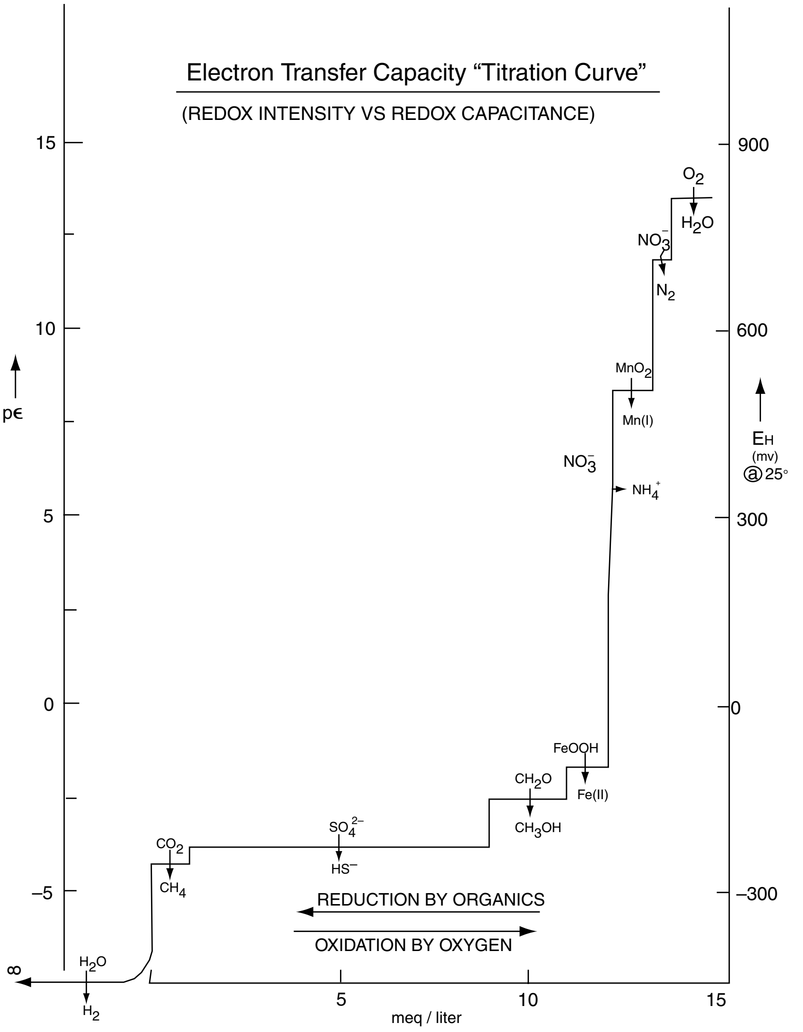


FIGURE 11 Electron transfer capacity "Titration Curve." In a system with excess organic material, the redox intensity falls as the electron acceptors are successively reduced. This diagram was constructed from the redox intensities and initial concentrations (expressed as electron equivalents) listed in Table 3. The reactions indicated on the curve proceed sequentially during the stagnation period in the deeper waters of a polluted lake, vertically downward or temporally in sediments, sequentially after starting an anaerobic digester and chronologically in ground water contaminated with organic nutrients.

appears likely that the operating p pε value does not differ greatly from the equilibrium value. The diagram also indicates the possibility of formation of elemental sulfur within a narrow pε range at neutral and low pH.

The major feature of the carbon system, shown in Figure 10(e), is simply the interconvertibility of C(IV) to C (−IV): i.e., the reduction of carbon dioxide bicarbonates and carbonates to methane and the reverse oxidation reactions. Formation of solid carbon is thermodynamically possible close to $p\varepsilon = -4$ but its inclusion does not change other relationships. Other carbon compounds exist under equilibrium conditions only at very small concentrations ($<10^{-9}$ M). The existence of myriad synthetic and biochemical organic compounds at ambient pε levels is due to their exceedingly slow rates of equilibration to CO_2 or CH_3 or, in the case of the complicated organic structures in living systems, to the constant input of energy.

Microbial Mediation and Free Energy of Redox Reactions

To survive and hence reproduce, microorganisms must not only capture a significant fraction of the thermodynamically available energy but must also acquire this energy at a rate compatible with the maintenance of life. Thus the salient capability is that of power production per unit biomass and therefore kinetics (or the rate of movement toward equilibrium) should be considered as well as thermodynamics. To provide for such energy production microorganisms have developed highly efficient and specific biological compounds (enzymes) which catalyze energy yielding reactions and cell constituent producing processes.

Organisms do not oxidize organic substrates or reduce O_2 or SO_4^{2-}; they only mediate those reactions which are thermodynamically possible, or more specifically, the electron transfer occurring in these reactions. The pε range in which certain oxidation or reduction reactions are possible may be estimated by calculating the equilibrium concentrations of the relevant species as a function of pε. Since, for example, SO_4^{2-} can be reduced only below a given pε or redox potential, an equilibrium model can characterize the pε ranges in which reduction of sulfate is possible and is not possible. Such models are graphically presented in Figure 11, where all the reactions are amenable to microbial mediation. These diagrams manifest the use of pε as a parameter that characterizes the ecological milieu in a restrictive fashion.

The data in Table 4 permit the calculation of combinations of the listed half reactions to give complete redox reactions. Those that are thermodynamically possible are always accompanied by a decrease in free energy. The free energy change of the complete redox reaction, $\Sigma\Delta G$, is easily calculated from the pε values through rearrangement of Eq. (34):

$$\sum \Delta G = -2.3\,RT \times [(n\mathrm{p}\varepsilon)_{\text{for reduction}} \sum (n\mathrm{p}\varepsilon)_{\text{for oxidation}}]. \tag{42}$$

If $-\Delta G$ is negative, the reaction can occur. Combinations that lead to energetically possible reactions are given in Table 3. All of these reactions are mediated by microorganisms.

REFERENCES

1. American Chemical Society: Equilibrium concepts in natural water systems, *Advances in Chem. Series,* **67,** Washington (1967).
2. Berner, R.A., The Benthic Boundary Layer from the viewpoint of a geochemist, in *The Benthic Boundary Layer,* I.N. McCave, Ed., Plenum, New York, 1976, pp. 33–55.
3. Berner, R.A., *Early Diagenesis; A Theoretical Approach,* Princeton University Press, Princeton, 1980.
4. Brewer, P.G., Minor elements in seawater, in *Chemical Oceanography,* J.P. Riley and G.
5. Broecker, W.S., *Chemical Oceanography,* Harcourt Brace Jovanovich, New York, 1974.
6. Broecker, W.S., A kinetic model for the chemical composition of seawater, *Quatern. Res.,* **1,** 188–207 (1971).
7. Butler, J.N., *Ionic Equilibrium a Mathematical Approach,* Addison-Wesley, Reading, Mass. (1964).
8. Conway, B.E., *Annual Review of Physical Chemistry,* **17,** Annual Reviews Inc., Palo Alto Calif. (1966).
9. Dorsey, N.E., *Properties of Ordinary Water-Substance,* ACS monograph No. 81, Reinhold Publishing Corp., New York (1940).
10. Drever, J.I., *The Geochemistry of Natural Waters,* Prentice Hall, Englewood Cliffs, 1981.
11. Eisenberg, D. and W. Kauzmann, *The Structure and Properties of Water,* Oxford University Press, England, 1969.
12. Fletcher, N.H., *The Chemical Physics of Ice,* Cambridge University Press, England, 1970.
13. Frank, H.S. *Science,* **169,** 635 1970.
14. Freze, R.A. and J.A. Cherry, *Groundwater,* Prentice Hall, Englewood Cliffs, NJ, 1979.
15. Garrels, R.M. and F.T. Mackenzie, Origin of the chemical composition of some spring and lakes, in *Equilibrium Concepts in Natural Water Systems,* Advances in Chemistry Series, No. 67, American Chemical Society, Washington, DC, 1967, pp. 222–242.
16. Garrels, R.M. and C.L. Christ, *Solutions, Minerals and Equilibria,* Harper and Row, New York, 1965.
17. Hem, J.D., *Study and Interpretation of the Chemical Characteristics of Natural Water,* US Geological Survey Water Supply Paper, No. 1473, Washington, DC, 1970.
18. Hoigne, J. and H. Bader, Combination of ozone/UV and Ozone/hydrogen peroxide; Formation of Secondary Oxidants, in *Proc. 8th World Conference on Ozone, Zurich, 1987.*
19. Hoigne, J., The Chemistry of Ozone in Water, in *Process Technologies for Water Treatment,* Plenum, New York, 1988.
20. Holland, H.D., *The Chemistry of the Atmosphere and Oceans,* Wiley-Interscience, New York, 1978.
21. Horne, R.A., *Marine Chemistry, The Structure of Water and the Chemistry of the Hydrosphere,* Wiley-Interscience, New York, 1969.
22. Horne, R.A., *Water and Aqueous Solutions, Structure, Thermodynamics, and Transport Processes,* Wiley-Interscience, New York, 1972.
23. Horne, R.A., *Surv. Progr. Chem.,* **4** (1968) 2.
24. Hush, N.S., *Reactions at Electrodes,* Wiley-Interscience New York (1971).
25. Kamb, Structure of the ices, in Horne, R.A., *Water and Aqueous Solutions,* Wiley-Interscience, New York (1972).
26. Kavanau, J.L., *Water and Solute-Water Interactions,* Holden-Day, San Francisco, 1964.
27. Keyes, F.G., *J. Chem. Phys.* **17,** 923, 1949.
28. King, T., *Water: Miracle of Nature,* The Macmillan Company, New York, 1953.
29. Klein, H.P., Ozone in Water Treatment Processes, in *Process Technologies for Water Treatment,* Plenum, New York, 1988.
30. Krindel, P. and I. Eliezer, *Coord. Chem. Rev.,* **6,** 217, 1971.
31. Lerman, A., Geochemical Processes: *Water and Sediment Environments,* Wiley-Interscience, New York, 1979.

32. Martin, J.M. and M. Meybeck, Elemental mass-balance of material carried by major world rivers, *Marine Chem.,* **7,** 173, 1979.
33. Narten, A.H. and H.A. Levy, *Science,* **165,** 447, 1969.
34. Pimental, G.C. and A.L. McClellan, *The Hydrogen Bond,* W.H. Freeman and Co., San Francisco, 1960.
35. Punzi, V. and B.M. Nebens, The Chemistry of Seawater Chlorination. *Advances in Environmental Science and Engineering,* Vol. 5, Gordon and Breach, New York, 1986.
36. Riley, J.P. and G. Shirrow (eds.), *Chemical Oceanography,* 2 Volumes, Academic Press, London, 1965.
37. Riley, J.P. and R. Chester, *Introduction to Marine Chemistry,* Academic Press, London, 1971.
38. Sayles, F.L., The Composition and Diagenesis of Interstitial Solutions—Fluxes across the Seawater-Sediment Interface, *Geochim. Cosmochim. Acta,* **43,** 527–545, 1979.
39. Skinner, B.J., *Earth Resources,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1969.
40. Stumm, W. and J. J. Morgan, *Aquatic Chemistry: An Introduction Emphasizing Chemical Equilibria to the Chemistry of Natural Waters,* Wiley-Interscience, New York, 2nd ed., 1981.
41. "Water II" Topic of *Journal Colloid Interface Science,* **36,** No. 4, 1971.
42. White, G.C., Handbook of Chlorination, Van Nostrant Reinhold, New York, 1989.

WERNER STUMM (DECEASED)
Swiss Federal Institute of Technology, Zürich

MARTIN FORSBERG

STEVEN GHERINI
Harvard University

WASTES OF INDUSTRY: see INDUSTRIAL WASTE MANAGEMENT

WATER—DESALTING: see DESALINATION

WATER FLOW

PROPERTIES OF FLUIDS

The fluid properties most commonly encountered in water flow problems are presented in the following paragraphs. The International System of units is used throughout the discussion unless otherwise stated to the contrary.

The unit of *mass, m,* is the kilogram (kg). A mass of one kg will be accelerated by a force of one newton at the rate of 1 m per sec^2.

The *density, ρ,* of a fluid is its mass per unit volume and is expressed in kilograms per cubic meter.

The *specific weight, γ,* is the weight per unit volume and denotes the gravitational force on a unit volume of fluid and is expressed in newtons per cubic meter.

Fluid density and specific weight are related by the expression:

$$\rho = \frac{\gamma}{\mathbf{g}} \tag{1}$$

in which **g** is the acceleration due to gravity.

The *specific gravity* of a fluid is found by dividing its density by the density of pure water at 4°C.

The relative shearing force required to deform a fluid gives a measure of the *viscosity* of the fluid. An increase in temperature causes a decrease in viscosity of a liquid and *vice versa.* Consider the space between two parallel plates (Figure 1) which is filled with fluid; the bottom plate remains at rest while the upper plate moves with velocity V under an applied force. The velocity of the fluid particles will range from V at the top boundary to zero at the bottom as they will assume the same velocity as the boundary in which they are in contact. Experiments have demonstrated that the shear stress, τ, is directly proportional to the rate of deformation, du/dy. Mathematically, this can be written as:

$$\tau = \frac{\mathrm{d}u}{\mathrm{d}y} \cdot \tag{2}$$

Equation (2) is known as Newton's equation of viscosity. The constant of proportionality, μ, in newton-second per square meter (N-s/m^2), is termed the coefficient of viscosity, the dynamic viscosity or the absolute viscosity.

The *kinematic viscosity, v,* is defined as the ratio of the coefficient of viscosity to the density and is expressed in

$$v = \frac{\mu}{\rho} \cdot \tag{3}$$

A more proper term for *surface tension, σ,* would be surface energy. Surface tension is a liquid surface phenomenon and is caused by the relative forces of cohesion, the attraction of liquid molecules for each other, and adhesion, the attraction of liquid molecules for the molecules of another liquid or solid. Surface tension has the units of newtons per meter (N/ m). When a liquid surface is in contact with a solid, a contact angel θ, greater than 90° results with depression of the liquid surface if the liquid does not "wet" the tube such as mercury and glass. If the solid boundary has a greater attraction for a liquid molecule than the surrounding liquid molecules, then the contact angle is less than 90°C and the liquid is said to "wet" the wall leading to a capillary rise as in the case of water and glass.

Table 1 gives the values of the fluid properties discussed in the preceding paragraphs for a few common fluids.

PRESSURE FLOW

Friction Formulae

Darcy-Weishbach's Equation The Darcy-Weishbach formula was first proposed empirically but later found by dimensional reasoning to have a rational basis:

$$h_f = \frac{fLV^2}{2\mathbf{g}D} \tag{4}$$

in which f = friction factor, L = pipe length, V = mean velocity, D = diameter, h = head loss, **g** = acceleration due to gravity.

Equation (4) was derived for circulation sections flowing full and the equation itself is dimensionally homogeneous. It can be extended to other cross-sections provided these shapes are not too different from circular; in this case, the equation has to be transformed by using the hydraulic radius, R, instead of the diameter, D:

$$h_f = \frac{fLV^2}{8\mathbf{g}R}, \tag{5}$$

where $r = D/4$ for flow at full bore. The use of Darcy's equation in the form given by Eq. (5) is sometimes extended to open channel flow.

The determination of the friction factor, *f,* depends on the flow regime, that is, whether the flow is laminar, critical, transitional, smooth, turbulent or rough fully turbulent.

Laminar Flow Consider the mean pipe velocity, *V,* as given by Hagen-Poiseuille's equation for laminar flow:

$$V = \frac{\gamma SD}{32\mu} \tag{6}$$

in which S = energy slope.

Combining Eq. (4) with Eq. (6) and noting that $S = H_f/L$, $\gamma = \rho \mathbf{g}$, and $\nu\mu/\rho$, the friction factor is given by:

$$f = \frac{64}{R_e}, \tag{7}$$

where $R_e = \nu D/\gamma$ is the Reynolds number. Eq. (7) can be used for all pipe roughness as the friction factor in laminar flow is independent of the wall protuberances and is inversely proportional to the Reynolds number. The energy loss varies directly as the mean pipe velocity in laminar flow which persists up to a Reynolds number of about 2000.

The velocity profile, which has a parabolic distribution, can be obtained from Hagen-Poiseuille's equation. The velocity, *u,* at any radius, *r,* of the pipe of diameter, *D,* is given by:

$$u = \frac{\gamma S}{4\mu}\left(\frac{D^2}{4} - r^2\right). \tag{8}$$

At the centre line, the velocity is a maximum:

$$u_{\max} = \frac{\gamma SD^2}{16\mu} \tag{9}$$

The mean velocity is:

$$u_{\text{mean}} = (u_{\max})/2 = \frac{\gamma SD^2}{32\mu}. \tag{10}$$

Critical Flow From a Reynolds number of about 2000 and extending to 4000 lies a critical zone where the flow may be either laminar or turbulent. The flow regime is unstable and no equation adequately describes it.

Smooth Turbulent For pipes fabricated from hydraulically smooth materials such as copper, plexiglass and glass, the flow is smooth turbulent for a Reynolds number exceeding 4000. The von Karman-Nikuradse smooth pipe equation is:

$$\frac{1}{\sqrt{f}} = 2\log_{10}(R_e\sqrt{f}) - 0.80. \tag{11}$$

Equation (11) indicates that the friction factor depends on the fluid properties and deceases with increasing Reynolds number.

Rough Fully Turbulent Nikuradse experimented with pipes artificially roughened with uniform sand grains. The results were fitted to the theory of Prandtl–Karman to give the well known rough-pipe equation:

$$\frac{1}{\sqrt{f}} = 2\log_{10}(D/\varepsilon) + 1.14. \tag{12}$$

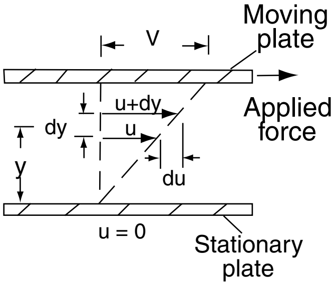


FIGURE 1 Fluid shear.

TABLE 1
Fluid properties

Fluid	Temperature °C	Mass density kg/m^3	Specific weight kN/m^3	Dynamic viscosity N-s/m^3	Kinematic viscosity m^2/s	Surface tension N/m
Water	0	1000	9.81	1.75×10^{-3}	1.75×10^{-6}	0.0756
—	5	1000	9.81	1.52×10^{-3}	1.52×10^{-6}	0.0754
—	10	1000	9.81	1.30×10^{-3}	1.30×10^{-3}	0.0742
Mercury	20	13,570	133.1	1.56×10^{-3}	1.15×10^{-7}	0.514
Sea water	20	1028	10.1	1.07×10^{-3}	1.04×10^{-6}	0.073

Equation (12) states that beyond a certain Reynolds number, when the flow is fully turbulent, the friction factor is influenced only by the relative roughness,*ε/D* and independent of the Reynolds number.

Transition Flow Most commercial pipe flows do not follow either the smooth pipe or rough pipe equations. Colebrook and White proposed a transitional flow equation which would be asymptotic to both:

$$\frac{1}{\sqrt{f}} = -2\log_{10}\left(\frac{e/D}{3.7} + \frac{2.51}{R_e\sqrt{f}}\right). \tag{13}$$

Equation (13) approaches the smooth pipe equation for low and the rough pipe equation for high values of the Reynolds number respectively. Unlike Nikuradse's ε, which represents the actual height of the sand grains, the ε of Colebrook—White's equation is not an actual roughness dimension but a representative height describing the roughness projections. It is referred to as the equivalent sand-grain diameter since the friction loss it represents is the same as the equivalent sand-grain diameter; Table 2 gives experimentally observed values:

TABLE 2
Equivalent sand-grain diameter

Pipe material	ε (mm)
Riveted steel	9.14
Rough concrete	3.05
Smooth concrete	0.31
Steel	0.05

Moody Diagram (Moody, 1944.) The Moody Diagram (Figure 2) summarises and solves graphically the four friction factor equations Eqs. (7), (11), (12), (13) as well as delineating the zones of the various flow regimes. The line separating transitional and fully turbulent flow is given by Rouse's equation:

$$\frac{1}{\sqrt{f}} = \frac{R_e}{200}\frac{\varepsilon}{D}. \tag{14}$$

Mannning's Equation The Manning equation, although originally developed for open channel flow, has often been extended for use in pressure conduits. The equation is usually favored for rough textured material (rough concrete, unlike rock tunnels) and cross-sections that are not circular (rectangular, horseshoe). It is most commonly given in the form:

$$v = \frac{1}{N}R^{2/3}S^{1/2}, \tag{15}$$

in which N = roughness coefficient. Equation (15) can also be transformed to:

$$h_f = 19.6N^2\frac{L}{R^{4/3}}\frac{V^2}{2\mathbf{g}}, \tag{16}$$

$$S = \frac{Q^2N^2}{A^2R^{4/3}}. \tag{17}$$

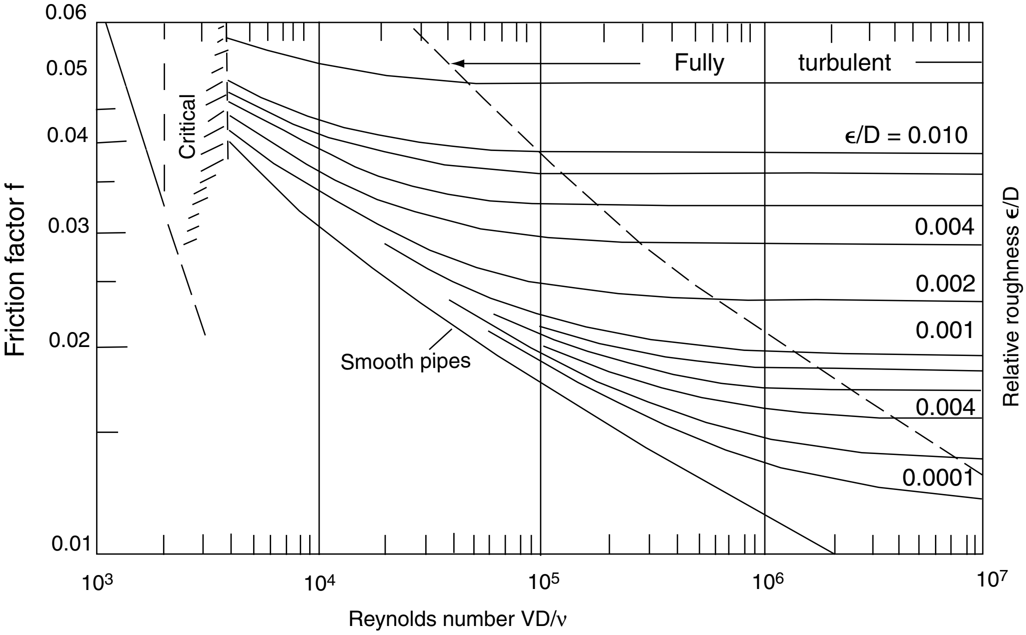


FIGURE 2 Pipe friction factors.

The dimension of the roughness coefficient, N, is frequently taken as $L^{1/6}$ as the equation by itself is not dimensionally homogenous. Table 3 provides values of Manning's N for the more widely used pipe materials.

Hazen-William's Equation This equation is used mainly in sanitary engineering.

$$V = 0.36CR^{0.63}S^{0.54}, \tag{18}$$

where C = roughness coefficient. Typical values of the Hazen-William's C is given in Table 4.

Energy Losses Due to Cross-Sectional Changes, Bends and Valves

Cross-Sectional Changes

Expansion Energy loss in an expansion is principally a form loss:

$$h_{\text{exp}} = \left[1 - \left(\frac{D_1}{D_2}\right)^2\right]^2 \frac{V_1^2}{2\mathbf{g}}, \tag{19}$$

in which D = diameter of the conduit, and subscripts 1 and 2 denote upstream and downstream values. From Eq. (19), if D_2 is very large compared to D_1, such as the discharge into a reservoir, the entire velocity head is lost.

TABLE 3
Normal values of Manning's N

Material	N
Brass	0.010
Corrugated metal	0.024
Glass	0.010
Concrete, unfinished	0.014
Vitrified clay	0.014
Steel	0.012
Cement	0.012
Brick	0.013

TABLE 4
Hazen-William's C

Material (new)	C
Cast iron	130
Welded steel	119
Riveted steel	110
Concrete	130
Wood-stave	120
Vitrified clay	110

Contraction The contraction loss equation can be expressed in terms of the downstream velocity, V_2, as:

$$H_{\text{con}} = \left(\frac{1}{C_c} - 1\right)^2 \frac{V^2}{2\mathbf{g}}. \tag{20}$$

Typical values of the coefficient of contraction, C_c, are given in Table 5.

Entrance Loss The head loss at the entrance of a conduit can be compared to that of a short tube:

$$h_{\text{ent}} = \left(\frac{1}{C^2} - 1\right)\frac{V^2}{2\mathbf{g}} \tag{21}$$

$$h_{\text{ent}} = K_{\text{end}} \frac{V^2}{2\mathbf{g}}, \tag{22}$$

in which C = coefficient of discharge, K = entrance loss coefficient. Typical values of C and K are given in Table 6.

Transition In gradual contractions and expansions, the lead losses are calculated in terms of the difference of velocity heads in the upstream and downstream pipes:

$$\text{Gradual contraction: } h_{tc} = K_{te}\left(\frac{V_2^2}{2\mathbf{g}} - \frac{V_1^2}{2\mathbf{g}}\right), \tag{23}$$

$$\text{Gradual expansion: } h_{tc} = K_{te}\left(\frac{V^2}{2\mathbf{g}} - \frac{V^2}{2\mathbf{g}}\right). \tag{24}$$

K_{tc} values vary from 0.1 to 0.5 for gradual to sudden contractions. Values of K_{tc} range from 0.03 to 0.80 for flare angles of 2° to 60°.

TABLE 5
Coefficient of contraction

A_2/A_1	C_c
0.25	0.64
0.50	0.68
0.75	0.78
1.00	1.00

TABLE 6
Values of C and K_{ent}

Type of entrance	C	K_{ent}
Circular bellmouth	0.98	0.05
Square bellmouth	0.93	0.16
Fully rounded	0.95	0.10
Moderately rounded	0.89	0.25
Sharp cornered	0.82	0.50

Bends The effect of the presence of bends is to induce secondary flow currents which are responsible for the additional energy dissipation:

$$h_b = K_b \frac{V^2}{2\mathbf{g}}. \tag{25}$$

The bend loss coefficient, K_b, depends on the ratio of the bend radius, *r,* to the pipe diameter, *d,* as well as the bend angel. For a 90° bend and *r*/*d* ratio varying from 1 to 12, values of K_b range from 0.20 to 0.07.

Gates and Gate Valves The gate and gate valve loss can be expressed as:

$$h_g = K_g \frac{V^2}{2\mathbf{g}} \tag{26}$$

The value of the loss coefficient, K_g, for gates depends on a variety of factors. The value of K_g for the case having the bottom and sides of the jet suppressed ranges from 0.5 to 1.0. for typical values of K_g for gate valves see Table 7.

TABLE 7

K_g for gate values

Fully open	0.2
$\frac{3}{4}$ open	1.3
$\frac{1}{2}$ open	5.5
$\frac{1}{4}$ open	24.0

Exit Loss In general the entire velocity head is lost at exit and the exit loss coefficient, K_e is unity in the equation:

$$h_e = K_e \frac{V^2}{2\mathbf{g}}. \tag{27}$$

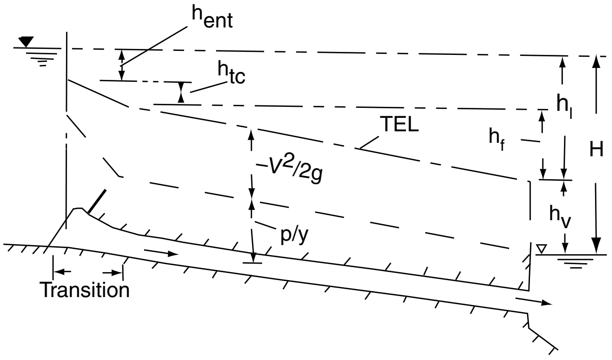


FIGURE 3 Energy relations.

Energy-discharge Relation

In pressure conduit flow, the water is transmitted through a closed boundary conveying structure without a free surface. Figure 3 illustrates graphically the various forms of energy losses which could take place within the conduit. The following energy relation can be written:

$$h_l = h_{\text{ent}} + h_{\text{tc}} + h_f \tag{28}$$

in which h_{end} = entrance loss, h_{tc} = transition loss, h_f = skin friction loss. If *H* denotes the total head required to produce the discharge and h_v represents the existing velocity head,

$$H = h_1 + h_v. \tag{29}$$

Writing Eq. (29) in terms of the velocity heads and their respective loss coefficients,

$$\begin{aligned} H = C_l \frac{V_2^2}{2\mathbf{g}} = \Bigg[& K_{\text{ent}} \frac{V_1^2}{2\mathbf{g}} + K_{\text{tc}} \left(\frac{V_2^2}{2\mathbf{g}} - \frac{V_1^2}{2\mathbf{g}} \right) \\ & + f \frac{LV_2^2}{2\mathbf{g}D_2} + K_v \frac{V_2^2}{2\mathbf{g}} \Bigg], \end{aligned} \tag{30}$$

where K_v = combined velocity head and exit loss coefficient. By the continuity equation:

$$A_1V_1 = A_2V_2 \tag{31}$$

and

$$\frac{V_1^2}{2\mathbf{g}} = \frac{A_2^2}{A_1^2} \frac{V^2}{2\mathbf{g}}.$$

Equation (30) could be expressed as,

$$\begin{aligned} H &= C_l \frac{V_2^2}{2\mathbf{g}} \\ &= \frac{V_2^2}{2\mathbf{g}} \left[K_{\text{ent}} \left(\frac{A_2}{A_1} \right)^2 + K_{\text{tc}} \left(1 - \frac{A_2^2}{A_1^2} \right) + \frac{fL}{2gD_2} + K_v \right] \end{aligned} \tag{32}$$

in which

$$C_l = \left[K_{\text{ent}} \left(\frac{A_2}{A_1} \right)^2 + K_{tc} \left(1 - \frac{A_2^2}{A_1^2} \right) + \frac{fL}{2\mathbf{g}D_2} + K_v \right] \tag{33}$$

$$V_2 = \frac{1}{C_l^{1/2}} \sqrt{2\mathbf{g}H} \tag{34}$$

$$\begin{aligned} Q = A_2V_2 &= \frac{A_2}{C_1^{1/2}} \sqrt{2gH} \\ &= CA_2\sqrt{2\mathbf{g}H}, \end{aligned} \tag{35}$$

in which $C = 1/C_l^{1/2}$ is the discharge coefficient. Equation (35) can be readily extended to multiple conduits in parallel.

Pipe Networks

Introduction The Hardy Cross method is most suitably adapted to the resolution of pipe networks. The statement of the problem resolves itself into:

1) the method of balancing heads is directly applicable if the discharges at inlets and outlets are known,
2) the method of balancing flows is very suitable if the heads at inlets and outlets are known.

It is assumed that:

a) sizes, lengths and roughness of pipes in the system are given,
b) law governing friction loss and flow for each pipe is known,
c) equations for losses in junctions, bends, and other minor losses are known. These relations are most conveniently expressed in terms of equivalent lengths of pipes.

The objectives of the analysis are:

a) to determine the flow distribution in the individual pipes of the network,
b) to compute the pressure elevation heads at the junctions.

In applying the Hardy Cross Method, two sets of conditions have to be satisfied:

a) the total change in pressure head along any closed circuit is zero:

$$\sum H = 0, \tag{36}$$

b) the total discharge arriving at any nodal point equals the total flow leaving it:

$$\sum Q = 0. \tag{37}$$

For the pressure head change in any closed path, the clockwise positive sign convention is used.

For the discharge continuity requirement at a nodal point, the inward flow positive sign convention is adopted.

The friction head loss equation is used in the form:

$$H = rQ^2. \tag{38}$$

Using Darcy's formula:

$$H = \frac{fLV^2}{2\mathbf{g}D} = \left(\frac{8fL}{\mathbf{g}\pi^2 D^5}\right)Q_2$$

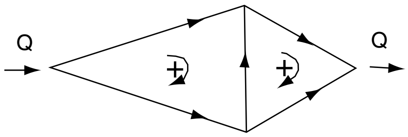


FIGURE 4 Pipe network.

and

$$r = \frac{8fL}{\mathbf{g}\pi^2 D^5}.$$

Method of Balancing Heads Based on the condition required by Eq. (36), the following equations for any closed pipe loop results (Figure 4):

$$\sum H = \sum r(Q_0 + \Delta Q)^2 = 0, \tag{39}$$

where Q_0 = assumed flow in the circuit for any one pipe, ΔQ = required flow correction. Expanding Eq. (39) and approximating by retaining only the first two terms, the flow correction ΔQ, can be expressed as:

$$\Delta Q = \frac{-\sum rQ_0^2}{\sum 2rQ_0}. \tag{40}$$

Method of Balancing Flows Utilising the continuity requirement at a pipe junction as given by Eq. (37), the head correction, ΔH, at anodal point is given by the equation:

$$\Delta H = \frac{\sum \left(\frac{H}{r}\right)^{1/2}}{\sum \frac{1}{2H}\left(\frac{H}{r}\right)^{1/2}}. \tag{41}$$

In both Eq. (40) and (41), the proper sign conventions must be used in the numerators.

OPEN CHANNEL FLOW

Introduction

Open channel flow refers to that class of water discharge in which the water flows with a free surface. The stream flow is said to be *steady* if the discharge does not vary with time. If the discharge is time dependent, the water flow is termed *unsteady. Uniform* flow refers to the case in which the mean velocities at any cross-section of the stream are identical; if these mean velocities vary from one cross-section to another, the flow is considered *non-uniform.* Steady uniform flow requires the conveyance section of the stream channel to be prismatic. Where the water surface profile is controlled principally by channel friction, this phenomenon is known as *gradually varied flow.* For the type of flow in which the water surface changes substantially within a very short channel length due to a sudden variation in bed slope or cross-section, this category is referred to as *rapidly varied flow.*

Open channels fabricated from concrete are often rectangular or trapezoidal in shape. Canals excavated in erodible material have trapezoidal cross-sections. Although sewer pipes are closed sections, they are still considered as open channels so long as they are not flowing full; these cross-sections are usually circular.

Channel Friction Equation

The most widely used open channel friction formula is the Manning equation as mentioned earlier in pressure flow:

$$Q = \frac{1}{N} A R^{2/3} S^{1/2} \tag{42}$$

$$S = \frac{Q^2 N^2}{A^2 R^{4/3}}. \tag{17}$$

Manning's equation in hydraulic engineering is used for fully turbulent flow and, as such, the values of Manning's N apply to this flow regime.

In a natural tortuous stream channel, the mean value of Manning's N can be obtained from the following considerations:

1) estimate an equivalent basic N_s, for a straight channel of that material,
2) select modifying values of N_m for non-uniform roughness, irregularity, variation in shape of cross-section, vegetation, and meandering,
3) sum the basic, N_s together with the modifying values to obtain the total mean N.

TABLE 8
Values of Manning's N

Basic N_S for straight channels	
Type of channel	N_s
Earth	0.010
Sand	0.012
Fine gravel	0.014
Rock	0.015
Coarse gravel	0.028
Cobbles and boulders	0.040
Modifying values of N_m	N_s
Irregularity	0.005 to 0.020
Changes in shape	0.005 to 0.020
Vegetation	0.005 to 0.100
Meander	0.10 N_s to 0.40 N_s

Normal values of Manning's N for straight channels and various modifying values are given in Table 8. The total mean N value for the channel is obtained from the relation:

$$N = N_s + \sum N_m. \tag{43}$$

Energy Principles

In deriving the energy relationships for open channel flow, the following assumptions are normally used:

1) a uniform velocity distribution over the cross-section is assumed, that is, the velocity coefficient, α, in the velocity head term, $\alpha V^2/2\mathbf{g}$, is taken as unity. In practice, the value of α depends on the shape of the stream channel and has an average value of about 1.02 which makes this assumption sufficiently valid.
2) streamlines are essentially parallel,
3) channel slopes are small.

Consider the water particle of mass, m, and of weight, W (Figure 5). The elevation and pressure energies of the particle are Wh_1; and Wh_2 respectively. Thus, the potential energy of the water particles is, $W(h_1 + h_2)$ and is independent of its elevation over the flow cross-section. As the kinetic energy is $WV^2/2\mathbf{g}$ the total energy of the water particles, e is:

$$e = W\left(h_1 + h_2 + \frac{V^2}{2\mathbf{g}}\right). \tag{44}$$

$$Z + D = h_1 + h_2 \tag{45}$$

and noting that the total flow passing the cross-section is γQ the total energy of the water passing the cross-section per second, E_t is given by:

$$E_t = \gamma Q\left(Z + D + \frac{V^2}{2\mathbf{g}}\right). \tag{46}$$

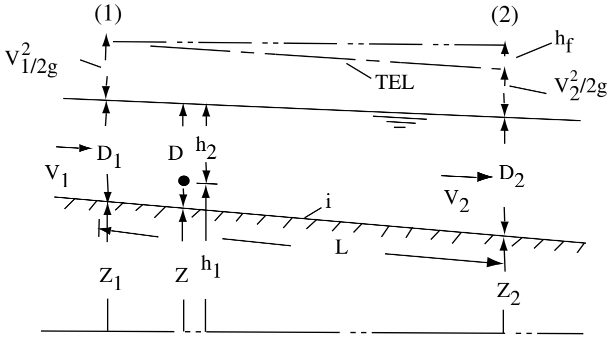


FIGURE 5 Energy principles.

Thus, the energy per unit weight of water passing the cross-section per second, *H* is:

$$H + Z + D + \frac{V^2}{2\mathbf{g}}. \tag{47}$$

The term, $z + D + V^2/2\mathbf{g}$ is known as the total head or total energy level (TEL); the latter name is used here. The slope of the total energy level line is the energy gradient or friction slope and gives the rate of energy dissipation in the flow.

The energies at Sections 1 and 2 are related by the expression:

$$z_1 + D_1 + \frac{V_1^2}{2\mathbf{g}} = Z_2 + D_2 + \frac{V_2^2}{2\mathbf{g}} + \frac{Q^2N^2}{A^2R^4} \tag{48}$$

in which the Manning equation is used to calculate the friction slope and the mean values for the flow area, *A,* and the hydraulic radius, *R,* are to be used.

Flow Regimes

Critical Flow The specific energy, *E,* is defined as the total head referred to the channel bottom (Figure 6):

$$E = D + \frac{Q^1}{2\mathbf{g}A^2}. \tag{49}$$

Differentiating Eq. (47) with respect to *D* and equating the derivative to zero to obtain its minimum value,

$$\frac{dE}{dD} = 1 - \frac{Q^2}{\mathbf{g}A^3}\frac{dA}{dD} = 0. \tag{50}$$

Noting that $dA = TdD$, Eq. (50) becomes:

$$\frac{Q^2}{\mathbf{g}} = \frac{A^3}{T}. \tag{51}$$

Equation (51) is the fundamental equation for critical flow and is applicable to all shapes of cross-sections.

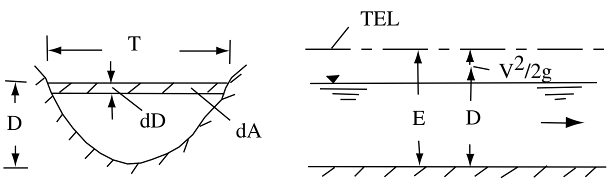


FIGURE 6 Derivation of critical flow.

If the mean depth of the flow section is defined as $D_m = A/T$, substitution of this relation into Eq. (49) would give the significant expressions:

$$\frac{V_c^2}{2\mathbf{g}} = \frac{D_m}{2} \tag{52}$$

and

$$\frac{V_c}{\sqrt{\mathbf{g}D_m}} = 1. \tag{53}$$

At critical flow, Eq. (52) demonstrates that the velocity head equals one-half the mean depth and Eq. (53) indicates that the Froude number equals unity.

Specific Energy Diagram for Rectangular Channel For a rectangular channel, $Q = qB$ in which q = discharge per unit width, B = channel width, and Eq. (49) becomes,

$$E = D + \frac{q^2}{2\mathbf{g}D^2}. \tag{54}$$

A plot of Eq. (54) for any given constant unit discharge gives Figure 7, which is known as the *specific energy diagram.* The

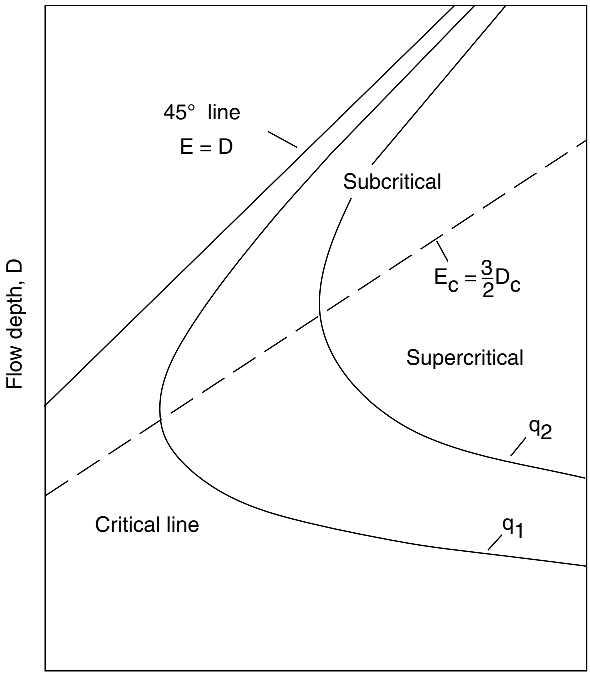


FIGURE 7 Specific energy diagram.

following flow regimes could be defined with reference to the specific energy diagram.

Subcritical flow denotes tranquil flow in which the Froude number and the mean velocity are less than unity and the celerity of the gravity wave respectively.

Critical flow represents the discharge phenomenon where: (1) for a constant specific energy, the discharge is a maximum; (2) the specific energy is a minimum for a constant discharge; (3) the critical velocity equals the celerity of a small gravity wave; (4) the Froude number equals unity; (5) the critical depth is also the depth of minimum pressure-momentum force.

Supercritical flow which is also known as shooting or rapid flow, is that state of water flow where the Froude number and mean velocity exceed unity and the speed of transmission of a surface wave respectively.

Based on the equations developed earlier for critical flow, particular formulae can be derived for a rectangular section:

$$D_c = \left(\frac{q_c}{\mathbf{g}}\right)^{1/3} \tag{55}$$

$$D_c = \frac{2}{3}(E_c) \tag{56}$$

$$D_c = \frac{V_c^2}{\mathbf{g}} \tag{57}$$

$$\frac{V_c}{\sqrt{\mathbf{g}D_c}} = 1. \tag{58}$$

Flow Transition The concept of the *normal depth* is an important parameter in the study of flow transition. For a given channel and any fixed discharge, uniform flow will occur at one unique depth. It is the depth attained in a long channel when the component of gravity force is just balanced by the frictional resistance of the channel.

When the normal and critical depth are equal, the flow is critical and the bed and energy slopes are the same. The channel bed then has a *critical slope.* The bed slope is termed *mild* when the normal depth exceeds the critical depth; the bed slope is then less than the critical energy slope and the flow regime is subcritical. When the normal depth lies below the critical depth and, hence, the bed slope is greater than the critical energy slope, the channel slope is considered to be *steep* and the supercritical flow regime prevails.

When water makes a transition from a channel with a mild slope to another with a steep slope, or vice versa, the flow passes through the critical depth close to the junction of the two channels. The section in which the water depth is critical defines a channel control. The weir acts as a control when water flows over it as critical depth is attained there.

Hydraulic Jump A hydraulic jump occurs when supercritical flow makes a transition to subcritical flow. A common occurrence of a hydraulic jump takes place at the base of a chute spillway. Figure 8 shows the energy momentum and depth relations for a hydraulic jump and also defines the symbols to be used.

In developing the equations for the hydraulic jump, the following assumptions are used:

1) the bed slope is considered small and neglected,
2) frictional resistance along the bed and sides of the channel are omitted.

Consider the control volume between Sections (1) and (2). Applying the impulse-momentum principle:

$$F_1 - F_2 = \frac{\gamma}{\mathbf{g}}Q(V_2 - V_1) \tag{59}$$

or

$$F_1 + \frac{\gamma}{\mathbf{g}}QV_1 = F_2 + \frac{\gamma}{\mathbf{g}}QV_2 \tag{60}$$

in which F_1 and F_1 denote the hydrostatic forces at sections 1 and 2 respectively. The term $(F + \gamma/\mathbf{g}QV)$ is given the name pressure-momentum force. Let A = flow area, $\bar{y}$ = distance of centroid of flow area from surface; then they hydrostatic force = $\gamma A\bar{y}$ and Eq. (60) can be written as:

$$A_1\bar{y}_1 + \frac{Q^2}{\mathbf{g}A_1} = A_2\bar{y}_1 + \frac{Q^2}{\mathbf{g}A_2}. \tag{61}$$

Equation (60) states the condition for the formation of a hydraulic jump and suggests a graphical solution. A plot of Eq. (60) for any fixed discharge is shown in Figure 8. For any given up-stream supercritical water depth, which is usually known such as at the toe of a spillway, the subcritical hydraulic jump depth or *sequent* depth can be obtained from the graph.

For a rectangular cross-section channel and utilising the continuity relation:

$$Q = V_1A_1 = V_2A_2. \tag{62}$$

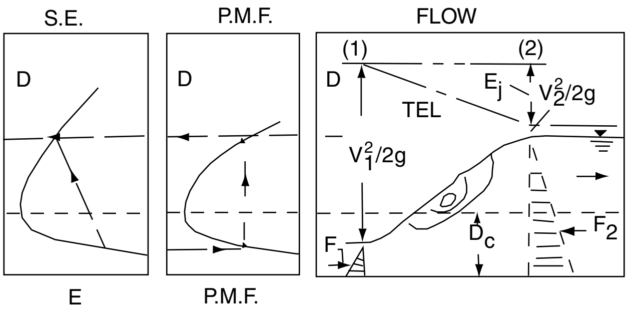


FIGURE 8 Hydraulic jump relations.

Equation (6) can be written as:

$$\frac{D_2}{D_1} = \frac{1}{2}(\sqrt{1+8F_1^2}-1), \tag{63}$$

where $F_1 = V_1/\sqrt{(\mathbf{g}D_1)}$ is the upstream Froude number.

The energy loss E_j across the hydraulic jump on a horizontal floor can be obtained by coming Eqs. (61), (62) and the energy equation:

$$D_1 + \frac{V_1^2}{2\mathbf{g}} = D_2 + \frac{V_2^2}{2\mathbf{g}} + E_j \tag{64}$$

to give:

$$E_j = \frac{(D_2 - D_1)^3}{4D_1D_2}. \tag{65}$$

The head loss E_j is graphically shown in the specific energy and flow diagrams (Figure 8).

Surface Water Profiles

Non-uniform Differential Equation Using the notation given in Figure 9, the energy relations can be expressed as:

$$iL + D + \frac{V^2}{2\mathbf{g}} = S\mathrm{d}L + (D + \mathrm{d}D) + \left(\frac{V^2}{2\mathbf{g}} + \mathrm{d}\frac{V^2}{2\mathbf{g}}\right) \tag{66}$$

$$\mathrm{d}L = \frac{\mathrm{d}(D + V^2/2\mathbf{g})}{(i - S)} \tag{67}$$

$$\frac{\mathrm{d}E}{\mathrm{d}L} = (i - S). \tag{68}$$

For a finite length, ΔL, Eq. (67) becomes:

$$\begin{aligned}\Delta L &= (L_2 - L_1)\\ &= \left(\frac{(D_1 + V_1^2/2\mathbf{g}) - (D_2 + V_2^2/2\mathbf{g})}{i - Q^2N^2/A^2R^{4/3}}\right),\end{aligned} \tag{69}$$

where the Manning equation is used to calculate the energy slope. Flow computation must start at a control section where all the flow parameters are known. The calculation proceeds upstream for subcritical and downstream for supercritical flow. In Eq. (69), the solution of the reach length, ΔL, is direct if the immediately upstream depth, D_2, is given a value. If ΔL is given a value, D_2 has to be solved by trial. This method of computing surface water profiles is suitable for regular channels.

Classification of Flow Profiles Twelve distinct types of non-uniform profiles have been systematically classified (Figure 10).

1) Firstly, the curves are identified according to bed slopes as mild (M), steep (S), horizontal (H), critical (C) and adverse (A).
2) Secondly, numbers are assigned to flow regions. The numerical 1 refers to actual flow depths exceeding both critical (D_c) *and normal* (D) depths. For flow depths less than both critical and normal, the number 3 is affixed to it. The numeral 2 is for depths intermediate between critical and normal.

Water Profiles in Irregular Channels The river channel has to be divided into panels (Figure 11) with the side panels conveying overbank flow. Let Q = total flow, Q_c = central channel discharge, Q_l = left overbank flow,

Q_r = right overbank flow. The continuity condition requires that:

$$Q = Q_c + Q_l + Q_r, \tag{70}$$

By Manning's Equation:

$$Q = \frac{1}{N_c}A_cR_c^{2/3}S^{1/2} + \frac{1}{N_l}A_lR^{2/3}S^{1/2} + \frac{1}{N_r}A_rR_r^{2/3}S^{1/2}. \tag{71}$$

The energy slope, S, has been taken as the same for Q_c, Q_l, Q_r; this assumption seemed to be justified in practice. Due to different channel roughness, vegetative and other obstructions, Manning's N for the three flow panels would

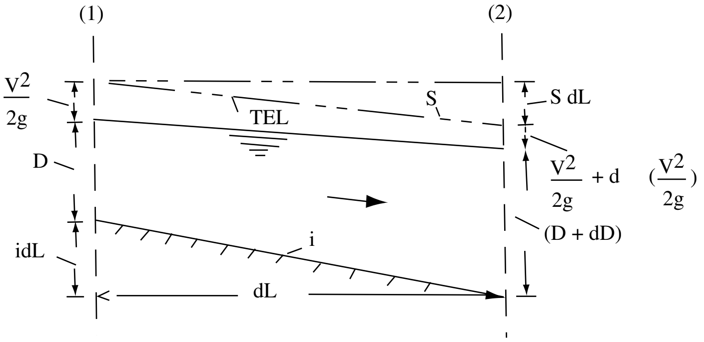


FIGURE 9 Non-uniform flow derivation.

