

McGraw-Hill Nanoscience and Technology Series

# NANO / MICROSCALE HEAT TRANSFER



ZHUOMIN M. ZHANG

---

# **NANO/MICROSCALE HEAT TRANSFER**

---

## ABOUT THE AUTHOR

---

Zhuomin M. Zhang has taught at the University of Florida (1995–2002) and the Georgia Institute of Technology (since 2002). Professor Zhang is an ASME Fellow and has done cutting-edge research in the areas of micro/nanoscale heat transfer, with applications to optoelectronic devices and semiconductor manufacturing. He is a recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE), the ASME Heat Transfer Division Best Paper Award, and the AIAA Thermophysics Best Paper Award. Professor Zhang currently serves on the Editorial Board of the *International Journal of Thermophysics* and is an associate editor for the *Journal of Quantitative Spectroscopy & Radiative Transfer* and the *Journal of Thermophysics and Heat Transfer*.

---

# NANO/MICROSCALE HEAT TRANSFER

---

**Zhuomin M. Zhang**

*Georgia Institute of Technology  
Atlanta, Georgia*



New York Chicago San Francisco Lisbon London Madrid  
Mexico City Milan New Delhi San Juan Seoul  
Singapore Sydney Toronto

Copyright © 2007 by The McGraw-Hill Companies, Inc. All rights reserved. Manufactured in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

0-07-150973-9

The material in this eBook also appears in the print version of this title: 0-07-143674-X.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please contact George Hoare, Special Sales, at [george\\_hoare@mcgraw-hill.com](mailto:george_hoare@mcgraw-hill.com) or (212) 904-4069.

#### TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. (“McGraw-Hill”) and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill’s prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED “AS IS.” McGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

DOI: 10.1036/007143674X

*To my wife Lingyun*

*This page intentionally left blank*

---

# CONTENTS

---

**Preface**    *xiii*

**List of Symbols**    *xvii*

---

## **Chapter 1. Introduction**

**1**

- 1.1 Limitations of the Macroscopic Formulation / 2
- 1.2 The Length Scales / 3
- 1.3 From Ancient Philosophy to Contemporary Technologies / 5
  - 1.3.1 Microelectronics and Information Technology / 6
  - 1.3.2 Lasers, Optoelectronics, and Nanophotonics / 8
  - 1.3.3 Microfabrication and Nanofabrication / 10
  - 1.3.4 Probing and Manipulation of Small Structures / 12
  - 1.3.5 Energy Conversion Devices / 15
  - 1.3.6 Biomolecule Imaging and Molecular Electronics / 17
- 1.4 Objectives and Organization of This Book / 19
  - References / 22

---

## **Chapter 2. Overview of Macroscopic Thermal Sciences**

**25**

- 2.1 Fundamentals of Thermodynamics / 25
  - 2.1.1 The First Law of Thermodynamics / 26
  - 2.1.2 Thermodynamic Equilibrium and the Second Law / 27
  - 2.1.3 The Third Law of Thermodynamics / 31
- 2.2 Thermodynamic Functions and Properties / 32
  - 2.2.1 Thermodynamic Relations / 32
  - 2.2.2 The Gibbs Phase Rule / 34
  - 2.2.3 Specific Heats / 36
- 2.3 Ideal Gas and Ideal Incompressible Models / 38
  - 2.3.1 The Ideal Gas / 38
  - 2.3.2 Incompressible Solids and Liquids / 40
- 2.4 Heat Transfer Basics / 41
  - 2.4.1 Conduction / 42
  - 2.4.2 Convection / 44
  - 2.4.3 Radiation / 46
- 2.5 Summary / 51
  - References / 51
  - Problems / 52

---

## **Chapter 3. Elements of Statistical Thermodynamics and Quantum Theory**

**57**

- 3.1 Statistical Mechanics of Independent Particles / 58
  - 3.1.1 Macrostates versus Microstates / 59
  - 3.1.2 Phase Space / 59



- 3.1.3 Quantum Mechanics Considerations / 60
- 3.1.4 Equilibrium Distributions for Different Statistics / 62
- 3.2 Thermodynamic Relations / 67
  - 3.2.1 Heat and Work / 67
  - 3.2.2 Entropy / 67
  - 3.2.3 The Lagrangian Multipliers / 68
  - 3.2.4 Entropy at Absolute Zero Temperature / 68
  - 3.2.5 Macroscopic Properties in Terms of the Partition Function / 69
- 3.3 Ideal Molecular Gases / 71
  - 3.3.1 Monatomic Ideal Gases / 71
  - 3.3.2 Maxwell's Velocity Distribution / 73
  - 3.3.3 Diatomic and Polyatomic Ideal Gases / 75
- 3.4 Statistical Ensembles and Fluctuations / 81
- 3.5 Basic Quantum Mechanics / 82
  - 3.5.1 The Schrödinger Equation / 82
  - 3.5.2 A Particle in a Potential Well or a Box / 84
  - 3.5.3 A Rigid Rotor / 86
  - 3.5.4 Atomic Emission and the Bohr Radius / 88
  - 3.5.5 A Harmonic Oscillator / 90
- 3.6 Emission and Absorption of Photons by Molecules or Atoms / 92
- 3.7 Energy, Mass, and Momentum in Terms of Relativity / 94
- 3.8 Summary / 96
  - References / 96
  - Problems / 96