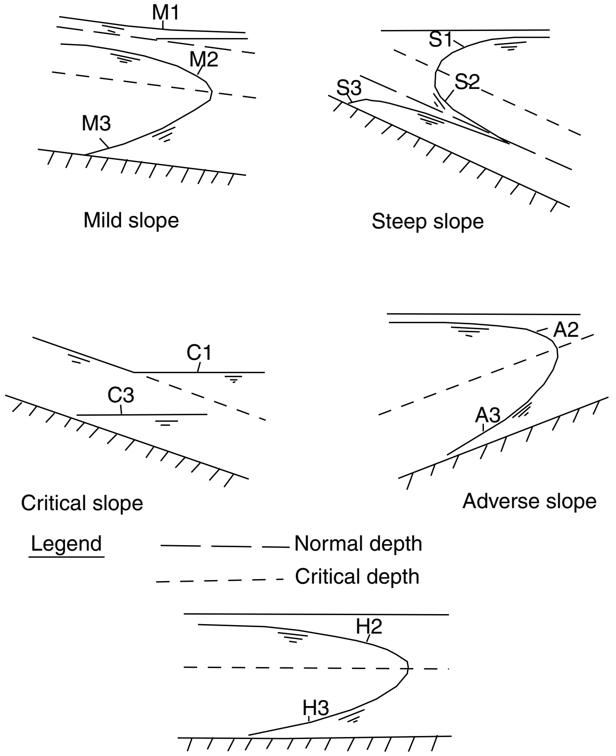


FIGURE 10 Surface water profiles.

not have the same values. As the energy gradient, S, is common to the panels in Eq. (71), water level versus energy slope curves can be plotted for any selected discharge for water level computations. The energy equation for the step method of surface profile calculation can conveniently be written in the form:

$$WL_2 + \frac{1}{2\mathbf{g}}\frac{Q^2}{A_2^2} = WL_1 + \frac{1}{2\mathbf{g}}\frac{Q^2}{A_1^2} + \bar{S}L_{1,2}. \tag{72}$$

If in practice, the change in kinetic head, $(Q^2/2\mathbf{g}A_2^2 - Q^2/2\mathbf{g}\,A_1^2)$ is small and could be removed from Eq. (72), this would greatly simplify the work. In determining the wetted perimeter for the calculation of the hydraulic radius, only the water-channel contact lines are relevant and the water-water contact lines between panels are excluded.

FLOW IN ERODIBLE CHANNELS

Introduction

Flow in erodible channels can be divided into two types, namely, canal and river flows. Blench describes canal flow as possessing there degrees of freedom due to its ability to adjust itself with respect to its flow depth, bed slope, and side widths which are taken to be the dependent variables. River flow, in addition to having the three degrees of freedom of canals, has a fourth degree by virtue of its ability to meander.

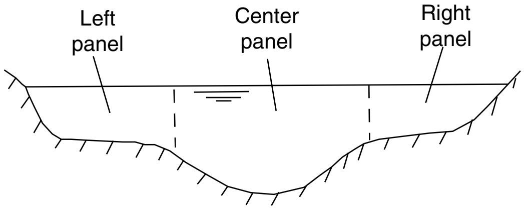


FIGURE 11 River channel division.

It is assumed that constant maintenance of canals suppresses the canal's tendency to meander. The water-sediment flow is usually regarded as the independent parameter.

In the concept of flow in mobile channels where the transport of sediment is an integral part of the system, two philosophies have emerged. Based on the work of Lindley (1919), Lacey (1952), Inglis (1949), and Blench (1953, 1966) in India and Pakistan, the regime theory has evolved. On the other hand, the United States Bureau of Reclamation under the direction of Lane (1952, 1953) developed the tractive force method.

Regime Theory for Canals

The regime theory postulates that for given water-sediment flow and bed material, there exists a regime channel which determines uniquely the flow area, cross-sectional shape and bed slope. The regime channel is considered a stable channel which on the average will neither silt nor scour. The flow occupying the regime channel is the dominant discharge and it is also variously referred to as the formative, regime, or bank-full discharge.

Lacey's Equations Based on extensive flow observations of the canals in India, Lacey (1952) proposed a set of formulae for alluvial channels with sandy mobile beds with the discharge ranging from 25 cfs to 2500 cfs, the bed material size varying from 0.2 mm to 0.6 mm and with the quantity of solids conveyed being less than 50 ppm:

Wetted perimeter: $$P = 2.67\,Q^{1/2} \tag{73}$$

Flow area: $$A = \frac{1.25Q^{5/6}}{f^{1/3}} \tag{74}$$

Bed slope: $$S = \frac{0.0054f^{5/3}}{Q^{1/6}} \tag{75}$$

Silt factor: $$f = 8d_{\text{inch}}^{1/2}. \tag{76}$$

From the above equations, the following two equations can be derived:

Mean velocity: $$V = 0.895_r^{1/6}Q^{1/6}$$

TABLE 9
Values of the bed factor

Material	$(F_s)_{\max}$	Remarks
Very sandy loam banks	0.1	Erosion if > $(F_s)_{\max}$
Silty clay loam	0.2	Erosion if > $(F_s)_{\max}$
Very cohesive banks	0.3	Erosion if > $(F_s)_{\max}$

Flow depth:
$$D = \frac{0.473Q^{1/3}}{f^{1/3}}. \tag{77}$$

Blench's Equations Blench (1953, 1966), using the concept that canals possessing three degrees of freedom must, therefore, have three basic equations to describe their motion, presented the stream equations for small bed load as:

$$V^2 D = F_b \tag{78}$$

$$\frac{V^2}{B} = F_s \tag{79}$$

$$\frac{V^2}{\mathbf{g}DS} = 3.63(VB/v)^{1/4}. \tag{80}$$

In Eq. 78, F_b is bed factor and the equation itself expresses the statement that channels with similar water-sediment flows tend towards the same Froude number in relation to a suitable depth. Equation (79) describes the scouring action on the hydraulically smooth sides and defines the side Factor, F_s. The dissipation of energy per unit mass of water per unit time in the channel is given by Eq. (80) in which S is the energy gradient.

For appreciable bed load, Eq. (80) becomes:

$$\frac{V^2}{\mathbf{g}DS} = 3.63\left(1+\frac{C}{233}\right)\left(\frac{VB}{v}\right)^{1/4}, \tag{81}$$

where C is the bed load charge in parts per hundred thousand by weight of fluid discharge.

The bed factor, F_b, for sand of subcritical flow is given by the empirical equations:

$$F_b = F_{b0}(1+0.12C) \tag{82}$$

$$F_{b0} = 1.9\sqrt{d_{mm}} \tag{83}$$

in which F_{b0} = zero bed factor and is the value of F_b when C tends to zero, d_{mm} = median bed material size by weight in millimeter. As a guide to the value of the side factor, F_s, the following table has been suggested by Blench (1966).

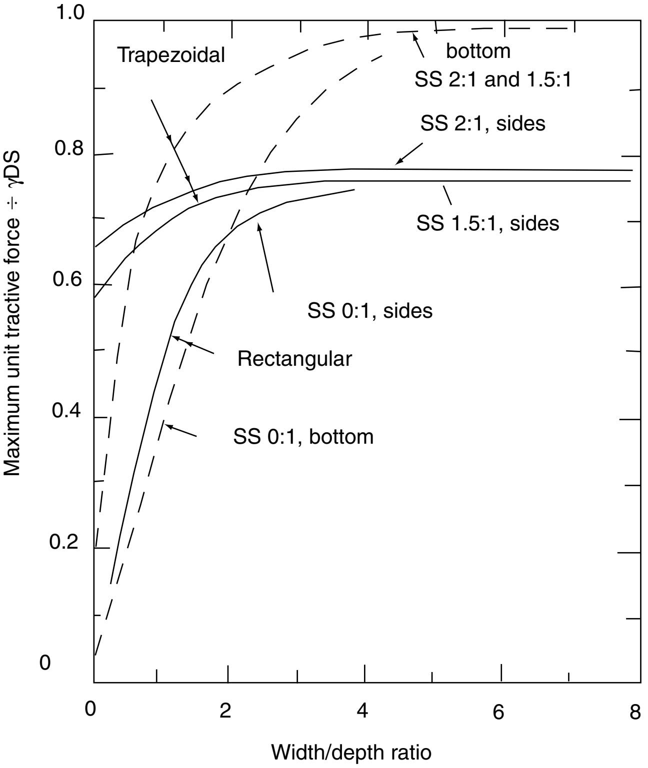


FIGURE 12 Maximum tractive force.

Tractive Force Method for Canals

Unit Tractive Force The stability of an erodible channel depends on (a) the resistance of the material lining the bottom and sides against the erosive force of the stream and (b) the ability of the stream to transport the sediment load without giving rise to significant deposition.

The shear or drag force exerted by the water on the bed and sides of the channel is termed the tractive force. The average unit tractive force, τ, in uniform flow is the component of the gravity force acting on the water parallel to the channel bottom per unit area, thus:

$$\tau = \gamma RS. \tag{84}$$

For wide channels, the flow depth can replace the hydraulic radius:

$$\tau = \gamma DS. \tag{85}$$

The distribution of tractive force has been investigated by the United States Bureau of Reclamation (Lane, 1952; 1953, Olsen and Florey, 1951, 1952). The maximum values of the unit tractive force for the bottom and sides of rectangular and trapezoidal cross-sections are given in Figure 12.

Tractive Force Ratio A soil particle of effective area, A_e, resting on the side of a channel is acted on by the tractive

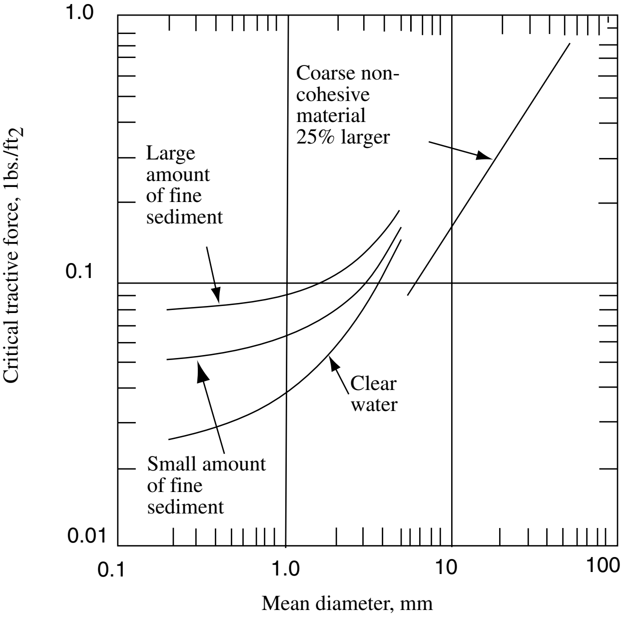


FIGURE 13 Critical tractive force for canals.

force, $A_e\tau_s$, in the direction of the flow and the gravity force component, $W_s \sin\varphi$ which attempts to cause the particle to roll down the side slope, where τ_s = unit tractive force on the side of the channel, W_s = submerged weight of the particle, φ = angle of the channel side. The resultant of these two forces, F, is:

$$F = (W_s^2 \sin^2\varphi + A_c^2\tau_s^2)^{1/2}. \qquad (86)$$

The motion of the article is resisted by its frictional force (R):

$$R = W_s \cos\varphi \tan\theta, \qquad (87)$$

where tan θ is the coefficient of friction and θ is the angle of repose of the material.

Equating Eq. (86) and (87) for the condition of impending motion and solving for τ_s:

$$\tau_s = \frac{W_r}{A_e}\cos\varphi\tan\theta\left(1-\frac{\tan^2\varphi}{\tan^2\theta}\right). \qquad (88)$$

A similar equation can be written for the case of a particle on a level bed when motion is impending, thus:

$$\tau_l = \frac{W_r}{A_e}\tan q, \qquad (89)$$

where τ_1 denotes the unit tractive force on the level bed.

The tractive force ratio, K, is defined as, τ_s/τ_l is obtained by dividing Eq. (87) to Eq. (88) and simplifying:

$$K = (1-\sin^2\varphi/\sin^2\theta)^{1/2}. \qquad (90)$$

From Eq. (89) it can be seen that the tractive force ratio, K is a function of the side slope and angle of repose of the material only.

Critical Tractive Force The permissible tractive force is the maximum unit tractive force that will not cause significant scour of the material lining the channel bed on a level surface. It is often found from laboratory observations and is known as the critical tractive force. It is influenced by the amount of organic matter and fine suspended sediment in the water. The effect of the fine sediment is to increase the allowable critical tractive force. Figure 13 shows curves of permissible tractive forces as recommended by the United States Bureau of Reclamation.

River Engineering

In river flows, a greater number and range of factors have to be considered in addition to those parameters used in the analysis of canals. These variables include bigger size bed materials, large suspended and bed sediment loads, unsteady and a wide variation of flood flows, meandering and braiding, large changes in stream channel cross-sections, obstructions to flow, and other factors involved. An analysis of river engineering is, therefore, beyond the scope of this chapter. Readers are recommended to consult the works of Blench (1966), Shen (1971, 1972), Inglis (1949) and Leopold, Wolman and Miller (1964). More specialized treatment of sediment transport, bedforms and stream geometry can be found in the publications of Einstein (1972), Leopold and Maddock (1953), Richardson and Simons (1967), Yalin (1971), Kennedy (1963), Christensen (1972), and Ackers (1964). Standard texts which cover the subject more formally include those of Graf (1971), Henderson (1966), Raudkivi (1967) and Leliavsky (1955).

FLOW WITH AN ICE COVER

A river flowing with an ice cover has, in addition to the bed and side frictional forces, the shear resistance imposed by a buoyant boundary represented by the floating ice cover. Chee and Haggag (1984) have developed equations concerning floating boundary stream flow which are reproduced here. The essential concepts and assumptions are first discussed.

A channel with a buoyant cover can be divided into two subsections as shown in Figure 14. The flow in subsection (1) is influenced by the bed and sides while subsection (2) is controlled by the cover. The two subsections are divided by a separation surface which represents the locus of no shear and maximum velocity. The equations of energy, continuity,

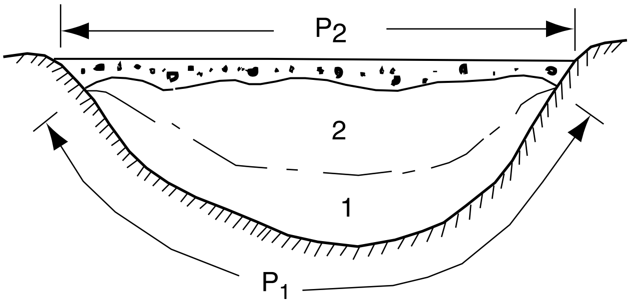


FIGURE 14 Ice covered channel.

and momentum are applicable to each subsection individually as well as to the entire channel cross-section.

Composite Roughness Equation

The Reynolds form of the Navier-Stokes equation was used to develop the shear distribution and the velocity profile was obtained using the Prandtl-von Karman mixing length theory. In addition, a channel momentum equation and the flow resistance formula of Manning were utilised to derive the relationship for the division surface separating the two flow subsections as

$$(0.66)\frac{R^{1/6}}{N_q \mathbf{g}^{1/2}} = \frac{(\lambda^{1/2}-1)}{1-N_1/N_2\lambda^{2/3}}\left[\alpha+(1-\alpha)\lambda\right]^{1/6} \qquad (91)$$

in which R = hydraulic radius of entire channel, N_1, N_2 = Manning's roughness for the bed and cover respectively, $\mathbf{g}$ = acceleration due to gravity, $\lambda = R_1/R_2$ is hydraulic radius ratio of the bed subsection to the cover subsection, $\alpha = P_1/P$ is the wetted perimeter ratio of the entire channel to the bed subsection. The division surface is found by solving for λ using Eq. (91).

The complete roughness, N, of an ice-covered channel is given by

$$\frac{N_1}{N} = \left[1+(1-\alpha)\lambda\right]^{-5/3}\left[\alpha+(1-\alpha)\frac{N_1}{N_2}\lambda^{5/3}\right]. \qquad (92)$$

REFERENCES

1. Ackers, P., Experiments on small streams in alluvium, proceedings, *American Society of Civil Engineers Journal,* Hydraulics Division, **90,** no. HY4, pp. 1–37, July 1964.
2. Albertson, M.L., J.R. Barton, and D.B. Simons, *Fluid Mechanics for Engineers,* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1961.
3. Allen, J. and S.P. Chee, The resistance to the flow of water round a smooth circular bend in an open channel proceedings, Institution of Civil Engineers, London, **23,** pp. 423–424, November 1962.
4. Blench, T., Regime theory equations applied to a tidal river estuary, proceedings, Minnesota International Hydraulics Convention I.A.H.R. IAHR and ASCE Joint Meeting, pp. 77–83, September 1953.
5. Blench, T., Mobile-bed fluviology, University of Alberta, 1966.
6. Chee, S.P. and M.R.I. Haggag, Flow resistance of ice-covered streams, Canadian Journal of Civil Engineering, Vol. II, No. 4, pp. 815–823, December 1984.
7. Chow, Ven Te, *Open-Channel Hydraulics,* McGraw-Hill Book Company, Inc., New York, 1959.
8. Christensen, B.A., Incipient motion on cohesionless banks, Chapter 4 in Shen, Hsieh Wen, *Sedimentation,* Colorado, 1972.
9. Einstein, H.A., The bed load function for sediment transportation in open channel flows, technical bulletin No. 1026, US Department of Agriculture, Soil Conservation Service, September 1950. Also reprinted in Appendix B in Shen, Hsieh Wen, *Sedimentation,* Colorado, 1972.
10. Graf, W.H., *Hydraulics of Sediment Transport,* McGraw-Hill Book Company, New York, 1971.
11. Henderson, F.M., *Open Channel Flow,* The Macmillan Company, New York, 1966.
12. Inglis, C.C., The behavior and control of river and canals, research publication No. 13, Central Waterpower Irrigation and Navigation Research Staton, Poona, 1949.
13. Kennedy, J.F., The mechanics of dunes and antidunes in erodible bed channels, *J. Fluid Mechanics,* **16,** part 4, 1963.
14. King, H.W., *Handbook of Hydraulics,* McGraw-Hill Book Company, Inc., New York, 1954.
15. Lacey, G., Flow in alluvial channels with sandy mobile beds, Proceedings, Institution of Civil Engineers, London Vol. 9, pp. 145–164, 1952.
16. Lane, E.W., Progress report on studies on the design of channels by the Bureau of Reclamation, Proceedings, American Society of Civil Engineers, Irrigation and Drainage Division, Vol. 79, pp. 280-1–280-31, September 1953.
17. Lane, E.W., Progress report on results on design of stable channels, US Bureau of Reclamation, Hydraulic Laboratory Report No. Hyd-352, June 1952.
18. Leliavsky, Serge, *An Introduction to Fluvial Hydraulics,* Constable and Co. Ltd., London, 1955.
19. Leopold, L.B. and T. Maddock, Jr., The hydraulic geometry of stream channels and some physiographic implications, US Geological Survey, Professional paper 252, 1953.
20. Leopold, L.B., M.G. Wolman, and J.P. Miller, *Fluvial Process in Geomorphology,* W.H. Freeman and San Francisco, 1964.
21. Lindley, E.S., Regime channels, Minutes or Proceedings, Punjab Engineering Congress, Lahore, India, Vol. 7, pp. 63–74, 1919.
22. Moody, L.F., Friction factors for pipe flow, *Trans. ASME.,* **66,** No. 8, 1944.
23. Morris, H.M., *Applied Hydraulics in Engineering,* The Ronald Press Company, New York, 1963.
24. Olsen, O.J. and Q.L. Florey (compilers), Stable Channel Profiles, Hydraulic Laboratory Report No. Hyd-325, US Bureau of Reclamation, 1951.
25. Olsen, O.J. and Q.L. Florey (compilers), Sedimentation studies in open channels: boundary shear and velocity distribution by membrane analogy, analytical and finite difference methods, US Bureau of Reclamation, Laboratory Report No. Sp-34, 1952.
26. Raudkivi, A.J., *Loose Boundary Hydraulics,* Pergamon Press, Oxford, 1967.
27. Richardson, E.V. and D.B. Simons, Resistance to flow in sand chanels, Twelfth Congress, IAHR, **1,** pp. 141–150, 1967.
28. Rouse, H., editor, *Engineering Hydraulics,* John Wiley and Sons, Inc., New York, 1950.
29. Shen, Hsieh Wen, Editor, *River Mechanics,* I and II, Colorado, 1971.
30. Shen, Hsieh Wen, Editor, *Sedimentation: Symposium to honour Professor H.A. Einstein,* Colorado, 1972.
31. Streeter, S.L., *Fluid Mechanics,* McGraw-Hill Book Company, New York, 1966.
32. United States Bureau of Reclamation, Design of small dams, US Printing Office, 1961.
33. Yalin, M.S., On the formation of dunes and meanders, paper C-13, proceedings Fourteenth Congress, IAHR, 3, 1971.

S.P. CHEE
University of Windsor

WATER: PROPERTIES, STRUCTURE, AND OCCURRENCE IN NATURE

THE CYCLE OF WATER

The Origin of the Hydrosphere

The assumption that the elemental composition originally present in all regions of the solar system was similar to that currently present in the sun furnishes a paradox in the marked deficiency of the unreactive noble gases, such as neon and argon, on the planet earth. Thus it is assumed that these gases, as well as those of comparable molecular weight such as water, nitrogen, oxygen, and carbon dioxide, escaped from the earth's surface before the planet cooled down. Thus the accumulation of the water necessary to form the hydrosphere by condensation of any primordial envelope of steam is inconsistent with the conditions of the primitive earth.

Rather, degassing, after the formation of the earth, is considered to be the mechanism through which the "excess volatiles" of water, nitrogen, chlorine, and so on accumulated on the earth's surface. According to Rubey, if only 1% of the water emerging from hot springs and related phenomena is truly juvenile water, then this process would have been sufficient to produce the oceans and hence the hydrosphere.

The Hydrologic Cycle and Solar Energy

The dynamic equilibrium or steady state existing between evaporation, precipitation, percolation, and runoff comprises the hydrologic cycle.

Except for the small amount of radioactive energy and heat transfer from within the earth, the sun provides the energy to run this complex cycle by evaporating surface water and heating the humid air, which then results in stirring of the weather systems. Four-fifths of the solar energy radiating toward the earth actually reaches the earth's surface. This incoming energy totals 1.6×10^{18} watts, about one-third of which recharges the hydrologic cycle by vaporization. Under normal conditions it requires about 540 calories per gram to evaporate water.

Evaporation

Of the 3.61×10^{20} grams per year evaporated from the sea, approximately 90% is returned directly to the ocean in the various forms of precipitation. The remaining 10% is transported in the atmosphere and eventually released on the land masses, as illustrated in Figure 1.

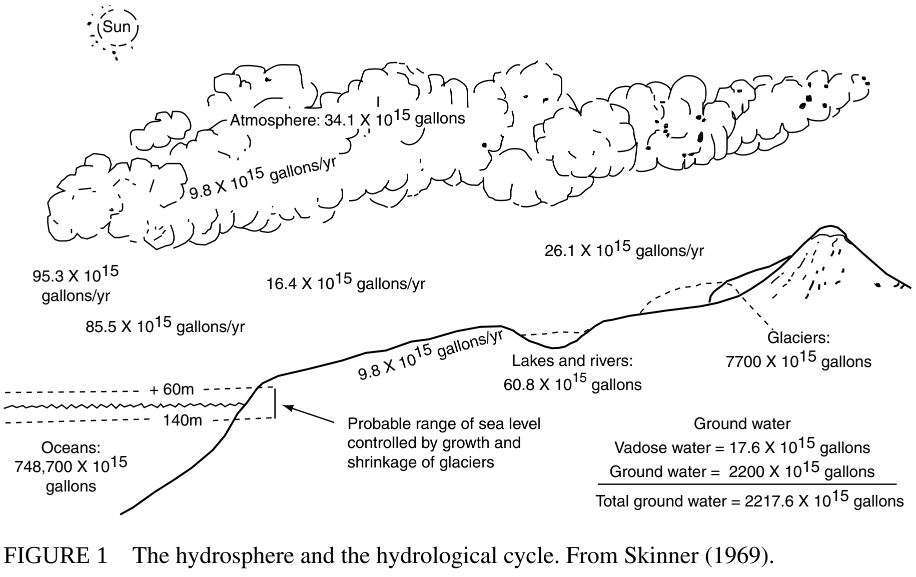


FIGURE 1 The hydrosphere and the hydrological cycle. From Skinner (1969).

Evaporation from the land, 6.21×10^{19} gms per year, is dependent on the soil and climatic conditions. It is strongly influenced by the relative humidity. For a given atmospheric water content, the relative humidity increases and the evaporation decreases as the temperature falls. Wind enhances evaporation by continually removing the vaporized water and thus locally lowering the relative humidity. With all other factors equal, sea water containing 3.5% salts will evaporate 3.5% more slowly than fresh water. Barometric pressure has little direct effect on the rates of evaporation.

Transpiration

Water loss from land is enhanced by the transpiration process by which plants and other photosynthetic organisms draw solutions containing essential growth nutrients from the soil to provide for maintenance and growth. The excess fluid is exuded to the atmosphere through the mesophyll cells. From 97 to 99% of the water which enters plants from soil is lost by evaporation from the leaves. The amount of dry matter produced in graphs per 1000 grams of water transpired is called the transpiration efficiency. Odum has estimated typical efficiencies to be 2 or less, with drought-resistant crops such as sorghum and millet having only slightly higher values. An acre of red maple trees in a soil with ample moisture may lose water sufficient to cover the acre with 0.7 meter of water during a single growing season. Transpiration is also affected by the factors governing evaporation, as well as light and soil conditions.

Atmospheric Water

The water in the atmosphere amounts to 1.29×10^{19} grams or approximately one-tenth of that in all fresh water lakes and is responsible for the pressure differences which create weather. It has been estimated that without this atmospheric water blanket, radiational cooling would reduce the earth's average surface temperature from 15°C to about 0°C. Low humidity desert environments often experience temperatures of 65°C during the day and below freezing at night. Under ordinary conditions the water vapor in the air allows only 20% of the earth's surface heat to radiate back through the atmosphere directly. As previously mentioned it is through the atmosphere that fresh water evaporated from the sea is transported over the global land masses.

Precipitation

Precipitation of atmospheric moisture occurs only when the temperature of an air mass is lowered to or below the point of saturation, the dew point. Air temperature is lowered by combinations of the following phenomena:

1) heat loss by infrared radiation,
2) adiabatic cooling as large air masses expand under reduced pressure after they have risen or been lifted to higher altitudes and
3) temperature decrease due to mixing with cooler air masses.

During the condensation of water the heat of vaporization, 540 cals. per gram is released. Various mechanisms of coalescence account for the different forms of precipitation such as rain, snow, hail and sleet. Condensation at ground level is responsible for the formation of dew, frost, and rime or hoarfrost.

The unequal worldwide distribution of precipitation, which varies from 0 to 10 meters per year, is determined to a large extent by the movement of large air masses or weather systems as directed by the profiles of mountain ranges.

The total precipitation, its annual distribution and the balance between precipitation and evapo-transpiration are the chief limiting factors that determine the resulting climax biotic communities for terrestrial ecosystems. The following average annual precipitations are general biotic determinants: 0 to 10 inches, deserts; 10 to 30 inches, grass lands, savanna or open woodland; 30 to 50 inches, dry forest; and greater than 50 inches, wet forest. Thus, the supply of water not only severely limits man's own agricultural, domestic and industrial activities but also dictates the products of nature. As ecologists have pointed out, the availability of water, more than any other factor, determines the ultimate population of a geographic province.

The influence on the earth is affected not only by the quantity of water which reaches the land masses from the sea but also by the natural processes which follow. Some precipitation is intercepted directly by the flora and soon evaporates. Some rainfall may never reach the ground. For example, in hot deserts rain on its downward flight through warm dry air often completely re-evaporates.

Runoff and Ground Water

Precipitation reaching the surface at rates greater than those of infiltration, evaporation and surface retention, will flow overland in drainage systems which may eventually reach the sea. This overflow is termed runoff. Although the 8.33×10^{21} grams of ground water which exists is a massive quantity, it moves too slowly to act as a significant transport process in the over-all hydrologic cycle. In fact, rapid withdrawal from the ground at a given site can cause serious local depletion.

Water from the hydrocycle which infiltrates into the ground water systems extracts soluble mineral matter from soil, sand and rock. According to King, "Water partakes of and is conditioned by the character of the land on which it falls, and through which it flows."

Hydrologic Weathering

These surface and subsurface waters weather natural geologic formations through physical, chemical and biological means. Physical weathering of formations is the initial mechanical weathering action which, through cycles of freezing and thawing, mechanical stress reversals, and swelling and contracting, exposes rock and mineral surfaces to chemical weathering processes.

Chemical weathering results from changes in the chemical environments surrounding the minerals; i.e., from environments conducive to their formation to those in which they are no longer thermodynamically stable. Specifically, these are chemical conditions which subject the minerals to hydrolysis, oxidation, and reaction with aqueous carbon dioxide. The minerals also react with weak organic acids, which can be considered biological weathering, since these acids result from the microbially mediated decay of biomass. Whereas surface runoff streams erode, transport and eventually deposit much of their suspended weathering products in river beds on the ocean, the dissolved species generally remain in solution.

Global Ice

Most of the earth's fresh water exists as ice in glaciers and in ice sheets, which together comprise about 2% of the earth's total water and which cover about 10% of the earth's land masses with thicknesses of up to several thousand feet. These solid forms may well be the most potent of the agencies that have changed and shaped the surface of the earth, according to King. Their mass causes the land to sink and their physical force performs erosion, transport, and deposition of mineral matter.

Ocean Currents

Water eventually returns to the sea with the wastes of nature such as dissolved minerals and suspended matter and, ever more significantly, also the wastes of man. To the sea belongs 97% of the planet's water. The sea is not a stagnant reservoir; instead it contains large "rivers" such as the Gulf Stream, which convey huge masses of water and perhaps more important great quantities of heat. For example, if the temperature difference between the warm current of the Gulf Stream and the cold receiving water is 20°F then each cubic mile of water can transfer about 2×20^{14} BTU.

Determinants of Physical–Chemical Properties

The properties of water and its interactions with other substances are in a class by themselves. The chief contributant to the anomalous physical characteristics of water and to many of its chemical properties is its powerful intermolecular forces, with which it attracts not only its own molecules, but also those of many other substances. The main intermolecular attractive force of water is hydrogen bonding which in turn results from the nature of the chemical bonds holding the atoms of the individual water molecule together.

Hydrogen Bonds

Hydrogen bonding is a special case of dipole-dipole interaction resulting from the molecular polarization when the hydrogen atom is covalently bonded to a highly electronegative atom. Such a condition produces a large dipole moment by displacing the electrons of the shared bond, which includes the sole electron of the hydrogen atom, toward the negative atom and gives a strong partial ionogenicity to the bond and a strong positive character to the hydrogen atom. This leaves an intense unshielded electric field on the other side of the hydrogen nucleus which can attract the unshared electrons of the negative atoms of other molecules.

STRUCTURAL AND PHYSICAL PROPERTIES

Electronic Structure of Water

A model of the single isolated water molecule with its electronic cloud is shown in Figure 2. The three nuclei are surrounded by ten electrons, with the two 1s electrons of the oxygen atom confined to the vicinity of its nucleus. The other eight electrons make up the four orbitals which point to the vertices of a somewhat distorted tetrahedron. Two of the four orbitals are directed along the OH bonds, and the so-called lone pair electrons constitute the remaining two orbitals which project above and below the H–O–H molecular plane. These lone pair electrons provide regions of negative electrification, which can attract the positively charged nuclei of hydrogen atoms in other water molecules thereby giving rise to hydrogen bonds. The charge distribution within the molecule results in the relatively large dipole moment of 1.84×10^{-18} esu.

Molecular Structure of Water

The planar H–O–H angle in the water molecule is 104°31, and the OH bond length is 0.9572 Angstrom. The symmetric group is C_2v: it has twofold axis of rotational symmetry, C_2, (the line bisecting the H–O–H angel) and a plane of reflection, v passing through this axis and normal to the plane of the molecule.

Precise calculations indicate that the average electron distributions forming the covalent OH bonds do not lie directly along the lines form the oxygen nucleus to the hydrogen nuclei, but are bent inward. Furthermore, the normal mode analysis of vibrational spectrum of the two OH bonds shows that the equilibrium length of one bond is dependent on the

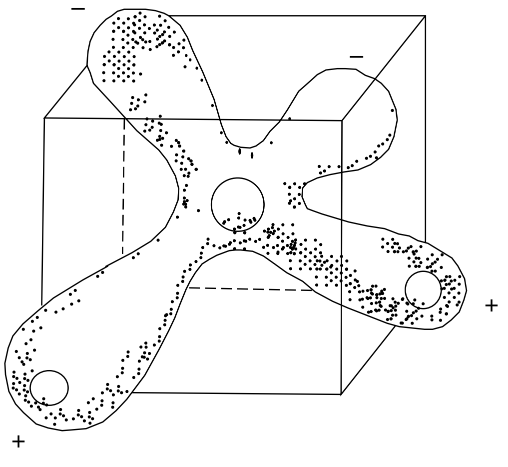


FIGURE 2 Single isolated water molecule with associated electronic cloud. Note the two regions of negative electrification which project above and below the H–O–H plane. From Horne (1970).

length of the other bond. The interdependence of the two OH bonds is also observed in their different dissociation energies. At 0°K, the dissociation energies. At 0°K, the dissociation energy of H–OH is 117.8 kilocalories per mole or 5.11eV whereas these respective values for H–O are 101.5 and 4.40. Thus the energy of formation of a water molecule from the individual atoms is −219.3 kilocalories per gram-mole.

The Vapor Phase of Water

In the vapor phase, water is mostly monomeric with the occurrence of an occasional dimer and, perhaps, a very rare trimer. At low pressure and low temperature the molecules do not interact to any appreciable extent.

The Solid Phases of Water

The structure of the solid phases of water is relatively well known compared to that of the liquid phase. Kamb (1972) recognizes 13 distinct ice phases by including among the distinctions certain "temperature dependent order-disorder" phenomena. The transition points and the densities of these phases are listed in Table 1.

Ordinary Ice

Ordinary ice, ice I, also known as I_h ($I_{hexagonal}$), is the only solid phase stable at normal low pressures. The presence in each water molecule of the two lone pairs of electrons on the oxygen atom and the two protons in approximately tetrahedral geometry create the opportunity for four intermolecular hydrogen bonds resulting in a coordination number of four for the ice structure. The arrangement of the oxygen atoms is isomorphous with tridymite, one of the crystalline forms of silica, and with wurtzite, a crystalline form of zinc sulfide. As shown by X-ray diffraction, ice I is a hexagonal lattice with each oxygen atom surrounded tetrahedrally by four other oxygen atoms at a distance of 2.76A. The oxygen atoms form what many authors call "puckered" hexagonal rings. Each ring layer is a mirror image of its neighbors. The hexagonal arrangement of ice I gives rise to an open structure thereby explaining the low density of ordinary ice, 0.9168 g/cm^3 at 0°C, relative to its melt, 0.99987 g/cm^3 at 0°C.

The OH bond length has been determined to be 0.99 Å by studies of stretching vibrations. As Figure 3 illustrates, a hydrogen atom occupies one of two possible positions on each O–O line, amounting therefore to four hydrogen atoms around each oxygen consistent with the tetrahedral spacing of the oxygen atoms. These allowable positions are 1.05Å from each molecule along the O–O axis. Early theoretical work by Pauling based on the residual entropy in ice at 0°C predicted that over a sufficiently long time interval, the hydrogen atom between the two oxygen atoms occupies each of these positions half of the time. This was later confirmed by neutron diffraction experiments.

The Polymorphs of Ice

The other crystalline forms of ice can be produced at various pressures and temperatures. The P–V–T surface illustrated in Figure 4 defines the regions of stability for the various polymorphs. Ice I*c* (I_{cubic}) has the same unambiguous half-hydrogen distribution as I_h, and thus the same configurational entropy. At about −90°C its cubic crystalline structure readily converts to the hexagonal form, ice I_h, at a rate dependent on its previous history.

Vitreous ice is so named since it is considered to be a glass. This amorphous structure is formed by the condensation of vapor below −150°C. Warming above this temperature transforms it irreversibly to I*c*, and then to I*h* upon further heating.

TABLE 1
The ice phases

			Limits of temperature range		
Phase	Density g cm^{-3}	Pressure kb	Lower °C	Upper °C	Transition at upper limit
I	0.92	0	−273°	0°	Melts
Ic.	0.92	0	−150°(?)	−90°	inv. to I
vit.	0.94	0	−273°(?)	−120°	inv. to Ic
II	1.21	3	−273°	−30°	trans. to III
III	1.15	3	−90°(?)	−20°	melts
IV	1.29	5	−40°(?)	−25°(?)	melts
V	1.28	5	−150°(?)	−10°(?)	melts
VI	1.38	14	−150°(?)	+50°	melts
VII	1.57	25	+2°	+110°	melts
VIII	1.63	25	−273°	+2°	trans. to VII
IX	1.19	3	−273°	−150°	trans. to III
X*	~1.4	—	−273°	−150°(?)	trans. to VI
XI*	~1.3	5	−273°	−150°(?)	trans. to V

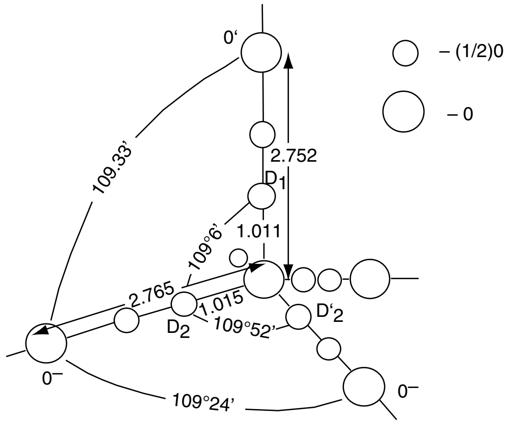


FIGURE 3 Structural relationships in one tetrahedron of the ice structure. The distances and angles are those at −50°C. From Krindel (1971).

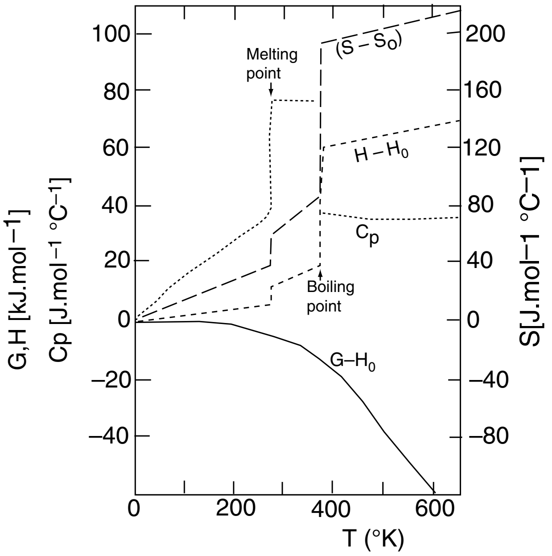


FIGURE 4 Enthalpy, free energy, and isopiestic heat capacity of H_2O at 1 atmosphere pressure, calculated from heat capacity measurements.

Ice II though VII are formed at high pressure but can subsequently exist at atmospheric pressure. Ice II is produced by a transition from ice III at about −100°C. The oxygen atoms of ice II, whose rhombohedral unit cell contains 12 water molecules, are apparently in the same tetrahedral positions as those in ice III, and the volume change of transition is very small but the resulting spatial ordering yields a much lower dielectric constant. Ices, II, III, V, and VI, all of which contain distorted hydrogen bonds, have internal energies only several tenths of a kilocalorie per gram-mole greater than ice I. In comparison, the internal energy of liquid water at 0°C is 1.88 kilocalories per gram-mole greater than that of ice I.

The crystal lattice of ice V is a monoclinic unit cell with 28 eater molecules. Ices IV, VII and VIII have extremely high densities of 1.38, 1.57, and 1.63 grams/cm^3, respectively. These high densities result from interpenetrating but not interconnecting structures. According to Fletcher "dense packing is achieved by placing water molecules upon two separate, completely hydrogen bonded four-coordinated lattices which are allowed to interpenetrate so that the molecules of one lattice occupy the cavities of the other." Thermodynamic criteria indicate that ice VI is probably metastale. Ice VII forms at 22 to 200 kilobars; only one additional allotrope, ice VIII, has been identified in this high pressure range.

Tables 2 and 3 from Kamb (1972) and Eisenberg (1969) summarize the properties of the ice polymorphs.

The Structure of Liquid Water

Liquid water is a highly structured liquid in which the tetrahedral condition observed in ice is still evident. The structure of liquid water still remains a greater enigma than highly complex molecules such as DNA and hemoglobin. Nevertheless, newly acquired data of various properties have increased the constraints operative in the hypothesis of models for the liquid structure.

$$H_T = H_0 = \int_0^T C_p dT + \Delta H_{pc} \qquad S_T - S_0 = \int_0^T \frac{C_p}{T} dT + \Delta S_{pc}$$

$$G_T - G_0 = H_0 - TS_T$$

The subscript "pe" indicates phase change. Thermodynamic constants for phase change are given in Table 4. (Figure after Eisenberg and Kauzmann, 1969; data from Dorsey 1941).

An acceptable model must explain anomalous properties of liquid water such as the following listed by Krundel and Eliezer (1971):

1) The density maximum at 4°C above the melting point.
2) A liquid specific heat approximately twice that of the solid.
3) The extraordinary large heat of fusion.
4) The increase in the coefficient of thermal expansion with increasing pressure (or, equivalently, the decrease in compressibility with temperature rise) between 0°C and 45°C compared with a decrease in this coefficient under same conditions for normal liquids.
5) The singular decrease in viscosity with increasing pressure (in the temperature range 20°C) to a minimum near 1000kg/cm^2, above which the normal increase is exhibited.
6) The exceptional dependence on pressure of the dielectric constant, the self-diffusion coefficient and other thermodynamic and transport properties.

TABLE 2

Structural characteristics of ice polymorphs[a]

Ice	I	Ic	II	III	V	VI	VII	VIII
Number of nearest neighbors	4	4	4	4	4	4	8[b]	8[b]
Distance of nearest neighbors (Å)	2.74	2.75[c]	2.75−2.84	2.76−2.80	2.76−2.87	2.81	2.86[d]	2.86[d]
Distance of closest non-H-bonded neighbour (Å)	4.49	4.50[c]	3.24	3.47	3.28, 3.46	3.51	2.86[d]	2.86[d]
O–O–O angles (deg)	109.5±0.2	109.5°	80°−128°	87°−141°	84°−135°	76°−128°	109.5°	109.5°
Hydrogen positions	Disordered	Disordered	Ordered	Disordered above −40°C	Disordered	Disordered	Disordered	Ordered

[a] Entries, except where noted, refer to −175°C and 1 atm pressure. Data for ices I and Ic are from Lonsdale (1958); data for the high-pressure polymorphs are from Kamb and Datta (1960), Kamb (1964), Kamb (1965 *a, b*), Kamb (1967), and Kamb *et al.* (1967).

[b] 4 are hydrogen-bonded to central molecule.

[c] At −130°C

[d] At 25 kbar. In quenched ice VII at atmospheric pressure the nearest-neighbor distance is 2.95 Å (Bertie *et al.,*1964)

The model must also be consistent with the interpretations from infrared, nuclear magnetic resonance, Raman, and X-ray spectra of liquid water.

Eisenberg has outlined the sequence by which most of the current theories have been developed. Experimental evidence and some intuition support the postulation of a model after which it is translated into mathematical terms, often by expressing as a simple partition function containing several variable parameters. The number of adjustable parameters has typically ranged from 9 to 18. The thermodynamic expressions derived from the partition function are fitted to the empirical measurements by varying the parameters. That the model furnishes a good correlation is not necessarily proof of its validity, due to the inductive nature of the method.

The different water models may be roughly divided into two classes: the continuum models (also called homogeneous or uniform models) and the mixture models (also called solution models). Models in both classes achieve partial success in explaining the properties of water.

The models of People and Frank—Wen, representative of the continuum and the mixture classes, respectively, are summarized here. Eisenberg and Kauzmann give a comprehensive review of the various models.

The People Distorted Bond Model

The continuum theories ascribe essentially complete hydrogen bonding to water, at least at low temperature. Associated with the hydrogen bonds is a distribution of angles, distance, and energies. The average bond energy varies with temperature due to the changes in the distributions of bond lengths and bond angle distortions.

People's model (1951) postulates that when ice melts, the hydrogen bonds become more flexible and disorder is introduced into the structure by the bending and stretching of the bonds rather than by breaking them. Thus the water molecules in the liquid phase, like those in the ice, retain a coordination number of four. All the molecules have essentially the same environments, thereby giving rise to the continuum. The model allows for a continuous spectrum of induced hydrogen bond disorder. This model is strongly supported by Raman and infrared spectral studies of liquid water. As Fletcher (1970) points out, "the uniform model of Pople stands or falls on the energy required to bend the hydrogen bond joining two molecules which are already participating in hydrogen bonds to other molecules." Pople and Lennard-Jones propounded that a significant percentage of broken hydrogen bonds in liquid water in thermodynamically impossible since the hydrogen bond energy between two water molecules is 4.5kcal/mole whereas the latent hat of fusion is 1.4kcal/mole and the average thermal energy, *RT,* 0°C is 0.5 kcal/mole.

The Frank–Wen Flickering Cluster Model

The flickering cluster model described water as a mixture of hydrogen bonded clusters surrounded by monomeric water. Frank and Wen (1957) treat the formation of hydrogen bonds in water as a cooperative phenomenon, with a sizeable covalent contribution to the energy of the bond. The existence of a hydrogen bond on one of the hydrogen atoms of water promotes the tendency of the other hydrogen to form a similar bond to another neighboring molecule. According to Krindel and Eliezer (1971), "Frank and Wen considered that hydrogen bonded molecules are mutually polarized, the resulting charge separation being such that the free ends of the associated molecules have a higher affinity for further hydrogen bonding than the ends of non-hydrogen bonded molecules." This susceptibility to further hydrogen bonding gives the process an element of positive feedback until the disaggregating thermal energies from the additional molecules supply negative compensation. Thus the formation of one

TABLE 3

Crystallographic properties of ice polymorphs[a]

Ice	I	Ic	II	III	V	VI	VII	VIII
Crystal system Space group[b] Unit cell	Hexagonal P6/mmc	Cubic F43 m	Rhomboherdral R3	Tetragonal $P4_12_12$	Monoclinic A2/a	Tetragonal P_2/nmc	Cubic Im3m	Cubic Im3m
dimensions (A)[c]	*a*4.48 *c*7.31	*a*6.35	*a*7.78 *a*113.1°	*a*6.73 *c*6.83	*a*9.22, *b*7.54 *c*10.35, β109.2°	*a*6.27 *c*5.79	—	*a*3.41
N. molecules/unit cell	4	8	12	12	28	10	2	2
Density at −175°C, 1 atm (g cm^{-3})	0.94	—	1.17	1.14	1.23	1.31	—	1.50
Density at (*T*°*C*, *P* kbar) in region of stability (gcm^{-3})	0.92 (0°, 1)	0.93 (−130°, 1)	1.18 (−35°, 2.1)	1.15 (−22°, 2.0)	1.26 (−5°, 5.3)	1.34 (15°, 8)	~1.65 (25°, 25)	−1.66 (−50°, 25)

[a] Data for ices I and Ic are from Lonsdale (1958); data for the high-pressure polymorphs are from Kamb (1965*a*, *b*), Kamb.

[b] The space groups for ice VI, VII, and VIII are not entirely certain.

[c] For 1 atm and −175°C, except for ice Ic which refers to −130°C.

TABLE 4
Thermodynamic constants for phase changes of H_2O (Eisenberg and Kauzmann, 1969)

	Fusion[a]	Vaporization[a]	Sublimation[b]
Temperature °K	273.15	273.15	273.16
ΔC_p (j mol^{-1}°C^{-1})	37.28	−41.93	—
ΔC_p (kJ mol^{-1})	6.01	40.66	51.06
ΔS (J mol^{-1}°C^{-1})	22.00	108.95	186.92
ΔV (cm^3 mol^{-1})	−1.621	3.01×10^4	—
ΔE (kJ mol^{-1})	6.01	37.61	48.79

[a] At 1 atm.

[b] At ice I-liquid-vapor triple point.

hydrogen bond tends to a cluster while the breaking of one hydrogen bond leads to the dissociation of a cluster, thereby resulting in clusters flickering into and out of existing with half-lives estimated at 10^{-10} to 10^{-11} second.

Though the model does not define the details of the organization of the water molecules, the clusters are thought to be compact and highly hydrogen bonded. The existence of a single dielectric relaxation time and a structural relaxation time approximating the neutron residence time appear to rule out the presence of Frank–Wen clusters in pure water according to Davis and Jarzynski.

Thermodynamics and Physical Properties

The isopiestic (1 bar) heat capacity of water, and the thermodynamic functions derived from the heat capacity, are given in Figure 4. The heat capacity of liquid water is almost twice that of ice at 0°C or steam at 100°C. Whereas the heat capacity of ice and steam is primarily due to vibrational energy, there is in liquid water an additional configurational heat capacity related to the energy involved in alterations in the structure of water. The thermodynamic constants for the phase changes of water are given in Table 4. The relatively low heat of fusion of water is evidence that the hydrogen bonding in the ice structure remains intact to some degree during fusion; the extremely high heat of vaporization reflects the disappearance of hydrogen bonding upon vaporization (Westall and Stumm, 1980).

P–V–T Relationships The density of water reaches a maximum at 4°C above the melting point (Figure 5). This extraordinary maximum may be explained as the result of two counteracting phenomena: as the temperature of liquid water is increased, (i) the open four-coordinated structure of water is further broken down, decreasing the volume; (ii) the amplitude of anharmonic intermolecular vibrations increases, increasing the volume. Below 4°C the first effect predominates.

The pressure-temperature curve for water is given in Figure 6. The triple point, at which ice, liquid water, and water vapor coexist at equilibrium, is a convenient reference state: T = 0.01°C, P = 6.11 mbar.

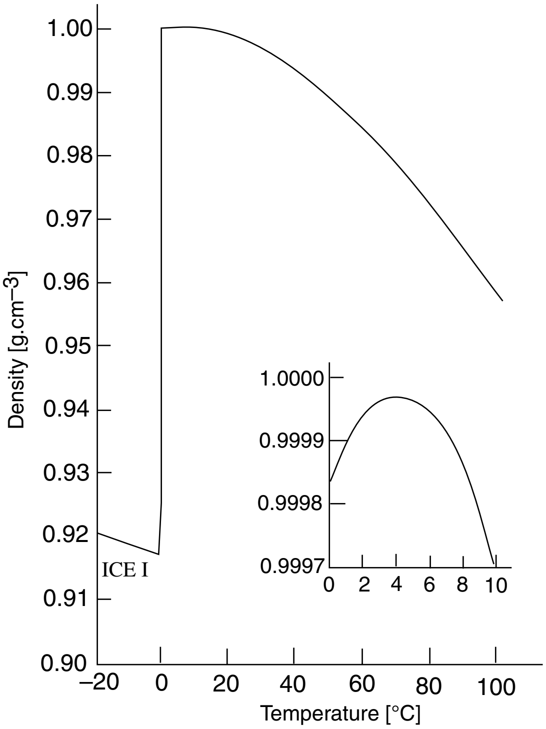


FIGURE 5 The density (at 1 atm) of ice and liquid water as a function of temperature. The inset shows the density in the domain of its maximum. (Data for ice from Eisenberg and Kauzmann, 1969; data for water from Kell, 1967).

Dielectric constant The dielectric properties of water result from (1) electronic and atomic polarization of molecules (formation of induced dipoles) and (2) the orientation of permanent dipoles. The high value of the dielectric constant of water is chiefly the result of the orientation of permanent dipoles, whereby not only the orientation of individual molecules, but also the mutual orientation of neighboring molecules, is important. For comparison, the dielectric constant of ice I is 91 at 0°C, and that of liquid water is 88 at the same temperature. As the temperature is increased, the degree of orientation of the molecules, and the dielectric constant, decreases (Figure 7).

The preceding discussion dealt with the dielectric properties of water in a static electric field. In a high frequently alternating electric field, the permanent dipoles can no longer reorient with the variable electric field; then the observed dielectric constant is due solely to the induced dipoles. For liquid water the value of this high frequently dielectric constant, ε, is between 4.5 and 6.0.

Viscosity The shear viscosity of water can be defined by the equation

$$\tau_{zx} = -\mu \frac{dv_x}{d_x}$$

where τ is the shear stress (force in the x-direction per unit area normal to z) to maintain the velocity gradient dv_x/dz in the z-direction, in the fluid of viscosity μ. The viscosity of water approximately doubles in the range between 25°C and 0°C (Figure 8).

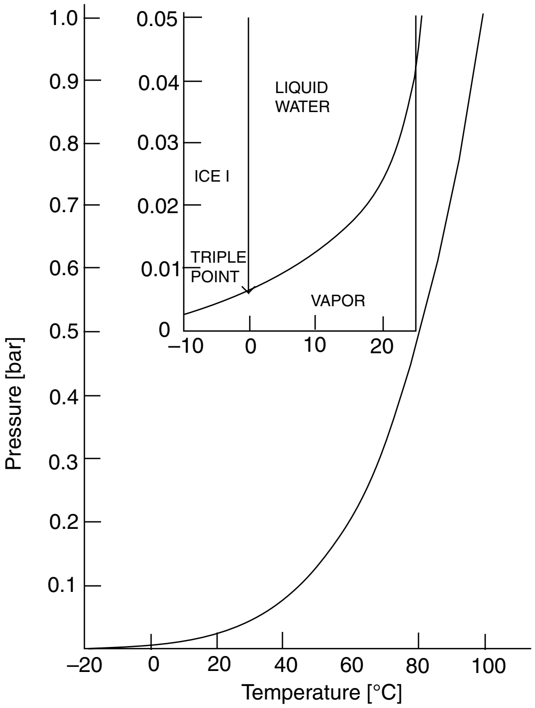


FIGURE 6 The vapor pressure of ice and liquid water as a function of temperature. The inset shows the vicinity of the triple point. (Data from Eisenberg and Kauzmann, 1969.)

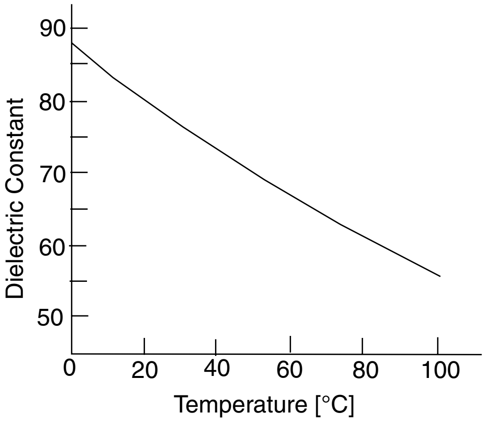


FIGURE 7 The dielectric constant of water as a function of temperature. (Data from Malmberg and Maryott, 1956.)

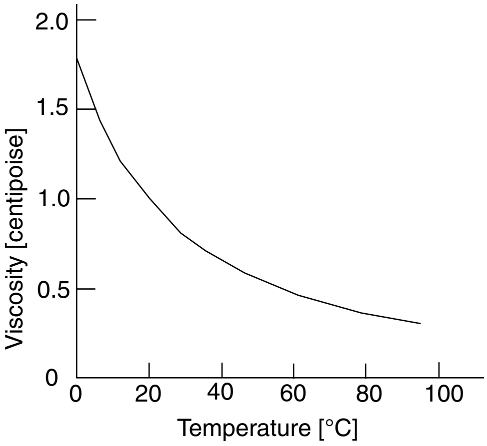


FIGURE 8 The shear viscosity of water as a function of temperature. (1 centipoise = 10^{-3} N s m^{-2}). (Data from Stokes and Millis, 1965.)

CHEMICAL PROPERTIES OF NATURAL WATERS

Natural waters acquire their chemical characteristics by dissolution of and by chemical reactions with solids, liquids, and gases with which they have come into contact during the various parts of the hydrologic cycle. Waters vary in their chemical composition, but these variations are at least partially understandable if the environmental history of the water and the chemical reactions of the rock-water-atmosphere systems are considered. Dissolved mineral matter is provided by the crustal materials of the earth; water disintegrates and dissolves mineral rocks by weathering. Gases and volatile substances participate in these processes. As a first approximation, seawater may be interpreted as the result of a gigantic acid-base titration: acids of volcanoes (sulfur dioxide, hydrochloric acid, hydrogen sulfide) reacting with bases of rocks (oxides, carbonates, silicates). The composition of fresh water similarly may be represented as resulting from the interaction of the carbon dioxide of the atmosphere with mineral rocks. The earth may be regarded as a heat engine which extracts energy from solar radiation in order to drive winds, ocean currents, hydrogeochemical cycles (cycles of water and rocks) and life cycles. Water, acting as a transport medium as well as a chemical reagent, gradually converts primary rocks into soil, sediments, and sedimentary rocks (Figure 9).

Weathering Processes

Dissolution reactions due to weathering take place because many constituents of the earth's crust are thermodynamically unstable in the presence of water and the atmosphere. Most important among the weathering reactions is the incongruent

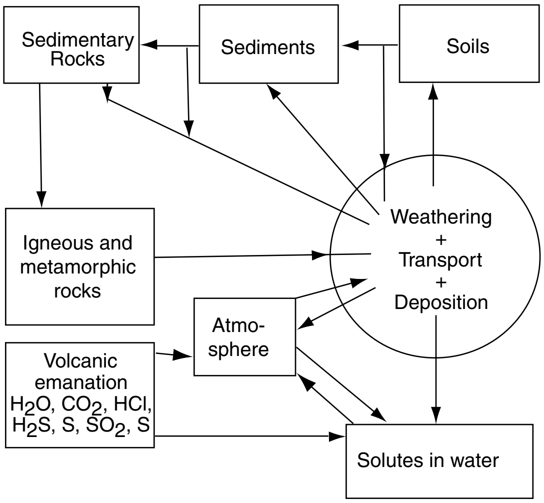


FIGURE 9 Interaction between the cycles of water and rocks.

(partial) dissolution of aluminium silicates which may be represented schematically as follows:

$$\mathrm{Me-Al-silicates(s)} + \mathrm{H_2CO_3}^{*} + \mathrm{H_2O} =$$
$$n\mathrm{HCO_3}^{-} + \mathrm{H_4SiO_4} + \mathrm{Me}^{n+} + \mathrm{Al-silicate(s)}, \text{(a)}$$

where Me is a cation such as Na^+ or Ca^{2+}. Table 5 shows examples of this type as well as weathering reactions of limestone (calcite).

In essence, primary minerals are converted into secondary minerals. The secondary minerals are frequently structurally ill-defined or X-ray amorphous. The structural breakdown of aluminium silicates is accompanied by a release of cations and usually of silicic acid. As a result of such reactions solution increases during the weathering process, the solid residue has a higher acidity than the original aluminium silicate.

Because $H_2CO_3^*$ is usually the acid in the attack of water on the primary silicates, HCO_3^- is the predominant anion in most fresh waters. Because of respiration by organisms, the CO_2 composition in soils can be a few hundred times larger than that in the atmosphere. Correspondingly ground waters tend to contain higher concentrations of HCO_3^- and other solutes. Minerals of the kaolinite group are the main alteration products of the weathering of feldspar. In addition to kaolinites, montmorillonites and micas are also found as intermediates or end products. Mica (illite) has been identified as an intermediate in the decomposition of potassium feldsparorthoclase.

The dissolution kinetics of feldspars may cause them to reach their final products through temporally stable intermediates. Virtually all bonds must be broken in the tetrahedric aluminium framework before the structure can be rearranged to a lattice having six coordination bonds around the aluminium as in kaolinite. Understandably, the rate of reaction is very slow. Physical weathering processes, that is, processes which by physical or mechanical means alter the size and hence the specific surface area of the minerals, enhance reaction rates since the weathering processes take place only at the rock-water interface.

TABLE 5
Examples of typical weathering reactions[a]

I. Congruent dissolution reactions

$SiO_2(s) + 2H_2O = H_4SiO_4$

quartz

$CaCO_3(s) + H_2CO_3^* = Ca^{2+} + 2HCO_3^-$

calcite

$Al_2O_3 \cdot 3H_2O(s) + 2H_2O = 2Al(OH)_4^- + 2H^+$

gibbsite

$Ca_5(PO_4)_3(OH)(s) + 4H_2CO_3^* = 5Ca^{2+} + 3HO_4^{2-} + 4HCO_3^- + H_2O$

apatite

II. Incongruent dissolution reactions

$NaAlSi_3O_8(s) + H_2CO_3^* + 9/2H_2O$

albite (Na-feldspar) =

$Na^+ + HCO_3^- + 2H_4SiO_4 + 1/2Al_2Si_2O_5(OH)_4(s)$

kaolinite

$7NaAlSi_3O_8(s) + 6H_2CO_3^* + 20H_2O$

albite (Na-feldspar) =

$6Na^+ + 6HCO_3^- + 10H_4SiO_4 + 3Na_{0.33}Al_{2.33}Si_{3.67}O_{10}(OH)_2(s)$

Na-montmorillonite

$Al_2Si_2O_5(OH)_4(s) + 5H_2O = 2H_4SiO_4 + Al_2O_3 \cdot 3H_2O$

kaolinite gibbsite

$CaMg(CO_3)_2(s) + Ca^{2+} = Mg^{2+} + 2CaCO_3(s)$

dolmite calcite

[a] After Stumm and Morgan (1981).

The composition of water is obviously influenced by these weathering reactions. While Ca^{2+}, HCO_3^-, H^+, and perhaps Mg^{2+} may be controlled by dissolution carbonate rocks, the source of Na^+, K^+, and H_4SiO_4, and possibly also of Ca^{2+} and Mg^{2+}, is the silicate minerals that make up 70% or more of the rocks in contact with underground waters and streams.

Figure 10 gives in an abbreviated periodic table the important elements typically found in fresh and sea water, their dominant forms of occurrence and their representative concentrations. A recent survey of the frequency distribution of various constituents of terrestrial waters by Davies and DeWiest (Figure 11) shows that many of the aquatic constituents exhibit little natural variation in their concentrations.

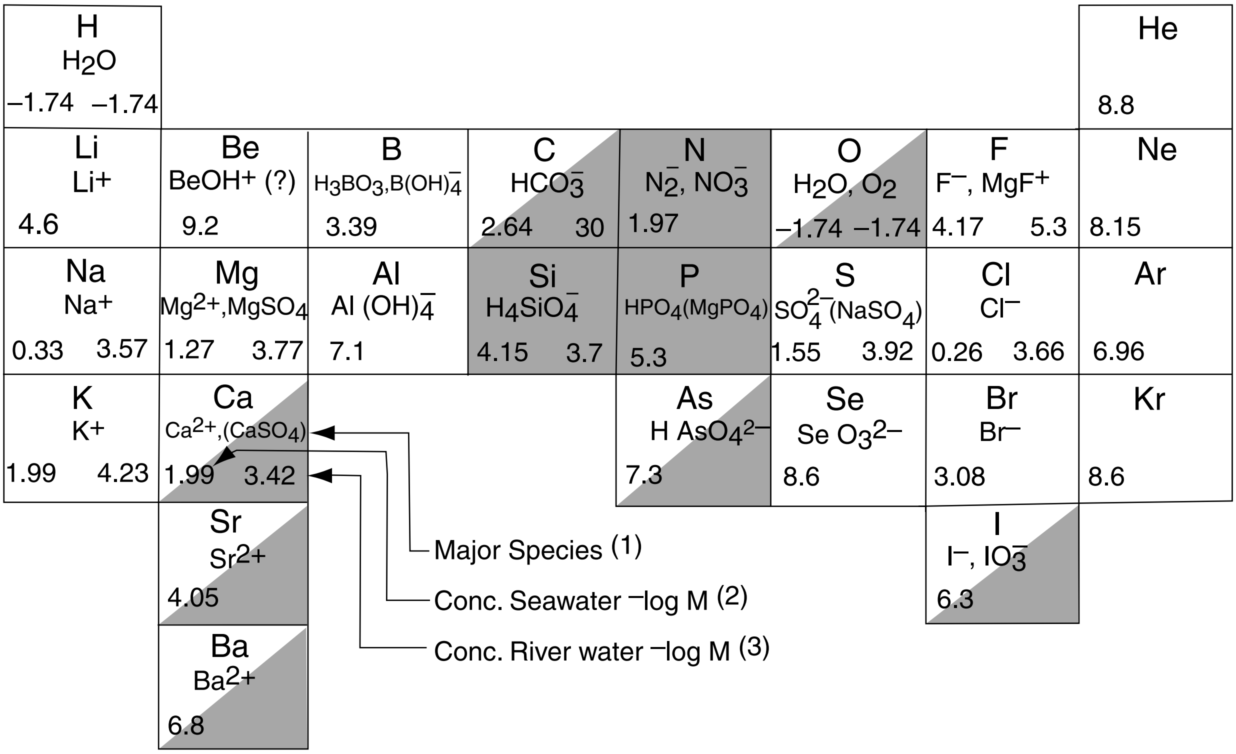


FIGURE 10 Some of the more important elements in natural waters, their form of occurrence and concentrations. Elements whose distribution is significantly affected by biota are shaded. P, N, and Si (fully shaded) often become depleted in surface waters. a) Species in parentheses are major ion pairs in sea water. b) Concentrations (M = mol/1) valid for seawater from P.G. Brewer in *Chemical Oceanography,* Volume 1, J.P. Riley and G. Skirrow, eds, Academic Press, London 1975. c) From A.D. Livingstone, Chemical Components of River and Lakes, US Geol. Survey Paper, 440 G, 1963.

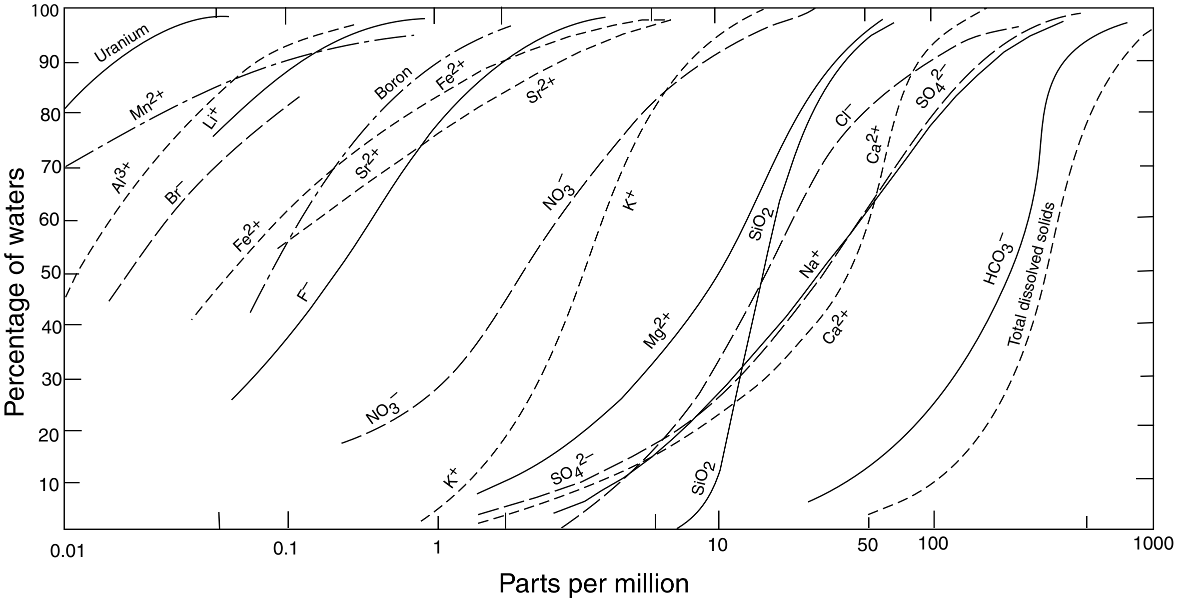


FIGURE 11 Cumulative distribution of inorganic species terrestrial waters. The curves represent the percentage of waters with concentrations below that indicated on the horizontal axis. Data are mostly from the United States from various sources. S.N. Davies and R.C.M. Dewiest, *Hydrogeology,* Wiley, New York, 1966.

Because the relationship between geological environment and water composition depends upon so many interdependent variables, and because there is no substantial variation in each genetic type of water, no reliable generalizations are possible with such a small number of examples. A few qualitative characteristics are apparent from Table 6. As expected, ground waters have higher ionic strengths than surface waters. This results from the difference in CO_3 partial pressure to which these two types of waters are exposed. In comparing waters Nos. 5, 6 and 7, the influence of calcium plagioclase, which is less stable than the silicic rocks, becomes apparent. Sandstone beds are important aquifers and as sedimentary rocks may often contain carbonates besides grains of quartz. Thus the pH of sandstone waters tends to be higher than that of waters from igneous rocks. Shales are laminated sediments whose constituent particles have sizes similar to those of clay. Most of these fine-grained sediments were deposited in saline environments. The relatively high HCO_3^- concentrations of slate ground waters may be due to high CO_2 concentrations resulting from oxidation of organic carbon. Waters from limestone and dolomite contain substantial quantities of Ca^{2+}, Mg^{2+}, and HCO_3^-, the concentrations of which are usually controlled by carbonate equilibria.

The Composition of the Ocean–Atmosphere System

Goldschmidt (1933) suggested that for each liter of presently existing sea water, 600 grams of igneous rocks had reacted with about 1 kilogram of volatile substances (H_2O, HCL, CO_2, etc.). Concomitantly formed during this process were about 600 grams of sediments and three liters of air.

As seen in Figure 10, the solute proportion introduced by rivers differs from that in seawater; in other words, simple concentration of freshwater does not produce ocean water. An important group of constituents of the ocean-atmosphere system (Cl^-, N_2, rare gases) is controlled largely by degassing of the earth. The concentration of O_2 is controlled by the biological reactions of photosynthesis and respiration. Mineral reactions control the concentrations of major cations.

Many constituents (e.g., CO_2, O_2, NO_3^-, HPO_4^-, and H_4SiO_4) are involved in biological cycles in the sea; their distribution and redistribution depend upon the physical circulation of the water and the settling and mineralization of plankton.

Water as a Life Preservation System; Photosynthesis and Respiration

Photosynthesis may be viewed conceptually as an energy consuming process leading to products of high oxidation intensity (organic matter). Respiration is the reverse reaction, the combination of the products of photosynthesis with the release of energy. The photosynthesis/respiration reaction may be written (simplified) as

$$CO_2 + H_2O \quad \underset{\text{respiration}}{\overset{\text{photosynthesis}}{}} \quad CH_2O + O_2.$$

In the aquatic environment the reaction may be written more precisely to include trace elements and to reflect the constitution of algal protoplasm.

$$106CO_2 + 16NO_3^- + HPO_4^2 + 122H_2O + 18H^+ \underset{R}{\overset{P}{}}$$
$$C_{106}H_{263}O_{110}N_{16}P_1 + 138O_2.$$

The relationship between photosynthesis/respiration and the nutrient cycle is seen in Figure 12.

The distribution of chemical species in the waters and in the sediments are strongly influenced by an interaction of mixing cycles and biological cycles. Radio-isotope measurements may often be used to establish the time scale of some of these processes. Similarly, evaluation of the fractionation of stable

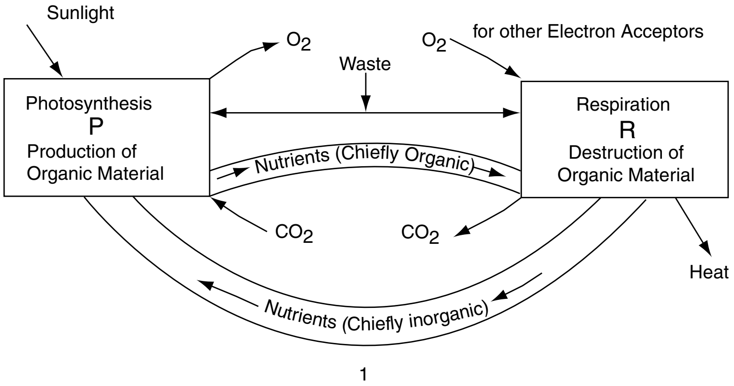


FIGURE 12 Photosynthesis and the nutrient cycle. Solar energy and inorganic nutrients including inorganic runoff are converted by photosynthesis into chemical energy stored within the organic material of algae and plants. Heterotrophic consumers and bacteria mobilize this energy by degrading the organic nutrients including organic waste during respiration and other oxidative process. A steady state of continuous circulation of biochemical materials is indicative of unpolluted water.

TABLE 6
Examples of natural waters

	1	2	3	4	5	6	7	8	9	10	11
Type[a]	Stream	Stream	Stream	Lake Eric	Ground water	Ground water	Ground water	Ground water	Ground water	Ground water	Closed Basin lake
Types of rocks being drained	Granite	Quartzite	Sand-stone[b]		Granite	Gabbro plagioclase	Sand-stone	Shale	Limestone	Dolomite	Soda lake
pH	7.0	6.6	8.0	7.7	7.0	7.8	8.0	7.3	7.0	7.9	9.6
pNa	4.0	4.6	4.3	3.4	3.4	3.0	3.3	2.6	3.0	3.5	0.0
pK	4.7	5.1	4.8	4.3	4.0	4.5	4.0	4.2	3.7	—	1.7
pCa	4.0	4.3	3.1	3.0	3.5	3.1	3.0	2.5	2.7	2.8	4.5
pMg	4.6	5.1	4.0	3.4	3.8	3.2	3.5	2.5	3.4	2.8	4.6
$p\mathrm{H_4SiO_4}$	3.8	4.2	4.1	4.7	3.2	3.0	3.9	3.5	3.7	3.4	2.8
$p\mathrm{HCO_3}$	3.6	4.0	2.9	2.7	2.9	2.5	2.6	2.1	2.3	2.2	0.4
pCl	5.3	5.8	5.3	3.6	4.0	3.5	3.7	4.0	3.2	3.3	0.3
$P\mathrm{SO_4}$	4.5	4.7	3.7	3.6	4.2	4.0	3.2	2.2	3.4	4.7	2.0
—log (ionic strength)	3.5	3.8	2.7	2.5	2.8	2.4	2.4	1.7	2.2	2.2	0.0

Note: $pX = -\log[X]$.

[a] 1–3: "Small Streams in New Mexico," *US Geol. Surv. Bull.*, **1535F** (1961). 4: J. Kramer, *Geochim. Cosmochim. Acta*, **29** 921 (1965). Types 5–10 are from *US Geol. Surv. Bull.*, **440F** (1963). 5: Granite McCormick Co. (Table 1). 6: Harrisburg (Table 2). 7: Home Wood (Table 3). 8: Cuyahoga (Table 5). 9: Edwards limestone (Table 6). 10: Precambrian dolomite (Table 7). 11: "Albert Summer Lake Basin, Oregon," *North. Ohio Geol.*, J.L. Rau, Ed., 1966, p. 181.

[b] With slate and limestone beds.

Ref: Stumm, W. and J. Morgan, *Aquatic Chemistry*, Wiley-Interscience, New York, 1981.

isotopes aids the quantitative interpretation of biogeochemical and environmental processes and cycles (Broeker, 1974).

Organic Carbon

The cycle of carbon in natural wastes is especially important because it is interrelated with the cycles of all other elements. Furthermore, the organic material produced in the oceans constitutes one of man's most significant marine resources.

Most of the carbon in the earth's crust has cycled through organisms and plants; as a result it has been incorporated into unstable but long-lived biochemical structures. Information on the catabolic pathway of organic life substances permits prediction of the type of organic substances to be found in natural waters. Table 7 gives a greatly condensed and simplified survey of the decomposition products of life substances and includes an abbreviated list of specific organic compounds reportedly found in natural bodies.

In fresh water bodies one typically encounters concentrations of a few milligrams carbon per liter, but occasionally (e.g., in bog or swamp waters) concentrations may be as high as 50 mg C per liter. In the oceans the concentrations of organic carbon range from 0.5 to 1.2 mg C per liter, with the higher values occurring in the surface waters. Particulate organic carbon, including planktonic organisms, generally accounts for

TABLE 7
Naturally occurring organic substances

Life substances	Decomposition intermediates	Intermediates and products typically found in nonpolluted natural water
Proteins	$\text{Polypeptides} \rightarrow \underset{\text{aminoacids}}{RCH(NH_2)COOH} \rightarrow \begin{cases} RCOOH \\ RCH_2OHCOOH \\ RCH_2OH \\ RCH_3 \\ RCH_2NH_2 \end{cases}$	$NH_4^+, CO_2, HS^-, CH_4, HPO_4^{2-}$, peptides, amino acids, urea, phenols, indole, fatty acids, mercaptans
Polynucleotides	Nucleotides → purine and pyrimidine bases	CO_2, CH_4, aliphatic acids, acetic, lactic, citric glycolic, malic, palmitic, steric, oleic acids, carbohydrates, hydrocarbons
Lipids $\left.\begin{matrix}\text{Fats}\\ \text{Waxes}\\ \text{Oils}\end{matrix}\right\}$ *Hydrocarbons*	$\underset{\text{fatty acids}}{RCH_2CH_2COOH} + \underset{\text{glycerol}}{CH_2OHCHOHCH_2OH} \rightarrow \begin{cases} RCH_2OH \\ RCOOH \\ \text{Shorter chain acids} \\ RCH_3 \\ RH \end{cases}$	
Carbohydrates $\left.\begin{matrix}\text{Cellulose}\\ \text{Starch}\\ \text{Hemicellulose}\end{matrix}\right\}$	$C_x(H_2O)_y \rightarrow \left\{\begin{matrix}\text{monosaccharides}\\ \text{oligosaccharides}\\ \text{chitin}\end{matrix}\right\} \rightarrow \left\{\begin{matrix}\text{hexoses}\\ \text{pentoses}\\ \text{glucosamine}\end{matrix}\right\}$	HPO_4^{2-}, CO_2, CH_4, glucose, fructose, galactose, arabinose, ribose, xylose
Lignin	$(C_2H_2O)_x$ → unsaturated aromatic alcohols → polyhydroxy carboxylic acids	
Porphyrins and Plant Pigments $\left.\begin{matrix}\text{Chlorophyll}\\ \text{Hemin}\\ \text{Carotenes and}\\ \text{Xantophylls}\end{matrix}\right\}$	Chlorin → pheophytin → hydrocarbons	Phytane Pristane, carotenoids Isoprenoid, alcohols, ketones, acids Porphyrins
breakdown intermediates, e.g., *Complex substances formed from*	Phenols + quinines + amino compounds Amino compounds + breakdown products of carbohydrates	→ Melanins, melanoidin, gelbstoffe → Humic acids, fulvic acids, "tannic" substances

From Stumm and Morgan (1981).

about 10% of the total organic carbon. In much of the ocean there is a "soluble" organic carbon concentration of around 500 micrograms C per liter, with the higher values occurring in the surface concentration of some 50 μg C per liter, all ultimately derived from a surface phytoplankton population equivalent to a concentration of about 5 μg C per liter. Below a depth of 400 to 600 meters the concentration variations in organic carbon become very slight. This suggests that materials present in deep waters are resistant to further metabolic decomposition.

WATER QUALITY CRITERIA AND WATER POLLUTION CONTROL

In a broad sense, pollution has been defined as an alteration of man's surroundings in such a way that they become unfavorable to him. This implies that pollution is not solely caused by contaminants or pollutants added to the environment but can also result from other direct or indirect consequences of man's action. Anything that produces a maladjustment of the ecological balance constitutes pollution.

Every ecosystem has a rather delicate balance of inputs and outputs. Ecological unbalance results, for example, when the rate of extraction of an element (e.g., by mining) by man exceeds the natural rates of cycling or when the output of a product (e.g., crops or forests) is maximized.

Disturbance of the Balance between Photosynthesis and Respiration

Pollution of surface waters frequently results from a disturbance in the balance between photosynthesis (P) and respiration (R). (Reactions (c) and (d)). As Figure 13 illustrates, this balance is upset by adding either an excess of organic wastes or an excess of inorganic algal nutrients (e.g., phosphorus and nitrogen compounds). In the first case (P ! R), heterotrophic processes tend to dominate and dissolved oxygen may become exhausted due to the biochemical oxygen demand and ultimately NO_3^-, SO_4^{2-}, and CO_2 may become reduced to N_2(g), NH_4^+, HS^- , and CH_4(g). In the second case, the incipient result is progressive accumulation of autotrophic biomass (e.g., algae and aquatic plants). Eventually this biomass decomposes enhancing the rate of respiration, which again may reduce dissolved oxygen levels. In either case the initial perturbation is followed by a readjustment, such as a shift in biological species, that finally leads to a new balance.

The P–R balance may also be distributed by a localized, physical separation of the P and R functions. For example, in a stratified lake an excessive production of algae and oxygen in the upper layers may be paralleled by anaerobic conditions (i.e., lack of dissolved oxygen) at the bottom because much of the photosynthetic oxygen escapes to the atmosphere and does not become available to heterotrophs in the deeper water layers.

When P–R, the organic material is decomposed as fast as it is produced photosynthetically. Such a balance,

$$\text{Import}_{\text{org.material}} + \text{P} \; ; \; \text{Export}_{\text{org.material}} + R,$$

is necessary to maintain a water in an aesthetically pleasing condition.

The Pollutants

Much of our concern in the future should be with the fact that most of the energy dissipation by our industrial society for agriculture, landscaping, urban construction and other interventions of civilization ultimates causes a simplification of the ecosystem: specifically, a reduction of the food web and a shortening of the food chains. Generally the less complex a natural

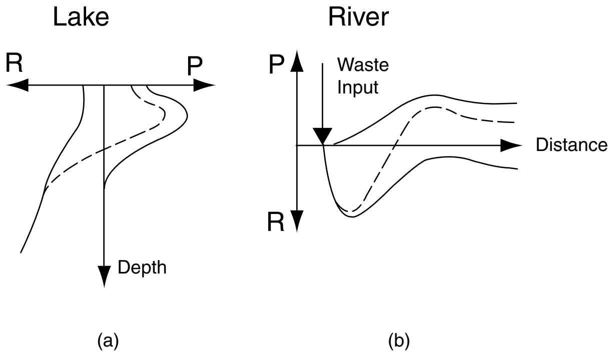


FIGURE 13 Separation of photosynthesis and respiration in polluted lakes and rivers. The separation of photosynthesis and respiration causes pollution by impeding the circulation of chemical constituents with consequent localized increases in photosynthetic or heterotrophic microbial populations. The dashed curves represent the net disturbance of the P–R balance and are indicative of the net rate of oxygen production or consumption.

a) Vertical separation of the P–R functions in lakes can be induced by inorganic nutrients, chiefly nitrogen and phosphorus, which fertilize surface algal growth. Upon settling, the algae stimulate bacterially mediated oxygen utilization and other redox reactions of lower pϵ. The resultant anaerobiosis stifles higher trophic levels.

b) Longitudinal separation of the P–R functions in rivers can be caused by localized injection of organic waste, which stimulates increased bacterial activity and concomitant oxygen consumption. The resultant mineralized nutrients enhance algal growth downstream which upon decay exerts an oxygen demand still further along the river.

ecosystem, that is the smaller the number of species (less diversity), the more liable it is to perturbations and to catastrophe.

Most of the conscious and reasonably successful efforts of environmental and sanitary engineers have been and may well continue to be directed towards the removal of organic carbon of *biological* wastes. For many natural waters, however, other pollutants are ecologically more harmful and may become even more so in the future. As shown earlier, inorganic fertilizing elements are often more serious pollutants than biological wastes for most of our lakes, estuaries, and littoral waters. While biological wastes and inorganic fertilizers cannot yet cause substantial alterations (besides local and temporal transients) in the composition and biology of the oceans, the invasion of the seas by man-made chemicals (e.g., persistent petrochemicals) may cause substantial modification of ecological interrelations in the marine environment.

Conventional biological waste treatment expedites the breakdown of those organic compounds that would eventually be degraded in receiving waters whose self-purification capacities are not overloaded. Such treatment is relatively inefficient with regard to many synthetic organic chemicals. Thus, for many inland waters more emphasis upon the elimination of refractory carbons is necessary.

The inability of some rivers and lakes to readjust rapidly to their initial conditions after even minor shock loadings (e.g., of pesticides or excess quantities of organic wastes) is frequently indicative of a loss of diversity (heterogeneity) in the aquatic biological communities. While many refractory chemicals, at the low concentrations encountered, do not impair the dissolved oxygen balance and do not interfere with the apparent health and survival of fish, they may prevent some fish from returning to their streams of birth for breeding. Often an environmental stress and a decrease in community diversity is signaled by the disappearance of a species, for example, a salmon in a particular stream.

Because of the continuous increase in industrial energy generation, heat will become one of the most significant water pollutants of the future. Woodwell (1969), has shown that many types of environmental disturbances (chronic irradiations, exposure to SO_2, pollution by pesticides, hydrocarbons, heavy metals, excess nutrients) cause similar and roughly predictable gross changes in ecosystems. In each case there is a shift away from complex arrangements of specialized species toward monotone, away from tight nutrient cycles toward loose ones with terrestrial systems becoming overloaded, away from stability toward instability.

Global Aspects of Water Pollution

A comparison of the fluxes of materials resulting from geologic processes and weathering with those of substances introduced by civilization (e.g., agricultural and industrial wastes) shows the relative importance of the role of man as a geological agent and manipulator.

The energy fixed by all plants on the surface of the earth is approximately 8×10^{17} kilocalories per year. Man's metabolism (3×10^{15} kilocalories per year) plays a minor role in the physiology of the biosphere. Man in his capacity to manipulate nature dissipates from 10 to 20 times (in developed countries, 50 to 100 times) as much energy as he requires for his metabolism. Much of the energy utilized by our industrialized society causes localized simplification of the ecosystem and tends to accelerate the cycling of individual elements.

Some of the fluxes caused by man's manipulation are of similar magnitude to those which do not involve man. The combustion of fossil fuels may introduce some elements into the atmosphere and subsequently into the hydrosphere at rates comparable to those of natural processes.

Strong Acids in Rainwater

Rainwaters often contain strong acids. These acids are formed from atmospheric pollutants (HCl, HNO_3, H_2SO_4) (Figure 14).

Chemical Management of Water and Waste Water Quality

Assigning a level of quality to water is relative and depends on the user of the water. For example, different quality criteria apply to drinking waters, to waste water effluents and to receiving waters. Table 8 summarizes some of the objectives of chemical unit processes typically used in water and waste treatment. Water treatment schemes often include several of these processes since generally no single one can achieve the desired water quality objectives. Pertinent chemical reactions in the management of water and waste water are listed in Table 9.

Precipitation reactions are used to remove objectionable ionic species from solution. Hard water containing calcium and magnesium ions, which form insoluble films with soap, is softened by precipitating the ions as calcium carbonate and magnesium hydroxide. Iron and manganese which cause undesired turbidity and stain surface are removed by oxidation and precipitation. Phosphate, a critical aquatic nutrient, can be precipitated as aluminium, iron or calcium solids.

Ion exchange involves the interchange of ions between a solid polymer resin and solution and is used in water treatment for softening and demineralization.

Coagulation is a physico-chemical process involving the agglomeration of colloidal particles for subsequent removal by sedimentation and filtration. The addition of chemical coagulants enhances the mutual attachment of colloidal particles by reducing the repulsive forces acting between these particles (destabilization). In water and waste water treatment practice the hydrolyzed metal ions of Al and Fe(III) salts are commonly used as coagulants. Initial particle contact due to thermal motion (perikinetic coagulation) is normally augmented by the introduction of velocity gradients in solution by slow mixing (orthokinetic coagulation).

The coagulation process may be aided by the addition of macromolecular polymers (polyelectrolytes) which bridge dispersed colloidal particles. Such destabilization is frequently called flocculation.

Gaseous chlorine is one of the economically most useful disinfectants for municipal and industrial waters. During

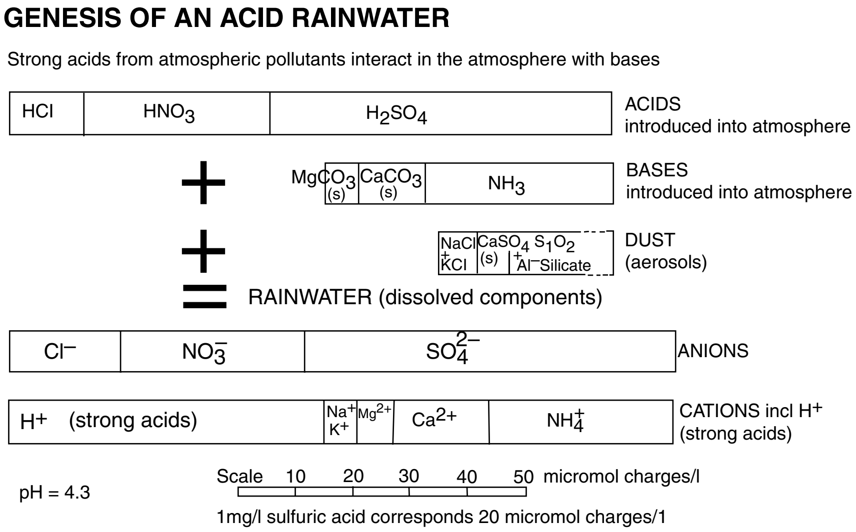


FIGURE 14 Strong acids in rainwater. The acid-base reaction involved in the genesis of a typical acid rainwater. Acids formed from atmospheric pollutant react in the atmosphere with bases and dust particles. The resulting rainwater contains an excess of strong acids. H_2SO_4 originates mostly from S in fossil fuels; after combustion the SO_2 formed is oxidised to SO_3 which gives, with H_2O, H_2SO_4; HNO_3 originates from NO and NO_2. These molecules are formed in the combustion of fossil fuels and to a large extent in the combustion of the automobile engine. For each molecule of NO one of HNO_3 is formed, for example, $NO + O_3 \rightarrow NO_2 + O_2$; $3NO_3 + H_2O \rightarrow 2HNO_3 + NO$. HCl may largely originate from the combustion of Cl-bearing polymers, for example, polyvinyl chloride, in refuse incinerations. Most bases in the atmosphere are often of natural origin. Atmospheric dust may contain carbonates (calcite and dolomite). NH_3 is released from many soils (together with urea). (From Stumm and Morgan, 1981.)

TABLE 8

Chemical unit processes in water and waste water treatment

Substances to be removed	Precipitation	Filtration (chemically aided)	Coagulation flocculation	Disinfection by chlorination	Ion exchange	Activated carbon adsorption	Aeration
Algae	−	+ + +	+ + +	+	−	−	−
Bacteria	−	+ +	+ +	++ +	−	−	−
Colloids turbidity	−	+ + +	+ + +	−	−	−	−
Natural color	−	+ +	+ +	−	+	+ +	−
Odor, taste	−	+	+	+	−	+ + +	+
Organic macromolecules	−	+ +	+ +	−	−	+	−
Organic solutes	−	−	−	−	+	+ +	−
Ca^{2+}, Mg^{2+}	+ + +	−	−	−	+ + +	−	−
Phosphate	+ + +	−	−	−	+	−	−
Inorganic nitrogen compounds	−	−	−	+	+	−	+
Heavy metal ions	+ + +	−	−	−	+ +	−	−
Dissolved Fe, Mn	+ +	+	−	+	+	−	+++

Effectiveness: + + + very effective, + + effective, + marginally effective, − ineffective.

the chlorination process diatomic chlorine gas hydrolyses in solution to form hypochlorous acid and hydrochloric acid. Hypochlorous acid is a stronger disinfectant in its undissociated form (HOCI). Aqueous chlorine, a powerful oxidizing agent, will react with reduced compounds by the formation of chloramines by reaction with aqueous ammonia. (Also see Water Chemistry.)

Water pollution control and improvement of water quality cannot solely depend on water purification and waste treatment. Various physical and biological means of stream management must be directed towards restoring an ecological balance. Because unlimited growth in energy dissipation is incompatible with maintenance of ecological stability and high quality of life, the ecological constraints demand the alteration of social and economic activity toward a situation where resources and materials are recycled.

TABLE 9

Reactions pertinent to the chemical management of water and waste water quality

(Reagent used are underlined)

Precipitation softening

a) Carbonate hardness (temporary or non-incrustant hardness)

$$Ca^{2+} + 2HCO_3^- + \underline{Ca(OH)_2} = 2CaCO_3(s) + 2H_2O$$

$$Mg^{2+} + 2HCO_3^- + \underline{2Ca(OH)_2} = Mg(OH)_2(s) + 2CaCO_3(s) + 2H_2O$$

b) Non-carbonate hardness (permanent or incrustant hardness)

$$Ca^{2+} + SO_4^{2-} + \underline{Na_2CO_3} = CaCO_3(s) + 2Na^+ + SO_4^{2-}$$

$$\begin{aligned} Mg^{2+} + SO_4^{2-} + Ca(OH)_2 + \underline{Na_2\mathit{CO}_3} &= Mg(OH)_2(s) + CaCO_3(s) \\ &\quad + 2Na^+ + SO_4^{2-} \end{aligned}$$

Iron and manganese removal

$$2Fe^{2+} + \underline{\frac{1}{2}O_2(g)} + 5H_2O = 2Fe(OH)_3(s) + 4H^+$$

$$Mn^{2+} + \underline{\frac{1}{2}O_2(g)} + H_2O = MnO_2(s) + 2H^+$$

$$Fe^{2+} + HCO_3^- + \underline{OH^-} = FeCO_3(s) + H_2O$$

Phosphate removal

$$\underline{AI^{3+}} + H_nPO_4^{n-3} = AIPO_4(s) + nH^+$$

$$\underline{Fe^{3+}} + H_nPO_4{}^{n-3} = FePO_4(s) + nH^+$$

$$10Ca^{2+} + 6H_nPO_4{}^{n-3} + \underline{(n+2)OH^-} = Ca_{10}(PO_4)_6(OH)_2(s) + nH_2O$$

Ion exchange

(Brackets, { }, are used to designate a positively or negatively charged polymer network of cation or anion exchange resin with many exchange sites)

$$\{HR\} + Na^+ = \{NaR\} + H^+$$

$$2\{NaR\} + Ca^{2+} = \{CaR\} + 2Na^+$$

$$\{RCl\} + OH^- = \{ROH\} + Cl^-$$

$$2\{ROH\} + SO_4{}^{2-} = \{R_2SO_4\} + 2OH^-$$

Formation of coagulants and flocculants

$$\underline{nFe^{3+}} + mH_2O = Fe_n(OH)_m{}^{3n-m} + mH^+$$

$$\underline{nAl^3} + mH_2O = Al_n(OH)_m{}^{3n-m} + mH^+$$

$$mNa_2SiO_3 + (n-2m)H + mH_2O = H_n(SiO_4)_m^{n-4m} + 2mNa^+$$

Polyelectrolytes; flocculation aids

$\underline{CH_3(CH_2CHCOO^-)_nCH_3}$ (polyacrylate, anionic)

$\underline{CH_3(CH_2CHC_5H_5NCH_3^+)_nCH_3}$ (polyvinylpyridinium, cationic)

Chlorination

$$Cl_2(g) + H_2O = HOCl + H^+ + Cl^-$$

$$NH_3 + HOCl = NH_2Cl + H_2O$$

$$NH_2Cl + HOCl = NHCl_2 + H_2O$$

REFERENCES

1. American Chemical Society, Equilibrium concepts in natural water systems, *Advances in Chem. Series,* **67** Washington (1967).
2. Broeker, W.S., *Chemical Oceanography,* Harcourt, Brace, Janovich, Inc., New York, 1974.
3. Butler, J.N., *Ionic Equilibrium, a Mathematical Approach,* Addison-Wesley, Reading, Mass, 1964.
4. Conway, B.E., *Annual Review of Physical Chemistry,* **17**, Annual Reviews, Inc., Palo Alto Calif., 1966.
5. Dorsey, N.E., *Properties of Ordinary Water-Substances,* ACS monograph No. 81, Reinhold Publishing Corp., New York, 1940.
6. Eisenberg, D. and W. Kauzmann, *The Structure and Properties of Water,* Oxford University Press, England, 1969.
7. Fletcher, N.H., *The Chemical Physics of Ice,* Cambridge University Press, England, 1970.
8. Frank, H.S., *Science,* **169**, 635, 1970.
9. Garrels, R.M., and C.L. Christ, *Solutions, Minerals and Equilibria,* Harper and Row, New York, 1965.
10. Goldberg, E.D., *The Health of the Oceans,* UNESCO Press, Paris, 1976.
11. Goldschmidt, *Fortschr. Mineral Krist. Petr.,* **17**, 112, 1933.
12. Horne, R.A., *Marine Chemistry, The Structure of Water and the Chemistry of the Hydrosphere,* Wiley-Interscience, New York, 1969.
13. Horne, R. A., *Water and Aqueous Solutions, Structure, Thermo-dynamics, and Transport Processes,* Wiley-Interscience, New York, 1972.
14. Horne, R.A., *Surv. Prog. Chem.,* **4**, 1968.
15. Hush, N.S., *Reactions at electrodes,* Wiley-Interscience, New York, 1971.
16. Kamb, *Structure of the ices* in Horne, R.A., *Water and Aqueous Solutions, Wiley-Interscience,* 1972.
17. Kavanau, J.L., *Water and Solute-Water Interactions,* Holden-Day, San Francisco, 1964.
18. Kell, G.S., *J. Chem. Eng. Data,* **12**, 66, 1967.
19. Keyes, F.G., *J. Chem. Phys.,* **17**, 923, 1949.
20. King, T., *Water: Miracle of Nature,* The Macmillan Company, New York, 1953.
21. Krindle, P. and I. Eliezer, *Coord. Chem. Rev.,* **6**, 217, 1971.
22. Malmberg, C.M., A.A. Maryott, *J. Res. Natn. Bur. Stand,* **56**, 1, 1956.
23. Narten, A.H. and H.A. Levy, *Science,* **165**, 447, 1969.
24. Pimental, G. C. and A.L. McClellan, *The Hydrogen Bond,* W.H. Freeman and Co., San Francisco, 1960.
25. Richards, F.A. in *Chemical Oceanography,* J.P. Riley and G. Skirrow, Eds., Academic Press, New York, Chap. 13, 1965.
26. Riley, J.P. and G. Shirrow (eds.), *Chemical Oceanography,* 2 Volumes, Academic Press, London, 1965.
27. Riley, J.P. and R. Chester, *Introduction to Marine Chemistry,* Academic Press, London, 1971.
28. Skinner, B.J., *Earth Resources,* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1969.

WERNER STUMM (DECEASED)
Swiss Federal Institute of Technology

STEVEN GHERINI
Harvard University

MARTIN FORSBERG
Harvard University

WATER REUSE

A commonly accepted definition of water reuse is that the water that will be further utilized will still be under the control of the first user. An early example of water reuse involved the sale of secondary wastewater effluent from the Back River Treatment Plant in Baltimore to the Bethlehem Steel Sparrows Point Plant. The Sparrows Point Plant was served by wells, which were close to brackish Baltimore Harbor. Excessive pumping of the wells led to saltwater intrusion, and low saltwater concentrations made the well water unsuitable for some operations. Steel rollers pit when applied water contains 5 mg/l chloride ion. Wolman recommended that the treated wastewater effluent be used for some rough operations, such as quenching. An 8-mile pipeline was installed to carry the effluent to the industrial plant. Drawdown on the wells was reduced significantly, and the city of Baltimore received revenue for a previously discarded product.

Water reuse can be regarded as a conservation strategy. Here the discussion will deal with reclamation of used water. Desalination, or desalting, is covered in another chapter (see "Desalination"). Guidelines and design standards have been developed for water reclamation and water reuse as alternatives to disposal of surface receiving waters. Table 1 lists water-reuse applications. It should be noted that other uses will be found.

All water is reused. It may be used and used again in its journey from deposition as rain to eventual mixing with the sea. Available water is not always near the point at which it is required. Economics of supply is important, and pumping cost may dictate the source chosen.

In many cases, water for industrial processes must be of much higher quality than that used for drinking. In water-poor areas reused water has long been accepted, even for drinking. In general, however, public-health authorities have not looked favorably on this use. A poll found that 50% of the people surveyed did favor reclaimed water for drinking.

Water usually is not immediately available for reuse. With more stringent regulations governing effluent quality coming into effect, these higher-quality discharges may become more practical for water reuse.

Conventional primary and secondary wastewater treatment alone are not normally sufficient to prepare wastewater for any but the coarsest uses. Further, or tertiary, treatment will be required. Secondary, or biological, treatment will remove about 90% of influent suspended solids and BOD (biochemical oxygen demand). BOD and COD (chemical oxygen demand) do not themselves affect suitability of the effluent for reuse but are indications of the unstable nature of the effluent. These will be reduced in the natural decomposition cycle.

Some compounds are not removed in conventional wastewater-treatment processes. These substances, termed "refractory," range from simple inorganic compounds to large synthetic organics such as pesticides, herbicides, and surfactants. In an essentially closed system, high concentrations of unwanted materials would soon result from continued recycling. In Table 2, components of typical tap water are compared with those of a typical secondary effluent.

Stream quality or other water sources might be improved enough by removal of BOD and suspended solids. In some cases it is possible to save large amounts of water by process changes and improved housekeeping, and water reuse might not be necessary.

If used water is not yet suitable for reuse after conventional treatment, it may be blended with clean water and the mixture rendered acceptable. However, in many cases tertiary treatment will be required. A common definition of tertiary treatment is any treatment in addition to secondary treatment by which further removals are achieved. A classification of wastewater-treatment processes is presented in Table 3.

Chlorination, ozonation, and application of ultraviolet light are not considered tertiary processes. Chemical precipitation, by itself, is primary treatment. However, applied after secondary treatment, it can give a high-quality effluent. An insoluble compound is formed and the resulting precipitate sweeps other material from suspension. Neutralization of charge on colloids and agglomeration of particles, including microorganisms, give a clear effluent. Phosphates are removed by addition of soluble aluminum or iron salts. Lime is usually added to produce hydroxide floc.

Filtration may be necessary for further particulate removal after chemical precipitation. In this case filtration is not straining out of the suspended matter. The primary mechanism is adsorption of the suspended matter onto the surface of the granular filter media. Sand, anthracite, and garnet are commonly used filter materials. A mixed-media filter has a greater effective depth than does a comparable sand filter. When a filter clogs, it may be backwashed by water alone or a mixture of water and air. A filter acts as a fluidized bed during backwashing.

Diatomaceous-earth filtration is also practiced. A layer of diatomaceous earth is placed over a relatively loose septum to

TABLE 1

Type of Reuse	Applications
Urban public access (unrestricted)	Parks, playgrounds, schools, residences, toilet flushing, air conditioning, fire protection, construction, fountains, and ponds
Urban public access (controlled)	Golf courses, cemeteries, and highway medians
Agriculture (food crops)	Irrigation needs of food crops the type of reuse water is further classified as food crops consumed raw or processed
Agriculture (nonfood crops)	Irrigation needs for crops used as fodder, fiber and seed crops, pastureland, commercial nurseries, and sod farms
Recreation (no restriction)	Recreational activities with no imposed restriction on body contact
Recreation (restricted)	Noncontact recreational activities such as fishing and boating
Environmental uses	Creating wetlands, enhancing natural wetlands, and maintaining or supplementing stream flows
Industrial uses	Cooling system makeup and washdown
Groundwater recharge	Infiltration basins, percolation ponds, and injection wells for aquifer recharge
Indirect potable reuse	Discharge of highly treated reclaimed water into surface well or groundwater for potable water use

TABLE 2

Component	Tap Water, mg/l	Secondary Effluent, mg/l
BOD	2	20
COD	2	100
Alkylbenzenesulfonates (ABS)	0.4	6.8
Cations		
Na^+	50	125
K^+	3	13
NH_4^+	0.1	16
Ca^{++}	42	60
Mg^{++}	16	19
Anions		
Cl^-	66	143
NO_2^-	0.15	1.5
NO_3^-	5	12
HCO_3^-	198	296
CO_3^{--}	0.1	1.0
SO_4^{--}	56	84
SiO_2^{--}	29	43
PO_4^{--}	8.1	25
Hardness (as $CaCO_3$)	158	235
Alkalinity	164	242
Total dissolved solids	382	700
pH	8.0	7.4

TABLE 3

Process	Substances Treated
Biological	
Conventional secondary	BOD, COD, suspended solids, microorganisms
Modifications of conventional	BOD, COD, suspended solids, microorganisms,
Secondary	nutrients
Anaerobic denitrification	Nitrates
Algal harvesting	Nitrates and phosphates
Chemical	
Ammonia stripping	Ammonia nitrogen
Ion exchange	Nutrients
Electrodialysis	Salts
Chemical precipitation	Suspended solids, phosphates
Physical	
Activated carbon adsorption	Organic matter, suspended solids
Sedimentation	Suspended solids, microorganisms
Filtration	Suspended solids, microorganisms
Microstrainers	Very small particles
Reverse osmosis	Salts
Distillation	Salts
Foam separation	Detergents
Vapor-compression evaporation	Concentration of wastewater
Wasteheat evaporation	Concentration of wastewater
Steam stripping	Hydrogen sulfide, ammonia

screen out suspended solids down to 0.5 to 1.0 µm. This type of filter clogs rapidly and must be backwashed frequently.

Both vacuum and pressure filters are used for tertiary treatment. About 85% of influent turbidity is removed in a vacuum filter. Application rates in the range of 0.5 to 1.0 gpm/ft^2 are common for both types of filters. These low rates make both vacuum and pressure filters unsuitable in large installations.

Microstrainers are rotating drum screens with extremely fine stainless steel or polyester fabric having openings in the range of 15 to 60 µm. Flowing liquid enters the effluent end of the horizontal drum and passes through the mesh to the effluent end. The mesh rotates with the drum while trapping solids. A spray of treated effluent washes the retained solids into a hopper, from which they are lead to final disposal. Ultraviolet radiation inhibits microbial growth. At intervals of 7 to 28 days the mesh is washed with chlorinated water for control of slime growth. Wash water is about 3 to 5% of the total treated water. Microstrainer removal efficiencies are 30 to 55% of applied BOD and 40 to 60% of suspended solids.

Preparation of effluents of exceptional quality involves "effluent polishing." Plant effluent is collected in a sump, pumped through tube settlers, and then sent through a mixed-media filter bed. Filter effluent is collected in storage tanks. These tanks serve the dual roles of reservoir for filter backwash water and chlorine contact chamber. The tube-settler–mixed-media combination removes phosphate when preceded by coagulation and flocculation. These units are suitable for production of a high-quality effluent when the applied influent has a suspended-solids concentration as high as 2000 mg/l (0.2%).

Hard detergents such as ABS (alkybenzenesulfonate) are not removed by conventional biological treatment. ABS has been replaced almost completely in household detergents by linear compounds, but foam continues to be a problem. Foam-separation techniques have been applied for removal of refractories such as organic hydrates and nitrogenous compounds. Air is pumped through spargers, producing small bubbles. Rising bubbles collect suspended solids and surface-active substances. Foam collects at the top and is removed and collapsed, yielding a waste concentrate. In some installations waste flows downward, while gas spargers at the bottom give countercurrent flow.

Activated carbon has long been used for removal of tastes, odors, and color. It has not been suitable for treatment of waters with high concentrations of organic matter because the surface area available for adsorption is soon exhausted at the influent end of a packed bed. However, a fluidized bed of granular activated carbon gives a more uniform concentration gradient throughout the bed. Removal rates of 70% for BOD and 80% for COD are possible with activated carbon adsorption. Carbon is regenerated in multistage hearth furnaces in a stream-air mixture. About 5% of the carbon is lost with each regeneration.

Hard waters are softened by ion exchange. Calcium and magnesium ions are exchanged for a cation in the resin, usually sodium, and 100% of sulfate, 95% phosphate, 85% nitrate, and 45% COD are removed by ion exchange. Color and organic matter are removed by cationic and anionic mixed beds, but organic matter tends to foul the beds. Cost and frequency of regeneration are disadvantages.

Cross-flow membrane-separation technologies of microfiltration, ultrafiltration, reverse osmosis, and nanofiltration may be defined on the basis of pore size or removal function. Membranes are commonly rated by their molecular-weight cutoff, the maximum molecular weight of a compound that will pass through the membrane. Microfiltration removes suspended submicron material but does not remove truly dissolved material. Ultrafiltration removes nonionic solute. These may be organic macromolecules. Microfilters and ultrafilters may be regarded as sieves. Reverse osmosis involves a solution or suspension flowing under pressure through a membrane and the product being withdrawn on the low-pressure side. This process can treat dissolved-solid concentrations of 1 mg/l to 35,000 mg/l and particles ranging in size from less than 1 nm to 600 nm. The treated material must be nonfouling. Frequently, it is necessary to pretreat in order to avoid membrane fouling. Recent advances have been mostly in membrane development. Pilot studies for particular applications may be required. Energy costs can be quite significant in process economics. Mass transfer in reverse osmosis comes about due to pressure difference across a membrane. Nanofiltration rejects dissolved ionic contaminants, but to a lesser degree than in reverse osmosis.

In electrodialysis, direct current is applied to a series of alternating anionic and cationic membranes. Anions pass through anion-permeable membranes but are prevented from migrating through cationic permeable membranes. Electrodialysis deals only with ionic species, but reverse osmosis can be shared with electrodialysis.

Vapor-compression evaporation and waste-heat evaporation remove water from contaminants instead of removing contaminants from water. Application of these processes is most economical when waste heat is readily available, such as industrial installations. Water thus produced is of high quality. However, organic substances that steam-distill or form azeotropes may appear as contaminants. A distinct advantage is concentration of residue. Transportation costs are thus reduced. Use of excess heat for these processes can reduce the amount of cooling water required.

Oxidation ponds are shallow ponds into which wastewater is led. Also called sewage lagoons, these ponds have long been in use in Europe but are not popular in North America. Public resistance has been on aesthetic grounds. A secondary effluent can be further purified in these ponds, and the water produced is suitable for many purposes. Algal action and surface aeration make this an aerobic process. Depth is an important consideration. The pond must be shallow enough to allow penetration of sunlight, but it must be deep enough to keep from being choked by aquatic weeds. Four feet is a common depth. Some installations have aerators at the pond center.

Water reuse is an art as well as a science. The processes presented here are commonly used in preparation of wastewater for immediate reuse. However, the particular situation under study must dictate the economics of reuse. The ultimate level of reuse is the dominant factor in deciding if reuse is practical.

REFERENCES

Cartwright, P.S., Membranes for process water reuse. *Chemical Engineering,* pp. 38–42. June 2004.

Cecil, L.K., Water reuse. *The encyclopedia of environmental science and engineering.* 2nd Ed., Vol. 3. Gordon and Breach. New York. 1983.

Cooper, W.J., Ed. *Chemistry in water reuse.* Ann Arbor Science. Woburn, MA.[2] 1981.

Dean, R.B. and E. Lund, *Water reuse: problems and solutions.* Academic Press. New York. 1981.

Delyannis, E., Desalination. *The Encyclopedia of Environmental Science and Engineering.* 4th Ed., Vol. 1. Gordon and Breach. New York. 1998.

Guidelines for Water Reuse. U.S. Environmental Protection Agency. Washington, D.C. 1992.

Hairston, D., Desalting process water. *Chemical Engineering,* pp. 27–30. November 2004.

Impelluso, A. and J.R. Pfafflin, Water reuse. *The Encyclopedia of Environmental Science and Engineering.* 4th Ed., Vol. 2. Gordon and Breach. New York. 1998.

Pfafflin, J.R., Water reuse. *The Encyclopedia of Chemical Technology.* Vol. 26. Wiley-Interscience. New York. 1983.

Reynolds, T.D., *Unit Operations and Processes in Environmental Engineering.* Wadsworth. Belmont, CA. 1982.

Wolman, A., *Sewage Works J.* **20**(1), 1. 1948.

PRASANNA RATNAWEER
Open University
Sri Lanka

WATER TREATMENT

INTRODUCTION

Water, of course, is used for many purposes associated with human activity. In its natural state it occurs in and on the ground in subsurface and surface reservoirs. The quality and reliability of a source of water will vary considerably, both in time and space. This means that characteristics (chemical, physical, and biological) will differ greatly depending upon the location and type of source. It also means that a given source may vary over the seasons of the year.

Thus, in the selection of a water source, consideration is usually given to the use to which the water will ultimately be put so as to minimize the cost of treatment. Simultaneously consideration must be given to the reliability of the source to provide an accurate and constant source of supply. It will be seen later in this section that a groundwater supply may enjoy the benefit of requiring little or no treatment, while a surface supply such as a river, pond or lake may require considerable and perhaps seasonally varying treatment. However, a surface supply is visible and therefore more reliable whereas a groundwater supply may just disappear with no warning or notice.

In certain areas, freshwater is so scarce that the source must be accepted and choices are not available. The history of water treatment dates back to the early Egyptian civilizations where the murky waters of the Nile River were held in large open basins to allow the mud to settle out. The earliest archeological records of a piped water supply and wastewater disposal system date back some five thousand years to Nippur of Sumaria.[1] In the Nippur ruins there exists an arched drain with an extensive system of drainage from palaces and residences to convey wastes to the outskirts of the city. Water was drawn through a similar system from wells and cisterns.

The earliest records of water treatment appear in the Sanskrit medical lore and Egyptian wall inscriptions.[2] Writings from about 2000 BC describe how to purify "foul water" by boiling in copper vessels, exposing to sunlight, filtering through charcoal and cooling in an earthenware vessel. There is little concerning water treatment in the Old Testament, but Elisha under instruction from the Lord "healed" certain waters so that "there shall not be from thence any more death or barren land." This "healing" was accomplished when Elisha "went forth unto the spring of the waters and cast salt in there …" It is not clear if this "salt" was a fertilizer to help grow crops or if it was some chemical to render the water safe.

Settling was first introduced as a modification of decanting apparatus used for water or wine. This apparatus was pictured on the walls of the tombs of Amenhotep II and Rameses II in the 15th and 13th Centuries BC. An engineering report on water supply was written by the then water commissioner for Rome in AD 98. He described an aqueduct with a settling basin.

In 1627 the experiments of Sir Francis Bacon were published just after his death, and were the first to describe coagulation as well as sedimentation and filtration as a means of treating drinking water. The first filtered supply of water for an entire town was built in Paisley, Scotland in 1804. Starting with a carted supply, a piped distribution system was added in 1807.[2]

However it was not until 1854, in London, that it was demonstrated that certain diseases could be transmitted by water. Dr. John Snow suggested that a cholera outbreak in a certain area in London resulted directly from the use of the Broad Street pump, and was in fact the source of infection in the parish of St. James. Dr. Snow recommended that the use of the pump should be discontinued and the vestrymen of the parish agreeing, the disease subsequently abated in that area.

The discovery was all the more incredible as the germ theory of disease, defined by Pasteur and subsequently postulated by Koch, had not at that time been clarified. Subsequently disinfection of water by addition of chlorine was introduced on a municipal scale. This step, together with an adequate and sanitary distribution system, probably did more to reduce the deaths due to typhoid and cholera and any other single item.

In 1854, cholera claimed a mortality of 10,675 people in London, England. In 1910, the death rate from typhoid fever in the City of Toronto, Canada, was 40.8 per 100,000. By 1931 it had fallen to 0.5 per 100,000. These improvements all related to the extensive water purification and sterilization techniques which are being introduced to municipal water treatment systems during that period.[3]

In general, the treatment processes of water can be subdivided into three groups: physical, chemical, and biological processes. The biological processes are generally reserved for waters grossly contaminated with organic (putrescible) carbon such as sewage or industrial waste waters. These waters are not normally considered as suitable for drinking supplies, but undoubtedly as demand for water increases *all* available sources will have to be examined. However, for the present purposes we will consider that the biological stabilization of originally polluted waters will be dealt with under the section on wastewater treatment. It should, however, be realized that there is a very fine line between treated wastewater discharged into a water body and the use of that water body as a source for drinking water and the treatment of the wastewater before discharge. Clearly in those areas where wastes are still not treated prior to release, a water treatment plant is essentially dealing with the treatment of diluted wastewater.

We must therefore determine the significance of water quality before we examine the types of treatment necessary to achieve this quality. Water quality very much depends upon the use for which the water was intended. For example, industrial boiler feed water requires a very low hardness because the hardness tends to deposit on the pipes in the boiler system and reduces the efficiency of the heat transfer. However, if the hardness of the boiler feed water is zero, the water tends to be very corrosive and this of course is also very undesirable for a boiler system.

If the water is to be used for a brewery or a distillery, a number of other chemical parameters are important. If the water is to be used for cooling then clearly the temperature is one of the most important parameters.

In the past the methods for setting standards for water supplies was very much a hit and miss affair and relied pretty well upon the philosophy of "If no one complains, all is well." Clearly, that is not a very satisfactory criterion. There are a number of drinking water standards or objectives published by various nations of the world, such as the World Health Organization International and European Drinking Water Standards (1963 and 1961), the US Public Health Service Drinking Water Standards (1962) and Objectives (1968). These standards are established on the principle that water in a public water supply system must be treated to the degree which is suitable for the highest and best use. The highest and best use for water of course is human consumption. This can frequently be argued as a rather unnecessary quality when one considers that much water which is processed in a municipal plant is used for watering lawns, washing cars and windows. However, the difficulty in ensuring that a second-class, perhaps unsafe water supply is not used as a potable supply is extremely difficult. It will be found that very few cities have a dual water supply representing a drinking water system and a non-potable system.

A few large cities, particularly when they are adjacent to large standing bodies of water, occasionally have a fire water supply system where the water is taken untreated from the lake or river and pumped under high pressure through a system connected only to fire hydrants and sprinklers.

Thus, assuming that natural water requires some kind of treatment in order to achieve certain predetermined standards, and the process of treating these waters can be subdivided into physical and chemical processes, the remainder of this section will deal with the physical and chemical methods of treating water for municipal or industrial use.

WATER SOURCES

The magnitude of the problem of supplying water to the major cities of the world is in fact a huge engineering problem. According to a US Department of Commerce estimate, the cities of the United States in 1955 with a total population of 110 million produced and distributed 17 billion gallons of water daily to their domestic, commercial, and industrial consumers. Of this, 12.88 billion gallons were from surface water sources which usually, it will be seen, require more elaborate treatment, whereas the remaining 4.12 billion gallons came from groundwater sources—only a small proportion of which would require treatment.[4] The most voluminous source of water is the oceans. It is estimated that they contain about 1060 trillion acre-feet.[5] Clearly this water is of little value as a potable source, but it certainly remains the main reservoir in the hydrologic cycle.

It can be seen from Figure 1 that evaporation is the first step in the purification of ocean water, and this requires the full energy of the sun in order to accomplish. Precipitation, percolation, and runoff are all parts of the cycle of water which is without a beginning or an ending. Of the water which falls upon the earth, part of it directly runs off to the nearest stream or lake, and part of it infiltrates down to the groundwater table and percolates through the groundwater, also into a stream or lake. Transpiration takes place through the leaves of green plants, and evaporation takes place from the groundwater, where it surfaces through swamps, lakes or rivers, and of course from the ocean. Of the water that soaks into the ground, part of it is retained in the capillary voids near the surface. Thus it can be said that the potential sources of water for society consist of wells, which are drilled or dug down to the groundwater table and withdraw water from that level; springs, which are natural outcroppings of groundwater table through rocks or ground; rivers, where the groundwater table has naturally broken through the ground and flown in a certain direction sufficiently to gouge out a channel for the water to flow in; lakes, where large bodies of water gather usually somewhere along a river system; and finally the ocean, if not other sources are available and the ocean is close by. The benefit derived from the costly treatment required to desalinate the ocean under these circumstances is outweighed by the necessity of having a fresh water source at any cost.

There are new water sources which exist deep in the earth's crust. These sources are rarely considered, due to the high salt and sulphur content which is frequently found in them.

The recycling of used water of course is a further source which may be tapped directly. It can be seen from the hydrologic cycle that all water is being continually reused, but the direct recycling of municipal treated sewage into the potable treatment plant is being considered in some water-scarce areas.

Some of the advantages and disadvantages which might be listed for the various sources of water are as follows:

1) Wells provide usually an extremely pure source of potable water. Rarely is any treatment required of this water, certainly before it is safe to drink, although certain industrial uses may require the removal of some of the soluble salts such

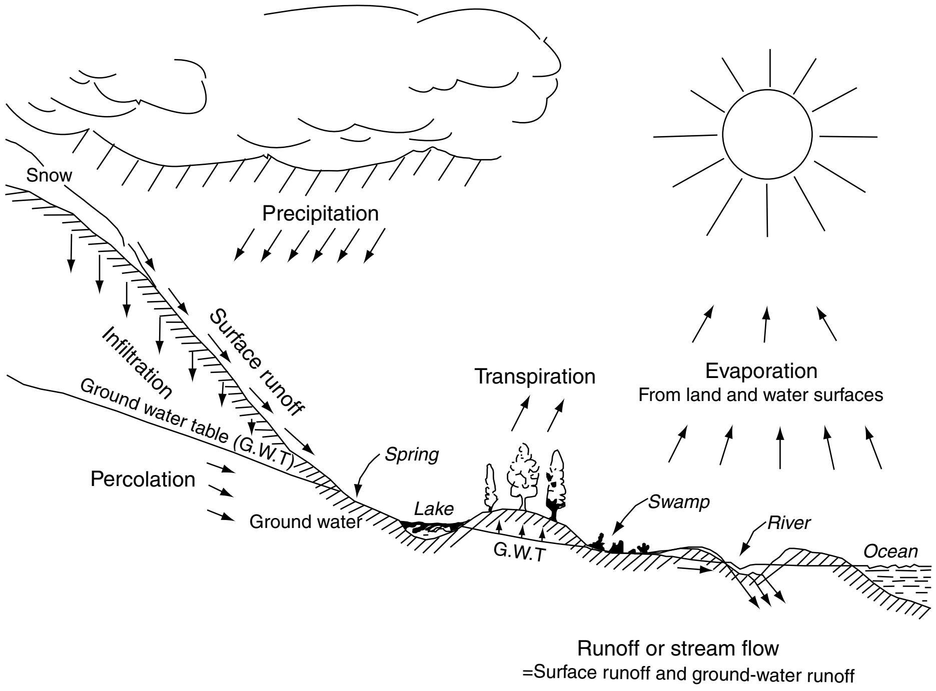


FIGURE 1 Hydrologic cycle (Fair and Geyer, Water Supply and Wastewater Disposal).

as hardness. One of the major disadvantages of wells is that they cannot be observed and therefore must be considered as somewhat unreliable. Frequently it has been experienced in the country, if a drought has persisted for a few days or a few weeks (depending on the environment) and the well has been pumped unusually hard, that the well will run dry. There is never very much warning of this kind of occurrence and therefore for a municipal supply it has the distinct disadvantage of being considered somewhat unreliable.

2) Springs are similarly unreliable, and have a further disadvantage in that they require a rather elaborate engineering system to capture them and concentrate them into one manageable system. Also, springs require rather a large protected area to ensure that man does not pollute this environment, thereby rendering the springwater unsafe.
3) Rivers tend also to be a little unreliable, although they do have the advantage that they can be observed and to some extent controlled through dams and other waterflow structures. Thus it can be seen, if the water level is falling, that a municipality may wish to impose water use restrictions to conserve water until such time as further augmentation of the supply is received through the hydrologic cycle. One of the major problems with a river source is that there is a considerable variation in the quality of the water. During the high flow flood period, there is frequently a considerable amount of silt and organic material which is washed off the ground, whereas at other times of year the water may be relatively clean and require remarkably little treatment prior to distribution. This of course means that water treatment facilities must be installed to deal with the worst possible condition, and at other times of the year it may not in fact be necessary and therefore the equipment lies idle.
4) Lakes and manmade reservoirs, due to the nature of flow through them, have a certain stability both from the point of view of quantity and quality. Undoubtedly, water coming from a lake or a reservoir would require far more elaborate treatment than would water from a well. However, the extreme reliability and the predictability of supply may well outweigh the considerations of cost of treatment. This of course is subject to an economic feasibility study.
5) Oceans. A good deal of attention is currently being focused on the desalination of ocean water, and some attention will be paid to this subsequently in this section. It should, however, be remembered that the ocean is only, economically available to these communities which are immediately adjacent to the ocean. This leaves a very large area of hinterland in most continents which does not have access to the sea. Thus the

desalting of sea water as a major water source has a restricted application to small islands and those coastal stretches of countries where fresh water reserves are either not available or not reliable.

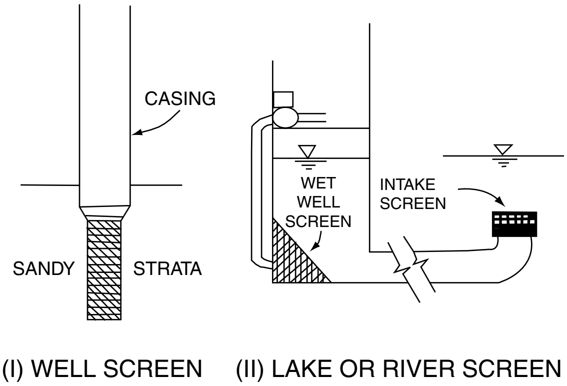
6) Recycled water. A considerable amount of research has been undertaken in the United States and elsewhere for the renovation of treated wastewater for the purposes of returning it directly into the potable supply. Some rather complex chemical and physical processes are required to make this a satisfactory process, and the details of many of these processes will be described subsequently in the next section of this chapter.

PHYSICAL TREATMENT

The items of treatment described under this section will be only those which alter the physical properties of the water or represent a unit process which is physical in nature. All of the processes described may be used individually, collectively or in any combination, in order to accomplish a predetermined water quality.

Screens

Whatever the source of water, it is necessary to insert some kind of screen in the system in order to prevent the passage of solids into the subsequent steps of water treatment. If the source of water is simply a well, the screens tend to be simply designed to prevent the admission of sand from the water-bearing strata into the pumping system. Where water supply is drawn from rivers or lakes, the intakes usually have to be screened and built of corrosion-resistant materials in order to prevent the admission of fish or logs or any other undesirable solids into the system. Intake screens are usually provided with openings approximately equal to one and one-half to two times the area of the intake pipe. The purpose of this is to ensure that the velocity through the screens is sufficiently low to prevent jamming of the screens. On occasion other screens are required as a backup system within the water treatment plant.



(I) WELL SCREEN (II) LAKE OR RIVER SCREEN

FIGURE 2

In some locations where it is found that seasonally algal blooms become a nuisance, a new type of screening known as microstraining has been introduced. Microstrainers are a very fine weave of stainless steel wire with apertures sufficiently small to prevent the passage of the microscopic algae which is normally found in an algal bloom. Such a screening system is normally only required on a seasonal basis and in certain locations where these problems are prevalent. Microstraining is conducted at such a very small diameter orifice that it is sometimes considered to be a part of a filtration process.

Coagulation

Although the basis of coagulation is in fact chemical treatment and will be discussed in the next section, the coagulation process itself (sometimes referred to as flocculation) is accomplished by a physical process involving the gentle agitation of the fluid which allows the small suspended particles to collide and agglomerate into heavier particles or flocs and settle out. Flocculation or coagulation is the principle used in the removal of turbidity from water. It will be shown subsequently that colloidal or very finely divided material will not settle very rapidly. Various processes have been employed to accomplish flocculation. Some of these are; diffused air, baffles, transverse or parallel shaft mixers, vertical turbine mixers, to mention but a few.

The most common type of flocculator used today is the paddle type, the other methods having shown some disadvantage such as being too severe for the fragile floc, or being too inflexible, or being too costly to operate. Horizontally mounted paddles, either located transverse or parallel to the floor, consist of a shaft with a number of protruding arms on which are mounted various blades. The shaft rotates at a very slow rate of 60 to 100 rpm, causing a very gentle agitation which results in the flocculation of the particles. The time required for the flocculation process is very carefully controlled and strongly related to the dosage of chemical which is used. The chemicals used and the chemistry of this process will be described later.

Prior to the flocculation step which has just been described, occurs a flash mixing step when the chemicals are added and mixed very rapidly at high speed to get uniform distribution of the chemical in the stream. A variety of devices are used for this rapid mixing operation; frequently one of the most common includes the low lift pumps which are usually located adjacent to the intake where the water is lifted up into the treatment plant. Here of course the chemicals must be pumped into the pump casing at a higher pressure than the pump is producing, and the mixing takes place in the casing of the pump.

Other devices frequently used are venturi flumes, air jets, paddles, turbines, propellers, the latter being one or the most favored and most widely used of the rapid mixing devices. It usually is composed of a vertical shaft driven by a motor

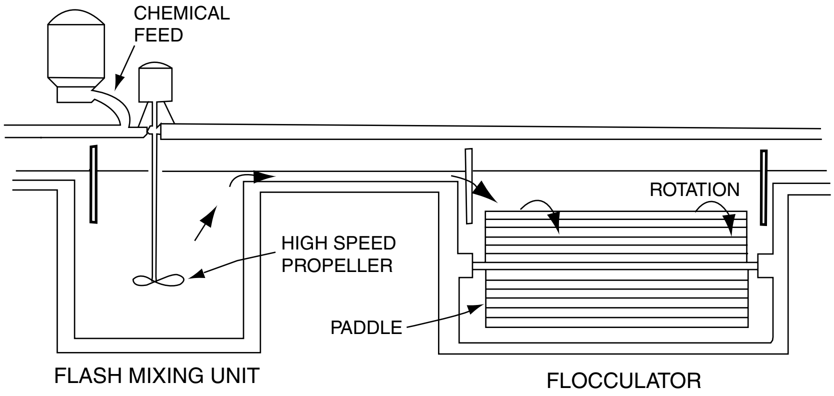


FIGURE 3

on which one or more propeller blades are mounted. Baffles are frequently used to reduce the vortexing about the propeller shaft. Vortexing hinders the mixing operation. Detention periods are usually of the order of one to five minutes, usually at the lower end of this range. Considerable study has been done on the baffling arrangement in a flash mixing unit, and a variety of arrangements have been shown to be successful.

Sedimentation

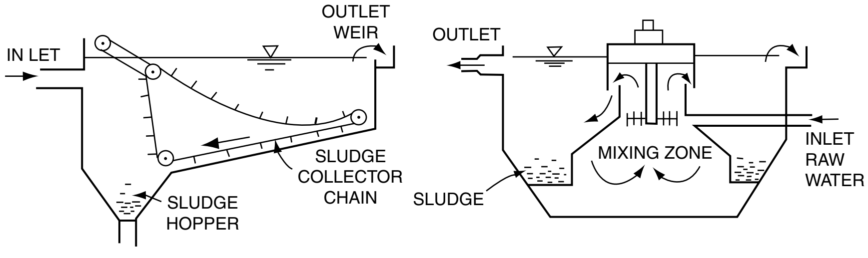
Sedimentation or settling may be accomplished by a variety of means and mechanisms, depending on the material which is to be settled from the liquid.

Discrete settling This type of sedimentation is primarily concerned with the settling out of non-flocculent discrete particles in a fairly dilute system. The primary feature of this type of settling is that the particles do not flocculate and therefore their settling velocity and particle size remain the same throughout the period of settling. It will be seen later that this is quite different from other forms of settling.

The particles in discrete settling will accelerate until the fluid/drag reaches equilibrium with the driving force acting on the particle. In other words, the resistance of the water is equal to the accelerating force of gravity of the particle. When this velocity is reached, it will not increase. This is known as the terminal settling velocity, and it is normally achieved quite rapidly. The loading rate which is used frequently for the design of a settling tank is known as the overflow rate and may be expressed in cubic feet per square foot per day based on the area. It can be seen that cubic feet per square foot per day is in fact the same as feet per day, or in fact a simple velocity. This velocity is defined as the settling velocity of the particles which are removed in this ideal basin if they enter at the surface.

Overflow rates or surface loadings of 150 gallons per day per square foot of tank surface are not unusual where the settling and sand, silt or clay are being accomplished by plain sedimentation.

Flocculent settling The primary difference between this type of settling and the previous one is that in a flocculent system the larger particles subsiding at a slightly higher rate



LONGITUDINAL SETTLING TANK

CIRCULAR COMBINATION SETTLING FLOCCULATOR
(Accelator by Infilco)

FIGURE 4

will overtake and coalesce with smaller particles to form even larger particles, which in turn increase the overall settling rate. Clearly, the greater the liquid depth, the greater will be the opportunity for this type of contact. There is no mathematical relationship which can be used to determine the general effect of flocculation on sedimentation, and empirical data is still required by studying individual laboratory cases. As a result, in flocculent settling the removal of suspended matter depends not only on the clarification rate but also on the depth.

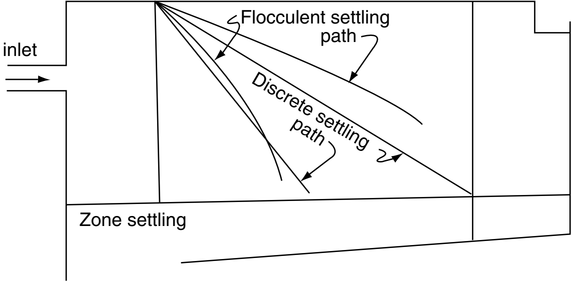
This is one of the significant differences between non-flocculent and flocculent settling.

Zone settling The previous two types of settling described have one property in common, and that is that they both deal with dilute suspensions. Zone settling, on the other hand, deals with very concentrated suspensions where it is assumed that one particle will in fact interfere with the settling rate of another particle. It is clear that in the type of discrete settling, where the particles are somewhat non-reactive and usually quite dense such as sand, the difference between dilute suspensions and concentrated or hindered suspensions is less apparent, so the zone settling phenomenon is usually considered for the flocculating materials. When the particles reach the vicinity of the bottom of the settling tank, a more concentrated suspension zone will be formed and the settling particles will tend to act in concert and reduce the overall rate of subsidence.

It can clearly be seen that in a water treatment plant, particularly if coagulation is applied to remove turbidity, all three types of settling will occur and any settling tank which is designed must take into account all three types (Figure 5).

Filtration

As described earlier, it has been found even in the early Egyptian days that passing water through sand resulted in a reduction in suspended and colloidal matter, and resulted in a further clarification of the water. Water which is on occasion extremely turbid should, of course, first of all be treated by some coagulation or settling or combination of both. However, water which is normally not too turbid may be directly applied to filters or water which has previously been treated by sedimentation and/or coagulation may also be applied to filters to provide the final polishing and the production of clear, aesthetically acceptable water.



Combined Settling Pattern

FIGURE 5

The filtration process actually consists of three phenomena occurring simultaneously (Figure 6).

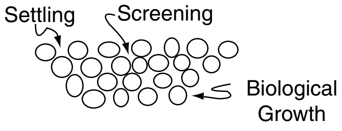
Settling takes place in the small settling basins which are provided between the particles. Screening takes place where particles which are larger than the interstices will be retained simply physically because they cannot pass through. And finally, a biological action takes place through bacterial growth which may occur on the particles of the filter which may occur on the particles of the filter which grow at the expense of the soluble organic carbon passing through in the water. This latter phenomenon is not a very satisfactory way of removing organic carbon, because it does tend to plug up the filter fairly rapidly and reduce its effectiveness.

Filters have been developed through the ages through a series of steps which are mainly related to their operating characteristics or the material which is used as a filtering medium.

Slow sand filter The slow sand filter is, as it suggests, a process whereby water is allowed to pass very slowly through the system at rates of 2.5 to 7.5 million gallons per acre per day.

Although this type of filter has been used traditionally and has been very effective in the past, it has certain operating disadvantages in that it cannot readily be cleaned. While some of these filters are still in use in some parts of the Orient, in Europe and North America, where labor tends to be more costly, other types of filters have been developed. When the difference in water level between the outlet and the water over the filter becomes too great, the filter is taken out of service and the top inch or two of sand is removed from the bed and may or may not be replaced with fresh sand before the filter is put back into operation (Figure 7).

Rapid sand filter A far more popular and common process for the filtration of water is the rapid sand filter. Instead of sitting on a sand bed of approximately three feet, as is the case in the slow sand filter, the bed is twelve to thirty inches thick and supported on a layer of gravel or other coarse grain, heavy material six to eighteen inches thick. Filtration rates on the rapid sand filter are of the order of three to four gallons per square foot per minute. Occasionally



Filtration Phenomena

FIGURE 6

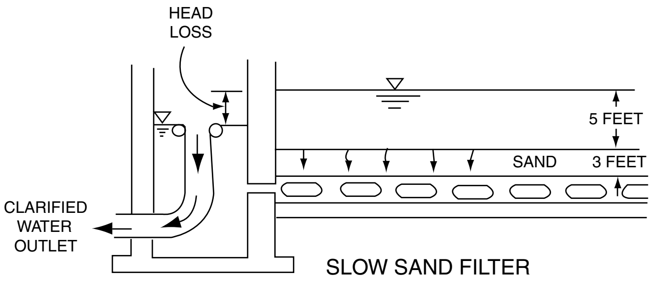


FIGURE 7

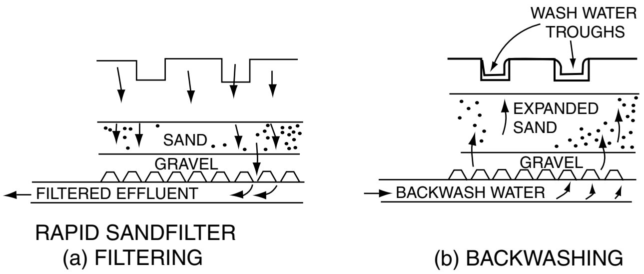


FIGURE 8

a plant is designed to operate at two gallons per square foot per minute, but provided for an overload when necessary (Figure 8).

The cleaning of the rapid sand filter, instead of throwing the filter out of service, is accomplished by simply backwashing. This is accomplished by passing clean water backwards through the filter at a high velocity. This velocity should not be greater than the terminal settling velocity of the smallest particle of sand which is in the filter which is not to be washed over the side. Through this mechanism the sand bed is expanded and the sand is lifted and floated while the particles rub mechanically against one another and wash off the foreign material. The dirty water is washed away in drains. After this has been conducted for a few moments, the filter is allowed to go back into service and the head loss is now smaller so the rate of flow through the filter is increased once more.

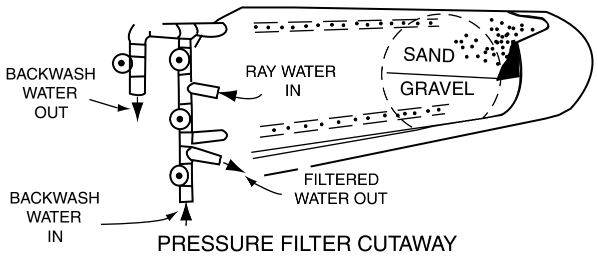


FIGURE 9

Pressure filters Whereas the rapid sand filter is indeed a gravity filter, a pressure filter is somewhat the same type of system only pressure is applied to the water to pass it through the filter. The most common household unit nowadays would be the swimming pool filter, where the water is pumped vertically through the sand and the filter, and when the head loss through the filter becomes excessive as registered on the pressure gauge, the operator will reverse the flow through the filter, accomplishing the backwash described above (Figure 9).

Diatomaceous earth filter Diatomaceous earth is the silicious residue of the bodies of diatoms which were deposited in past geological ages and now form extensive beds where they are mined. The earth is processed and ground, and the silica particles are extremely irregularly shaped and thus provide a very good porous coating. The diatomaceous earth filter was developed by the army for field use to remove certain chlorine-resistant organisms responsible for dysentery.

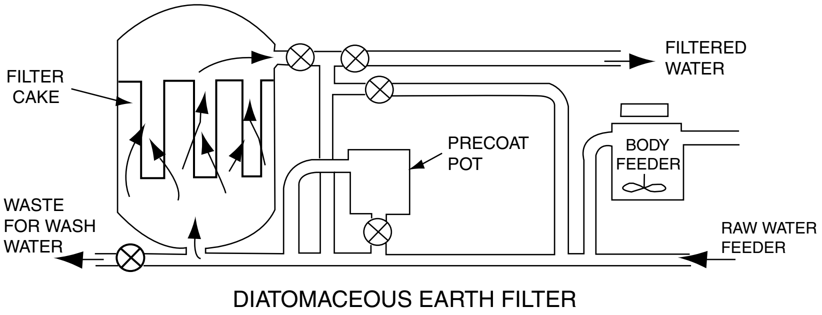


FIGURE 10

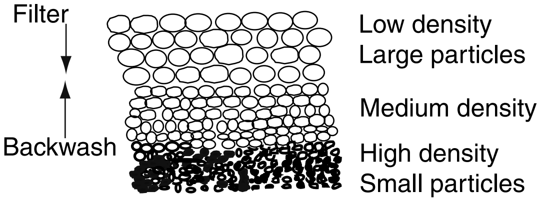


FIGURE 11

The filter medium is supported on a fine metal screen or a porous material. There are three steps in the filtration cycle. There are three steps in the filtration cycle. First of all, the deposit of a pre-coat, which is a thin layer of diatomite deposited on the filter element. The second step is the actual filtration and the body feed addition. The reason why body feed is continually added to the filter is to reduce the amount of clogging that occurs at the surface. This also permits significantly longer filter runs. The third step, when the pressure drops or the filtration rate reaches such a low very thin film over or under the source of irradiation. Commercial equipment is currently being developed for the individual water supply of the small household or institution, and is gaining some acceptance in some quarters. The irradiation of water by ultra-violet light of suitable wave-lengths for a proper period of time will kill bacteria, spores, molds, and viruses and in fact all microorganisms. The bactericidal wave-lengths extend from about 2000 to 2950 Å (angstrom units) with a maximum effect around 2540 Å.

CHEMICAL TREATMENT

The unit operations of chemical coagulation, precipitation, ion exchange and stabilization all produce change in the chemical quality of the water. Some of these are aimed at the removal of the suspended and colloidal substances, others are aimed at the removal of dissolved substances. Finally, some chemicals are simply added for their own sake, but these will not be discussed in this section.

To understand some of the basic chemistry of the treatment processes, it is first of all essential to understand a phenomenon known as chemical equilibrium and reaction velocities. An analogy might be considered as the physical equilibrium between ice and water.

$$\text{Ice} \underset{\text{Add Heat}}{\overset{\text{Remove heat}}{\leftrightharpoons}} \text{Water}$$

If the ice-water system is maintained at 0°C, then molecules of water are transferred from the solid to the liquid state and back again at the same rate. The addition of heat or the removal of heat from the system will result in the equilibrium moving in one direction or the other. The same principles might be applied to what is known as ionic equilibrium, which, like molecular equilibria, are subject to a shift under given stresses.

As an example, we might consider pure water

$$H_2O \rightleftharpoons H^+ + OH^-.$$

Certain stresses will give rise to an increase in hydrogen ion concentration (H^+). The expression of this shift is a reduction in pH, whereas an increase in the OH^- concentration brings about an increase of pH. One of the most important equilibria which exists in natural waters is the relationship between carbon dioxide and carbonate ion, which is shown in the following four equilibrium expressions.

$$CO_2\ (\text{gas}) \rightleftharpoons CO_2\ (\text{solution}) \quad (1)$$

$$CO_2\ (\text{solution}) + H_2O \rightleftharpoons H_2CO_3 \quad (2)$$

$$H_2CO_2 \rightleftharpoons H^+ + HCO_3^- \quad (3)$$

$$HCO_3^- \rightleftharpoons H^+ + CO_3^-. \quad (4)$$

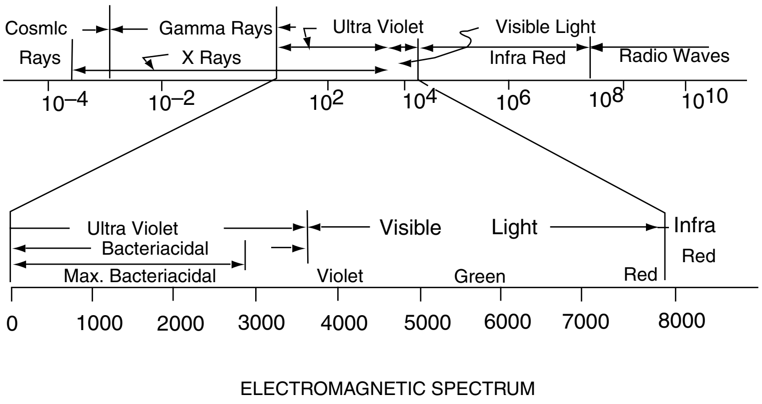


FIGURE 12

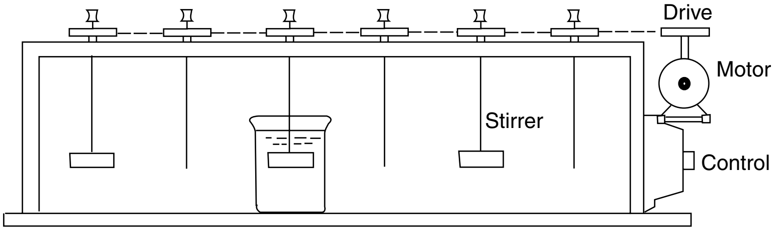


FIGURE 13 Jar test equipment—coagulant dosage varied in each jar to determine optimum concentration.

The equilibrium of the first of these equations is purely physical, since the solubility of gas and water is determined by the pressure of that gas and the temperature and a number of other physical parameters.

Coagulation

The principle function of chemical coagulation is known as destabilization, aggregation, and binding together of colloids. Alum, or aluminum sulphate, ($Al_2(SO_4)_3 \cdot 18H_2O$) is one of the most common coagulants which may be added to a water system. Such a coagulant possesses tiny positive charges and therefore has the ability to link together with negatively charged color or turbidity particles by mutual coagulation. Alum also reacts with the natural alkalinity (carbonate-bicarbonate system) of the water to produce a precipitate which is usually thought to be aluminum hydroxide. If the reaction takes place with natural alkalinity, it may be expressed as follows:

$$\begin{gathered}Al_2(SO_4)_3 \cdot XH_2O\ 3Ca(HCO_{32}) \rightarrow 2Al(OH)_3 + 3CaSO_4 \\ + XH_2O + 6CO_2.\end{gathered}$$

In the event that there is insufficient natural alkalinity for this to occur, then calcium oxide (lime) may be added to create the same effect. Because this system is very poorly understood, the optimum dosage required in practice has to be done by trial and error through a series of tests known as jar tests.

In these jar tests, the flash mixing and flocculation steps described previously are stimulated at various concentrations of alum and the clarification which takes place and the reduction of turbidity and the rate at which the floc settles are all observed in order to determine the optimum dosage of coagulant. If too much coagulant is added, then the colloidal system which is primarily negatively charged will become supersaturated by the aluminum system which is primarily positively charged and the suspension will become restabilized and this can be observed by conducting jar tests over a wide range of concentrations of coagulant.

The reason why alum is so generally used is that it is highly effective over a wide pH range in waters of vastly different chemical make-up. Other materials such as ferrous sulphate are occasionally used to increase the settling rate of plankton and thus increase the time of the filter run, making the filter process more efficient.

Precipitation

There are two important processes which are associated with precipitation in the treatment of water. One is the reduction of hardness (calcium and magnesium) and the other is the reduction of iron and manganese.

Water Softening The lime-soda-ash process involves the addition of $Ca(OH)_2$ and Na_2CO_3 to water. The reactions which occur are as follows:

$$Ca(HCO_3)_2 + \underset{\text{Lime}}{Ca(OH)_2} \rightarrow \underline{2CaCO_3^+} + 2H_2O \tag{1}$$

$$Mg(HCO_3)_2 + Ca(OH)_2 \rightarrow MgCO_3 + \underline{CaCO_3} + 2H_2O \tag{2}$$

$$MgCO_3 + Ca(OH)_2 \rightarrow \underline{Mg(OH)_2} + \underline{CaCO_3} \tag{3}$$

$$CaSO_4 + \underset{\text{Soda Ash}}{Na_2CO_3} \rightarrow \underline{CaCO_3} + Na_2SO_4. \tag{4}$$

In this reaction it can be seen that the lime is added to precipitate the carbonate hardness, while the soda ash provides the carbonate ion to precipitate the non-carbonate hardness.

Precipitation of Iron and Manganese Normally, iron and manganese are only highly soluble if they are in their ferrous (Fe^{2+}) and manganous (Mn^{2+}) forms. Normally, these two metals will only occur in this form if there is an absence of dissolved oxygen. However, on occasions when the water is particularly acid, such as might occur in mine drainage areas, the metals may remain in solution even though a very high dissolved oxygen is present. Under these circumstances, aeration is frequently sufficient to drive off the surplus carbon dioxide, increase the pH and bring about a natural precipitation of these materials in their ferric and manganic form. In order to catalyze or accelerate this reaction, the water is frequently caused to trickle over coke or crushed stone, or to flow upward through some contact material. This allows deposits of iron and manganese to accumulate on the surfaces and catalyze the further precipitation of ferric and manganic oxides.

If the pH of the system is forced to values higher than 7.1, the positively charged ferric hydroxide particles may be adsorbed on the negatively charged calcium carbonate particles and a stable colloidal suspension may result. Iron and manganese are objectionable constituents of water supplies because they impart a brown colour to laundry goods and frequently will stain household plumbing fittings.

Precipitation of iron and manganese can also be satisfactorily accomplished by using the lime-soda-ash process as described above for softening.

Ion Exchange

Ion exchange units are most frequently used for softening waters, but are also used by certain industries for the production of de-ionized water. This is quite common in the brewery industry, where an attempt is made to strip the water down to its most pure constituents so that water in one part of the world is similar to water in other parts of the world. Following de-ionization, breweries and often distilleries will reconstitute the water so that the water used for the production of a certain type of beer will be the same all over the continent and not have the variations which were characteristic of beers when native waters were used for their production.

The chemistry of the ion exchange process is shown below, where a cation resin which will exchange the sodium (Na^+) for the calcium and magnesium (Ca^{2+}, Mg^{2+}). When the resin is saturated with calcium and magnesium, a regeneration is required such as is used in household water softening units, when a very strong brine solution is forced back through the resin and in turn displaces the calcium and magnesium into the backwash line and restores the sodium on the resin for further softening.

$$\text{Softening } Na^2 + \begin{matrix} Ca \\ Mg \end{matrix} \left[\begin{matrix} (HCO_3)_2 \\ SO_4 \\ Cl_2 \end{matrix} \quad \rightarrow \quad \begin{matrix} Ca \\ \\ Mg \end{matrix} \right] \qquad (1)$$

$$R + \left[\begin{matrix} 2NaHSO_4 \\ Na_2SO_4 \\ 2NaCl \end{matrix} \right.$$

$$\text{Regeneration } \left. \begin{matrix} Ca \\ Mg \end{matrix} \right] R + 2NaCl \rightarrow Na_2R + \left. \begin{matrix} Ca \\ Mg \end{matrix} \right] Cl_2 \qquad (2)$$

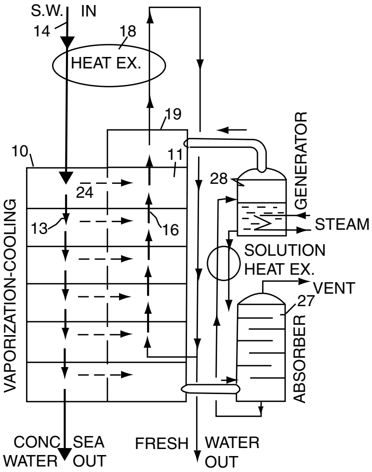


FIGURE 14 Conventional multistage flash evaporation – MSF evaporating and cooling of hot feed brine (vertical arrows down) on left side at successively lower pressures after heating to highest temperature in prime heater (PH) at top; vapors (horizontal arrows) from MSF, passing to preheat the sea water by condensation-heating on right side; fresh water condensate passing stagewise from top to discharge at bottom; additional sea water coolant (dotted line) rejecting heat in lower stages, withdrawing of vapors from prime heater to be condensed in half-stage (dashed lines) increasing the production of fresh water.

Desalination

Although the principles of desalination were fully known in Julius Caesar's time, the energy requirements of this process are presently so high that these will be usually considered as a last resort after all other water sources have been explored. Water quality is frequently referred to as fresh, brackish, sea water or brine. Fresh water normally contains less than 1000 mg/liter of dissolved salts, while brackish water ranges from 1000–35,000 mg/liter of dissolved salts. Sea water contains 35,000 mg/liter of dissolved salts, whereas brine contains very much more from salt water by a semi-permeable membrane, the fresh water will tend to flow into the salt water to equalize the concentration of salts on both sides of the membrane. Bearing in mind that the membrane will not allow the salts to pass back, it is clear that a certain pressure which is known as the osmotic pressure is forcing the fresh water through to the brine side of the membrane. If a force greater than this osmotic pressure is applied on the sea water side, then fresh water will flow backwards through the semi-permeable membrane at a rate proportional to the incremental pressure over the osmotic pressure. In practice, quite high pressures are required in order to get a useful volume of water to pass through the membrane—such pressures as 40–100 kg per square centimeter. This has been shown to work for waters of fairly high dissolved solids, but the structural properties of the membranes must be fairly well developed and of course the membranes must be very well supported. Membranes used for this type of process are frequently cellulose acetate or some derivatives thereof. The power requirement for this process is considerably less than electrodialysis, but it is a

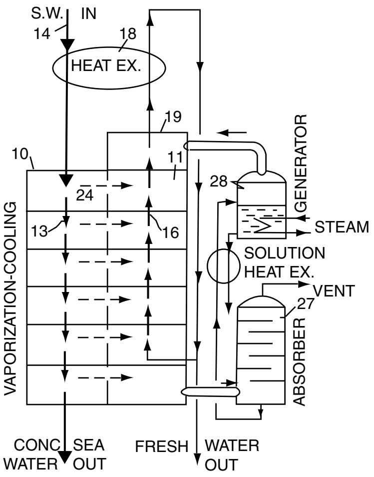


FIGURE 15 Vapor reheat with vapor recompression by absorption (from US Patent 3,288,686). Vapor reheat MSF, vapors formed in low pressure stage pass to Absorber, 27, and are absorbed in the hydrophilic liquid from solution heat exchanger, which cycles through heat exchanger to evaporator or generator, 28, for concentration at higher pressure. Vapors leave at higher pressure to half-stage, 19, to supply prime heat to evaporator.

process which offers considerable promise for the desalination of brackish waters.

The Vacuum Freezing Process Cooled saline water passes into a low pressure chamber where flash evaporation brings about further cooling. Close to freezing ice crystals will form with a brine coating, and washing of these crystals will yield fresh water from the washed ice. Small plant application appears to be feasible at the moment. Considerable development work on this was done by the Technion Institute in Israel.

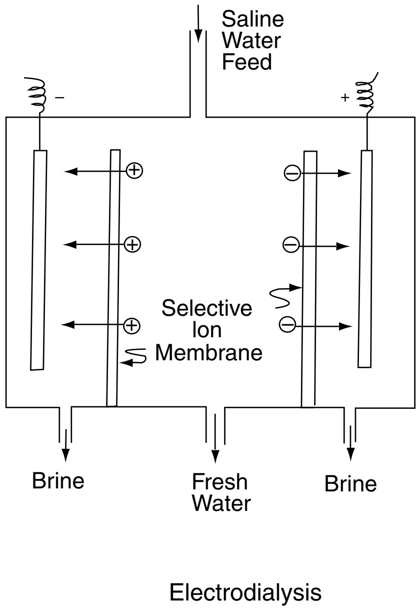


FIGURE 16

A Secondary Refrigerant Process This involves the use of a second hydrocarbon refrigerant, such as butane, that will not mix with water. A great deal of care must be exercised

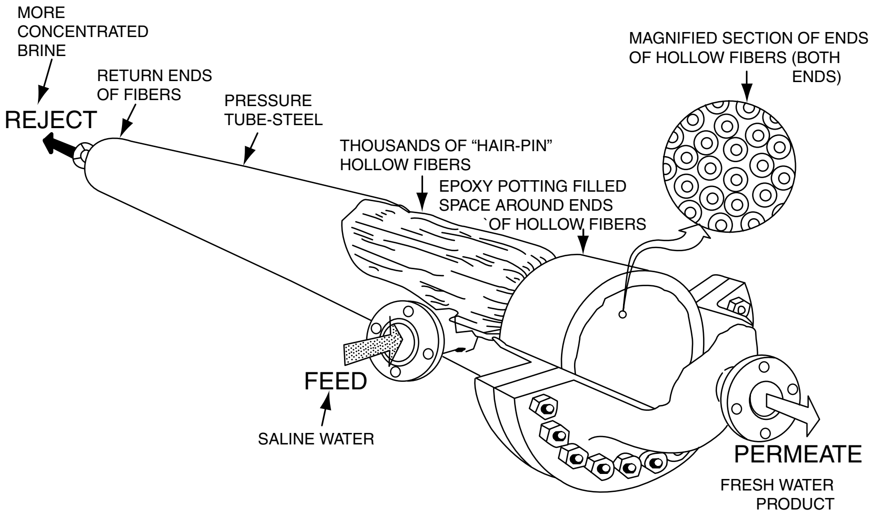


FIGURE 17

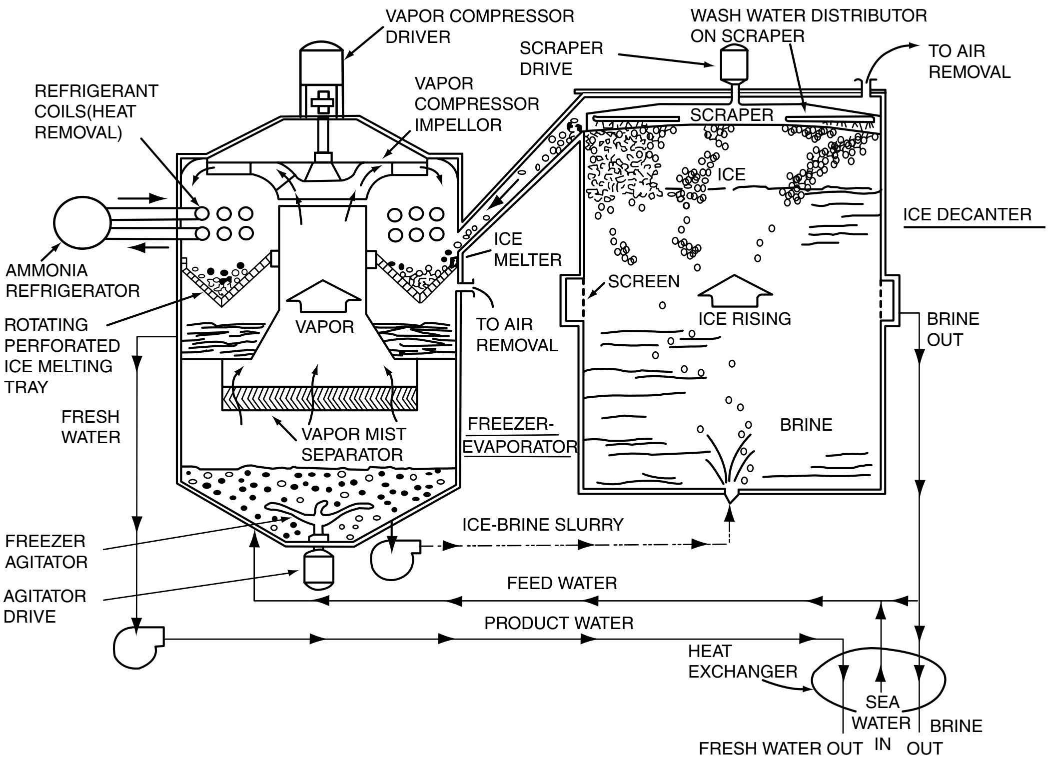


FIGURE 18 Vacuum freezing process

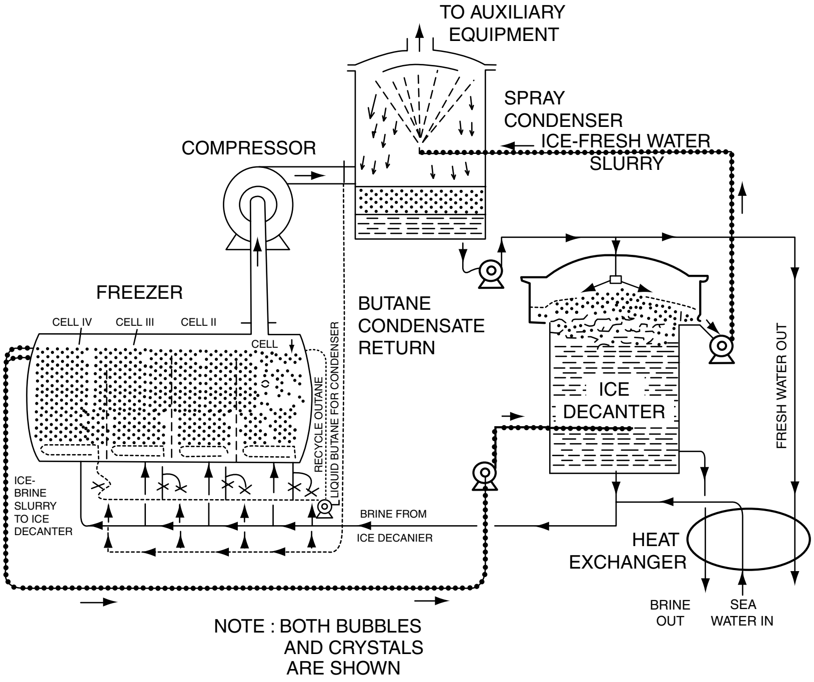


FIGURE 19 Secondary refrigerant freezing process

here to ensure that the product is stripped of butane; otherwise serious explosion hazards exist. It has a low energy input and has minimized corrosion and scale-forming properties. However, so far this process has not been examined on a very large-scale basis.

Other Processes There are a number of other processes which are being considered for desalination, such as solar distillation, but so far this has been restricted to use on a very small scale such as survival kits. Ion exchange, such as was described previously, has some potential here but it is very much limited down to approximately 3000 mg/liter solids. A hydrate process has been considered where propane is added to form a hydrate and react with the water, leaving the salt behind. Then the propane hydrate is decomposed to recover the propane and the water; this one is rather difficult to handle.

Disinfection

As mentioned earlier, water has long since been identified as a means of distributing pathogenic organisms among society. The purpose therefore of disinfecting water supplies is to prevent the spread of water-borne disease by destroying pathogenic organisms. Most of the physical and chemical treatment processes described previously will remove most of the micro-organisms to some extent. However, very small numbers of microorganisms which are viable and pathogenic are all that are required to bring about disastrous epidemics. Thus, disinfection is considered to be a necessary final step before treated water is delivered to a municipal system. This may not be the case in certain industrial supplies.

A physical process for disinfection was previously described using ultra-violet irradiation. Other forms of chemical disinfectant are the halogens such as chlorine, bromine, iodine, and the powerful, unstable oxidant, ozone. In North America chlorination is the most common of the disinfectant processes used, for two reasons. Firstly, it is fairly simple to handle, can be manufactured inexpensively in bulk and delivered to the site, can be applied under fairly controlled conditions, and can maintain a measurable residual in the water supply to indicate safety at all points on a water distribution system. The first attempt at continuous chlorination of a public water supply was made in England during 1904, and subsequently in 1908 in Jersey City, New Jersey, USA. There are certain disadvantages of chlorination, in that a high residual chlorine will bring about a taste which is unacceptable to many people; and chlorine furthermore will react with certain micro-constituents of water, such as phenols, to bring about substantial odors (chlorophenols) quite out of proportion to the concentration of the causative chemicals. The addition of chlorine to water releases a group of substances, all of which have some disinfecting properties. The substances so released are:

1) hypochlorite ion(OCl);
2) hypochlorous acid (HOCl);
3) monochloramine (NH_2Cl);
4) dichloramine ($NHCl_2$);
5) nitrogen trichloride (NCl_3);
6) organic compounds containing chloride; and
7) chlorine dioxide (ClO_2).

Hypochlorite ion and hypochlorous acid are known collectively as free available chlorine residuals. The following substances are known as chloramines: NH_2Cl, $NHCl_2$, NCl_3, and organic chlorine compounds. The chloramines are brought about by the reaction of hypochlorous acid with ammonia

$$NH_3 + HOCl \rightarrow NH_2Cl + H_2O$$
$$NH_2Cl + HOCl \rightarrow NHCl_2 + H_2O$$
$$NHCl_2 + HOCl \rightarrow NCl_3 + H_2O.$$

The process which brings about the various chloramines are shown above. Chlorination is applied in a series of different forms as follow:

Superchlorination This process represents the addition of very high concentrations of chlorine which are intended to oxidize not only the pathogenic and potential pathogenic microorganisms in the system, but also to oxidize those organic compounds which might bring about taste and odor. Following superchlorination, a step involving dechlorination which involves the addition of sulphur dioxide, sodium bisulphite, or sodium sulphite or some similar reducing agent. The bisulphite is frequently used in practice because it is cheaper and more stable.

If there is any amount of ammonia naturally present in the water, a strange phenomenon will occur such as shown above in the graph. On the initial part of the graph, labelled 1 to 2, the ratio (molar) of chlorine to ammonia is less than one and the residual chlorine is essentially all monochloramine. In the next section, between 2 and 3, the oxidation of ammonia and reduction of chlorine continue until the complete oxidation reduction occurs at point 3. At this point once again, all the residual chlorine is in the

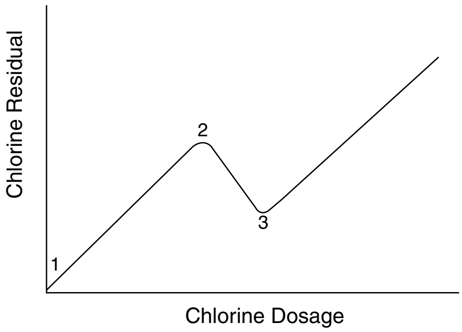


FIGURE 20

form of monochloramine. Beyond point 3, all chlorine added remains in solution. This phenomenon is known as break-point chlorination, and in order to ensure satisfactory disinfecting properties chlorination must go beyond the break-point 3 on the curve.

In certain instances, such as when phenols are present in the water, small concentrations of free available chlorine will combine with the phenols, forming chlorphenols which produce a distinctive taste and odor at very, very small concentrations. When this occurs, chlorine dioxide is frequently used as a disinfecting agent as this does not react with the phenols but in fact destroys them.

Other substances which are used for disinfecting purposes are the other halogens such as bromine and iodine, although these are not commonly used in water supplies. Occasionally they have been used in swimming pools for similar purposes.

Ozone is a particularly effective disinfectant, but it has certain disadvantages in that it must be manufactured on the site, using fairly sizeable and expensive capital equipment. No residual can be maintained due to the instability of the substance, and the methods of detection are rather imperfect. However to compensate for this, the disinfecting properties of ozone are considerably greater than chlorine.

VOC REMOVAL

Treatment of Volatile Organic Compounds Found in Groundwater Sources

With the advent of technological advances in testing of water supplies and concerns regarding possible contamination of groundwater sources, many water supply systems have focused on the treatment of Volatile Organic Compounds (VOC's). VOC's are man-made chemicals, some of which have been shown to be carcinogenic. VOC's are generally found in industrialized settings where substances such as cleaning fluids, degreasers or solvents have been disposed of improperly.

The treatment of VOC's utilizing conventional water treatment techniques involving flocculation, sedimentation and filtration are relatively ineffective at reducing VOC concentrations. VOC's may be treated by either packed tower aeration (air stripping) or granular activated carbon (GAC) absorption. Details of the two treatment techniques are as follows:

Packed Tower Aeration (Air Stripping)

Aeration is the process where air and water are brought into contact for the purposes of transferring volatile substances from water to air. This process is commonly referred to as air stripping. Air stripping basically involves the transfer of dissolved gas molecules from the liquid phase to the gas phase. There are two major factors which determine the removal efficiency of various volatile compounds by air stripping; 1) the ratio of concentration of VOC's in the gaseous phase to the concentration of VOC's in the aqueous phase at equilibrium, and 2) the rate at which equilibrium is obtained.

Numerous types of aeration devices have been used where air stripping can occur. Some of these alternatives involve diffused aeration, spray aeration and water fall aeration.

In packed towers or stripping towers, water flows downward by gravity and air is forced upward. The tower is filled with various forms of packing material which serves to continuously disturb the liquid flow, creating and improving the air-to-water interface. Packed towers typically have void volumes in excess of 90 percent which allows for a large liquid-air interface and minimizes the pressure drop through the column, an operating cost consideration. Packed towers, which are currently in service, have provided VOC removals in the 95–99.9 percent range. A schematic of a typical airstripping facility is shown on Figure A.

There are three major design factors controlling the mass transfer of VOC's from water to air.

Packing depth—is the primary factor influencing removal efficiency. Increasing the packing depth will increase the removal efficiency of the tower.

Tower diameter—controls the liquid loading rate as measured in gallons per minute per square foot, (GPM/sq. ft.). The lower the liquid loading rate, the greater the removal efficiency due to the increased air-to-water interaction zone.

Air-to-water ratio—is the most influential parameter with respect to removal efficiency. Generally, the removal efficiency increases as the air-to-water ratio is increased.

In stripping towers, packing materials are used to provide high void volumes together with high surface area. The water flows downward by gravity and air is forced upward. The raw, untreated water is evenly distributed on the top of the packing with either spray or distribution trays and the air is forced through the tower by either blowers or induced draft fans.

Many options exist for packings involving a variety of shapes and materials. Packings are available in plastic, metal and ceramic. Plastics are best suited for water treatment because of their durability and low cost.

Since the mass transfer of VOC's is basically accomplished by passing significant quantities of air through a fixed quantity of water, the air-to-water ratio can be varied by either, i) increasing the diameter of the column, or ii) increasing the air blower capacity. Hence, an optimum balance of tower diameter and blower size must be evaluated. Given a specific water loading rate and a packing selection, the air-to-water ratio determines the height of the stripping tower required to provide the specified removals.

Various liquid loading rates are evaluated to optimize the tower diameter versus air pressure drops. Once the tower diameter is determined, a cost analysis comparing capital and operating cost is determined. A matrix of air-to-water ratios and depth of packing is then developed to determine the optimum design.

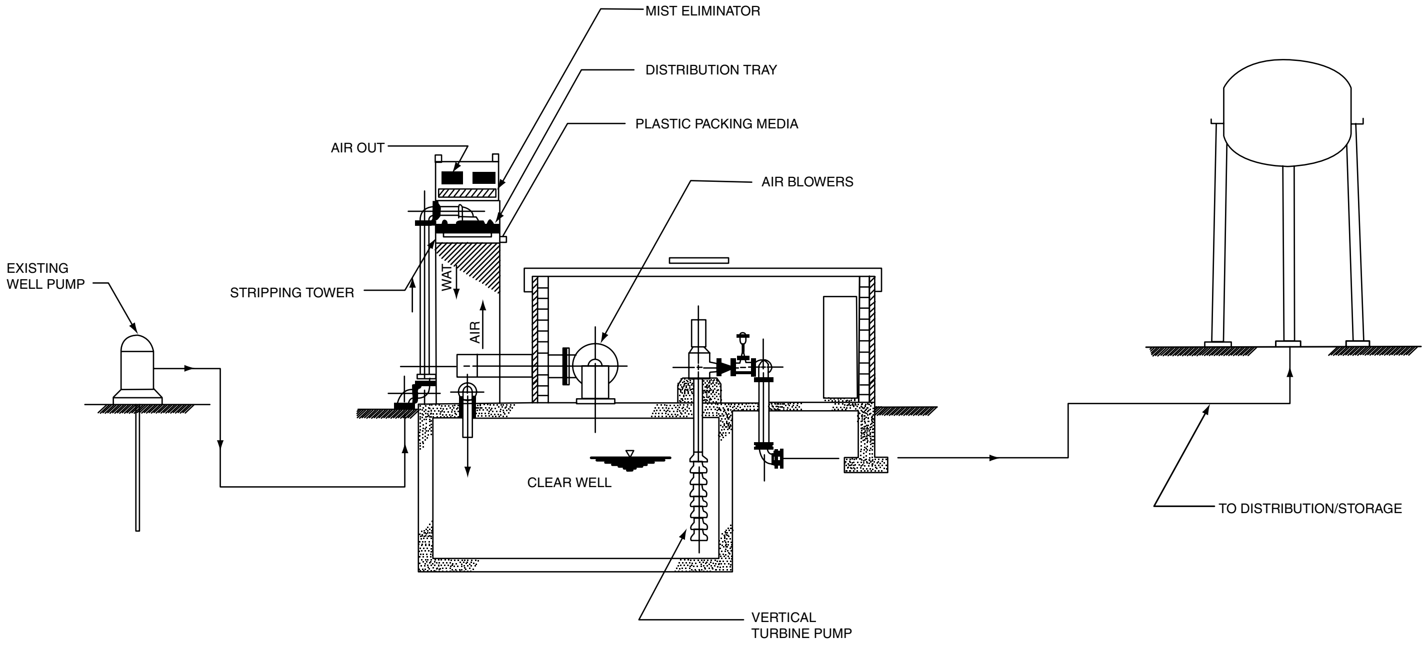

MIST ELIMINATOR
DISTRIBUTION TRAY
PLASTIC PACKING MEDIA
AIR OUT
AIR BLOWERS
EXISTING
WELL PUMP
STRIPPING TOWER
WAT
AIR
CLEAR WELL
TO DISTRIBUTION/STORAGE
VERTICAL
TURBINE PUMP

FIGURE A

Granular Activated Carbon (GAC)

The removal of VOC's through adsorption involves passing the contaminated water through a medium of adsorbent, such as activated carbon, where the VOC's will adhere (stick) to its surface. Adsorbates which could possibly be used to remove VOC's from groundwater include granular activated carbon (GAC) and powdered activated carbon (PAC). GAC exhibits a wide range of effectiveness in adsorbing various compounds and generally tends to adsorb high-molecular weight compounds more readily than low-molecular weight substances such as VOC's. However, GAC is currently the best available adsorbent for the removal of VOC's.

Powdered activated carbon has been used traditionally for the removal of trace organics associated with causing taste and odors in drinking water. PAC typically requires coagulation and sedimentation facilities to be effective and is not normally used for groundwater treatment.

GAC has a spectrum of effectiveness like aeration; however, the process is more complicated and water quality can have an influence on performance. The adsorption of VOC's can be affected by the amount of background organic carbon, generally measured as total organic carbon. High background organic content can result in lower adsorption capacity.

GAC contractors also require regeneration or replacement when the material becomes saturated with contaminants. The life of the GAC is dependent upon the concentrations of the contaminants present, the flow rate through the media and the required effluent concentrations. GAC contractors have a reported removal efficiency of 99%. A schematic of a typical GAC facility is shown in Figure B.

A combination of aeration-adsorption can also be a highly effective method of reducing VOC levels to very low concentrations. This combination is quite attractive when several different types of contaminants are present. However, the corresponding cost of treatment increases dramatically.

VOC Treatment with Granular Activated Carbon (GAC)

GAC treatment can employ either a gravity or pressure system. The gravity disrepair is generally used in surface water treatment plants and operates in a manner similar to a gravity sand filter.

In groundwater treatment systems, a pressure disrepair (contactor) is generally used and involves a pressurized vessel which can accommodate flow rates at high pressures and allow direct discharge to an existing distribution system.

Numerous GAC contactors are currently in use for the removal of VOCs and a significant amount of data is available on this form treatment.

GAC contactors would involve vertical steel pressure vessels which would allow the raw water to enter the top of the vessel and pass downward through the carbon bed.

The treated water is collected at the bottom of the vessel utilizing a header-lateral arrangement or a bottom plate with nozzles. The collected water would then be disinfected and discharged to the distribution system for consumption.

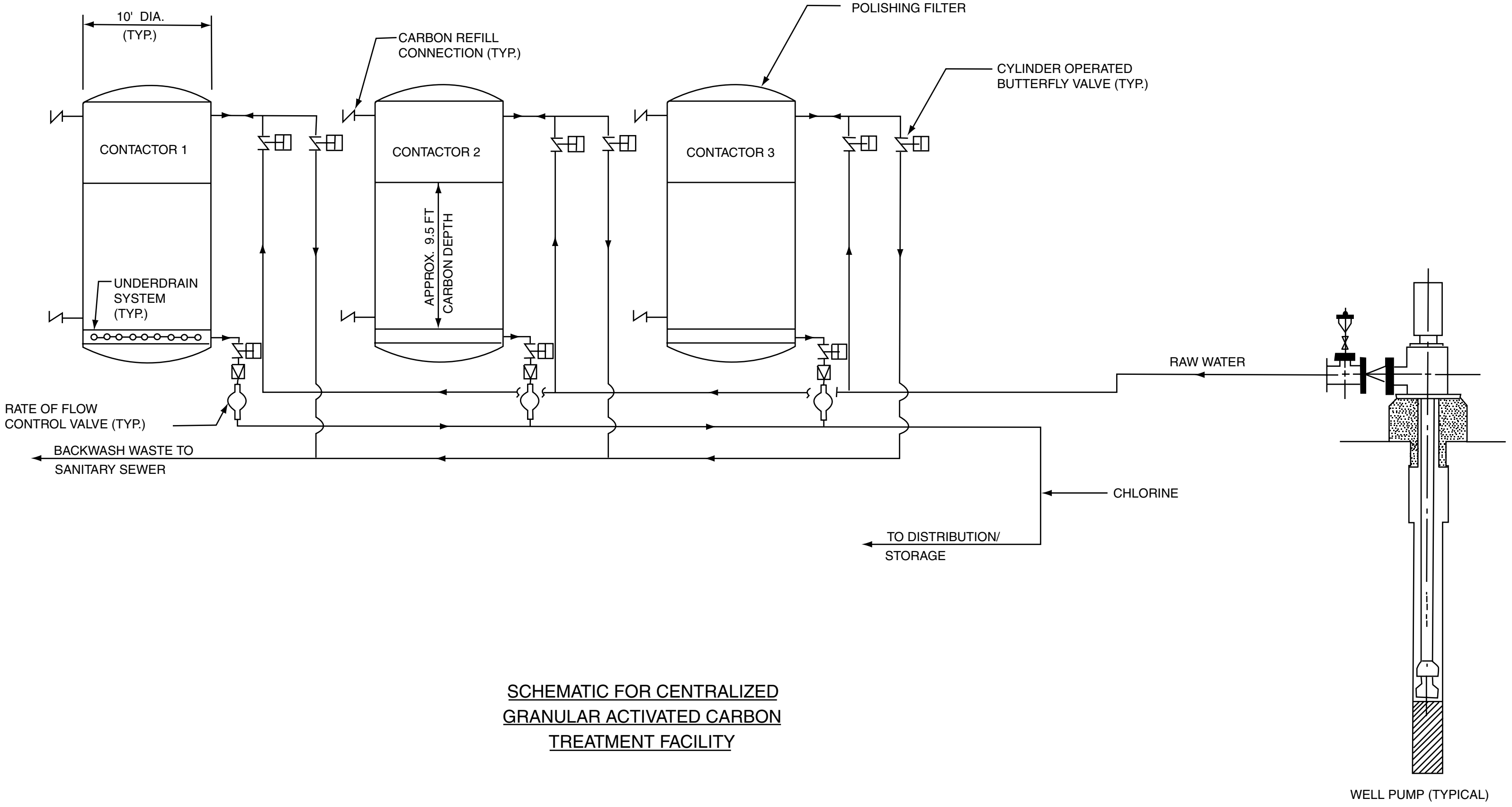
Since the contactor is essentially a filter, the vessel would be equipped with backwashing facilities. The carbon filters would not require frequent backwashing. The contactor backwash waste would be disposed of by discharging to the nearby holding pond and subsequently to the sanitary sewer system. Connections would also be provided to readily remove the spent carbon and to readily install the new material.

As a general rule, aeration is most effective with low-molecular weight, highly volatile substances, while adsorption works best with high-molecular weight compounds with a low solubility. The selection of the treatment alternative is based on many factors such as the contaminant(s), concentrations of contaminants, groundwater quality, site constraints, pumping system configuration as well as other factors.

Air stripping facilities, by their nature, require the existing pumping facilities to be modified as well as the need to install a second pumping system. Existing pumping systems must be modified to produce less head (pressure) in order to direct water to the stripping tower. The tower will dissipate the energy provided by the well pump as the water passes through the tower and into the clearwell below. From the clearwell, the water must be repumped to the water system for use. As such, the economic feasibility of an air-stripping facility must account not only for the capital and operating expenses of the stripping facility, but must also account for modifications to existing well pump(s) and the costs associated with repumping the water supply for use.

Unlike an air-stripping facility, a pressurized carbon contactor does not utilize mechanical equipment as part of the treatment process. Most often, existing pumping facilities may remain unchanged if a small head loss (which would result in a slightly reduced flowrate) can be tolerated. After water is pumped through the contactor, it is discharged directly to the distribution system for use. As such, GAC facilities have a major advantage over air strippings due to ease of operation and the ability to discharge directly to existing systems without repumping.

There is a move toward privatization of public water (and wastewater) operations. There are advantages and disadvantages associated with this. The Contractor assumes responsibility for operating results. If a plant has a staff that is too large or inadequately trained, a for-profit operation can be expected to introduce greater efficiency in operation. At times, public office holders have introduced privatization to show a better municipal financial picture than is the actual case. The Contractor may not practice proper maintenance or may try to operate with a staff that is too small. This may be reflected in poor operating results and the public agency charged with oversight can be expected to take action if the public health is threatened. Each case is unique and each decision to privatize must be evaluated taking into account all pertinent factors.



SCHEMATIC FOR CENTRALIZED GRANULAR ACTIVATED CARBON TREATMENT FACILITY

NOTE: CONCEPTUAL SCHEMATIC ARRANGEMENT SHOWN IS FOR PARALLEL OPERATION. ADDITIONAL VALVING WOULD BE REQUIRED FOR SERIES OPERATION.

FIGURE B

REFERENCES

1. Durant, Will, 1954, *Our Oriental Heritage*, Simon and Schuster Inc., New York, p. 132.
2. Baker, M.N., 1949, The quest for pure water, American Water Works Association, New York, pp. 13, 6, 11.
3. Brown, J.R. and D.M. McLean, 1967, Water-borne diseases: An historical review, *Medical Services Journal Canada*, **23**, no. 8, pp. 1011–1026.
4. Estimates of the water and sewerage industries and utilities division, Jan. 1956, Business Service Bulletin 136, Business and Defense Services Administration, US Department of Commerce.
5. Ackerman, E.A. and G.O. Lof, 1959, *Technology in American Water Development*, Johns Hopkins Press, Baltimore.
6. Clark. J.W. and Warren Viessman, 1965, *Water Supply and Pollution Control*, International Textbook Company, Scranton.
7. Fair, G.M. and J.C. Geyer, 1954, *Water Supply and Wastewater Disposal*, Wiley, New York.
8. Steel, E.W., 1960, *Water Supply and Sewerage*, 4th Ed., McGraw-Hill Book Company Inc., New York.
9. Michaelides, G. and R. Young, 1986, Provisions in Design and Maintenance to Protect Water Quality from Roof Catchments, *Advances in Environmental Science and Engineering*, Vol. 5, Gordon and Breach, New York.
10. Glaze, W.H., 1987, Drinking-water Treatment with ozone, *Environ. Sci. Tech.*, **21**.
11. Hoigne, J. and H. Bader, 1988, The Formation of Trichlormethane (Chloropicrin) and Chloroform in a Combined Ozonation/Chlorination of a Drinking Water, *Water Res.*, **22**.
11. Kogelschatz, M., 1988, *Advanced Ozone Generation*, in Process Technologies for Water Treatment, Plenum, New York.
12. White, G.C., 1989, *Handbook of Chlorination*, 2nd Ed., Van Nostrand Reinhold, New York.

PHILIP H. JONES (DECEASED)
Griffith University

MARK A. TOMPECK
Hatch Mott MacDonald

WATER SUPPLY AND SANITATION: see COMMUNITY HEALTH

WATER TRANSPORT: see HYDROLOGY; WATER FLOW

APPENDIX

SELECTED PROPERTIES

TABLE 1
Conversion factors

To Convert From	To	Multiply by
Angstrom units	Centimeters	1×10^{-8}
Angstrom units	Microns	1×10^{-4}
Atmospheres	Centimeters of mercury at 0°C	76
Atmospheres	Dynes per square centimeter	1.01325×10^{6}
Atmospheres	Feet of water at 39.1°F	33.899
Atmospheres	Grams per square centimeter	1,033.3
Atmospheres	Inches of mercury at 32°F	29.921
Atmospheres	Millimeters of mercury	760
Atmospheres	Pounds per square inch	14.696
Atmospheres	Bars	1.0133
Barrels, oil	Gallons (US)	42
Boiler horsepower	Btu per hour	33,479
Btu	Calories (gram)	252
Btu	Foot-pounds	777.98
Btu	Joules (absolute)	1,054.8
Btu per minute	Foot-pounds per second	12.96
Calories, gram	Btu	3.968×10^{-3}
Calories, gram	Foot-pounds	3.087
Calories, gram	Joules	4.185
Candles (International)	Lumens (International) per steradian	1.000
Candlepower (spherical)	Lumens	12.566
Candles per square inch	Lamberts	0.48695
Centimeters	Angstrom units	1×10^{8}
Centimeters	Feet (British or US)	0.032808
Centimeters	Inches (US)	0.393700
Centimeters of mercury at 0°C	Atmospheres	0.013158
Centimeters of mercury at 0°C	Dynes per square centimeter	1.33322×10^{4}
Centimeters of mercury at 0°C	Kilograms per square meter	135.95
Centimeters of mercury at 0°C	Pounds per square foot	27.845
Centimeters per second	Feet per minute	1.9685
Centimeters per second	Miles per hour	0.02237
Centipoises	Pound mass per foot-second	0.672×10^{-3}
Centipoises	Kilogram mass per meter-second	1×10^{-3}

(*continued*)

TABLE 1
Conversion factors (*continued*)

To Convert From	To	Multiply by
Cubic feet (US)	Gallons (US)	7.481
Cubic feet (US)	Liters	28.316
Dynes per square centimeter	Millimeters of mercury	7.5006×10^{-4}
Dynes per square centimeter	Pounds per square foot	0.0020886
Dynes per square centimeter	Pounds per square inch	1.4504×10^{-5}
Ergs	Btu (mean)	9.4805×10^{-11}
Ergs	Calories, gram (mean)	2.3889×10^{-8}
Ergs	Joules	1×10^{-7}
Ergs per square centimeter	Dynes per centimeter	1
Feet (US)	Centimeters	30.48006096
Feet per minute	Centimeters per second	0.508001
Feet per minute	Meters per second	0.00508001
Feet per minute	Miles per hour	0.011364
Foot-candles	Lumens incident per square foot	1
Foot-candles	Lumens per square meter	10.764
Gallons (US)	Barrels (US, liquids)	0.031746
Gallons (US)	Liters	3.78533
Gallons (US)	Cubic feet	0.13368
Grains	Grams	0.064798918
Grains	Milligrams	64.798918
Grains	Pounds (avoirdupois)	1/7000
Grains per cubic foot	Milligrams per cubic meter	2.288×10^{3}
Grains per gallon (US)	Parts per million	17.118
Grams	Grins	15.4324
Gravity	Centimeters per square second	980.665
Gravity	Feet per square second	32.174
Kilometers	Centimeters	1×10^{5}
Kilometers	Feet	3280.83
Kilometers	Miles (US)	0.6213699495
Liters	Cubic feet	0.035316
Liters	Gallons (US)	0.26417762
Liters	Ounces (US, fluid)	33.8143
Liters	Quarts (US, liquid)	1.056681869
Lumens per square foot	Foot-candles	1
Meters	Feet (US)	3.280833333
Meters	Inches (US)	39.3700
Micrograms per cubic meter	Grains per cubic foot	4.370×10^{-7}
Microns	Inches	3.937×10^{-5}
Microns	Centimeters	1×10^{-4}
Miles (US, statute)	Feet	5280
Miles (US, statute)	Yards	1760
Miles per hour	Centimeters per second	44.7041
Miles per hour	Feet per minute	88
Mile per hour	Feet per second	1.4667
Miles per hour	Meters per minute	26.82
Millibars	Atmospheres	98.692
Millibars	Millimeters of mercury	1.329×10^{-5}
Millibars	Dynes per square centimeter	1.000×10^{9}
Milligrams	Grains	0.01543236

TABLE 1
Conversion factors (*continued*)

To Convert From	To	Multiply by
Milligrams	Ounces (avoirdupois)	3.52739×10^{-5}
Milligrams per cubic meter	Grains per cubic foot	0.4370×10^{-3}
Milliliters	Cubic centimeters	1.000027
Milliliters	Cubic inches	0.061025
Millimeters	Inches (US)	0.0393700
Millimeters of mercury at 0°C	Atmospheres	0.00131579
Millimeters or mercury at 0°C	Dynes per square centimeter	1333.22
Millimeters of mercury at 0°C	Grams per square centimeter	1.3595
Millimicrons or Micronmillimeters	Angstrom units	10
Millimicrons	Meters	1×10^{-9}
Ounces (avoirdupois)	Grams	28.349527
Ounces (avoirdupois)	Grains	437.5
Ounces (US, fluid)	Milliliters	29.5729
Parts per million	Per cent	1×10^{4}
Per cent	Parts per million	1×10^{-4}
Pints (US, dry)	Liters	0.550599
Pints (US, liquid)	Cubic inches	28.875
Pints (US, liquid)	Liters	0.473168
Pounds (avoirdupois)	Grains	7000
Pounds (avoirdupois)	Grams	453.5924277
Pounds (avoirdupois)	Tons (long)	4.464286×10^{-4}
Pounds (avoirdupois)	Tons (metric)	4.5359243×10^{-4}
Pounds (avoirdupois)	Tons (short)	5×10^{-4}
Pounds per hour	Grams per second	0.12599
Pounds per hour	Tons per stack day	1.2×10^{-2}
Quarts (US, dry)	Cubic centimeters	1101.23
Quarts (US, dry)	Cubic inches	67.2006
Quarts (US, dry)	Cubic feet	0.038889
Quarts (US, dry)	Liters	1.10120
Quarts (US, liquid)	Milliliters	946.358
Quarts (US, liquid)	Cubic inches	57.749
Quarts (US, liquid)	Liters	0.946333
Square centimeters	Square feet	0.0010764
Square centimeter	Square inch	0.15500
Square feet (US)	Square meters	0.09290341
Square feet (US)	Acres	2.29568×10^{-5}
Square inches (US)	Square yards	1/1296
Square kilometers	Square miles (US)	0.3861006
Square meters	Square yards (US)	1.195985
Square miles	Acres	640
Square miles	Square feet	2.78784×10^{7}
Square miles	Square yards	3.0976×10^{6}
Square miles	Square meters	2589.998
Square yards (US)	Square meters	0.83613
Tones per stack day	Pounds per hour	83.3
Yards (US)	Meters	0.91440183
Yards (US)	Miles	5.68182×10^{-4}

TABLE 2
Viscosity and density of water calculated from international critical tables, 1928 and 1929

Temperature, °C	Density ρ, γ (grams/cm^3), also *s*	Absolute viscosity μ, centipoises[a]	Kinematic viscosity γ, centistokes[b]
0	0.99987	1.7921	1.7923
1	0.99993	1.7320	1.7321
2	0.99997	1.6740	1.6741
3	0.99999	1.6193	1.6193
4	1.00000	1.5676	1.5676
5	0.99999	1.5188	1.5188
6	0.99997	1.4726	1.4726
7	0.99993	1.4288	1.4288
8	0.99988	1.3872	1.3874
9	0.99981	1.3476	1.3479
10	0.99973	1.3097	1.3101
11	0.99963	1.2735	1.2740
12	0.99952	1.2390	1.2396
13	0.99940	1.2061	1.2068
14	0.99927	1.1748	1.1756
15	0.99913	1.1447	1.1457
16	0.99897	1.1156	1.1168
17	0.99880	1.0876	1.0888
18	0.99862	1.0603	1.0618
19	0.99843	1.0340	1.0356
20	0.99823	1.0087	1.0105
21	0.99802	0.9843	0.9863
22	0.99780	0.9608	0.9629
23	0.99757	0.9380	0.9403
24	0.99733	0.9161	0.9186
25	0.99707	0.8949	0.8975
26	0.99681	0.8746	0.8774
27	0.99654	0.8551	0.8581
28	0.99626	0.8363	0.8394
29	0.99597	0.8181	0.8214
30	0.99568	0.8004	0.8039
31	0.99537	0.7834	0.7870
32	0.99505	0.7670	0.7708
33	0.99473	0.7511	0.7551
34	0.99440	0.7357	0.7398
35	0.99406	0.7208	0.7251
36	0.99371	0.7064	0.7109
37	0.99336	0.6925	0.6971
38	0.99299	0.6791	0.6839
39	0.99262	0.6661	0.6711

[a] 1 centipoise = 10^{-2} (grams mass)/(cm)(sec).

[b] 1 centistoke = 10^{-2} cm^2/sec.

TABLE 3A

Composition of some typical natural waters (after Baldwin, 1948)[1]

	Na	K	Ca	Mg	Cl	SO_4	CO_3	Total (g/liter)
Soft fresh water	0.016	—	0.010	—	0.019	0.007	0.012	0.065
Hard fresh water	0.021	0.016	0.065	0.014	0.041	0.025	0.119	0.30
Sea water	10.7	0.39	0.42	1.31	19.3	2.69	0.073	34.9

TABLE 3B

Relative abundance of ions (by weight) in natural waters

	Na	K	Ca	Mg	Cl	SO_4	CO_3
Soft fresh water	100	—	62.5	3.3	119.0	43.8	75.0
Hard fresh water	100	76.0	310.0	66.8	195.0	119.0	567.0
Sea water	100	3.6	3.9	12.1	181.0	20.9	0.7

TABLE 4

Saturation values of dissolved oxygen in fresh- and seawater exposed to an atmosphere containing 20.9% oxygen under a pressure of 760 mm of mercury

	Dissolved oxygen (mg/l) for stated concentrations of chloride, mg/l					
Temperature, °C	0	5000	10,000	15,000	20,000	Difference per 100 mg/l chloride
0	14.62	13.79	12.97	12.14	11.32	0.0165
1	14.23	13.41	12.61	11.82	11.03	0.0160
2	13.84	13.05	12.28	11.52	10.76	0.0154
3	13.48	12.72	11.98	11.24	10.50	0.0149
4	13.13	12.41	11.69	10.97	10.25	0.0144
5	12.80	12.09	11.39	10.70	10.01	0.0140
6	12.48	11.79	11.12	10.45	9.78	0.0135
7	12.17	11.51	10.85	10.21	9.57	0.0130
8	11.87	11.24	10.61	9.98	9.36	0.0125
9	11.59	10.97	10.36	9.76	9.17	0.0121
10	11.33	10.73	10.13	9.55	8.98	0.0118
11	11.08	10.48	9.92	9.35	8.80	0.0114
12	10.83	10.28	9.72	9.17	8.62	0.0110
13	10.60	10.05	9.52	8.98	8.46	0.0107
14	10.37	9.85	9.32	8.80	8.30	0.0104
15	10.15	9.65	9.14	8.63	8.14	0.0100
16	9.95	9.46	8.96	8.47	7.99	0.0098
17	9.74	9.26	8.78	8.30	7.84	0.0095
18	9.54	9.07	8.62	8.15	7.70	0.0092
19	9.35	8.89	8.45	8.00	7.56	0.0089
20	9.17	8.73	8.30	7.86	7.42	0.0088

(*continued*)

TABLE 4 (*continued*)

Temperature, °C	Dissolved oxygen (mg/l) for stated concentrations of chloride, mg/l 0	5000	10,000	15,000	20,000	Difference per 100 mg/l chloride
21	8.99	8.57	8.14	7.71	7.28	0.0086
22	8.83	8.42	7.99	7.57	7.14	0.0084
23	8.68	8.27	7.85	7.43	7.00	0.0083
24	8.53	8.12	7.71	7.30	6.87	0.0083
25	8.38	7.96	7.56	7.15	6.74	0.0082
26	8.22	7.81	7.42	7.02	6.61	0.0080
27	8.07	7.67	7.28	6.88	6.49	0.0079
28	7.92	7.53	7.14	6.75	6.37	0.0078
29	7.77	7.39	7.00	6.62	6.25	0.0076
30	7.63	7.25	6.86	6.49	6.13	0.0075

Calculated by G. C. Whipple and M. C. Whipple from measurements of C. J. J. Fox. For other barometric pressures the solubilities vary approximately in proportion to the ratios of these pressures to the standard pressures.

TABLE 5
Standard atmosphere

Altitude (m)	Temperature (°C)	Pressure (mb)	Density (kg/m^3)
0	15.0	1013.2	1.2250
500	11.8	954.6	1.1673
1,000	8.5	898.8	1.1117
1,500	5.2	845.6	1.0581
2,000	2.0	795.0	1.0066
2,500	−1.2	746.9	0.9569
3,000	−4.5	701.2	0.9092
3,500	−7.7	657.8	0.8634
4,000	−11.0	616.6	0.8194
4,500	−14.2	577.5	0.7770
5,000	−17.5	540.5	0.7364
5,500	−20.7	505.4	0.6975
6,000	−24.0	472.2	0.6601
6,500	−27.2	440.8	0.6243
7,000	−30.4	411.0	0.5900
7,500	−33.7	383.0	0.5572
8,000	−36.9	356.5	0.5258
8,500	−40.2	331.5	0.4958
9,000	−43.4	308.0	0.4671
9,500	−46.7	285.8	0.4397
10,000	−49.9	265.0	0.4140
10,500	−53.1	245.4	0.3886
11,000	−56.4	227.0	0.3648

TABLE 5

Standard atmosphere (*continued*)

Altitude (m)	Temperature (°C)	Pressure (mb)	Density (kg/m^3)
11,100	−56.5	223.5	0.3593
11,500	−56.5	209.8	0.3374
12,000	−56.5	194.0	0.3119
13,000	−56.5	165.8	0.2666
14,000	−56.5	141.7	0.2279
15,000	−56.5	121.1	0.1948
16,000	−56.5	103.5	0.1665
17,000	−56.5	88.5	0.1423
18,000	−56.5	75.6	0.1216
19,000	−56.5	64.7	0.1040
20,000	−56.5	55.3	0.0889
25,000	−51.6	25.5	0.0401
30,000	−46.6	12.0	0.0184
35,000	−36.6	5.6	0.0085
40,000	−22.8	2.9	0.0040
45,000	−9.0	1.5	0.0020
50,000	−2.5	0.8	0.0010
60,000	−17.4	0.225	0.000306
70,000	−53.4	0.055	0.000088
80,000	−92.5	0.010	0.000020
90,000	−92.5	0.002	0.000003

TABLE 6

Dynamic viscosity of gases in absolute CGS units

Temperature °C	Air	N_2	O_2	CO_2	CO	H_2	CH_4	C_2H_4	C_2H_6	Water vapour
−160	78.1	77.1	87.9	—	74.4	46.0	44.9	—	—	—
−140	92.0	90.0	103	—	87.4	51.5	52.8	—	—	—
−120	105	103	117	—	99.9	56.8	60.5	—	—	—
−100	117	114	131	—	112	61.8	68.1	62	—	—
−80	129	126	144	—	124	66.5	75.1	68	—	—
−60	140	136	157	108	135	71.1	82.5	74	—	—
−40	151	147	169	118	145	75.6	89.3	81	—	—
−20	162	157	181	127	156	79.9	95.9	88	—	—
0	172	166	192	137	166	84.1	102	94	86	—
20	181	175	203	146	175	88.2	109	101	92	95.6
40	191	184	213	156	185	92.2	115	108	98	103
60	200	193	223	165	194	96.1	121	115	104	110
80	209	201	233	173	202	100.0	127	122	109	117
100	217	209	243	182	211	104	133	128	115	125
150	238	229	266	203	231	113	147	142	129	143
200	257	247	288	222	251	122	160	156	142	161
250	275	265	309	241	269	130	173	169	155	179
300	293	282	329	259	287	139	184	181	166	197
350	309	298	348	276	304	147	196	192	176	215

(*continued*)

TABLE 6 (*continued*)

Temperature °C	Air	N_2	O_2	CO_2	CO	H_2	CH_4	C_2H_4	C_2H_6	Water vapour
400	325	313	367	393	320	154	207	202	184	233
450	340	328	385	309	336	162	217	212	195	251
500	355	342	402	324	352	169	227	222	204	269
550	369	355	419	339	367	177	237			287
600	383	369	436	354	382	184	246		230[a]	306
650	396	381	452	368	396	191	256			327
700	409	393	468	382	410	198	265		249[a]	348
800	433	417	500	408	437	211	283		269[a]	387
900	457	440	530	434	464	223	300		283[a]	424
1000	479	461	559	459	490	235	316		299[a]	456
Sutherland[a] Equation factors K = $10^{-7} \times C$										
	150	143	176	158	135	66	—	—	—	—
	124	110	131	240	102	72	—	—	—	—

Data from *Technical Data on Fuels,* ed. Spiers[2].

[a] See *Fluid Mechanics,* R.L. Daugherty and A.C. Ingersoll, McGraw-Hill, 1954, p. 8.[3]

TABLE 7

Kinematic viscosity of gases in stokes

μ/κ = Kinematic Viscosity—Stokes

Temperature °C	Air	N_2	O_2	CO_2	CO	H_2	CH_4	C_2H_4	C_2H_6	Water vapour
−160	0.0249	0.0252	0.0251	—	—	0.212	—	—	—	—
−140	0.0346	0.0348	0.0348	—	—	0.279	—	—	—	—
−120	0.0455	0.0458	0.0457	—	—	0.354	0.0466	—	—	—
−100	0.0575	0.0577	0.0579	—	—	0.436	0.0599	—	—	—
−80	0.0707	0.0709	0.0712	—	—	0.523	0.0738	0.0374	—	—
−60	0.0849	0.0849	0.0855	—	0.0839	0.618	0.0894	0.0456	—	—
−40	0.100	0.100	0.101	0.0506	0.0991	0.718	0.106	0.0545	—	—
−20	0.116	0.116	0.117	0.0596	0.115	0.824	0.124	0.0644	—	—
0	0.133	0.133	0.134	0.0693	0.133	0.936	0.143	0.0749	0.0635	—
20	0.151	0.151	0.152	0.0796	0.151	1.05	0.163	0.0860	0.0730	5.53*
40	0.170	0.169	0.171	0.0905	0.169	1.18	0.184	0.0981	0.0833	2.01*
60	0.190	0.188	0.191	0.102	0.189	1.30	0.206	0.111	0.0941	0.845*
80	0.210	0.208	0.211	0.114	0.209	1.44	0.230	0.125	0.105	0.400*
100	0.231	0.229	0.233	0.126	0.231	1.58	0.254	0.140	0.117	0.208
150	0.287	0.283	0.289	0.160	0.287	195	0.318	0.177	0.148	0.272
200	0.340	0.343	0.349	0.196	0.348	2.35	0.388	0.216	0.184	0.344
250	0.410	0.406	0.414	0.235	0.413	2.78	0.462	0.258	0.221	0.424
300	0.478	0.474	0.483	0.277	0.482	3.23	0.540	0.302	0.260	0.512
350	0.549	0.545	0.557	0.321	0.555	3.72	0.622	0.350	0.299	0.608
400	0.623	0.618	0.633	0.367	0.632	4.23	0.711	0.399	0.342	0.713
450	0.701	0.695	0.714	0.416	0.713	4.77	0.801	0.449	0.385	0.825
500	0.782	0.774	0.798	0.467	0.797	5.33	0.898	0.502	0.430	0.946
550	0.865	0.856	0.886	0.521	0.885	5.92	0.999	—	—	1.07
600	0.951	0.942	0.977	0.576	0.976	6.53	1.100	—	—	1.22

Data from *Technical Data on Fuels,* ed. Spiers[4].

* Estimated (see *Fluid Mechanics,* R.L. Daugherty and A.C. Ingersoll, McGraw-Hill, 1954, pp. 9–12).[5]

TABLE 8
Densities of gases at standard conditions (0°C, 1 atm)

Gas	Formula	Mol. wt.	Density G./l.	Lb./cu. ft
Acetylene	C_2H_2	26.02	1.1708	0.0732
Air	—	—	1.2928	0.0808
Ammonia	NH_3	17.03	0.7708	0.0482
Argon	A	39.91	1.7828	0.1114
Bromine	Br_2	159.83	7.1388	0.4460
Butane	C_4H_{10}	58.08	2.5985	0.1623
Carbon dioxide	CO_2	44.0	1.9768	0.1235
Carbon monoxide	CO	28.00	1.2501	0.0781
Carbon oxychloride	$COCl_2$	98.91	4.5313	0.2830
Carbon oxysulfide	COS	60.06	2.7201	0.1700
Chlorine	Cl_2	70.91	3.2204	0.2011
Chlorine monoxide	Cl_2O	86.91	3.8874	0.2428
Cyanogen	C_2N_2	52.02	2.3348	0.1459
Ethane	C_2H_6	30.05	1.3567	0.0848
Ethyl chloride	C_2H_5Cl	64.50	2.8700	0.1793
Ethylene	C_2H_4	28.03	1.2644	0.0783
Fluorine	F_2	38.00	1.6354	0.1022
Helium	He	4.00	0.1769	0.0111
Hydrogen	H_2	2.016	0.0898	0.0056
Hydrogen chloride	HCl	36.47	1.6394	0.1024
Hydrogen fluoride	HF	20.01	0.9218	0.0576
Hydrogen iodide	HI	127.94	5.7245	0.3576
Hydrogen selenide	H_2Se	81.22	3.6134	0.2258
Hydrogen sulfide	H_2S	34.08	1.5392	0.0961
Hydrogen telluride	H_2Te	129.52	5.8034	0.3625
Krypton	Kr	82.90	3.6431	0.2275
Methane	CH_4	16.03	0.7167	0.0448
Methyl chloride	CH_3Cl	50.48	2.3044	0.1440
Neon	Ne	20.40	0.8713	0.0544
Nitric oxide	NO	30.01	1.3401	0.0837
Nitrogen	N_2	28.02	1.2507	0.0782
Nitrous oxide	N_2O	44.02	1.9781	0.1235
Nitrosyl chloride	NOCl	65.47	2.9864	0.1865
Oxygen	O_2	32.00	1.4289	0.0892
Phosphine	PH_4	34.05	1.5293	0.0955
Silicon fluoride	SiF_2	104.06	4.6541	0.2907
Sulfur dioxide	SO_2	64.06	2.9268	0.1828
Xenon	Xa	130.20	5.7168	0.3570

TABLE 9

C_p/C_V: Ratios of specific heats of gases at 1 atm pressure

Compound	Formula	Temperature,°C	Ratio of specific heats (γ) = C_p/C_v	Compound	Formula	Temperature, °C	Ratio of specific heats (γ) = C_p/C_v
Acetaldehyde	C_2H_4O	30	1.14	Hydrogen (*contd.*)			
Acetic acid	$C_2H_4O_2$	136	1.15	cyanide	HCN	65	1.31
Acetylene	C_2H_2	15	1.26	—	—	140	1.28
		−71	1.31	—	—	210	1.24
Air	—	925	1.36	iodide	HI	20–200	1.40
		17	1.403	sulfide	H_2S	15	1.32
		−78	1.408	—	—	−45	1.30
		−118	1.415	—	—	−57	1.29
Ammonia	NH_3	15	1.310	Iodine	I_2	185	1.30
Argon	A	15	1.668	Isobutane	C_4H_{10}	15	1.11
		−180	1.76	Krypton	Kr	19	1.68
		0–100	1.67	Mercury	Hg	360	1.67
Benzene	C_6H_6	90	1.10	Methane	CH_4	600	1.113
Bromine	Br_2	20–350	1.32	—	—	300	1.16
Carbon dioxide	CO_2	15	1.304	—	—	15	1.31
		−75	1N.37	—	—	−80	1.34
disulfide	CS_2	100	1.21	—	—	−115	1.41
monoxide	CO	15	1.404	Methylacetate	$C_4H_6O_2$	15	1.14
		−180	1.41	alcohol	CH_4O	77	1.203
Chlorine	Cl_2	15	1.355	ether	C_2H_6O	6–30	1.11
Chloroform	$CHCl_3$	100	1.15	Methylal	$C_3H_8O_2$	13	1.06
Cyanogen	$(CH)_2$	15	1.256	—		40	1.09
Cyclohezane	C_6H_{12}	80	1.08	Neon	Ne	19	1.64
Dichlorodifluoromethane	CCl_2F_2	25	1.139	Nitric oxide	NO	15	1.400
Ethane	C_2H_6	100	1.19	—	—	−45	1.39
		15	1.22	—	—	−80	1.38
		−82	1.28	Nitrogen	N_2	15	1.404
Ethyl alcohol	C_2H_6O	90	1.13	—	—	−181	1.47
ether	$C_4H_{10}O$	35	1.08	Nitrous oxide	N_2O	100	1.28
		80	1.086	—	—	15	1.303
Ethylene	C_2H_4	100	1.18	—	—	−30	1.31
		15	1.255	—	—	−70	1.34
		−91	1.35	Oxygen	O_2	15	1.401
Helium	He	−180	1.660	—	—	−76	1.415
Hexane(n-)	C_6H_{14}	80	1.08	—	—	−181	1.45
Hydrogen	H_2	15	1.410	Pentane(n-)	C_5H_{12}	86	1.086
		−76	1.453	Phosphorus	P	300	1.17
		−181	1.597	Potassium	K	850	1.77
bromide	HBr	20	1.42	Sodium	Na	750–920	1.68
chloride	HCl	15	1.41	Sulfur dioxide	SO_2	15	1.29
		100	1.40	Xenon	Xe	19	1.66

TABLE 10

Fuel-oil grades[a]

Grade of fuel oil[b]	Flash point, °F	Pour point, °F	Water and sediment, % by vol.	Carbon residue on 10% bottoms	Ash, % by wt	Distillation temperatures, °F: 10% point	Distillation temperatures, °F: 90% point	Distillation temperatures, °F: End point	Sayboltiscosity, sec: Universal at 100°F	Sayboltiscosity, sec: Universal at 100°F	Sayboltiscosity, sec: Furol at 122°F	Sayboltiscosity, sec: Furol at 122°F	Kinematic viscosity, centistokes: At 100°F	Kinematic viscosity, centistokes: At 100°F	Kinematic viscosity, centistokes: At 122°F	Kinematic viscosity, centistokes: At 122°F	Gravity, API	Corrosion at 122°F (50°C)[d]
	Min	Max	Max	Max	Max	Max	Max	Max	Max	Min	Max	Min	Max	Min	Max	Min		
No. 1: A distillate oil intended for vaporizing pot-type burners and other burners requiring this grade of fuel	100 or legal	0	Trace	0.15	—	420	—	625	—	—	—	—	2.2	1.4	—	—	35	Pass
No. 2: A distillate oil for general-purpose domestic heating for use in burners not requiring No. 1 fuel oil	100 or legal	20[c]	0.10	0.35	—	e	675	—	40	—	—	—	(4.3)	—	—	—	26	—
No. 4 An oil for burner installations not equipped with preheating facilities	130 or legal	20	0.50	—	0.10	—	—	—	125	45	—	—	(26.4)	(5.8)	—	—	—	—
No. 5: A residual-type oil for burner installations equipped with preheating facilities	130 or legal	—	1.00	—	0.10	—	—	—	—	150	40	—	—	(32.1)	(81)	—	—	—
No. 6: An oil for use in burners equipped with preheaters permitting a high-viscosity fuel	150	—	2.00[f]	—	—	—	—	—	—	—	300	45	—	—	(638)	(92)	—	—

---

[a] Recognizing the necessity for low-sulfur fuel oils used in connection with heat treatment, nonferrous metal, glass ceramic furnaces, and other special uses, a sulfur requirement may be specified as follows:

Grade of fuel oil	*Sulfur, (Max %)*
No. 1	0.5
No. 2	21.0
No. 4	No limit
No. 5	No limit
No. 6	No limit

Other sulfur limits may be specified only by mutual agreement between the purchaser and the seller.

[b] It is the intent of these classifications that failure to meet any requirement of a given grade does not automatically place an oil in the next lower grade unless in fact it meets all requirements of the lower grade.

[c] Lower or higher pour points may be specified whenever required by conditions of storage or use. However, these specifications shall not require a pour point lower than 0°F under any conditions.

[d] Report as passing when the copper test strip shows no gray or black deposit.

[e] The 10% point may be specified at 440°F maximum for use in other than atomizing burners.

[f] The amount of water by distillation plus the sediment by extraction shall not exceed 2.0%. The amount of sediment by extraction shall not exceed 0.50%. A deduction in quantity shall be made for all water and sediment in excess of 1.0%.

After *ASTM Standards on Petroleum Products and Lubricants,* November, 1953, p. 184.[7]

TABLE 11

The Nordwijk Declaration on Climate Change (Nov. 6, 1989)

(1) The composition of the earth's atmosphere is being seriously altered at an unprecedented rate due to human activity. Based on our current understanding, society is being threatened by man-made changes to the global climate.

(2) While there are still uncertainties regarding the magnitude, timing and regional effects of climate change due to human activity, there is a growing consensus in the scientific community that significant climate change and instability are most likely over the next century.

Predictions available today indicate potentially severe economic and social dislocations for future generations.

Assuming these predictions, delay in action may endanger the future of the planet as we know it.

(3) Fortunately, there is a growing awareness among the world population and their political leaders that action is needed. The basic principle of ecologically sustainable development has gained wide currency following the report of the World Commission on Environment and Development. This principle should be fundamental to efforts to tackle the problem of climate change and atmospheric pollution. The protection of the ozone layer is being addressed by the 1985 Vienna Convention on the Protection of the Ozone Layer and the 1987 Montreal Protocol on Substances that Deplete the Ozone layer. Further strengthening of control measures contained in the Protocol was called for all the London Conference on Saving the Ozone Layer in March 1989 and the first meeting of the parties to the Montreal Protocol at Helsinki in May 1989. A decision will be taken by the second meeting of the contracting parties to be held in London in 1990. The process aims at phasing out the production and consumption of chlorofluorocarbons (CFCs) controlled under the Montreal Protocol by the year 2000 by the developed countries. The should also phase out other controlled substances which deplete the ozone layer as soon as feasible. Developing countries should also phase out these substances as soon as possible after their technology and resource needs are met.

(4) Global warming is being addressed by the Intergovernmental Panel on Climate Change (IPCC), which was established by UNEP and WMO, and recognized by UN General Assembly Resolution 43/53 on Protection of global climate for present and future generations of mankind. The Hague Declaration of March 1989 put forward challenging ideas for international co-operation, and legal and institutional measures. The 15th session of the UNEP Governing Council and the XLI session of the WMO Executive Council in 1989 requested their executive heads to begin preparations for negotiations on a framework convention on climate; these negotiations should be initiated as soon as possible after the interim report of the IPCC is adopted. This interim report will be reviewed at the Second World Climate Conference in November 1990. The 1989 Economic Summit agreed that a framework convention on climate change setting out general principles was urgently required and that specific protocols containing concrete commitments could be fitted into the framework as scientific evidence requires and permits. The Economic Summit also strongly advocated common efforts to limit emissions of carbon dioxide and other greenhouse gases. The July 1988 declaration of the states, parties to the Warsaw Treaty, and the meeting of non-aligned countries in Belgrade in September 1989 also addressed the issue of climate change. The Tokyo Conference on Global Environment and Human Response Towards Sustainable Development was held in September 1989. The Langkawi Declaration on Environment issued by the Commonwealth Heads of Governments in October 1989 stated the need to take new action to address the serious deterioration in the environment, including climate change. Given this base it is now time for governments of all countries to commit themselves to the IPCC, to strengthen and to extend the process of addressing climate change.

(5) Measures to limit climate change will have other significant benefits such as reducing acidification, protecting the ozone layer, preserving biodiversity and other natural resources, preventing mean sea-level change and promoting sustainable development.

(6) The Conference recognizes the principle of the sovereign right of states to manage their natural resources independently. The Conference also reaffirms that global environmental problems have to be approached through international co-operation. Solving the external debt problem of developing countries, and establishing fair economic and commercial relationships between industrialized and developing countries would assist developing countries in creating appropriate conditions to protect the environment.

(7) Climate change is a common concern of mankind. All countries should now, according to their capabilities and the means at their disposal, initiate actions and develop and maintain effective and operational strategies to control, limit or reduce emissions of greenhouse gases. As a first step, they should take those actions which are beneficial in their own right. Industrialized countries, in view of their contribution to the increase of greenhouse gas concentrations, and in view of their capabilities, have specific responsibilities of different kinds: i) they should set an example by initiating domestic action, ii) they should support, financially and otherwise, the action by countries to which the protection of the atmosphere and adjustment to climate change would prove to be an excessive burden and iii) they should reduce emissions of greenhouse gases, also taking into account the need of the developing countries to have sustainable development. Developing countries establishing industrial facilities for the first time have a unique opportunity to include up-to-date technologies for controlling the emissions of greenhouse gases.

(8) For the long term safeguarding of our planet and maintaining its ecological balance, joint effort and action should aim at limiting or reducing emissions and increasing sinks for greenhouse gases to a level consistent with the natural capacity of the planet. Such a level should be reached within a time frame sufficient to allow ecosystems to adapt naturally to climate change, to ensure that food production is not threatened and permit economic activity to develop in a sustainable and environmentally sound manner. Stabilizing the atmospheric concentrations of greenhouse gases is an imperative goal. The IPCC will need to report on the best scientific knowledge as to the options for containing climate change within tolerable limits. Some currently available estimates indicate that this could require a reduction of global anthropogenic greenhouse gas emissions by more than 50 per cent. These estimates should be the subject of further examination by the IPCC.

(9) While striving to preserve the global environment, it is important to work at the same time to ensure stable development of the world economy, in line with the concept of "sustainable development." Effort and action should include: i) the phasing-out of CFCs controlled by the Montreal Protocol, which are responsible for about one fifth of projected global warming, by national action and international co-operation in the context of the Montreal Protocol. This includes financial assistance and transfer of technology and information. In this connection, it is important that the substitutes for CFCs also should not contribute significantly to the global warming problem, ii) action especially by industrialized countries to limit or reduce CO_2-emissions, iii) action to reduce deforestation, prevent soil erosion and desertification; increase afforestation, and sound forest management in the temperate as well as the tropical zones, iv) action to limit or reduce the emissions of all greenhouse gases other than CO_2 and their precursors and to increase the sinks for such substances and v) intensified efforts for technological breakthrough, for example with regards to renewable energy and removal and re-utilization of CO_2.

(10) The conference recommends that appropriate fora, including the IPCC, consider the necessity and efficiency of the introduction of the concept of CO_2-equivalence. This would provide a single parameter to describe the radiative effects of the various greenhouse gases, including CFCs. Such a concept, after taking into account other environmental considerations, creates a basis for negotiations in response measures for different greenhouse gases in the most cost-effective manner. The Conference further recommends the development of common definitions and the harmonization of methods to calculate CO_2-emissions.

(11) All countries should increase co-operation in developing new, environmentally sound technologies, to improve existing technologies and increasingly to use these technologies in order to limit climate change or adapt to it. Maximum use should be made of existing international organizations, institutions and mechanisms, governmental and non-governmental, for technology co-operation with and transfer to interested countries, especially developing countries. Factors that impede effective transfer of appropriate technologies should be identified and measures implemented to overcome these impediments.

(12) Progress in reducing atmospheric pollution depends not only on technical and economic issues but also on attitudinal and conceptual changes. All countries, especially industrialized countries, should recognize the need to make their socio-economic activities and life-styles environmentally sound. Improved dissemination of information and better training of personnel is needed, both at the national and international level. Public awareness programmes, including school curricula, should include the issue of climate change and its connection with the way individuals use energy and other natural resources. Wider public awareness can be supported by increased scientific evidence arising from systematic research and monitoring activities. The Conference calls upon the non-governmental organizations to participate, in co-operation with international, regional and national authorities, in the efforts that are needed to respond to the problems of global warming, more specifically in the field of education and awareness building.

(13) Many countries, especially developing countries will require assistance in identifying the causes of anthropogenic climate change, in establishing its extent and effect and also in responding to it. They will need help in acquiring, using, developing and maintaining technologies that are appropriate to their industrial, energy, transport, forestry and agricultural infrastructure. Industrialized countries will take steps to facilitate the transfer to developing countries of technologies to limit the global climate change through financial assistance and other mechanisms to overcome the incremental costs of acquiring and using these technologies. Furthermore, the capabilities of these countries should be increased so that they can develop appropriate technologies themselves. In this context the concept of assured access to appropriate technologies in relation to proprietary rights needs to be explored.

Given this stage of development of the issue of climate change, the Conference more specifically:

CARBON DIOXIDE (CO_2)

—(14) *Urges* all countries to take steps individually and collectively, to promote better energy conservation and efficiency and the use of environmentally sound energy sources, practices and technologies with no or minimum environmentally damaging characteristics. These policies should be reflected in short and long term energy policies and be pursued by all relevant sectors, including industry and transport, taking into account the need of developing countries for an adaptation period in order to enable them to meet their technological and other developmental needs. One direct means of allowing markets to incorporate the risk of climate change could be to ensure that the prices of all fuels reflect their full social, long run marginal and environmental costs and benefits.

—(15) *Agrees* that it is timely to investigate quantitative emission targets to limit or reduce CO_2 emissions and encourages the IPCC, in their interim report due in 1990, to include an analysis of target options.

—(16) *Recognizes* the need to stabilize, while ensuring stable development of the world economy, CO_2 emissions and emissions of other greenhouse gases not controlled by the Montreal Protocol. Industrialized nations agree that such stabilization should be achieved by them as soon as possible, levels to be considered by the IPCC and the Second World Climate Conference of November 1990. In the view of many industrialized nations such stabilization of CO_2 emissions should be achieved as a first step at the latest by the year 2000.

Urges all industrialized countries to support the process of IPCC through the investigation of the feasibility of achieving targets to limit or reduce CO_2 emissions including e.g. a 20 per cent reduction of CO_2 emission levels by the year 2005 as recommended by the scientific World Conference on the Changing Atmosphere in Toronto 1988.

Urges all industrialized countries to intensify their efforts in this respect, while ensuring sustainable development and taking into account the specific circumstances of individual countries.

—(17) *Agrees* that industrialized countries with, as yet relatively low energy requirements, which can reasonably be expected to grow in step with their development, may have targets that accommodate that development.

—(18) *Calls* on the IPCC to present the analysis and conclusions referred to above in the Second World Climate Conference in November 1990.

—(19) *Agrees* that developing countries endeavour to meet future targets for CO_2- emissions and sinks, with due regard to their development requirements and within the limits of their financial and technical capabilities. International co-operation, whenever available, would be a contributing factor for greater action. New processes or industries to be introduced should, as far as possible, incorporate technologies which are more energy-efficient and produce less pollution than present technologies.

—(20) *Agrees* that developing countries will need to be assisted financially and technically, including assistance with training, i.e. by strengthening relevant mechanisms to ensure that they will be in a position to manage, develop, and conserve their forest resources in a sustainable and environmentally sound manner. They will also contribute to combatting erosion and desertification. Recognition by the market of the total value of forests, including non-wood values, is a precondition for developing countries' being able to successfully use such financial and technical assistance for sustainable forest management.

—(21) *Agrees* to pursue a global balance between deforestation on the one hand and sound forest management and afforestation on the other. A world net forest growth of 12 million hectares a year in the beginning of next century should be considered as a provisional aim.

Requests the IPCC to consider the feasibility of achieving this aim. To this end; the world deforestation rate should be slowed inter alia through the suppression of acid rain and other pollutants of fires and through the reduction of pressures on biota. Sound forest management practices should be encouraged and at the same time vigorous forestry programmes should be developed in both temperate and tropical zones; biological diversity should be maintained; strategies addressing climate change issues through forest management and afforestation should be integrated with strategies addressing the sustainability of other forest based values resulting in full multiple-use plans where appropriate, but with due consideration of the people living in or dependent on forest land.

Welcomes the work of the Tropical Forestry Action Plan and the International Timber Trade Organization in pursuit of these goals.

CHLOROFLUOROCARBONS (CFCs)

—(22) *Welcomes* the commitment of the industrialized countries to amend the Montreal Protocol and to phase out the production and consumption of controlled chlorofluorocarbons by the year 2000, and of other controlled ozone depleting substances as soon as feasible.

Urges all countries to become Parties to the Vienna Convention for the Protection of the Ozone Layer and to the Montreal Protocol. To facilitate this broad participation suitable amendments of the Montreal Protocol should be considered urgently by the Parties to the Protocol.

Urges industrialized countries to use financial and other means to assist developing countries in phasing out their production and consumption of controlled substances as soon as possible, by providing them with sufficient means to enable them to meet their target date. The development of alternative technologies and products in developing countries should be promoted.

OTHER GREENHOUSE GASES

—(23) *Recommends* that the development and implementation of specific means of limiting the atmospheric concentrations of greenhouse gases other than CO_2 and CFCs should be energetically pursued, taking into consideration the special situation of developing countries.

MINISTERIAL MEETING

—(24) *Recognizes* the need to convene a Ministerial Conference to review the interim report of the IPCC. The conference endorses the plan of the organization by WMO, UNEP, UNESCO and ICSU of such a meeting as part of the Second World Climate Conference in November 1990.

FUNDING

—(25) *Recommends* that existing institutions for development and financial assistance including the Multilateral Development Banks, Bilateral Assistance Programmes, the relevant United Nations organizations and specialized agencies, and scientific and technological organizations should give greater attention to climate change issues within their environmental and other relevant programmes by providing expanded funding including concessional funding. In addition, regional and subregional co-operation should be reinforced and funded so as to address and implement the required action at that level.

—(26) *Recommends* that additional resources should, over time, be mobilized to help developing countries take the necessary measures to address climate change and that are compatible with their development requirements.

Further *recommends* that the scope of resources needed must be assessed. Such assessments should include inter alia country studies and the capabilities of existing institutions and mechanisms to meet the financing needs identified, similar to the approaches developed under the Montreal Protocol.

Further consideration should be given to the need for funding facilities including a clearinghouse mechanism and a possible new international fund and their relationship to existing funding mechanisms, both multilateral and bilateral. Such funding should be related to the implementation of a future climate convention and associated protocols. In the meantime the donor community is urged to provide assistance to developing countries to support actions addressing climate change.

—(27) *Recommends* that, initially, international funding be directed towards
(i) funding of a CFC phase-out in developing countries in the context of the Montreal Protocol;
(ii) promoting efficient use of energy, including appropriate end use technologies, increasing the use of non-fossil fuels and switching to energy sources with lower greenhouse gas emissions, and the use of renewable energy sources;
(iii) increased financial support for forest production and forest management improvement, for example through the Tropical Forestry Action Plan (TFAP), the Plan of Action to Combat Desertification, the International Tropical Timber Organization (ITTO) and other relevant international organizations;
(iv) assisting developing countries in planning how to address problems posed by climate change;
(v) supporting developing countries to enable their participation in the IPCC process and the other international meetings on this subject;
(vi) conducting research and monitoring;
(vii) arranging for technology transfer to and technology development in developing countries;
(viii) promoting public awareness, education and institutional and manpower development.

The use of financial resources could subsequently be extended inter alia to major energy sources with little or no environmentally damaging characteristics and for steps to reduce other global man-made emissions of greenhouse gases.

RESEARCH AND MONITORING

—(28) *Urges* all countries and relevant organizations to increase their climate change research and monitoring activities and to provide for adequate data bases on emissions. *Also urges* states to co-operate in, and provide increasing support for, international co-ordination of these activities building on international programmes such as the World Climate Programme and the International Geosphere-Biosphere Programme, and on the present roles of the UNEP, WMO, ICSU, IEA, UNESCO, IOC, and other competent international organizations and bodies. The enhancement and strengthening of operational aspects of their work should be examined.

Recommends that more research should be carried out by 1992 into the sources and sinks of the greenhouse gases other than CO_2 and CFCs, like methane (CH_4), nitrous oxide (N_2O) and tropospheric ozone (O_3), including further research on the effect of the ocean on the concentration of radiatively active gases in the atmosphere.

CLIMATE CHANGE CONVENTION

—(29) 1. *Urges* all countries to join and intensify the ongoing work within UNEP and WHO through the IPCC with respect to the compilation of elements for a framework convention on climate change so that negotiations upon it can start as soon as possible after the adoption of the interim report of the IPCC.
2. *Recommends* that such convention will be framed in such a way as to gain the adherence of the largest possible number and most suitably balanced spread of countries.

3. Agrees that to this end the framework convention and associated protocols should commit the parties inter alia to:
   —enhancement of research and systematic observation of climate, aimed at detecting and monitoring climate variations and change;
   —action to deal with greenhouse gas emissions and the effects of global warming;
   —address to particular financial needs of the developing countries in the access to and transfer of technology; and
   —strengthen sustainable forest management.
4. *Agrees* further that in developing the framework convention on climate change special attention should be given to ensuring that provision is made for appropriate decision making procedures and powers.
5. *Urges* all involved or to be involved in the negotiations to do their utmost to conclude these negotiations to ensure adoption of the convention as early as 1991 if possible and no later than at the Conference of the United Nations on Environment and Development in 1992.
6. *Considers* that in the preparation of the framework convention and protocols the relevant aspects of the Vienna Convention on the Protection of the Ozone Layer should be taken into account, as well as innovative approaches as may be required by the complex character of the problem.

—(30) *Recommends* that this declaration and the supporting papers be conveyed to the IPCC at the conclusion of this Conference for further consideration and action.

LIST OF PARTICIPATING COUNTRIES

Argentina
Australia
Austria
Bangladesh
Belgium

Bolivia
Brazil
Bulgaria
Cameroon
Canada

China
Colombia
Costa Rica
Cuba
Denmark

Egypt
Finland
France
Gambia
German Democratic Republic

Germany, Federal Republic of
Ghana
Greece
Guinea
Hungary

India
Indonesia
Ireland
Israel
Ivory Coast

Italy
Jamaica
Japan
Jordan
Kenya

Luxembourg
Malaysia
Mali
Malta
Mexico

Netherlands, Kingdom of the
New Zealand
Nigeria
Norway
Pakistan

Philippines
Poland
Portugal
Republic of Korea
Saudi Arabia

Senegal
Spain
Sweden
Switzerland
Thailand

Trinidad and Tobago
Tunisia
Turkey
United States of America
Union of Soviet Socialist Republics

United Kingdom of Great Britain and Northern Ireland
United Republic of Tanzania
Venezuela
Yugoslavia
Zaire

Zambia
Zimbabwe

Commission of European Communities and International Organizations

Commission of European Communities

Food and Agriculture Organization
International Atomic Energy Agency
International Energy Agency
Organization for Economic Cooperation and Development
United Nations Development Program

United Nations Education, Scientific and Cultural Organization
United Nations Industrial Development Organization
United Nations Environment Program
World Meteorological Organization
World Bank

TABLE 12

The following excerpts from the Kyoto Protocol of the United Nations Framework Convention on Climate Change are presented without comment:

The Parties to this Protocol,

Being Parties to the United Nations Framework Convention on Climate Change, hereinafter referred to as "the Convention",

In pursuit of the ultimate objective of the Convention as stated in its Article 2,

Recalling the provisions of the Convention,

Being guided by Article 3 of the Convention,

Pursuant to the Berlin Mandate adopted by decision 1/CP.1 of the Conference of the Parties to the Convention at its first session,

Have agreed as follows:

ARTICLE 1

For the purposes of this Protocol, the definitions contained in Article 1 of the Convention shall apply. In addition:

1. "Conference of the Parties" means the Conference of the Parties to the Convention.
2. "Convention" means the United Nations Framework Convention on Climate Change, adopted in New York on 9 May 1992.
3. "Intergovernmental Panel on Climate Change" means the Intergovernmental Panel on Climate Change established in 1988 jointly by the World Meteorological Organization and the United Nations Environment Programme.
4. "Montreal Protocol" means the Montreal Protocol on Substances that Deplete the Ozone Layer, adopted in Montreal on 16 September 1987 and as subsequently adjusted and amended.
5. "Parties present and voting" means Parties present and casting an affirmative or negative vote.
6. "Party" means, unless the context otherwise indicates, a Party to this Protocol.
7. "Party included in Annex I" means a Party included in Annex I to the Convention, as may be amended, or a Party which has made a notification under Article 4, paragraph 2(g), of the Convention.

ARTICLE 2

1. Each Party included in Annex I, in achieving its quantified emission limitation and reduction commitments under Article 3, in order to promote sustainable development, shall:

(a) Implement and/or further elaborate policies and measures in accordance with its national circumstances, such as:
  (i) Enhancement of energy efficiency in relevant sectors of the national economy;
  (ii) Protection and enhancement of sinks and reservoirs of greenhouse gases not controlled by the Montreal Protocol, taking into account its commitments under relevant international environmental agreements; promotion of sustainable forest management practices, afforestation and reforestation;
  (iii) Promotion of sustainable forms of agriculture in light of climate change considerations;
  (iv) Research on, and promotion, development and increased use of, new and renewable forms of energy, of carbon dioxide sequestration technologies and of advanced and innovative environmentally sound technologies;
  (v) Progressive reduction or phasing out of market imperfections, fiscal incentives, tax and duty exemptions and subsidies in all greenhouse gas emitting sectors that run counter to the objective of the Convention and application of market instruments;
  (vi) Encouragement of appropriate reforms in relevant sectors aimed at promoting policies and measures which limit or reduce emissions of greenhouse gases not controlled by the Montreal Protocol;
  (vii) Measures to limit and/or reduce emissions of greenhouse gases not controlled by the Montreal Protocol in the transport sector;
  (viii) Limitation and/or reduction of methane emissions through recovery and use in waste management, as well as in the production, transport and distribution of energy;

(b) Cooperate with other such Parties to enhance the individual and combined effectiveness of their policies and measures adopted under this Article, pursuant to Article 4, paragraph 2(e)(i), of the Convention. To this end, these Parties shall take steps to share their experience and exchange information on such policies and measures, including developing ways of improving their comparability, transparency and effectiveness. The Conference of Parties serving as the meeting of the Parties to this Protocol shall, at its first session or as soon as practicable thereafter, consider ways to facilitate such cooperation, taking into account all relevant information.

2. The Parties included in Annex I shall pursue limitation or reduction of emissions of greenhouse gases not controlled by the Montreal Protocol from aviation and marine bunker fuels, working through the International Civil Aviation Organization and the International Maritime Organization, respectively.

3. The Parties included in Annex I shall strive to implement policies and measures under this Article in such a way as to minimize adverse effects, including the adverse effects of climate change, effects on international trade, and social, environmental and economic impacts on other Parties, especially developing country Parties and in particular those identified in Article 4, paragraphs 8 and 9, of the Convention, taking into account Article 3 of the Convention. The Conference of the Parties serving as the meeting of the Parties to this Protocol may take further action, as appropriate, to promote the implementation of the provisions of this paragraph.

4. The Conference of the Parties serving as the meeting of the Parties to this Protocol, if it decides that it would be beneficial to coordinate any of the policies and measures in paragraph 1(a) above, taking into account different national circumstances and potential effects, shall consider ways and means to elaborate the coordination of such policies and measures.

ARTICLE 3

1. The Parties included in Annex I shall, individually or jointly, ensure that their aggregate anthropogenic carbon dioxide equivalent emissions of the greenhouse gases listed in Annex. A do not exceed their assigned amounts, calculated pursuant to their quantified emissions limitation and reduction commitments inscribed in Annex B and in accordance with the provisions of this Article, with a view to reducing their overall emissions of such gases by at least 5 per cent below 1990 levels in the commitment period 2008 to 2012.

2. Each Party included in Annex I shall, by 2005, have made demonstrable progress in achieving its commitments under this Protocol.

3. The net changes in greenhouse gas emissions by sources and removals by sinks resulting from direct human-induced land-use change and forestry activities, limited to afforestation, reforestation and deforestation since 1990, measured as verifiable changes in carbon stocks in each commitment period, shall be used to meet the commitments under this Article of each Party included in Annex I. The greenhouse as emissions by sources and removals by sinks associated with those activities shall be reported in a transparent and verifiable manner and reviewed in accordance with Articles 7 and 8.

4. Prior to the first session of the Conference of the Parties serving as the meeting of the Parties to this Protocol, each Party included in Annex I shall provide, for consideration by the Subsidiary Body for Scientific and Technological Advice, data to establish its level of carbonstocks in 1990 and to enable an estimate to be made of its changes in carbon stocks in subsequent years. The Conference of the Parties serving as the meeting of the Parties to this Protocol shall, at its first session or as soon as practicable thereafter, decide upon modalities, rules and guidelines as to how, and which, additional human-induced activities related to changes in greenhouse gas emissions by sources and removals by sinks in the agricultural soils and the land-use change and forestry categories shall be added to, or subtracted from, the assigned amounts for Parties included in Annex I, taking into account uncertainties, transparency in reporting, verifiability, the methodological work of the Intergovernmental Panel on Climate Change, the advice provided by the Subsidiary Body for Scientific and Technological Advice in accordance with Article 5 and the decisions of the Conference of the Parties. Such a decision shall apply in the second and subsequent commitment periods. A Party may choose to apply such a decision on these additional human-induced activities for its first commitment period, provided that these activities have taken place since 1990.

5. The Parties included in Annex I undergoing the process of transition to a market economy whose base year or period was established pursuant to decision 9/CP.2 of the Conference of the Parties at its second session shall use that base year or period for the implementation of their commitments under this Article. Any other Party included in Annex I undergoing the process of transition to a market economy which has not yet submitted its first national communication under Article 12 of the Convention may also notify the Conference of the Parties serving as the meeting of the Parties to this Protocol that it intends to use an historical base year or period other than 1990 for the implementation of its commitments under this Article. The Conference of the Parties serving as the meeting of the Parties to this Protocol shall decide on the acceptance of such notification.

6. Taking into account Article 4, paragraph 6, of the Convention, in the implementation of their commitments under this Protocol other than those under this Article, a certain degree of flexibility shall be allowed by the Conference of the Parties serving as the meeting of the Parties to this Protocol to the Parties included in Annex I undergoing the process of transition to a market economy.

7. In the first quantified emission limitation and reduction commitment period, from 2008 to 2012, the assigned amount for each Party included in Annex I shall be equal to the percentage inscribed for it in Annex B of its aggregate anthropogenic carbon dioxide equivalent emissions of the greenhouse gases listed in Annex A in 1990, or the base year or period determined in accordance with paragraph 5 above, multiplied by five. Those Parties included in Annex I for whom land-use change and forestry constituted a net source of greenhouse gas emissions in 1990 shall include in their 1990 emissions base year or period the aggregate anthropogenic carbon dioxide equivalent emissions by sources minus removals by sinks in 1990 from land-use change for the purposes of calculating their assigned amount.

8. Any Party included in Annex I may use 1995 as its base year for hydrofluorocarbons, perfluorocarbons and sulphur hexafluoride, for the purposes of the calculation referred to in paragraph 7 above.

9. Commitments for subsequent periods for Parties included in Annex I shall be established in amendments to Annex B to this Protocol, which shall be adopted in accordance with the provisions of Article 21, paragraph 7. The Conference of the Parties serving as the meeting of the Parties to this Protocol shall initiate the consideration of such commitments at least seven years before the end of the first commitment period referred to in paragraph 1 above.

10. Any emission reduction units, or any part of an assigned amount, which a Party acquires from another Party in accordance with the provisions of Article 6 or of Article 17 shall be added to the assigned amount for the acquiring Party.

11. Any emission reduction units, or any part of an assigned amount, which a Party transfers to another Party in accordance with the provisions of Article 6 or of Article 17 shall be subtracted from the assigned amount for the transferring Party.

12. Any certified emission reductions which a Party acquires from another Party in accordance with the provisions of Article 12 shall be added to the assigned amount for the acquiring Party.

13. If the emissions of a Party included in Annex I in a commitment period are less than its assigned amount under this Article, this difference shall, on request of that Party, be added to the assigned amount for that Party for subsequent commitment periods.

14. Each Party included in Annex I shall strive to implement the commitments mentioned in paragraph 1 above in such a way as to minimize adverse social, environmental and economic impacts on developing country Parties, particularly those identified in Article 4, paragraphs 8 and 9, of the Convention. In line with relevant decisions of the Conference of the Parties on the implementation of those paragraphs, the Conference of the Parties serving as the meeting of the Parties to this Protocol shall, at its first session, consider what actions are necessary to minimize the adverse effects of climate change and/or the impacts of response measures on Parties referred to in those paragraphs. Among the issues to be considered shall be the establishment of funding, insurance and transfer of technology.

ARTICLE 4

1. Any Parties included in Annex I that have reached an agreement to fulfill their commitments under Article 3 jointly, shall be deemed to have met those commitments provided that their total combined aggregate anthropogenic carbon dioxide equivalent emis-sions of the greenhouse gases listed in Annex A do not exceed their assigned amounts calculated pursuant to their quantified emission limitation and reduction commitments inscribed in Annex B and in accordance with the provisions of Article 3. The respective emission level allocated to each of the Parties to the agreement shall be set out in that agreement.

2. The Parties to any such agreement shall notify the secretariat of the terms of the agreement on the date of deposit of their instruments of ratification, acceptance or approval of this Protocol, or accession thereto. The secretariat shall in turn inform the Parties and signatories to the Convention of the terms of the agreement.

3. Any such agreement shall remain in operation for the duration of the commitment period specified in Article 3, paragraph 7.

4. If Parties acting jointly do so in the framework of, and together with, a regional economic integration organization, any alteration in the composition of the organization after adoption of this Protocol shall not affect existing commitments under this Protocol. Any alteration in the composition of the organization shall only apply for the purposes of those commitments under Article 3 that are adopted subsequent to that alteration.

5. In the event of failure by the Parties to such an agreement to achieve their total combined level of emission reductions, each Party to that agreement shall be responsible for its own level of emissions set out in the agreement.

6. If Parties acting jointly do so in the framework of, and together with, a regional economic integration organization which is itself a Party to this Protocol, each member State of that regional economic integration organization individually, and together with the regional economic integration organization acting in accordance with Article 24, shall, in the event of failure to achieve the total combined level of emission reductions, be responsible for its level of emissions as notified in accordance with this Article.

ARTICLE 5

1. Each Party included in Annex I shall have in place, no later than one year prior to the start of the first commitment period, a national system for the estimation of anthropogenic emissions by sources and removals by sinks of all greenhouse gases not controlled by the Montreal Protocol. Guidelines for such national systems, which shall incorporate the methodologies specified in paragraph 2 below, shall be decided upon by the Conference of the Parties serving as the meeting of the Parties to this Protocol at its first session.

2. Methodologies for estimating anthropogenic emissions by sources and removals by sinks of all greenhouse gases not controlled by the Montreal Protocol shall be those accepted by the Intergovernmental Panel on Climate Change and agreed upon by the Conference of the Parties at its third session. Where such methodologies are not used, appropriate adjustments shall be applied according to methodologies agreed upon by the Conference of the Parties serving as the meeting of the Parties to this Protocol at its first session. Based on the work of, *inter ali*a, the Intergovernmental Panel on Climate Change and advice provided by the Subsidiary Body for Scientific and Technological Advice, the Conference of the Parties serving as the meeting of the Parties to this Protocol shall regularly review and, as appropriate, revise such methodologies and adjustments, taking fully into account any relevant decisions by the Conference of the Parties. Any revision to methodologies or adjustments shall be used only for the purposes of ascertaining compliance with commitments under Article 3 in respect of any commitment period adopted subsequent to that revision.

3. The global warming potentials used to calculate the carbon dioxide equivalence of anthropogenic emissions by sources and removals by sinks of greenhouse gases listed in Annex A shall be those accepted by the Intergovernmental Panel on Climate Change and agreed upon by the Conference of the Parties at its third session. Based on the work of, *inter ali*a, the Intergovernmental Panel on Climate Change and advice provided by the Subsidiary Body for Scientific and Technological Advice, the Conference of the Parties serving as the meeting of the Parties to this Protocol shall regularly review and, as appropriate, revise the global warming potential of each such greenhouse gas, taking fully into account any relevant decisions by the Conference of the Parties. Any revision to a global warming potential shall apply only to commitments under Article 3 in respect of any commitment period adopted subsequent to that revision.

ARTICLE 6

1. For the purpose of meeting its commitments under Article 3, any Party included in Annex I may transfer to, or acquire from, any other such Party emission reduction units resulting from projects aimed at reducing anthropogenic emissions by sources or enhancing anthropogenic removals by sinks of greenhouse gases in any sector of the economy, provided that:
(a) Any such project has the approval of the Parties involved;
(b) Any such project provides a reduction in emissions by sources, or an enhancement of removals by sinks, that is additional to any that would otherwise occur;
(c) It does not acquire any emission reduction units if it is not in compliance with its obligations under Articles 5 and 7; and
(d) The acquisition of emission reduction units shall be supplemental to domestic actions for the purposes of meeting commitments under Article 3.

2. The Conference of the Parties serving as the meeting of the Parties to this Protocol may, at its first session or as soon as practicable thereafter, further elaborate guidelines for the implementation of this Article, including for verification and reporting.

3. A Party included in Annex I may authorize legal entities to participate, under its responsibility, in actions leading to the generation, transfer or acquisition under this Article of emission reduction units.

4. If a question of implementation by a Party included in Annex I of the requirements referred to in this Article is identified in accordance with the relevant provisions of Article 8, transfers and acquisitions of emission reduction units may continue to be made after the question has been identified, provided that any such units may not be used by a Party to meet its commitments under Article 3 until any issue of compliance is resolved.

ARTICLE 7

1. Each Party included in Annex I shall incorporate in its annual inventory of anthropogenic emissions by sources and removals by sinks of greenhouse gases not controlled by the Montreal Protocol, submitted in accordance with the relevant decisions of the Conference of the Parties, the necessary supplementary information for the purposes of ensuring compliance with Article 3, to be determined in accordance with paragraph 4 below.

2. Each Party included in Annex I shall incorporate in its national communication, submitted under Article 12 of the Convention, the supplementary information necessary to demonstrate compliance with its commitments under this Protocol, to be determined in accordance with paragraph 4 below.

3. Each Party included in Annex I shall submit the information required under paragraph 1 above annually, beginning with the first inventory due under the Convention for the first year of the commitment period after this Protocol has entered into force for that Party. Each such Party shall submit the information required under paragraph 2 above as part of the first national communication due under the Convention after this Protocol has entered into force for it and after the adoption of guidelines as provided for in paragraph 4 below. The frequency of subsequent submission of information required under this Article shall be determined by the Conference of the Parties serving as the meeting of the Parties to this Protocol, taking into account any timetable for the submission of national communications decided upon by the Conference of the Parties.

4. The Conference of the Parties serving as the meeting of the Parties to this Protocol shall adopt at its first session, and review periodically thereafter, guidelines for the preparation of the information required under this Article, taking into account guidelines for the preparation of national communications by Parties included in Annex I adopted by the Conference of the Parties. The Conference of the Parties serving as the meeting of the Parties to this Protocol shall also, prior to the first commitment period, decide upon modalities for the accounting of assigned amounts.

ARTICLE 8

1. The information submitted under Article 7 by each Party included in Annex I shall be reviewed by expert review teams pursuant to the relevant decisions of the Conference of the Parties and in accordance with guidelines adopted for this purpose by the Conference of the Parties serving as the meeting of the Parties to this Protocol under paragraph 4 below. The information submitted under Article 7, paragraph 1, by each Party included in Annex I shall be reviewed as part of the annual compilation and accounting of emissions inventories and assigned amounts. Additionally, the information submitted under Article 7, paragraph 2, by each Party included in Annex I shall be reviewed as part of the review of communications.

2. Expert review teams shall be coordinated by the secretariat and shall be composed of experts selected from those nominated by Parties to the Convention and, as appropriate, by intergovernmental organizations, in accordance with guidance provided for this purpose by the Conference of the Parties.”...............

For Articles 9-27, the reader is referred to
http://unfccc.int/resource/docs/convkp/kpeng.pdf

ARTICLE 28

The original of this Protocol, of which the Arabic, Chinese, English, French, Russian and Spanish texts are equally authentic, shall be deposited with the Secretary-General of the United Nations.

DONE at Kyoto this eleventh day of December one thousand nine hundred and ninety-seven.

IN WITNESS WHEREOF the undersigned, being duly authorized to that effect, have affixed their signatures to this Protocol on the dates indicated.

ANNEX A

Greenhouse gases

Carbon dioxide (CO_2)
Methane (CH_4)
Nitrous oxide (NO_2)
Hydrofluorocarbons (HFCs)
Perfluorocarbons (PFCs)
Sulphur hexafluoride (SF_6)

Sectors/source categories

Energy Fuel combustion Energy industries Manufacturing industries and construction Transport
Other sectors Fugitive emissions from fuels Solid fuels Oil and natural gas
Other Industrial processes Mineral products Chemical industry Metal production
Other production

Production of halocarbons and sulphur hexafluoride
Consumption of halocarbons and sulphur hexafluoride
Solvent and other product use Agriculture Enteric fermentation Manure management
Rice cultivation Agricultural soils Prescribed burning of savannas Field burning of agricultural residues
Other

ANNEX B

Party Quantified emission limitation or reduction commitment

(Percentage of base year or period)
Australia 108 Austria 92 Belgium 92 Bulgaria* 92 Canada 94 Croatia* 95 Czech Republic* 92
Denmark 92 Estonia* 92 European Community 92 Finland 92 France 92 Germany 92
Greece 92 Hungary* 94 Iceland 110 Ireland 92 Italy 92 Japan 94 Latvia* 92
Liechtenstein 92 Lithuania* 92 Luxembourg 92 Monaco 92 Netherlands 92 New Zealand 100
Norway 101 Poland* 94 Portugal 92 Romania*92 Russian Federation* 100 Slovakia* 92
Slovenia* 92 Spain 92 Sweden 92 Switzerland 92 Ukraine* 100
United Kingdom of Great Britain and Northern Ireland 92 United States of America 93

---

* Countries that are undergoing the process of transition to a market economy.

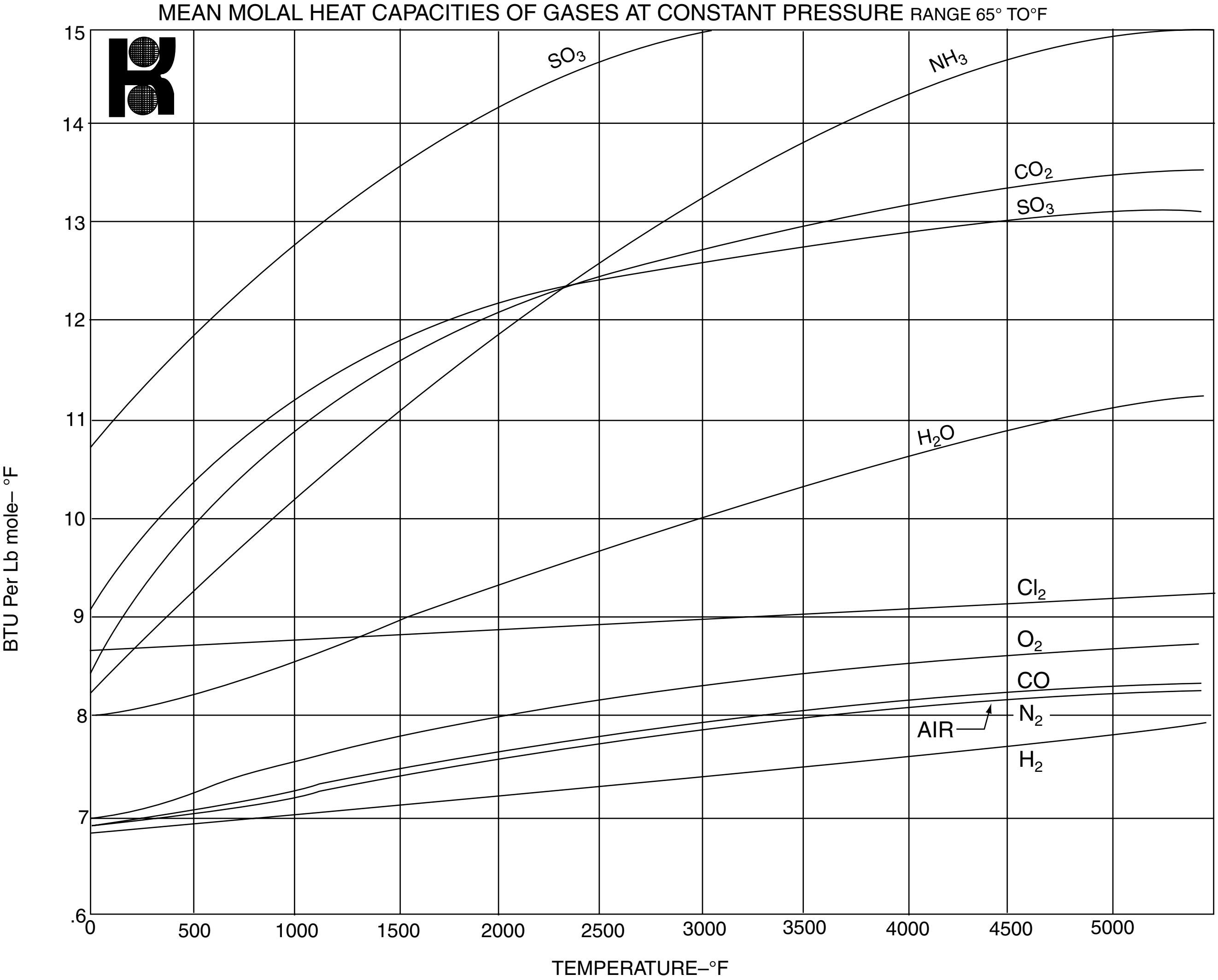


FIGURE 1 (Courtesy of Koch Engineering.)

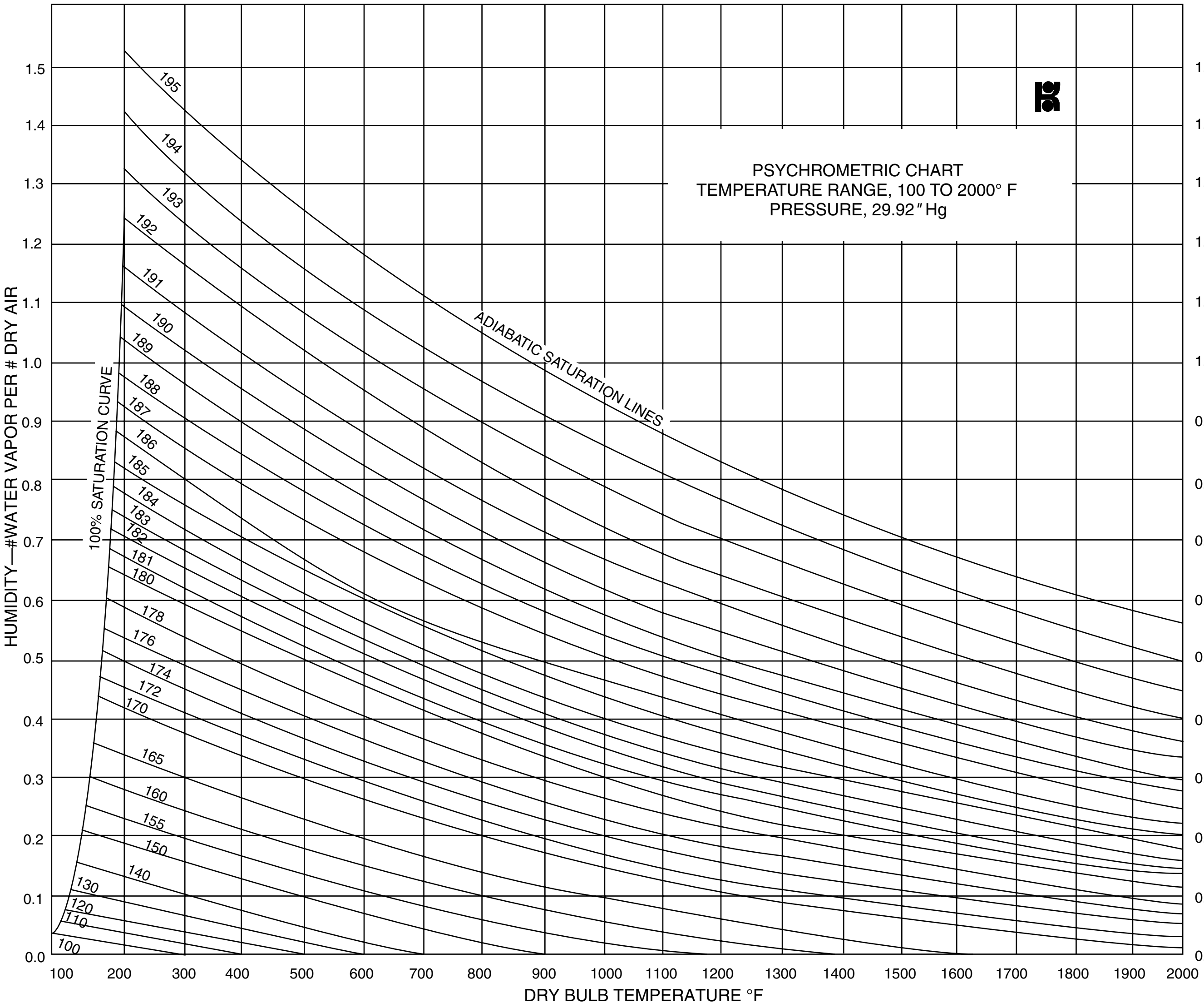


FIGURE 2 (Courtesy of Koch Engineering.)

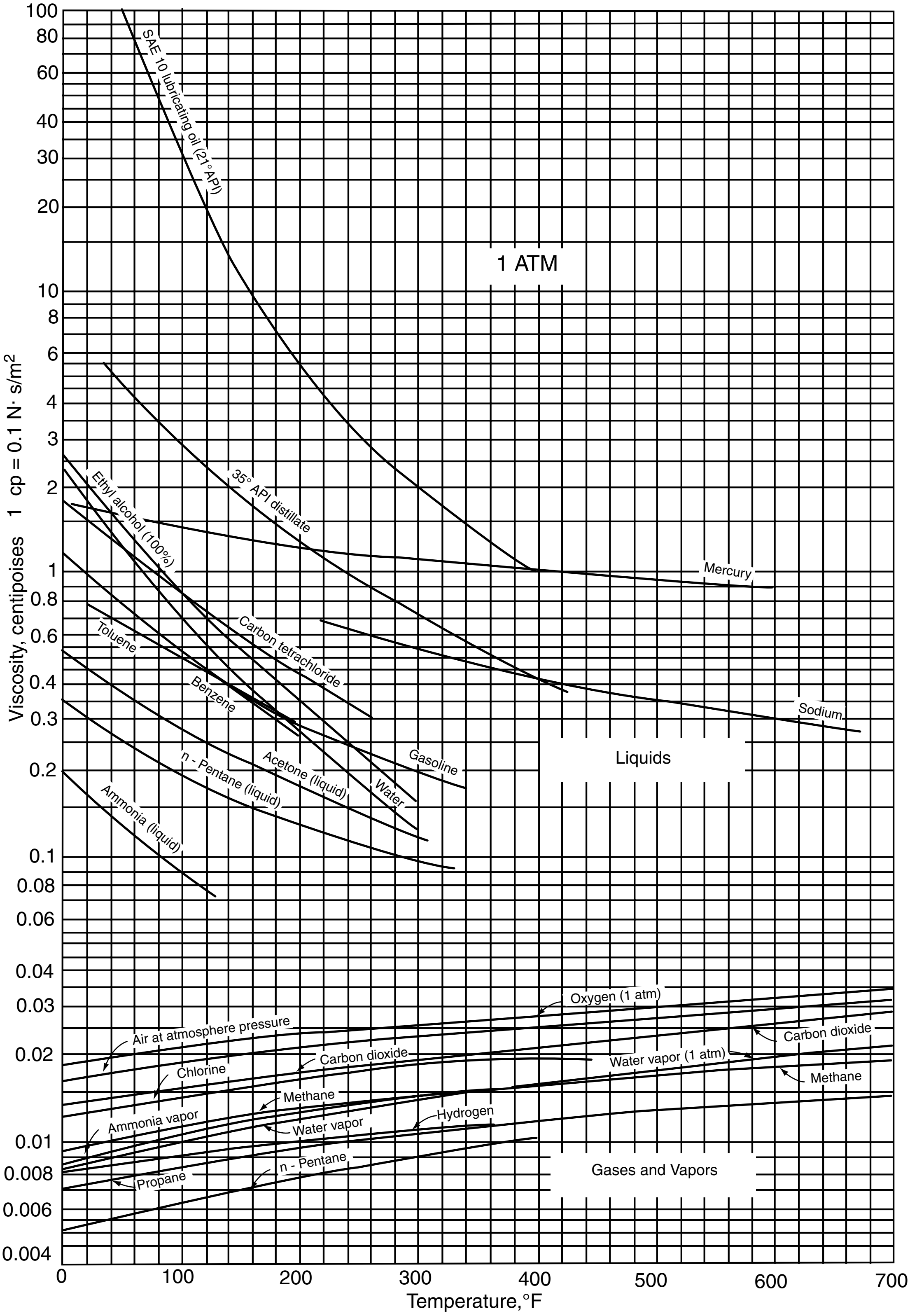


FIGURE 3 Viscosities of fluids as a function of temperature. Adopted from G. G. Brown and Associates, *Unit Operations,* John Wiley & Sons, New York, 1950.

ACRONYMS AND ABBREVIATIONS

AA—Atomic Adsorption
AAAS—American Association for the Advancement of Science
ALARA—As Low As Reasonably Achievable
ALARP—As Low As Reasonably Practicable
ALPCO—Association of Local Air Pollution Control officers (U.S.)
ALAT—Alamine Amino Transferase
ALK—Alkalinity
ALLEA—All European Academies
AM—Annual-Maximum
AMA—American Medical Association
AMD—Acid Mine Drainage
AMO—Atlantic Multidecadal Oscillation
AMP—Adenosine Monophosphate
AMU—Atomic Mass Unit
ANDRA—National Waste Disposal Organization (France)
ANL—Argonne National Laboratory (U.S.)
ANOVA—Analysis of Variance
ANP—National Petroleum Agency (Brazil)
ANPR—Advanced Notice of Proposed Rulemaking (U.S.)
ANSI—American National Standards Institute
ANS—American Nuclear Society
ANTLE—Affordable Near Term Low Emissions
ANWR—Alaska National Wildlife Refuge
AOAC—Association of Official Agricultural Chemists
AOC—Area of Concern (U.S.)
AOC—Assimilable Organic Carbon
AOCS—American Oil Chemists Society
AOGCM—Atmospheric-Ocean General Circulation Model
AONB—Area of Outstanding Natural Beauty (UK)
AOT—Advanced Oxidation Technology
AOX—Organochlorine
APC—Adenomatous Polyposis Coli
APC—Air Pollution Control
APCA—Air Pollution Control Association
APCS—Air Pollution Control System
APDC—Ammonium Pyrolidinodithiocarbonate
APEG—Airborne Particle Expert Group
APEM—Association of Professional Engineers of Manitoba
APHA—American Public Health Association
APHIS—Animal and Plant Health Inspection Service (U.S.)
API—American Petroleum Institute
APS—Aerodynamic Particle Size
APS—American Physical Society
AQCR—Air Quality Control Region (U.S.)
AQEI—Air Quality Exposure Index
AQMD—Air Quality Management District
ARAR—Applicable or Relevant and Appropriate Requirement
ARB—Air Resources Board
ARBRE—Arable Biomass Renewable Energy
ARDS—Acute Respiratory Distress Syndrome
ARM—Aluminized Red Mud
ARM—Availability, Reliability and Maintainability
ARMS—Agricultural Resources Management Study
AROI—Acceptable Range of Oral Intake
ARS—Absorption Refrigeration System
ARS—Agricultural Research Service (U.S.)
AS—Activated Sludge
AS—Australian Standard
ASBR—Anaerobic Sequencing Batch Reactor
ASCE—American Society of Civil Engineers
ASEE—American Society for Engineering Education
ASHRAE—American Society of Heating, Refrigerating and Air Conditioning Engineers
ASLB—Atomic Safety Licensing Board (U.S.)
ASME—American Society of Mechanical Engineers
ASPEN—Advanced Systems for Process Engineering
ASSE—American Society of Safety Engineers
AST—Aboveground Storage Tank
ASTM—American Society for Testing and Materials
ASTMS—Association of Scientific, Technical and Management Staffs
ASV—Anode Stripping Voltammetry
ATOC—Acoustic Thermometry of Ocean Climate
ATOFMS—Aerosol Time-of-Flight Mass Spectrometer
ATP—Adenosine Triphosphate
ATP—Advanced Technology Program (U.S.)
ATS—Automatic Tracking Service
ATSDR—Agency for Toxic Substances and Disease Registry (U.S.)
au—Absorbance Unit
AWEP—Association of Women in Environmental Professions
AWG—Association of Women Geoscientists
AWIS—Association for Women in Science
AWMA—Air and Water Management Association
AWRA—American Water Resources Association
AWSE—Association of Women in Science and Engineering (UK)
AWWA—American Water Works Association

BABFO—British Association for Biofuels and Oil
BACT—Best Available Current (Control) Technology
BAF—Biological Amplification Factor
BALF—Bronchoalveolar Fluid
BANANA—Build Absolutely Nothing Anywhere Anytime
BAP—Benzo-(a)-Pyrene
BAP—Biomass-Associated Product
Bara—Bar Absolute
BART—Best Available Retrofit Technology
BAT—Best Available Technology
BATNEEC—Best Available Technology Not Entailing Excessive Cost (UK)
BBADCP—Broad Band Acoustic Doppler Current Profiler
BBL—Barrels
BCC—Bioaccumulative Chemicals of Concern
BCDL—Beverage Container Deposit Legislation
BCF—Bio-Concentration Factor
BCT—Best Conventional Technology
BD—Butadiene
BDAT—Best Demonstrated Available Technology
BDO—Butanediol
BDOC—Biodegradable Organic Carbon
BDS—Biocatalytic Desulfurization
BEDI—Brownfields Economic Development Initiative
BEI—Biological Exposure Indices
BEP—Board for the Engineering Profession (UK)
BER—Board for Engineering Regulation (UK)
BESS—Battery Energy Storage System
BFSS—Bioremediation in the Field Search System
BGB—Brilliant Green Bile
BGS—British Geological Survey
BHA—Butylated Hydroxy Anisole
BHC—Benzene Hexachloride
BHC—Lindane
BHET—Bis-Hydroxyethylene Tetraphthalate
BHT—Butylated Hydroxy Toluene
BICER—Baikal International Center for Ecological Research
BID—Background Information Document
BIF—Boilers and Industrial Furnaces
BIOMOVS—Biospheric Model Validation Study
BLL—Blood Lead Levels
BLM—Bureau of Land Management (U.S.)
BMB—British Medical Bulletin
BMP—Best Management Practices
BN—Base Neutral
BNA—Base Neutral Acid Extractable
BNFL—British Nuclear Fuels PLC
BNL—Brookhaven National Laboratory (U.S.)
BNR—Biological Nutrient Removal
BOD—Biochemical Oxygen Demand
BOHC—British Occupational Hygiene Society
BOOS—Burner Out of Service
BPBR—Batch Packed Reactor Bed
BPD—Biocidal Products Directive (EU)
BPEO—Best Practicable Environmental Option
BPL—Beta Propiolacetone
BPM—Best Practicable Means
BPWTT Best Practicable Waste Treatment Technology
Bq—Becquerel
BRE—Building Research Establishment (UK)
BREEAM—Building Research Establishment Environmental Assessment Management
BRH—Bureau of Radiological Health (U.S.)
BRWM—Board on Radioactive Waste Management
BS—British Standard
BSE—Bovine Spongiform Encephalopathies
BSERT—Building Services Engineering Research and Technology (UK)
BSI—British Standards Institution
BSL-4—Biosafety Level 4
BSP—Biosafety Protocol
BT—Biotechnology
BTC—Breakthrough Curve
BTEX—Benzene, Toluene, Ethylbenzene and Xylene
BTF—Biotrickling Filter
BTU—British Thermal Unit
BTWC—Biological and Toxin Weapons Convention
BTX—Benzene-Toluene-Xylene
BV—Bed Volume
BWEA—British Wind Energy Association
BWR—Boiling Water Reactor

C—Celsius (Centigrade)
CA—Cellulose Acetate
CAA—Clean Air Act (UK) (U.S.)
CAAA—Clean Air Act Amendment (U.S.)
CAD—Computer Aided Design
CaFCP—California Fuel Cell Partnership
CAE—Computer Aided Engineering
CAfE—Community Action for Energy
CAFE—Corporate Average Fuel Economy
CAM—Computer Assisted Mapping
CAMP—Continuous Air Monitoring Program
C and S—Capture and Storage
CANDU—Canadian Heavy Water Moderated Reactor
CAP—Chemical Analysis Parameter
CAP—Coagulant Aid Polymer
CAP—Common Agricultural Policy (EU)
CARB—California Air Resources Board
CARL—Conservation and Recreation Lands (U.S.)
CASSIOPEE—Association of Western European Waste Disposal Organizations
CATNIP—Cheapest Available Techniques Not Involving Prosecution (UK)
CBA—Cost-Benefit Analysis
CBED—Convergent Beam Electron Diffraction
CBER—Center for Biologics Evaluation and Research
CBI—Confidential Business Information
CBO—Congressional Budget Office (U.S.)
CBOD—Carbonaceous Biochemical Oxygen Demand
CBR—Crude Birth Rate
CCAA—Canadian Clean Air Act
CCAR—Closed-Cycle Air Refrigeration
CCD—Charged-Coupled-Device
CCDT—Calcium Carbonate Deposition Test
CCGT—Combined Cycle Gas Turbine

CCI—Clean Coal Initiative
CCL—Climate Change Levy
CCL—Contaminant Candidate List (U.S.)
CCME—Canadian Council of Ministers of the Environment
CCMS—Committee on the Challenges of Modern Society
CCP—Combined Cool and Power
CCPE—Canadian Council of Professional Engineers
CCR—Council for Chemical Research
CCSM—Community Climate System Model
CCT—Clean Coal Technology
CDAA—2-Chloro-N, N-Diallyl Acetamide
CDC—Centers for Disease Control and Prevention (U.S.)
CDD—Chlorinated Dibenzo-p-Dioxin
CDEC—2-Chlorallyl-N, N-Diethyldithiocarbamate
CDER—Center for Drug Evaluation and Research
CDF—Chlorinated Dibenzofuran
CDI—Capacitive Deionization
CDI—Continuous Deionization
CDI—Critical Density Index
CDM—Clean Development Mechanism
CDRH—Center for Devices and Radiological Health
CDWG—Canadian Drinking Water Guidelines
CEA—Cumulative Environmental Assessment
CEA—Atomic Energy Commission (France)
CEAA—Canadian Environmental Assessment Act
CEAB—Canadian Engineering Accreditation Board
CEARC—Canadian Environmental Assessment Research Council
CEC—Cation Exchange Capacity
CEC—Commission of the European Communities
CECAB—Canadian Environmental Certification Approvals Board
CEES—Complex Ecological-Economic System
Cefic—European Chemical Industry Council
CEGB—Central Electricity Generating Board (UK)
CEM—Continuous Emissions Monitoring
CEMS—Continuous Emission Monitoring System
CEN—European Committee for Standardization
C ENG—Chartered Engineer (UK)
CEN/SCK—Nuclear Research Organization (Belgium)
CEPA—Canadian Environmental Protection Act
CEQ—Council on Environmental Quality (U.S.)
CERCLA—Comprehensive Environmental Response, Compensation and Liability Act (U.S.)
CERN—European Laboratory for Particle Physics
CESER—Countermeasures: Environmental and Socio-Economic Response
CETP—Common Effluent Treatment Plant
CFB—Circulating Fluidized Bed
CFBC—Circulating Fluidized Bed Combustion
CFBC—Circulating Fluidized Bed Combustor
CFC—Chlorofluorocarbons
CFC—1,1,2-Trichloro-1,2,2-Trifluoroethane
CFD—Computational Fluid Dynamics
CFR—Code of Federal Regulations (U.S.)
CFR—Continuous Flow Reactor
CFS—Canadian Forest Service
CFS—Cubic Feet per Second
CFSTR—Continuous Flow Stirred Tank Reactor
CFV—Clean Fuel Vehicle
CGCM—Coupled General Circulation Model
CGDF—Coal Gasification Development Facility
CG/HCCS—Coordinating Group for the Harmonization of Chemical Classification Systems
CGIAR—Cooperative Group of International Agricultural Research Centers
CGL—Comprehensive General Liability
CGMP—Current Good Manufacturing Practice
CGSB—Canadian General Standards Board
CHIP—Chemical Hazards Information Profile
CHL—Chlorophyll
CHP—Combined Heat and Power
CHPA—Combined Heat and Power Association
CHP/HP—Combined Heat and Power/Heat Pump
CHS—Contact Hypersensitivity
CI—Confidence Interval
CI—Cumulative Impact
Ci—Curie
CIA—Chemical Industries Association (UK)
CIA—Cumulative Impact Assessment
CIB—Conseil International du Baitment Pour Recherche, l'Etude et la Documentation
CIBSE—Chartered Institution of Building Services Engineers (UK)
CIEH—Chartered Institution of Environmental Health (UK)
CIH—Chartered Institute of Housing (UK)
CIP—Clean in Place
CISTI—Canadian Institute for Science and Technical Information
CITES—Convention on International Traffic in Endangered Species
CJD—Creutzfeldt-Jacob Disease
CL—Criterion Level
CLAIRE—Contaminated Land Applications in Real Environments
CLARINET—Contaminated Land Rehabilitation Networks for Environmental Technologies
CLD—Chemiluminescence Detector
CLER—Carbon-Limited Enrichment
CM—Combustible Matter
CMA—Chemical Manufacturers Association (U.S.)
CMC—Critical Micelle Concentration
CMD—Count Mean Diameter
CMDL—Climate Monitoring and Diagnostic Laboratory (U.S.)
CMPU—Chemical Manufacturing Process Unit
CMR—Complete Mix Reactor
CNC—Condensation Nucleus Counter
CNG—Compressed Natural Gas
CNI—Consiglio Nazionale Ingegneri (Italy)
CNISF—Conseil National des Ingenierus et Scientifiques de France
CNL—Compound Noise Level
CNL—Corrected Noise Level
CNP—National Research Council (Italy)

CNRS—Centre National de la Recherche Scientifique (France)
CNSIF—National Council of French Engineers and Scientists
COD—Chemical Oxygen Demand
COGEMA—Nuclear Waste Processing Organization (France)
COH—Coefficient of Haze
COHb—Carboxyhemoglobin
COM—Coastal Oceanographic Model
COMEAP—Committee on the Medical Effects of Air Pollution
CONCAWE—Conservation and Clean Air and Water in Europe
COP—Coefficient of Performance
COP—Conference of the Parties (UNFCCC)
CoPA—Control of Pollution Act (UK)
Co-PCB—Coplanar Polychlorinated Biphenyl
COPD—Chronic Obstructive Pulmonary Disease
COPUS—Committee on the Public Understanding of Science (UK)
COS—Carbonyl Sulfide
CO2eq—Carbon Dioxide Equivalent
COVRA—Nuclear Waste Disposal Organization (Holland)
CP—Coarse Particle
CPC—Condensation Particle Counter
CPCB—Central Pollution Control Board (India)
CPF—Cancer Potency Factor
CPI—Chemical Process Industries
CPSC—Consumer Product Safety Commission (U.S.)
CPVC—Chlorinated Polyvinyl Chloride
CQC—2,6-Dichloroquinonechlorimide
CR—Community Respiration
CR—Conditioned Response
CRADA—Cooperative Research and Development Agreement
CREATE—Centre for Research, Education and Training in Energy
CRM—Capture-Recapture Method
CRNL—Chalk River Nuclear Laboratories (Canada)
CRT—Continuously Regenerating Trap
CS—Conditioned Stimulus
CSB—Chemical Safety (and Hazards Investigation) Board
CSCE—Canadian Society for Civil Engineering
CSH—Calcium Silicate Hydrate
CSI—Common Sense Initiative (U.S.)
CSIR—Council for Scientific and Industrial Research (India) (South Africa)
CSIRO—Commonwealth Scientific and Industrial Research Organization (Australia)
CSMA—Chemical Specialties Manufacturers Association
CSO—Combined Sewer Overflow
CSOD—Carbonaceous Sediment Oxygen Demand
CSPE—Canadian Society for Professional Engineers
CSPI—Center for Science in the Public Interest
CSR—Center for Scientific Review (U.S.)
CSS—Combined Sewer System
CSSP—Council of Scientific Society Presidents
CST—Capillary Suction Time
CST—Carbon Storage Trust
CSTR—Continuous Flow Stirred Tank Reactor
CTG—Control Techniques Guideline
CTI—Climate Technology Initiative
CTR—Council for Tobacco Research
CVI—Children's Vaccine Initiative
CVC—Constant Volume Sampler
CWA—Chemical Warfare Agent
CWA—Clean Water Act (U.S.)
CWC—Chemical Weapons Convention
CWO—Catalytic Wet Oxidation
CWS—Coal-Water Slurry
CZARA—Coastal Zone Act Reauthorization Amendments
CZMA—Coastal Zone Management Act (U.S.)

DAD—Decide, Announce and Defend
DADA—Decide, Announce, Defend and Abandon
DAF—Dissolved Air Flotation
DAS—Data Acquisition System
dB—Decibel
dB (A)—Decibel A-Weighted
DBMCA—Dibromo-p-methyl-carboxyazo
DBMS—Database Management Systems
DBOOM—Design, Build, Own, Operate and Maintain
DBP—Disinfection By-Product
DBPC—Di Tertiary Butyl-P-Cresol
DBPP—Disinfection By-Product Precursor
DBT—Dibutyltin
DBT—Dimethyldibenzothiophene
DC—Delta Commission (Holland)
DCB—4,4′ -Dichlorobiphenyl
DCE—1,2–Dichloroethylene
DCER—Dwelling Carbon Emission Rate
DCL—Design Control Limit
DDA—4,4′ -Dichlorodiphenylacetic Acid
D/DBP—Disinfectant/Disinfection Byproduct Rule (U.S.)
DDD—1,1-Dichloro-2,2-Bis (p-chlorophenyl)-Ethane
DDE—1,1-Dichloro-2,2-Bis (p-chlorophenyl)-Ethylene
DDT—Dichlorodiphenyltrichloroethane
DDVP—2,2-Dichlorovinyl Dimethyl Phosphate
DE—Diatomaceous Earth
DEA—Diethanolamine
DEC—Department of Environmental Conservation
DEET—Diethyltoluamide
DEFRA—Department of the Environment, Food and Rural Affairs (UK)
DEG—Diethylene Glycol
DEP—Department of Environmental Protection
DER—Discrete Emission Reduction (U.S.)
DERC—Discrete Emission Reduction Credits (U.S.)
DES—Diethylstilbesterol
DETR—Department of the Environment, Transport and Regions (UK)
D/F—Dioxin/Furan
DFE—Design for Environment
DFB—Diffusion Fiber Bed
DHA—District Health Authority (UK)
DHEP—Di (2-Ethylhexyl) Phthalate

DHEW—Department of Health, Education and Welfare (U.S.)
DHHS—Department of Health and Human Services (U.S.)
DHSS—Department of Health and Social Security (UK)
DIAL—Differential Absorption Lidar
DIH—Division of Industrial Hygiene (U.S.)
DiP—Discussion in Principal
DL—Detection Limit
DL—Difference Limen
DMA—Differential Mobility Analyzer
DMDS—Dimethyldisulfide
DME—Dimethyl Ether
DMHRF—Dual Media High Rate Filtration
DMI—Danish Meteorological Institute
DMR—Discharge Monitoring Report
DMS—Dimethylsulfide
DMSO—Dimethyl Sulfoxide
DMT—Dimethyl Tetraphthalate
DNA—Deoxyribonucleic Acid
DNAPL—Dense Non-Aqueous Phase Liquid
DNB—Dinitrobenzene
DNIP—Di-Isononyl Phthalate
DO—Dissolved Oxygen
DOC—Dissolved Organic Carbon
DOD—Department of Defense (U.S.)
DOE—Department of Energy (UK) (U.S.)
DoE—Department of the Environment (UK)
DoH—Department of Health (UK)
DOL—Department of Labor (U.S.)
DOM—Dissolved Organic Matter
DOT—Department of Transport(ation) (Canada) (UK) (U.S.)
DPBA—Diphenylbenzamidine
DPE—1,1-Diphenylethylene
DPKO—Dipyridyl Ketone Oxime
DPW—Department of Public Works
DQI—Data Quality Indicator
DQO—Data Quality Objective
DRE—Destruction and Removal Efficiency
DRI—Development of Regional Impact
DS—Degree of Succession
DS—Dry Scrubber
DS—Dry Solids
DSCD—Danish Research Agency Committee on Scientific Dishonesty
DSM—Demand Side Management
DSR—Dissimilatory Sulfate Reducing
DTH—Delayed Type Hypersensitivity
DTI—Department of Trade and Industry (UK)
DTLR—Department for Transport, Local Government and the Regions (UK)
DTPA—Diethylenetriaminepentaacetic Acid
DWF—Dry Weather Flow
DWL—Derived Working Limit
DWT—Deadweight Ton

EA—Energy Amplifier
EA—Environmental (Agency) (Authority) (UK)
EA—Environmental Assessment
EA—Environmental Audit
EAA—Electrical Aerosol Analyzer
EAC—Extruded Activated Carbon
E and C—Engineering and Construction
EAGE—European Association of Geoscientists and Engineers
EANET—Acid Deposition Monitoring Network in East Asia
EAP—Emergency Action Plan
EAPS—Environmental Aspects in Product Standards
EARP—Environmental Assessment and Review Process
EASST—European Association for the Study of Science and Technology
EB—Ethylene Glycol Monobutyl Ether
EBDC—Ethylene Bisdithiocarbamate
EBI—European Bioinformatics Institute
EBRD—European Bank of Reconstruction and Development
EB/S—Ethylbenzene/Styrene
EC—Effective Concentration
EC—Electrical Conductivity
EC—Engineering Council (UK)
EC—European Code
EC—European Commission
EC—European Community
ECCP—European Climate Change Programme
ECEPA—Environmental Challenge for European Port Authorities
ECF—Elemental Chlorine Free
EC_{50}—Mean Effective Concentration
ECHS—Environmental Clearing House System
ECL—Exposure Control Limit
ECMWF—European Center For Medium Range Weather Forecasting
ECOSS—Environmental Coalition of South Seattle
ECTFE—European Fluorochemicals Technical Committee
EC2000—Engineering Criteria 2000
ECU—Extractable Copper
ED—Electrodialysis
ED—Exposure Dose
ED—Exposure Duration
ED—Extractive Distillation
EDC—Electron Donor Compound
EDC—Ethylenedichloride
EDF—Environmental Defense Fund
EDF—European Development Fund
EDI—Electrodeionization
EDI—Electronic Data Interchange
EDIP—Environmental Design of Industrial Products
EDR—Electrodialysis Removal
EDR—Electrodialysis Reversal
EDTA—Ethylenediaminetetraacetate
EDXRA—Energy Dispersive X-Ray Analysis
EEA—Economic Espionage Act (U.S.)
EEA—European Environment Agency
EEB—European Environment Bureau
EEBPP—Energy Efficiency Best Program
EEC—European Economic Community
EELS—Electron Energy Loss Spectroscopy

EEMD—Environmental and Energy Management Directorate (UK)
EES—Environmental Extension Service
EF—Emissions Factor
EFCE—European Federation of Chemical Engineers
EFCEA—European Federation of Consulting Engineers Associations
EFFS—European Flood Forecasting System
EFR—External Floating Roof
EG—Ethylene Glycol
EGA—Exhaust Gas Analyzer
EGL—Energy Grade Line
EGR—Exhaust Gas Recirculation
EGSB—Expanded Granular Sludge Bed
EGTC—Exhaust Gas Test Cell
EHC—Environmental Health Criteria
EHO—Environmental Health Officer (UK)
EHS—Environmental Health and Safety
EHS—Extremely Hazardous Substance
EI—Energy Institute (UK)
EIA—Energy Information Administration (U.S.)
EIA—Environmental Impact Assessment
EIA—Environmental Investigation Agency (UK)
EIC—Energy Industry Council
EIC—Engineering Institute of Canada
EIO-LCA—Economic Input-Output Life-Cycle Assessment
EIP—Eco-Industrial Park
EIS—Environmental Impact Statement
EJ—Environmental Justice
EKMA—Empirical Kinetic Modeling Approach
ELNES—Energy Loss Near Edge Structure
EMAS—Ecoaudit and Management Scheme (EU)
EMAS—Employment Medical Advisory Service (UK)
eM&T—Energy Monitoring and Targeting
EMB—Eosin Methylene Blue
EMC—Event Mean Concentration
EMEC—European Marine Energy Centre
EMEP—European Monitoring and Evaluation Programme
EMF—Electromagnetic Field
EMIS—Environmental Management Information System
EML—Environmental Measurement Laboratory (U.S.)
EMP—Emergency Management Plan
EMS—Environmental Management Systems
EMTIC—Emission Measurement Technical Information Center (U.S.)
EN—European Committee for Standardization
ENEA—Italian National Agency for New Technology, Energy and the Environment
EngC—Engineering Council (UK)
ENHPA—European Network of Health Protection Agencies
ENSO—El Nino-Southern Oscillation
EO—Ethylene Oxide
EO—Executive Order (U.S.)
EOS—Earth Observation (Observing) System
EPA—Energy Policy Act (U.S.)
EPA—Environmental Protection Act (UK)
EPA—Environmental Protection Agency (U.S.)
EPB—Extractable Lead
EPCRA—Emergency Planning and Community Right-to-Know Act (U.S.)
EPE—Environmental Performance Evaluation
EPER—European Pollutant Emissions Register
EPFT—Swiss Federal Institute of Technology
EPH—Extractable Petroleum Hydrocarbons
EPHA—European Public Health Alliance
EPM—Environmental Pressure Measurement
EPMA—Electron Probe Micro Analysis
EPNdB—Effective Perceived Noise Level
EPR—European Pressurized Reactor
EPR—Extended Producer Responsibility
EPS—Extracellular Polymeric Substance
EPSRC—Engineering and Physical Sciences Research Council (UK)
EPTC—S-ethyl, N,N-dipropylthiocarbamate
EPS—Engineering Process Safety
EQIP—Environmental Quality Incentive Program
ER—Environmental Review
ERA—Environmental Risk Assessment
ERC—Emissions Reductions Credits
ERDA—Energy Research and Development Administration (U.S.)
ERP—Emergency Response Plan
ERP—Enterprise Resource Planning
ERS—Environmental Remote Sensing
ERTS—Earth Resources Technology Satellite
ESA—Ecological Society of America
ESA—Endangered Species Act (U.S.)
ESD—Electrostatic Dissipation
ESEF—European Science and Environment Forum
ESEM—Earth Systems Engineering and Management
ESEM—Environmental Scanning Electron Microscope
ESEP—European Science Exchange Program
ESF—Exploratory Studies Facility
ESF—European Science Foundation
ESP—Electrostatic Precipitator
ESR—Electron Spin Resonance
ESR—Environmental Screening Review
EST—Energy Saving Trust (UK)
ESV—Equivalent Sample Volume
ETA—Ecologically Threatened Area
ETAN—European Technology Assessment Network
ETB—Engineering and Technology Board (UK)
ETBE—Ethyl Tert-Butyl Ether
ETC/S—European Topic Center on Soil
ETG—Emissions Trading Group
ETH—Swiss Federal Institute of Technology
ETO—Ethylene Oxide
ETR—Ecological Tax Reform (UK)
ETS—Emissions Tracking System
ETS—Emissions Trading Scheme
ETS—Environmental Tobacco Smoke
ETSU—Energy Technology Support Unit (UK)
ETT—Effluent Toxicity Treatability
ETU—Ethylene Thiourea
EU—European Union

EURADWASTE—Nuclear Waste Reports (CEC)
EURING—European Engineer
EUROMETAUX—European Association of Metals
EuroREX—European Renewable Energy Exchange
EV—Electric Vehicle
eV—Electron Volt
EWEA—European Wind Energy Association
EWG—Environmental Working Group
EZN—Extractable Zinc

F—Fahrenheit
FACA—Federal Advisory Committee Act (U.S.)
FAIR—Federal Agriculture Improvement and Reform Act (U.S.)
FAME—Fatty Acid Methyl Ester
FAO—Food and Agriculture Organization (UN)
FAP—Filter Aid Polymer
FAR—Fatal Accident Rate
FAZ—Fly Ash Based Zeolite
FB—Fractional Bias
FBC—Fluidized Bed Combustion
FBI—Family Biotic Index
FC—Fecal Coliforms
FC—Fuel Cell
FCC—Fluid Catalytic Cracker
f/cc—Fibers per Cubic Centimeter
FCCC—Framework Convention on Climate Change
FCM—Fuel Containing Mass
FCO—Foreign and Commonwealth Office (UK)
FD—Formamide
FDA—Food and Drug Administration (U.S.)
FDD—Fault Detection and Diagnostics
FEA—Finite Element Analysis
FEANI—European Federation of National Engineering Associations
FEG—Field Emission Electron Gun
FEH—Flood Estimation Handbook
FEI—Fire and Explosion Index
FEMA—Federal Emergency Management Agency (U.S.)
FERC—Federal Energy Regulatory Commission (U.S.)
FETC—Federal Energy Technology Center (U.S.)
FEV—Forced Expired Volume
FF—Fabric Filter
FF—Future Forests
FFA—Federal Facilities Agreement (U.S.)
FFA—Free Fatty Acid
FFCA—Federal Facilities Compliance Act (U.S.)
FFDCA—Federal Food, Drug and Cosmetic Act (U.S.)
FFF—Forum for the Future
FGD—Flue Gas Desulfurization
FGR—Flue Gas Recirculation
FHSA—Family Health Service Authority (UK)
FIA—Flame Ionization Analyzer
FIA—Flow Injection Analysis
FID—Flame Ionization Detector
FIFRA—Federal Insecticide, Fungicide and Rodenticide Act (U.S.)
FIRM—Flood Insurance Rate Map
FISEC—Foundation for Scientific and Educational Cooperation
FML—Flexible Membrane Liner
FoE—Friends of the Earth
FONSI—Finding of No Significant Impact
FP—Fine Particle
FPC—Federal Power Commission (U.S.)
FPD—Flame Photometric Detector
FP6—Sixth Framework Program (EU)
EQPA—Food Quality Protection Act (U.S.)
FR—Federal Register (U.S.)
FRC—Functional Residual Capacity
FRP—Fiberglass Reinforced Plastics
FS—Fecal Streptococci
FSA—Food Standards Agency (UK)
FSR—Final Safety Report
FSU—Former Soviet Union
FT—Fourier Transform
FTA—Fault Tree Analysis
FTC—Federal Trade Commission (U.S.)
FVC—Forced Vital Capacity
FWPCA—Federal Water Pollution Control Act (U.S.)
FWQA—Federal Water Quality Administration (U.S.)
FWS—Fish and Wildlife Service (U.S.)
FWS—Free Water Surface
FY—Fiscal Year

GAC—Granular Activated Carbon
GAMOT—Global Acoustic Mapping of Ocean Temperatures
GAO—General Accounting Office (U.S.)
GATT—General Agreement on Trade and Tariffs
Gb—Giga Barrels
GC—Gas Chromatograph
GCCI—Global Climate Change Initiative
GC-MS—Gas Chromatograph and Mass Spectrometry
GCM—General Circulation Model
GCM—Global-Circulation Model
GCP—Good Combustion Practice
GDI—Gasoline Direct Injection
GDP—Gross Domestic Product
GEF—Global Environment Facility
G8—Group of Eight
GEM—Genetically Engineered Microbe
GEMI—Global Environmental Management Initiative
GEP—Global Environmental Program
GEV—Generalized Extreme Value
GEV—Gumbel Extreme Value
GEWEX—Global Energy and Water Cycle Experiment
GFDL—Geophysical Fluid Dynamics Laboratory
GHG—Greenhouse Gas
GHS—Globally Harmonized System
GIS—Geographic Information Systems
GISS—Goddard Institute of Space Studies
GJ—Giga Joules
g.l.c.—Ground Level Concentration
GLM—Generalized Linear Model
GLP—Good Laboratory Practice
GLUE—Generalized Likelihood Uncertainty Estimation

GLUMRB—Great Lakes–Upper Mississippi River Board (U.S.)
GM—Genetically Modified
GMA—General Mining Act (U.S.)
GMO—Genetically Modified Organism
GNP—Gross National Product
GOR—Gain Output Ratio
gpcd—Gallons per Capita per Day
GPD—Gallons per Day
GPH—Gallons per Hour
GPM—Gallons per Minute
GPO—Government Printing Office (U.S.)
GPP—Gross Primary Production
GPS—Global Positioning System
GRAS—Generally Recognized as Safe
GRAS/E—Generally Recognized as Safe and Effective
GRI—Gas Research Institute (U.S.)
GRIP—Greenland Ice Core Project
GRR—Gross Reproduction Rate
GSA—General Services Administration
GSA—Geothermal Steam Act (U.S.)
G7—Group of Seven
GSF—Nuclear Research Organization (Germany)
GSH-Px—Glutathione Peroxidase
GtC—Gigatons Carbon
GTL—Gas to Liquid
GTN—Global Trend Network
GV—Guideline Value
GVF—Gas Void Fraction
GVWHO—World Health Organization Guideline Value
GWP—Global Warming Potential
Gy—Gray

ha—Hectare
HA—Humic Acid
HAA—Haloacetic Acids
HAc—Acetic Acid
HACCP—Hazard Analysis of Critical Control Point
HAD—Hazards Assessment Document
HAP—Hazardous Air Pollutant
HAZMAT—Hazardous Material
HAZOP—Hazard and Operability Study
HBD—2-Hydroxyisobutyl Amide
HBM—2-Hydroxyisobutric Acid Methyl Ester
HCFC—Hydrochlorofluorocarbon
HCS—Hazard Communication Standard
HDM—Home Dust Mite
HDPE—High Density Polyethylene
HDS—Hydrodesulfurization
HDTMABr—Hexadecyltrimethyl Ammonium Bromide
HECA—Home Energy Conservation Act (UK)
HEES—Home Energy Efficiency Scheme (UK)
HEI—Health Effects Institute
HEL—Highly Erodible Land
HEPA—High Efficiency Particulate Air
HERP—Human Exposure/Rodent Potency
HEST—Hazardous Element Sampling Train
HEV—Hybrid Electric Vehicle
HEW—Department of Health, Education and Welfare (U.S.)
HF—Hydrogen Fluoride
HFC—Hydrofluorocarbon
HFE—Hydrofluoroether
HFO—Heavy Fuel Oil
HGH—Human Growth Hormone
HGL—Hydraulic Grade Line
HGM—Hazardous Gas Monitor
HGMS—High Gradient Magnetic Separation
HHS—Department of Health and Human Services (U.S.)
HIPPS—High Performance Power Systems
HIUI—Health Institute of the Uranium Industry (Czechoslovakia)
HIV—Human Immuno-Deficiency Virus
HLB—Hydrophilic-Lipophilic Balance
HLLW—High Level Liquid Waste
HLW—High Level Radioactive Waste
HM—Heavy Metal
HMDE—Hanging Mercury Drop Electrode
HMIP—HM Inspectorate of Pollution (UK)
HMNII—HM Nuclear Installations Inspectorate (UK)
HMSO—HM Stationary Office
HON—Hazardous Organic Neshap (National Emissions Standards for Hazardous Air Pollutants) (U.S.)
HPA—Health Protection Agency (UK)
HPA—Hetero Polyanion
HPC—Heterotrophic Plate Count
HPLC—High Performance Liquid Chromatography
HPS—Health Physics Society
HQ—Health Quotient
HR—Humidity Ratio
HRA—Halogen Resistant Azole
HRGC—High Resolution Gas Chromatography
HRS—Hazard Ranking Score (U.S.)
HRSA—Health Resources and Services Administration (U.S.)
HRT—Hydraulic Residence Time
HRTEM—High-Resolution Transmission Electron Microscopy
HSC—Health and Safety Commission (UK)
H6CB—3,3′ 4,4,4′ ,5-Hexachlorinated Biphenyl
HSCT—High Speed Civil Transport
HSDB—Hazardous Substances Data Base
HSE—Health and Safety Executive (UK)
HSWA—Hazardous and Solid Waste Amendments (U.S.)
HSWA—Health and Safety at Work Act (UK)
HTE—Horizontal Tube Evaporator
HTF—Heat Transfer Fluid
HTFW—High Temperature Fluid Wall
HTH—High Test Hypochlorite
HTlc—Hydrotalcite-Like-Compounds
HTME—Horizontal Tube Multiple Evaporator
HTU—Hydro Thermal Upgrading
HTRW—Hazardous, Toxic and Radioactive Waste
HUD—Department of Housing and Urban Development (U.S.)

HVAC—Heating, Ventilating and Air Conditioning
HVOC—Halogenated Volatile Organic Compound
HWC—Hazardous Waste Combustion
HWE—Healthy Worker Effect
HWI—Hazardous Wastes Inspectorate (UK)
HWIR—Hazardous Waste Identification Rule (U.S.)
HWL—High Water Level

I/A—Innovative and Alternative
IA—Index of Agreement
IAEA—International Atomic Energy Agency
IAHE—International Association for Hydrogen Energy
IAP—Indoor Air Pollution
IAP—Inter-Academy Panel
IAQ—Indoor Air Quality
IARC—International Agency for Research on Cancer
IAWPRC—International Association on Water Pollution Research and Control
IAWQ—International Association on Water Quality
IBC—International Bulk Container
IBRD—International Bank for Reconstruction and Development (World Bank)
IC—Internal Combustion
ICAA—International Council of Chemical Associations
ICAR—Indian Council of Agricultural Research
ICBEN—International Commission on the Biological Effects of Noise
ICBG—International Cooperative Biodiversity Grants Program
ICD—International List of Diseases, Injuries and Causes of Death
ICE—Institution of Civil Engineers (UK)
IC_{50}—Mean Inhibitory Concentration
ICIDI—Independent Commission on International Development Issues
ICP—Integrated Contingency Plan (U.S.)
ICPD—International Conference on Population and Development
ICP-OES—Inductively Coupled Plasma Optical Emission
ICR—Industrial Cost Recovery
ICR—Information Correction Rule
ICRAM—Central Institute for Scientific and Technological Research Applied to the Sea (Italy)
ICRP—International Commission on Radiological Protection
ICSU—International Council of Scientific Unions
ICT—Information Communication Technology
ICT—International Critical Table
IDEA—International Development Association
IDF—Intensity Duration Frequency
IDGEC—Institutional Dimensions of Global Environmental Change
IDLH—Immediately Dangerous to Life and Health
IDNDR—International Decade for Natural Disaster Reduction
IDSE—Initial Distribution System Evaluation
IDWSSD—International Drinking Water Supply and Sanitation Decade
IE—Cationic Exchange
IE—Institute of Energy (UK)
IEA—International Energy Agency
IEC—International Electrotechnical Commission
IEE—Institute of Electrical Engineers (UK)
IEEE—Institute of Electrical and Electronics Engineers
IESWTR—Interim Enhanced Surface Water Treatment Rule (U.S.)
IFB—Impaction Fiber Bed
IFC—International Finance Corporation
IFGR—Induced Flue Gas Recirculation
IFP—Institut Francais du Petrole (France)
IFR—Internal Floating Roof
IGBP—International Geosphere-Biosphere Program
IGC—Index of Grassland Condition
IGCC—Integrated Gasification Combined Cycle
IGO—Intergovernmental Organization
IGS—Inert Gas System
IGT—Institute of Gas Technology
IGY—International Geophysical Year
IHS—Indian Health Service (U.S.)
I/I—Infiltration/Inflow
IIASA—International Institute for Applied Systems Analysis
IIHE—International Institute for Hydraulic and Environmental Engineering
IJC—International Joint Commission
IKSR—International Commission for the Protection of the Rhine Against Pollution
ILGRA—Interdepartmental Liaison Group on Risk Assessment (UK)
ILO—International Labor Organization
ILW—Intermediate Level Radioactive Waste
IMDG—International Maritime Dangerous Goods Code
IME—Institution of Mechanical Engineers (UK)
IMF—International Monetary Fund
IMI—Imidazolinone
IMIQ—Mexican Institute of Chemical Engineers
IMO—International Maritime Organization
IMR—Infant Mortality Rate
IMS-ICS—Interagency Incident Management System-Incident Command System (U.S.)
INEEL—Idaho National Engineering and Environmental Laboratory (U.S.)
INSA—Institut National des Sciences Appliques (France)
InstE—Institute of Energy
INTRAVAL—International Geosphere Transport Model Validation
INWAC—IAEA Advisory Committee
I/O—Input/Output
IOA—Input-Output Analysis
IOC—Inorganic Chemical
IOD—Immediate Oxygen Demand
IoE—Institute of Energy
IOH—Institute of Occupational Hygiene (UK)
IoP—Institute of Physics (UK)
IOP—Indian Ocean Dipole
IOPC—International Oil Pollution Compensation Fund
IOWG—International Orimulsion Working Group

I/P—Incidence and Prevalence
IP—Inhalable Particles
IP—Institute of Petroleum
IPA—Isopropanol
IPACT—International Pharmaceutical Aerosol Consortium for Toxicology Testing
IPAT—Impact, Population, Affluence, and Technology
IPC—Integrated Pollution Control (UK)
IPCC—Intergovernmental Panel on Climate Change
IPCS—International Program on Chemical Safety
IPM—Inhalable Particulate Matter
IPM—Integrated Pest Management
IPP—Independent Power Producer
IPPC—Integrated Pollution Prevention and Control (EC)
IPPR—Institute for Public Policy Research
IPTS—Institute for Prospective Technological Studies
IR—Inactivation Ratio
IR—Infrared
IRAA—Indoor Radon Abatement Act (U.S.)
IRB—Institutional Review Board
IRIS—Integrated Risk Information System
IRP—Inward Processing Release
IRPP—Institute for Research on Public Policy
IRPTC—International Register of Potentially Toxic Chemicals
IRR—Ionizing Radiations Regulations (UK)
IRUS—Intrusion Resistant Underground Structures
ISA—Ideologically Structured Action
ISA—International Seabed Authority
ISES—International Solar Energy Society
ISI—Indian Standards Institution
ISIE—International Society for Industrial Ecology
ISM—Industrial, Scientific, Medical
ISO—Independent System Operator
ISO—International Standards Organization
ISPE—International Society of Pharmaceutical Engineers
ISRM—In-Situ Redox Manipulation
ISSX—International Society for the Study of Xenobiotics
ISTAS—International Symposium on Technology and Society
ISV—*In Situ* Vitrification
IT—Information Technology
ITC—Interagency Testing Committee (U.S.)
ITD—Ion Trap Detector
ITF—Industry Technology Facilitator
ITM—Ionic Transport Membrane
IUCN—International Union for Conservation of Nature and Natural Resources
IUFoST—International Union of Food Science and Technology
IUGG—International Union of Geodesy and Geophysics
IUPAC—International Union of Pure and Applied Chemistry
IURE—Inhalation Unit Risk Estimate
IWC—International Whaling Commission
IWM—Integrated Waste Management
IWPCC—Interstate Water Pollution Control Compact (U.S.)
IWS—Ionizing Wet Scrubber
IX—Ion Exchange

J—Joule
JAEC—Japanese Atomic Energy Commission
JAERI—Japanese Atomic Energy Research Institute
JAPCA—Journal of the Air Pollution Control Association
JAWWA—Journal of the American Water Works Association
JBR—Jet Bubbling Reactor
JECFA—Joint Expert Committee on Food Additives
JGOFS—Joint Global Ocean Flux Study
JMPR—Joint Meeting on Pesticide Residues
JRSH—Journal of the Royal Society for the Promotion of Health
JTU—Jackson Turbidity Unit
JWPCF—Journal of the Water Pollution Control Federation

K—Kelvin
KASAM—National Council for Nuclear Waste (Sweden)
KAST—Korean Academy of Science and Technology
KT—Clearness Index
kV—Kilovolt
Kw—Kilowatt
Kwh—Kilowatt Hour

LAER—Lowest Achievable Emission Rate
LAI—Leaf Area Index
LANDSAT—Earth Resources Technology Satellite
LANL—Los Alamos National Laboratory (U.S.)
LAPIO—Low API Oil
LAS—Laser Aerosol Spectrometer
LAS—Linear Alkyl Benzene Sulfonate
LASER—Light Amplification by Stimulated Emission of Radiation
LBL—Lawrence Berkeley Laboratory
LBP—Length Between Perpendiculars
LCA—Life-Cycle Assessment
LCC—Life Cycle Costing
LC_{50}—Mean Lethal Concentration
LCI—Life-Cycle Inventory
LCIP—Low Carbon Innovation Programme
LCPD—Large Combustion Plant Directive (EU)
LCR—Lead and Copper Rule
LD—Lethal Dose
LDA—Laminar Directional Airflow
LDAR—Leak Detection and Repair
LDC—Less Developed Country
LDC—London Dumping Convention
LDF—Local Deposition Fraction
LDH—Lactate Dehydrogenase
LD_{50}—Mean Lethal Dose
LDI—Laser Desorption/Ionization
LDPE—Low Density Polyethylene
LEA—Low Excess Air Operation
LED—Light Emitting Diode
LEL—Lower Explosive Limit
LET—Linear Energy Transfer
LFG—Landfill Gas
LFL—Lower Flammability Limit
LIBS—Laser Induced Breakdown Spectrometry

LID—Low Impact Development
LIDAR—Light Detection and Ranging
LILW-LL—Low and Intermediate Level Waste—Long Lived
LILW-SL—Low and Intermediate Level Waste—Short Lived
LLDPE—Linear Low Density Polyethylene
LLNL—Lawrence Livermore National Laboratory (U.S.)
LLW—Low Level Radioactive Waste
LMC—Lime-Magnesium Carbonate
LMO—Living Genetically Modified Organism
LMx—Low Mix Burner
LNAPL—Light Non-Aqueous Phase Liquid
LNB—Low NOx Burner
LNG—Liquefied Natural Gas
LNT—Linear No-Threshold
LOD—Limit of Deposition
LOEC—Lowest Observable Effects Concentration
LOEL—Lowest Observed Effect Level
LOI—Loss on Ignition
LOOP—Locally Organized and Operated Partnerships
LOPA—Layers of Protection Analysis
LPC—Limiting Permissible Concentration
Lpcd—Liters per Capita per Day
LPG—Liquefied Petroleum Gas
LPS—Lipopolysaccharide
LRT—Long-Range Transport
LS—Low Sulfur
LSE—London School of Economics
LSI—Langlier Saturation Index
LSM—Land Surface Model
LSTK—Lump-Sump Turnkey
LSHTM—London School of Hygiene and Tropical Medicine
LTAR—Long Term Acceptance Rate
LTR—Long Range Transport Model
LTVE—Long Tube Vertical Evaporator
LUC—Land Use Classification
LULU—Locally Unacceptable Land Use
LUST—Leaking Underground Storage Tank
LVHV—Low Volume High Velocity
LWL—Low Water Level
LWR—Light Water Reactor

MAA—Mycosporine Amino Acid
MAB—Man and the Biosphere (UNESCO)
MAC—Maximum Acceptable (Allowable) Concentration
MAC—(EEC)—Maximum Acceptable Concentration (European Economic Community)
MACS—Miniature Acid-Condensation System
MACT—Maximum Achievable Control Technology
MAFF—Ministry of Agriculture, Fisheries and Food (UK)
MAH—Maleic Anhydride
MANEB—Trimangol 80, Poligram M, Plantineb 80 PM
MANOVA—Multivariate Analysis of Variance
MA7CD10—Minimum Average 7 Consecutive Day 10 Year Flow
MARAMA—Mid-Atlantic Regional Air Use Management Association (U.S.)
MARC—Major Accident Reporting System (EU)
MAS—Mobile Aerosol Spectrometer
MATC—Maximum Acceptable Toxicant Concentration
MB—Mixed Bed
MBR—Membrane Bioreactor
MBT—Monobutyltin
MBL—Marine Boundary Layer
MC—Microcarrier
MCA—Manufacturing Chemists Association
MCA—Multicriteria Analysis
MCDSS—Multicriteria Decision Support System
MCFC—Molten Carbonate Fuel Cell
MCI—Management Charter Initiative (UK)
MCL—Maximum Contaminant Level
MCLG—Maximum Contaminant Level Goal
MCM—Master Chemical Mechanism
MCPA—4,Chloro-2-Methylphenoxyacetic Acid
MCR—National Institute of Materials and Chemical Research (Japan)
MCRT—Mean Cell Residence Time
MCS—Multiple Chemical Sensitivity
MCTT—Multichambered Treatment Train
mdd—md × day^{-1} × dm^{-2}
MDL—Method Detection Limit
MDNR—Missouri Department of Natural Resources
MDS—Multidimensional Scaling
ME—Multiple-Effect Evaporator
MEA—Monoethanolamine
MEA—Multinational Environmental Agreement
MED—Modified Electrodialysis
MED—Multieffect Distillation
MEK—Methyl Ethyl Ketone
MEMS—Microelectro-Mechanical Systems
MEPA—Meteorology and Environmental Protection Administration (Saudi Arabia)
MESL—Marine Environmental Studies Laboratory (IAEA)
MEUC—Major Energy Users’ Council (UK)
MF—Microfiltration
M/F—Mass/Food Ratio
MFA—Material Flow Accounting
MFN—Most Favored Status
MGD—Million Gallons per Day
mg/L—Milligrams per Liter (ppm)
MHD—Magnetohydrodynamics
MIBK—Methyl Isobutyl Ketone
MICROMORT—One in a Million Chance of Death from an Environmental Hazard
MIE—Magnetic Ion Exchange
MIE—Minimum Ignition Energy
MIGA—Multilateral Investment Guarantee Agency
MIPAES—Microwave Induced Plasma Atomic Emission Spectroscopy
MIPS—Material Intensity per Service Unit
MIR—Maximum Incremental Reactivity
MIR—Maximum Individual Risk
MIRAGE—CEC Radionuclide Geosphere Migration Project
MJO—Madden-Julian Oscillation
MLD—Mean Lethal Dose
MLD—Million Liters per Day
MLSS—Mixed Liquor Suspended Solids

MLVSS—Mixed Liquor Volatile Suspended Solids
MM—Methyl Mercaptan
MMA—Methyl Methacrylate
MMC—Monopolies and Mergers Commission (UK)
MMH—Monoethyl Hydrazine
MMR—Material Morbidity Rate
MMR—Maternal Mortality Rate
MMT—Methylcyclopentadienyl Manganese Tricarbonyl
MoA—Ministry of Agriculture
MoD—Ministry of Defense (UK)
MOEE—Ministry of Environment and Energy (Canada)
MoEF—Ministry of Environment and Forests (India)
MONICA—Monitoring of Trends and Determinants in Cardiovascular Disease
MOX—Mixed Oxide
MPD—Maximum Permissible Dose
MPE—Maximum Permissible Exposure
MPN—Most Probable Number
MPP—Macroporous Polymer
MPPE—Macroporous Polymer Extraction
mppcf—Millions of Particles per Cubic Foot
MPRSA—Marine Protection, Research and Sanctuaries Act (Ocean Dumping Act) (U.S.)
MR—Magnetic Resonance
MR—Mass Removal
MRC—Medical Research Council (UK)
MRDL—Maximum Residual Disinfectant Level
MRDLG—Maximum Residual Disinfectant Level Goal
MRF—Material Recovery Facilities
MRF—Municipal Recycling Facilities
MRI—Mean Recurrence Interval
MRS—Monitored Retrievable Storage
MS—Management System
MS—Mass Spectrometry
MSAT—Mobile Source Air Toxics
MSDS—Material Safety Data Sheets (U.S.)
MSE—Mean Square Estimate
MSG—Monosodium Glutamate
MSHA—Mine Safety and Health Administration (U.S.)
MSL—Mean Sea Level
MSMR—Mean Standardized Mortality Rate
MSS—Multispectral Scanner
MSW—Municipal Solid Waste
MSY—Maximum Sustainable Yield
m.t.—Metric Ton
Mta^{-1}—Million Tons per Year (Annum)
MTB—Multiply Twinned Particles
MTBE—Methyl Tert-Butyl Ether
MTBF—Mean Time between Failures
MtCeq—Million Tons Carbon Equivalent
MTD—Maximum Tolerable Dose
MTHF—Methyltetrahydrofuran
MTI—Mixture Toxicity Index
Mtoe—Million Tons Oil Equivalent
mtoe—Million Tonnes Oil Equivalent
MVC—Mechanical Vapor Compression
MW—Megawatt
MWCO—Molecular Weigh Cutoff
MWT—Multiple Well Tracer Test

N—Newton
NA—Nutrient Agar
NAA—Neutron Activation Analysis
NAAQS—National Ambient Air Quality Standards (U.S.)
NACA—National Agricultural Chemicals Association (U.S.)
NADH—Nicotinamide Adenine Dinucleotide
NAD—Nitric Acid Dihydrate
NADP/NTN—National Atmospheric Deposition Program/National Trends Network
NADW—North Atlantic Deep Water
NAE—National Academy of Engineering (U.S.)
NAFO—North Atlantic Fisheries Organization
NAFTA—North American Free Trade Agreement
Nagra—Waste Disposal Organization (Switzerland)
Naics—North American Industry Classification System
NaNp—Sodium Napthalenide
NAO—National Audit Office (UK)
NAO—North Atlantic Oscillation
NAPAG—National Academics Policy Advisory Group (UK)
NAPAP—National Acid Precipitation Assessment Program (U.S.)
NAPCA—National Air Pollution Control Administration (U.S.)
NaPEG—Sodium Polyethylene Glycol
NAPL—Non-Aqueous Phase Liquid
NAS—National Academy of Sciences (U.S.)
NAS—National Audubon Society
NAST—National Assessment Synthesis Team
NASA—National Aeronautics and Space Administration (U.S.)
NAST—National Assessment Synthesis Team (U.S.)
NAT—Nitric Acid Trihydrate
NB—Nitrobenzene
NB—Nutrient Broth
NBOD—Nitrogenous Biochemical Oxygen Demand
NBS—National Bureau of Standards (U.S.)
NC—Number Concentration
NCAB—National Cancer Advisory Board (U.S.)
NCAR—National Center for Atmospheric Research (U.S.)
NCB—National Coal Board (UK)
NCC—Nature Conservancy Council (UK)
NCDC—National Climate Data Center
NCEA—National Center for Environmental Assessment (U.S.)
NCEES—National Council of Examiners for Engineers and Surveyors (U.S.)
NCEJN—North Carolina Environmental Justice Network
NCEP—National Centers for Environmental Prediction (U.S.)
NCHGR—National Center for Human Genome Research (U.S.)
NCHS—National Center for Health Statistics (U.S.)
NCI—National Cancer Institute (U.S.)
NCI-MS—Negative Chemical Ionization–Mass Spectrometry
NCP—National Contingency Plan (U.S.)

NCRP—National Council on Radiation Protection and Measurements (U.S.)
NCS—Notification of Compliance Status (U.S.)
ND—Not (None) Detected (Detectable)
NDA—New Drug Approval (Application)
NDA—Nuclear Decommissioning Authority (UK)
NDIR—Nondispersive Infrared
NEA—Nuclear Energy Agency (UK)
NEAT—National Environmental Achievement Award
NECI—Network Coordinating Institute
NEDO—New Energy and Industrial Technology Development Organization (Japan)
NEERI—National Environmental Engineering Research Institute (India)
NEL—No Effect Level
NEP—National Energy Policy (U.S.)
NEPA—National Environmental Policy Act (U.S.)
NEPA—National Environmental Protection Agency (PRC)
NERC—National Environmental Research Council (UK)
NERI—National Environmental Research Institute (Denmark)
NERI—Nuclear Energy Research Initiative
NERL—National Exposure Research Laboratory (U.S.)
NESCAUM—Northeast States for Coordinated Air Use Management (U.S.)
NESDIS—National Environmental Satellite Data and Information Center
NESHAP—National Emissions Standards for Hazardous Air Pollutants (U.S.)
NESS—National Environmental Satellite Service (U.S.)
NETA—New Electricity Trading Agreements (UK)
NETCEN—National Environmental Technology Centre (UK)
NETL—National Energy Technology Laboratory (U.S.)
NEXRAD—Next Generation Radar
NF—Nanofiltration
NFFO—Non-Fossil Fuel Obligation
NFIP—National Flood Insurance Program (U.S.)
NFMA—National Forest Management Act (U.S.)
NFPA—National Fire Protection Association (U.S.)
NG—No Guideline Value
NGO—Non-Government Organization (UN)
NGV—Natural Gas Vehicle
NHC—National Hurricane Center (U.S.)
NHEERL—National Health and Environmental Effects Research Laboratory (U.S.)
NHER—National Home Energy Rating (UK)
NHS—National Health Service (UK)
NIAID—National Institute of Allergy and Infectious Disease (U.S.)
NIBS—National Institute of Building Sciences (U.S.)
NICOLE—Network for Industrially Contaminated Land in Europe
NIEHS—National Institute of Environmental Health Sciences (U.S.)
NIH—National Institute of Health (U.S.)
NII—Nuclear Installations Inspectorate (UK)
NIMBY—Not in My Backyard
NIOSH—National Institute for Occupational Safety and Health (U.S.)
NIREX—Nuclear Industry Radioactive Waste Executive (UK)
NIST—National Institute of Standards and Technology (U.S.)
NLER—Nitrogen-Limited Enrichment
NL/h—Normal Liters per Hour
NLM—National Library of Medicine
NMFS—National Marine Fisheries Service (U.S.)
NMC—National Meteorological Centre (UK)
NMHC—Non-Methane Hydrocarbon
NMMAPS—National Morbidity, Mortality and Air Pollution Study (U.S.)
NMOC—Non-Methane Organic Carbon
NMR—Neonatal Mortality Rate
NMR—Nuclear Magnetic Resonance
NMSE—Normalized Mean of the Square of the Error
NMVOC—Non-Methane Volatile Organic Compound
NNI—Noise and Number Index
NNR—National Nature Reserve (UK)
NNR—Net Reproductive Rate
NOAA—National Oceanographic and Atmospheric Administration (U.S.)
NOAEL—No Observable Adverse Effect Level
NOEC—No Observable Effects Concentration
NOEL—No Observed Effect Level
NOHSCP—National Oil and Hazardous Substance Contingency Plan (U.S.)
NOM—Natural Organic Matter
NONHEL—Non-Highly Erodible Land
NOS—National Occupational Standard
NOS—National Ocean Survey (U.S.)
NO2, 3—Nitrite Plus Nitrate
NOx—Nitrogen Oxides
NOy—Reactive Nitrogen
NP—Neutralization Point
NPAA—Noise Pollution and Abatement Act (U.S.)
NPCA—National Parks and Conservation Association (U.S.)
NPDES—National Pollutant Discharge Elimination System (U.S.)
NPL—National Priority List (U.S.)
NPP—Net Primary Production (Productivity)
NPP—Nonprecipitated Phosphorous
NPP—Nuclear Power Plant
NPR—National Performance Review (U.S.)
NPRI—National Pollution Release Inventory (Canada)
NPRM—Notice of Proposed Rule Making (U.S.)
NPS—National Park Service (U.S.)
NPSH—Net Positive Suction Head
NRA—National Rivers Authority (UK)
NRC—Nuclear Regulatory Commission (U.S.)
NRC—National Research Council (Canada) (U.S.)
NRCS—Natural Resources Conservation Service
NRD—Natural Resources Damage
NRDA—Natural Resource Damage Assessment
NRDC—Natural Resources Defense Council

NREL—National Renewable Energy Laboratory (U.S.)
NRMRL—National Risk Management Research Laboratory (U.S.)
NRPB—National Radiological Protection Board (UK)
NRR—Net Reproduction Rate
NRT—National Response Team (U.S.)
NSERC—National Science and Engineering Research Council (Canada)
NSF—National Science Foundation (U.S.)
NFSOM—Near-Field Scanning Optical Microscopy
NSPE—National Society of Professional Engineers (U.S.)
NSPS—New Source Performance Standards (U.S.)
NSR—New Source Review
NTA—Nitrilotriacetic Acid
NTI—Nuclear Threat Initiative
NTN—National Trends Network
NTNU—Norwegian University of Science and Technology
NTP—National Toxicology Program (U.S.)
NTS—Not to Scale
NTSB—National Transportation Safety Board (U.S.)
NTTC—National Technology Transfer Center (U.S.)
NTU—Nephelometric Turbidity Unit
Nu—Nusselt Number
NURP—National Urban Runoff Program (U.S.)
NWP—Numerical Weather Prediction
NWR—National Wildlife Refuge (U.S.)
NWS—National Weather Service (U.S.)
NWTRB—Nuclear Waste Technical Review Board (U.S.)

OA—Overfire Air
OBO—Ore-Bulk-Oil
OCAG—Off-Site Consequences Analysis Guidance (U.S.)
OCCM—Office of Air Quality Planning and Standards Control Cost Manual (U.S.)
OCEF—Overseas Economic Cooperation Fund (Japan)
OCGCM—Ocean Coupled General Circulation Model
OCP—Organopesticides
OCPSF—Organic Chemicals, Plastics and Synthetic Fibers
OCS—Outer Continental Shelf
O.D.—Optical Density
ODA—Overseas Development Agency (UK)
ODP—Ocean Drilling Programme (Australia)
ODP—Ozone Depletion Potential
ODS—Ozone Depleting Substance
OECD—Organization for Economic Cooperation and Development
OEL—Occupational Exposure Limit
O8CDD—Octochloro Dibenzo-p-Dioxin
O8CDF—Dibenzofuran
OEM—Office of Emergency Management (U.S.)
OEM—Original Equipment Manufacturer
OERR—Office of Emergency and Remedial Response (U.S.)
OES—Occupational Exposure Standard (UK)
OES—Office of Endangered Species (U.S.)
OFA—Overfire Air
OHS—Occupational Hygiene Secretariat (UK)
OH & S—Occupational Health and Safety
OIE—Organisation Internationale des Epizooties
OIES—Oxford Institute for Energy Studies (UK)
OLED—Organic Light-Emitting Diode
OLR—Organic Loading Rate
O & M—Operation and Maintenance
OMB—Office of Management and Budget (U.S.)
OMG—Old Mixed Grade
OMT—Open Market Trading
OMTR—Open Market Trading Credits Rule (U.S.)
OMZ—Oxygen-Minimum Zone
ONBC—Overnight Nutrient Broth Culture
ONDRAF—Waste Disposal Organization (Belgium)
OP—Organophosphate
OPA—Oil Pollution Act (U.S.)
OPCW—Organization for Prevention of Chemical Warfare
OPEC—Organization of Petroleum Exporting Countries
OPETS—Organization for the Promotion of Energy Technologies (EU)
OP-FTIR—Open-Path Fourier Transform Infrared
OPLA—High Level Waste Disposal Program (Holland)
OPS—Office of Pipeline Safety (U.S.)
OR—Odds Ratio
ORD—Office of Research and Development (U.S.)
ORI—Office of Research Integrity (U.S.)
ORNL—Oak Ridge National Laboratory (U.S.)
ORP—Oxidation-Reduction Potential
ORSANCO—Ohio River Sanitation Commission (U.S.)
OSAT—On-Site Assistance Team (EC)
OSHA—Occupational Safety and Health Act (U.S.)
OSHA—Occupational Safety and Health Administration (U.S.)
OSPAR—Oslo-Paris Agreement
OSPM—Operational Street Pollution Model
OSPREY—Ocean Swell Powered Renewable Energy (UK)
OST—Office of Science and Technology (UK) (U.S.)
OSW—Office of Saline Water (U.S.)
OSWER—Office of Solid Waste and Emergency Response (U.S.)
OTA—Office of Technology Assessment (U.S.)
OTEC—Ocean Thermal Energy Conversion
OTR—Oxygen Transfer Rate
OTR—Ozone Transport Region (U.S.)
OUR—Oxygen Uptake Rate
OVA—Organic Vapor Analysis
OW—Oil in Water
Ox—Total Oxidants

P—Polystyrene
Pa—Pascal
PA—Polymide
PABA—Para-Amino-Benzoic Acid
PAC—Polyaluminum Chloride
PAC—Powered Activated Carbon
PAC1—Polyaluminum Chloride
PACT—Programme for Alternative Fluorocarbon Toxicity Testing
PAFC—Phosphoric Acid Fuel Cell
PAH—Polynuclear Aromatic Hydrocarbons

PAHO—Pan American Health Organization (WHO)
PAL—Plantwide Applicability Limits
PAN—Peroxyacetyl Nitrate
P and T—Pump and Treat
Paracom—Paris Commission
PARC—Pan African Rinderpest Campaign
PARCCS—Precision, Accuracy, Representativeness, Comparability, Completeness, Sensitivity
PARIS—Program for Assisting the Replacement of Industrial Solvents
PATH—Plan for Analyzing and Testing Hypotheses
PBL—Planetary Boundary Layer
PBMR—Pebble-Bed Modular Reactor
PBS—Package Boiler Simulator
PC—Polycarbonate
PC—Pulverized Coal
PCA—Plate Count Agar
PCA—Principal Component Analysis
PCB—Polychlorinated Biphenyl
PCBP—Polychloro Bi Phenylene
PCBz—Polychloro Benzene
PCCY—Polychlorinated Chrysene
PCD—Particle Charge Detector
PCD—Process Control Diagram
PCDD—Polychlorodibenzo-p-Dioxin
PCDET—Post-Completion Discrete Extraction Test (U.S.)
PCDF—Polychlorinated Dibenzofuran
PCDPE—Polychlorinated Diphenyl Ether
PCE—Perchlorethylene
PCE—Tetrachloroethylene
pCi/L—Picocuries per Liter
PCN—Polychlorinated Napthalene
PCP—Pentachlorophenol
PCPA—Post-Closure Performance Assessment (U.S.)
PCPY—Polychlorinated Pyrene
PCQ—Polychlorinated Quaterphenyl
PCQE—Polychlorinated Quaterphenyl Ether
PCR—Polymerase Chain Reaction
PCV—Positive Crankcase Ventilation
PDA—Potato Dextrose Agar
PDE—Partial Differential Equation
PDF—Probability Density Function
PDO—Pacific Decadal Oscillation
PE—Performance Evaluation
PE—Polyethylene
PE—Professional Engineer (U.S.)
PEA—Performance Evaluation Audit
PEACE—Pollution Effects on Asthmatic Children in Europe
PEC—Process Economic Program
PEELS—Parallel Acquisition Systems
PEFR—Peak Expiratory Flow Rate
PEI—Potential Environmental Impact
PEIs—Professional Engineering Institutions (UK)
PEL—Permissible Exposure Limit
PEM—Proton Exchange Membrane
PEMFC—Proton Exchange Membrane Fuel Cell
PEN—Polyethylene Napthalate
P ENG—Professional Engineer (Canada)
PEO—Professional Engineers Ontario
PER—Partial Exfiltration System
PES—Project Environmental Summary
PES—Public Electricity Supplier
PET—Polyethylene Terephthalate
PF—Phenol-Formaldehyde
PF—Pulverized Fuel
P5CB—3,3' ,4,4' ,5-Pentachlorinated Biphenyl
PFBC—Pressurized Fluidized Bed Combustion
PFC—Perfluoro (Carbon) Compounds
PFD—Process Flow Diagram
PFR—Plug Flow Reactor
PFRP—Process to Further Reduce Pathogens
PFS—Pulverized Fly Ash
PFT—Peak Flame Temperature
PHA—Process Hazard Analysis
PHA—Pulse Height Analysis
PHARE—CEC Program to Assist Eastern European Countries
PHB—Polyhydroxy Butyrate
PHC—Petroleum Hydrocarbons
PHS—Public Health Service (U.S.)
PI—Principal Investigator
PIC—Prior Informed Consent Procedure
PIC—Product of Incomplete Combustion
PIEL—Pharmacologically Insignificant Exposure Limit
PIOT—Physical Input-Output Table
PIRG—Public Interest Research Group (U.S.)
PIXIE—Particle Induced X-Ray Emission
PL—Public Law (U.S.)
PLA—Polylactic Acid
PLC—Programmable Logic Controller
PLL—Probable Loss of Life
PLM—Polarized Light Microscopy
PM—Particulate Matter
PM0.1—Ultrafine Particle
PM2.5—Particles Less than 2.5 Micrometers
PM10—Particles Less than 10 Micrometers
PMA—Phenyl Mercuric Acetate
PMMA—Polymethylmethacrylate
PMN—Phenyl Mercuric Nitrate
PMM—Polymorphonuclear Neutrophils
PMN—Polymorphonuclear Neutrophils
PMN—Premanufacture Notice (U.S.)
PMR—Proportionate Mortality Rate
PNC—Japanese Nuclear Power Corporation
PNdB—Perceived Noise Level
PNGV—Partnership for a New Generation of Vehicles
PNL—Pacific Northwest Laboratory (U.S.)
PNNL—Pacific Northwest National Laboratory (U.S.)
PODAAC—Physical Oceanography Distributed Active Archive
POE—Point of Entry
POHC—Principal Organic Hazardous Constituent
POL—Project Objectives Letter (U.S.)
PolyTHF—Polytetrahydrofuran
POM—Polycyclic Organic Matter

POM—Princeton Ocean Model
POP—Persistent Organic Pollutant
POT—Peaks-Over Threshold
POTW—Publicly Owned Treatment Works (U.S.)
POU—Point of Use
PP—Polypropylene
PP—Precautionary Principle
ppb—Parts per Billion
ppbv—Parts per Billion by Volume
PPE—Personal Protective Equipment
PPEQ—Pre-Project Environmental Questionnaire
ppm—Parts per Million (Milligrams per Liter) (Grams per Cubic Meter)
ppmv—Parts per Million by Volume
ppmw—Parts per Million by Weight
ppt—Parts per Trillion
ppth—Parts per Thousand
PPU—Pertinent Process Unit
Pr—Prandt1 Number
PRA—Paperwork Reduction Act (U.S.)
PRA—Probabilistic Risk Assessment
PRAS—Prereduced Anaerobically Sterilized
PRG—Preliminary Remediation Goals (U.S.)
PRP—Potentially Responsible Party (U.S.)
PRZ—Potential Repository Zone
PS—Polystyrene
PSA—Probabilistic Safety Assessment
PSAC—President's Science Advisory Committee (U.S.)
PSC—Polar Stratospheric Cloud
PSD—Prevention of Significant Deterioration (U.S.)
PSF—Peat-Sand Filter
PSF—Pounds per Square Foot
PSI—Pounds per Square Inch
PSM—Process Safety Management (U.S.)
PSR—Preliminary Safety Report
PSRP—Process to Significantly Reduce Pathogens
PTFE—Polytetrafluoroethylene
P3—Public Performance-Based Permitting Program
PTM—Photochemical Trajectory Model
PTR—Project Tracking Register
PTWI—Provisional Tolerable Weekly Intake
P2—Pollution Prevention
PU—Polyurethane
PULSAR—Phillips' Ultra Low Sulfur Atmospheric Residue
PV—Permanganate Value
PV—Photovoltaic
PV—Polyvinyl
PVA—Polyvinyl Acetate
PVA—Process Vulnerability Analysis
PVC—Polyvinyl Chloride
PVS—Physical Vapor Synthesis
PVT—Photovoltaic/Thermal
PWS—Public Water System
PZC—Point of Zero Charge

QA—Quality Assurance
QACC—Quaternary Ammonium Compounds
QAPP—Quality Assurance Project Plan (U.S.)
QC—Quality Control
QEP—Qualified Environmental Professional
QF—Qualified Facility
QIP—Quality Improvement Program
QPF—Quantitative Precipitation Forecast
QRA—Quantitative Risk Assessment
QSAR—Quantitative Structure Activity Relationship
QUANGO—Quasi Autonomous Non-Governmental Organization
QUARG—Quality of Urban Air Review Group

R—Rankine
R—Universal Gas Constant
r—Roentgen
RA—Rapid Appraisal
RAB—Registrar Accreditation Board
RACT—Reasonably Available Control Technology
Rad—Roentgen-Absorption-Dose
RADWASS—IAEA Publication on Radioactive Waste
RAF—Radiation Amplification Factor
RAMS—Regional Atmospheric Model System
RAP—Rapid Assessment Program
RAR—Reasonable Assumed Resource
RaSoS—Raman Sort Spectrometer
RBC—Rotating Biological Contactor
RBCA—Risk Based Corrective Action
RBE—Relative Biological Effectiveness
RBSL—Risk-Based Screening Levels
RCEP—Royal Commission on Environmental Pollution (UK)
RCM—Regional Climate Modelling
RCO—Regenerative Catalytic Oxidation
RCRA—Resource Conservation and Recovery Act (U.S.)
RCT—Reference Control Technology
R & D—Research and Development
RD & D—Research, Development and Demonstration
RDF—Refuse-Derived Fuel
RDS—Residue-Oil Hydrodesulfurization
Re—Reynolds Number
REACH—Registration, Evaluation, Authorization of Chemicals
RECLAIM—Regional Clean Air Incentives Market (U.S.)
REI—Regional Environmental Initiative
REL—Recommended Exposure Limit
REMA—Regulatory Environmental Modelling of Antifoulants
REO—Renewable Energy Obligation
RERF—Radiation Effects Research Foundation
Rem—Roentgen-Equivalent-Man
Rep—Roentgen-Equivalent-Physical
RF—Radio Frequency
RFBR—Russian Federation for Basic Research
RFCC—Residue-Oil Fluid Catalytic Cracking
REF—Resources for the Future
RFG—Reformulated Gasoline
RFLPS—Restriction-Fragment-Length Polymorphisms
RGH—Renewably Generated Hydrogen
RGR—Relative Growth Rate

RH—Relative Humidity
RHA—Regional Health Authority (UK)
RI—Remedial Investigation
RI—Return Interval
Ri—Richardson Number
RIA—Regulatory Impact Analysis
RIBA—Royal Institution of British Architects
RI-FS—Remedial Investigation and Feasibility Study (U.S.)
RIIA—Royal Institute of International Affairs (UK)
RL50—Residue Half Life
RMCL—Recommended Maximum Contaminant Level
RME—Reasonable Maximum Exposure
RMLT—Regression Models in Life Tables
RMP—Risk Management Program (Plan) (U.S.)
RMR—Required Mass Removal
RNA—Ribonucleic Acid
RNI—Rates of Natural Increase
RO—Renewables Obligation
RO—Reverse Osmosis
ROC—Renewable Obligation Certificate
ROD—Record of Decisions (Superfund) (U.S.)
ROFA 6—Residual Oil Fly Ash (No. 6 Fuel Oil)
RORO—Roll On Roll Off
RoSPA—Royal Society for the Prevention of Accidents (UK)
RP—Return Period
R/P—Reserve/Production
RPA—Radiation Protection Act (UK)
RPA—Renewable Power Association
RPE—Respiratory Protection Equipment
RPP—Radiation Protection Program (Canada)
RPS—Renewable Portfolio Standard
RQ—Reportable Quantity (U.S.)
RR—Relative Risk
RRRR—Reduce/Recovery/Recycle/Reuse
RRT—Relative Retention Time
RS—Royal Society (UK)
RSA—Radioactive Substances Act (UK)
RSC—Reactor Safety Commission (Germany)
RSC—Royal Society of Chemistry (UK)
RSD—Relative Standard Deviation
RSH—Royal Society of Health (UK)
RTD—Residence Time Distribution
RTDF—Remediation Technologies Development Forum
RTO—Regenerative Thermal Oxidation
RVP—Reid Vapor Pressure
RWMAC—Radioactive Waste Management Advisory Committee (UK)

SAED—Selected Area Electron Diffraction
SAGE—Stratospheric Aerosol and Gas Experiment
S and T—Science and Technology
SAR—Sodium Adsorption Ratio
SAR—Specific Absorption Rate
SAR—Structure-Activity Relationship
SARA—Superfund Amendments and Reauthorization Act (U.S.)
SARTOR—Standards and Routes to Registration (UK)
SASS—Source Assessment Sampling System
SAVEII—Specific Action for Vigorous Energy Efficiency (EU)
SBR—Sequencing Batch Reactor
SBR—Stillbirth Rate
SBS—Sick Building Syndrome
SBT—Segregated Ballast Tank
Sc—Schmidt Number
SC—Sierra Club
SCADA—Supervisory Control and Data Acquisition
SCAPS—Site Characterization and Analysis Penetrometer System
SCAQMD—South Coast Air Quality Management District
SCF—Standard Cubic Feet
SCF—Supercritical Fluid
SCOPE—Scientific Committee on Problems of the Environment
SCR—Selective Catalytic Reduction
SCS—Soil Conservation Service (U.S.)
SCW—Supercritical Water
SCWO—Supercritical Wet (Water) Oxidation
SD—Solar Distillation
SD—Standard Deviation
SD—Sustainable Development
SDBS—Sodium Dodecyl Benzene Sulfonate
SDI—Silt Density Index
SDR—Surplus Discrete Reductions (U.S.)
S & DSI—Stiff and Davis Stability Index
SDWA—Safe Drinking Water Act (U.S.)
SEA—Strategic Environmental Assessment
SEARR—South East Asia Rainforest Research Program
SEASAT—Earth Satellite for Sea Surveys
SEC—Site-Specific Energy Consumption
SEC—Size Exclusion Chromatography
SED—Soil Evacuation and Disposal Plan (U.S.)
SEE—Society of Environmental Engineers (UK)
SEER—Surveillance, Epidemiology, and End Results
SEFI—European Society for Engineering Education
SELCHP—South East London Combined Heat and Power (UK)
SEM—Scanning Electron Microscope
SEP—Supplemental Environmental Project
SERC—Scientific and Engineering Research Council (UK)
SERI—Solar Energy Research Institute
SET—Science, Engineering, Technology
SETAC—Society of Environmental Toxicology and Chemistry
SF—Safety Factor
SFIP—Sector Facility Indexing Project (U.S.)
SG—Sustained Growth
Sh—Sherwood Number
SHASE—Society of Heating, Air Conditioning and Sanitary Engineers (Japan)
SHE—Safety, Health and the Environment
SHE—Systeme Hydrologique European
SHP—Shaft Horsepower
SHPO—State Historic Preservation Officer (U.S.)

SHWP—Seismic Hazards Working Party
SI—International System of Units (Systeme International d'Unites)
SI—Saturation Index
SI—Statutory Instruments (UK)
SIC—Standard Industrial Classification (U.S.)
SIL—Safety Integrity Level
SIM—Selective Ion Monitoring
SIMS—Secondary Ion Mass Spectrometry
SIP—Site Implementation Plan (U.S.)
SIP—Sterilization in Place
SIPI—Scientists' Institute for Public Information
SIS—Susceptible-Infected-Susceptible
SKAPP—Scientific Knowledge and Public Policy
SKI—Radiation Protection Institute (Sweden)
SL—Surface Layer
SLAMM—Source Loading and Management Model
SLAPP—Strategic Lawsuit against Public Participation (Canada)
SLM—Supported Liquid Membrane
SMB—Simulated Moving Bed
SMCL—Secondary Maximum Contaminant Level
SMCRA—Surface Mining Control and Reclamation Act (U.S.)
SME—Solar Mesospheric Explorer
SMP—Soluble Microbial Product
SMPS—Scanning Mobility Particle Sizer
SMR—Standardized Mortality Ratio (Rate)
SMZ—Surface Modified Zeolite
SNAP—Significant New Alternatives Policy
SNARL—EPA Suggested No Adverse Response Level (U.S.)
SNCR—Selective Noncatalytic Reduction
SNL—Sandia National Laboratory (U.S.)
SNL—Scottish Nuclear Ltd
SNL—Sense Noise Level
SOA—Secondary Organic Aerosols
SOC—Soluble Organic Compound (Chemical)
SOC—Synthetic Organic Chemical
SOCMI—Synthetic Organic Chemical Manufacturing Industry
SOD—Sediment Oxygen Demand
SOFC—Solid Oxide Fuel Cell
SOI—Southern Oscillation Index
SoPHE—Society of Public Health Engineers
SOx—Sulfur Oxides
SPCC—Spill Prevention Control and Countermeasures
SPE—Solid Phase Extraction
SPE—Survivor Population Effect
SPFC—Solid Polymer Fuel Cell
SPI—Society of the Plastics Industry
SPM—Scanning Probe Microscopy
SPME—Solid-Phase Microextraction
SPOT—Systeme Pour l'Observation de La Terre (France)
SPSS—Sulfur Polymer Stabilization and Solidification
SRC—Science Research Council (UK)
SRF—State Revolving Fund (U.S.)
SR&O—Statutory Regulations and Orders (UK)
SRT—Solids Retention Time
S/S—Solidification/Stabilization
SS—Suspended Solids
SSA—Specific Surface Area
SSBY—Sewage Solids By-Product
SSD—Safe Separation Distance
SSHRC—Social Sciences and Humanities Research Council (Canada)
SSI—Nuclear Safety Inspectorate (Sweden)
SSIT—Society on Social Implications of Technology
SSOW—Source Separated Organic Waste
SSSI—Site of Special Scientific Interest (UK)
SST—Sea-Surface Temperature
SST—Supersonic Transport
SSTL—Site Specific Target Levels
STAPPA—State and Territorial Air Pollution Program Administrators (U.S.)
STEL—Short Term Exposure Limit
STEM—Standard Transmission Electron Microscope
STEP—Septic Tank Effluent Pump
STEP—Standard for Exchange of Production
STM—Scanning Tunneling Microscopy
STORM—Storage Treatment Overflow and Runoff Model
STP—Standard Temperature and Pressure
STS—Supercooled Ternary Solution
SU—Standard Unit
SUVA—Specific Ultraviolet Absorbance
Sv—Sievert
SVE—Soil Vapor Extraction
SVI—Sludge Volume Index
SVIN—Swiss Association of Women Engineers
SVOC—Semivolatile Organic Compound
SVR—Sludge Volume Ratio
SWAMP—Storm Water Assessment, Monitoring and Performance Program (Canada)
SWDA—Solid Waste Disposal Act (U.S.)
SWE—Society of Women Engineers (U.S.)
SWH—Solar Water Heater
SWJ—Sewage Works Journal
SWMM—Stormwater Management Model

TA—Total Alkalinity
TAC—Total Annual Cost
TADC—Tire-Derived Activated Carbon
TAEE—Tert-Amyl Ethyl Ether
TAME—Tert-Amyl Methyl Ether
TAO—Tropical Atmosphere Ocean
TAPPI—Technical Association of the Pulp and Paper Industry
TAW—Technical Advisory Committee on Water Defenses (Holland)
TB—Tracheobronchial
TBT—Tributyltin
TC—Total Coliforms
TCA—Trichloroethane
TCDD—2,3,7,8-Tetrachlorodibenzo-p-Dioxin
TCDF—Tetrachlorodibenzofuran
TCE—Trichloroethylene

TCER—Target Carbon Emission Rate
TCF—Totally Chlorine Free
TCI—Total Capital Investment
TCLP—Toxicity Characteristic Leaching Procedure (Potential) (U.S.)
TCN—Technical Cooperation Network
TCP—Trichloropropane
TCR—Total Coliform Rule
TCU—True Color Unit
TDH—Total Dynamic Head
TDI—Tolerable Daily Intake
TDI—Toluene-2,4-Diisocyanate
TDS—Total Dissolved Solids
TEA—Triethanolamine
TEB—Total Exchangeable Bases
TECSEC—Technical Secretariat
TEL—Tetraethyl Lead
TEL—Total Energy Line
TEM—Total Extractable Matter
TEOM—Tapered Elemental Oscillating Microbalance
TEQ—Toxic Equivalent
TFAP—Tropical Forestry Action Plan
TFE—Tetrafluoroethylene
T4CB—3,3' ,4,4' -Tetrachlorinated Biphenyl
THAA—Total Haloacetic Acids
THC—Thermohaline Circulation
THC—Total Hydrocarbon
THESUS—Thermal Solar European Power Station
THF—Tetrahydrofuran
THM—Chloro-Organics
THM—Thermo-Hydro-Mechanical
THM—Trihalomethane
ThOD—Theoretical Oxygen Demand
THORP—Thermal Oxide Reprocessing Plant
3P—Pollution Prevention Pays
TI—Tobacco Institute
TIC—Total Industry Control
TIE—Toxicity Investigation Evaluation
TIEL—Toxicologically Insignificant Exposure Limit
TIO—Technology Innovation Office (U.S.)
TISE—Take It Somewhere Else
TKN—Total Kjeldahl Nitrogen
TL—Threshold Limit
TLC—Total Lung Capacity
TL, m—Threshold Limit, median
TLV—Threshold Limit Value
TM—Thematic Mapper
TMDL—Total Maximum Daily Load
TMI—Three Mile Island
TMP—Transmembrane Pressure
TN—Total Nitrogen
TNCB—Trinitrochlorobenzene
TNO—Netherlands Organization for Applied Scientific Research
TNO—Total Number of Organisms
TNRCC—Texas Natural Resources Conservation Commission
TOC—Total Organic Carbon
toe—Tonnes Oil Equivalent
TOGA—Tropical Ocean Global Atmosphere
TOMS—Total Ozone Mapping Spectrometer
TOX—Tetradichloroxylene
TP—Total Phosphorous
TPA—Tetraphthalic Acid
TPH—Total Petroleum Hydrocarbons
TPL—Trust for Public Lands
TPP—Tripolyphosphate
TPU—Thermal Processing Unit
TPV—Third-Party Verification
TQ—Threshold Quantity
TQM—Total Quality Management
TRE—Total Resource Effectiveness
TRE—Toxicity Reduction Evaluation
TRI—Toxics Release Inventory (U.S.)
TRIP—Trade Related Intellectual Property
TRS—Total Reduced Sulfur
TS—Total Solids
TSA—Technical Systems Audit
TSCA—Toxic Substances Control Act (U.S.)
TSCATS—Toxic Substances Control Act Test Submission
TSDF—Treatment, Storage and Disposal Facility
TSDHW—Transportation, Storage and Disposal of Hazardous Wastes
TSE—Transmissible Spongiform Encephalopathies
TSP—Total Soluble Phosphorous
TSP—Total Suspended Particulates
TSS—Total Suspended Solids
TT—Treatment Technique
TTHM—Total Trihalomethane
TTNBS—Technology Transfer Network Bulletin System
TTO—Total Toxic Organics
TTO—Troposphere Ozone
TTT & O—Time, Temperature, Turbulence and Oxygen
TUC—Trades Union Congress (UK)
TVC—Thermal Vapor Compression
TVO/IVO—Finnish Nuclear Organization
TWA—Time Weighted Average
TWh—Tetrawatt Hours

UASB—Upflow Anaerobic Sludge Blanket (Bed)
UATI—Union of Technical Associations and Organizations
UBA—German Federal Environmental Agency
UC—Uniformity Coefficient
UCM—Unresolved Complex Mixtures
UCPC—Ultrafine Condensation Particle Counter
UCS—Unconditioned Stimulus
UEIP—Use and Exposure Information Voluntary Project (U.S.)
UEL—Upper Explosive Limit
UF—Ultrafiltration
UFL—Upper Flammability Limit
UKMO—UK Meteorological Office
UL—Underwriters Laboratories
ULCC—Ultra Large Crude Carrier
ULSD—Ultra Low Sulfur Diesel
ULSP—Ultra Low Sulfur Petrol

UMIST—University of Manchester Institute of Science and Technology
UN—United Nations
UNCED—UN Conference on Environment and Development
UNCED—UN Convention on Environmental Diversity
UNCLOS—UN Conference on the Law of the Sea
UNDP—UN Development Program
UNECE—United Nations Economic Commission for Europe
UNEP—UN Environment Program
UNESCO—UN Education, Scientific and Cultural Organization
UNFCCC—United Nations Framework Convention on Climate Change
UNICEF—UN Children's Fund
UNICHAL—International Union of Heat Distributors
UNIDO—UN Industrial Development Organization
UNU—United Nations University
UOD—Ultimate Oxygen Demand
UP—Ultrafine Particle
UPW—Ultrapure Water
URV—Unit Risk Value
USC—U.S. Code
USCG—U.S. Coast Guard
USCGS—U.S. Coast and Geodetic Survey
USDW—Underground Source of Drinking Water
USGS—U.S. Geological Survey
USP—U.S. Pharmacopoeia
USPHS—U.S. Public Health Service
UST—Underground Storage Tank
UV—Ultraviolet
UVR—Ultraviolet Radiation

V—Vinyl
VA—Volatile Acid
VA—Vulnerability Assessment
VBS—Volatile Biofilm Solids
VC—Vapor Compression
VCM—Vinyl Chloride Monomer
VDI—Verein Deutscher Ingenieure
VE—Visual Emissions
VER—Variable Energy Recovery
VFA—Volatile Fatty Acid
VFVA—Vacuum Flash Vapor Absorption
VHAP—Volatile Hazardous Air Pollutant
VITO—Flemish Institute for Technological Research
VLCC—Very Large Crude Carrier
VLF—Very Low Frequency
VLLW—Very Low Level Radioactive Waste
VO—Volatile Organics
VOC—Volatile Organic Compounds (Contaminants) (Chemicals)
VOHAP—Volatile Organic Hazardous Air Pollutant
VOST—Volatile Organic Sampling Train
VPH—Volatile Petroleum Hydrocarbons
VPP—Voluntary Protection Program
VSS—Volatile Suspended Solids
VTE—Vertical Tube Evaporator
VTOC—Volatile Toxic Organic Compound
v/v—Volume/Volume
VVC—Vacuum Vapor Compression

W—Watt
WAC—Weak Acid Cation
WAMAP—IAEA Waste Management Advisory Program
WAS—Waste Activated Sludge
WASP—Water Quality Analysis and Simulation Program
WATRP—IAEA Waste Management Assessment and Technical Review Program
WBGU—German Advisory Council on Global Change
WC—Water Closet
WCED—World Commission on Environment and Development
WCRP—World Climate Research Program
WCS—Wildlife Conservation Society
WCS—World Conservation Strategy
WDF—Waste Derived Fuel
WEAO—Water Environment Association of Ontario
WEC—Wave Energy Converter
WEC—World Energy Conference
WEC—World Energy Council
WEDO—Women and Environment Development Organization
WEF—Water Environment Federation
WEF—World Environment Federation
WEPSD—World Engineering Partnership for Sustainable Development
WES—Women's Engineering Society (UK)
WESP—Wet Electrostatic Precipitator
WFEO—World Federation of Engineering Organizations
WFP—World Food Program
WFS—World Food Survey
WFTO—World Federation of Technology Organizations
Wh—Watt Hour
WHMIS—Waste Hazardous Materials Information System
WHO—World Health Organization
WHRC—Woods Hole Research Center
WID—Waste Incineration Directive (EU)
WIPP—Waste Isolation Pilot Plant
WIRE—World Wide Information System for Renewable Energy
WISE—Women into Science and Engineering
WITT—Women in Trade and Technology (Canada)
WL—Working Level
WLFO—Wet Limestone Forced Oxidation
WLM—Working Level Month
WMO—World Meteorological Organization
WO—Water in Oil
WPCF—Water Pollution Control Federation
WPWP—Western Pacific Warm Pool
WQCV—Water Quality Control Volume
WQI—Water Quality Indices
WQM—Water Quality Management
WRA—Waste Regulation Authority (UK)

WRC—World Resources Council
WRDA—Water Resources Development Act (U.S.)
WRI—World Resources Institute
WRPA—Water Resources Planning Act (U.S.)
WRRA—Water Resources Research Act (U.S.)
WSC—World Solar Commission
WSI—Water-Steam Injection
WSP—World Solar Programme
WSR—World Standardized Rates
WSRA—Wild and Scenic Rivers Act (U.S.)
WSSP—World Solar Summit Process
WTF—Water Treatment Facility
WTO—World Trade Organization
WWF—Wet Weather Flow
WWF—World Wildlife Fund
WWI—World Watch Institute
WWTP—Wastewater Treatment Plant

XLPE—Crosslinked Polyethylene
XRD—X-Ray Diffraction
XRF—X-Ray Fluorescence

Y—Growth Yield Coefficient

ZEV—Zero Emission Vehicle
ZIP—Zero-Incident Performance
ZLD—Zero Liquid Discharge
ZPG—Zero Population Growth

INDEX