## Chapter 4. Kinetic Theory and Micro/Nanofluidics

101

- 4.1 Kinetic Description of Dilute Gases / 101
  - 4.1.1 Local Average and Flux / 102
  - 4.1.2 The Mean Free Path / 105
- 4.2 Transport Equations and Properties of Ideal Gases / 108
  - 4.2.1 Shear Force and Viscosity / 109
  - 4.2.2 Heat Diffusion / 110
  - 4.2.3 Mass Diffusion / 112
  - 4.2.4 Intermolecular Forces / 115
- 4.3 The Boltzmann Transport Equation / 116
  - 4.3.1 Hydrodynamic Equations / 117
  - 4.3.2 Fourier's Law and Thermal Conductivity / 119
- 4.4 Micro/Nanofluidics and Heat Transfer / 121
  - 4.4.1 The Knudsen Number and Flow Regimes / 122
  - 4.4.2 Velocity Slip and Temperature Jump / 124
  - 4.4.3 Gas Conduction—From the Continuum to the Free Molecule Regime / 129
- 4.5 Summary / 132
  - References / 132
  - Problems / 133

## Chapter 5. Thermal Properties of Solids and the Size Effect

137

- 5.1 Specific Heat of Solids / 137
  - 5.1.1 Lattice Vibration in Solids: The Phonon Gas / 137
  - 5.1.2 The Debye Specific Heat Model / 139
  - 5.1.3 Free Electron Gas in Metals / 143
- 5.2 Quantum Size Effect on the Specific Heat / 148
  - 5.2.1 Periodic Boundary Conditions / 148
  - 5.2.2 General Expressions of Lattice Specific Heat / 149
  - 5.2.3 Dimensionality / 149
  - 5.2.4 Thin Films Including Quantum Wells / 151
  - 5.2.5 Nanocrystals and Carbon Nanotubes / 153

- 5.3 Electrical and Thermal Conductivities of Solids / 154
  - 5.3.1 Electrical Conductivity / 155
  - 5.3.2 Thermal Conductivity of Metals / 158
  - 5.3.3 Derivation of Conductivities from the BTE / 160
  - 5.3.4 Thermal Conductivity of Insulators / 162
- 5.4 Thermoelectricity / 166
  - 5.4.1 The Seebeck Effect and Thermoelectric Power / 167
  - 5.4.2 The Peltier Effect and the Thomson Effect / 168
  - 5.4.3 Thermoelectric Generation and Refrigeration / 170
  - 5.4.4 Onsager's Theorem and Irreversible Thermodynamics / 172
- 5.5 Classical Size Effect on Conductivities and Quantum Conductance / 174
  - 5.5.1 Classical Size Effect Based on Geometric Consideration / 174
  - 5.5.2 Classical Size Effect Based on the BTE / 178
  - 5.5.3 Quantum Conductance / 182
- 5.6 Summary / 187
  - References / 187
  - Problems / 190

## Chapter 6. Electron and Phonon Transport

193

- 6.1 The Hall Effect / 193
- 6.2 General Classifications of Solids / 195
  - 6.2.1 Electrons in Atoms / 195
  - 6.2.2 Insulators, Conductors, and Semiconductors / 197
  - 6.2.3 Atomic Binding in Solids / 199
- 6.3 Crystal Structures / 201
  - 6.3.1 The Bravais Lattices / 201
  - 6.3.2 Primitive Vectors and the Primitive Unit Cell / 204
  - 6.3.3 Basis Made of Two or More Atoms / 206
- 6.4 Electronic Band Structures / 209
  - 6.4.1 Reciprocal Lattices and the First Brillouin Zone / 209
  - 6.4.2 Bloch's Theorem / 210
  - 6.4.3 Band Structures of Metals and Semiconductors / 214
- 6.5 Phonon Dispersion and Scattering / 217
  - 6.5.1 The 1-D Diatomic Chain / 217
  - 6.5.2 Dispersion Relations for Real Crystals / 219
  - 6.5.3 Phonon Scattering / 221
- 6.6 Electron Emission and Tunneling / 226
  - 6.6.1 Photoelectric Effect / 226
  - 6.6.2 Thermionic Emission / 227
  - 6.6.3 Field Emission and Electron Tunneling / 229
- 6.7 Electrical Transport in Semiconductor Devices / 232
  - 6.7.1 Number Density, Mobility, and the Hall Effect / 232
  - 6.7.2 Generation and Recombination / 236
  - 6.7.3 The  $p$ - $n$  Junction / 238
  - 6.7.4 Optoelectronic Applications / 240
- 6.8 Summary / 242
  - References / 242
  - Problems / 244

## Chapter 7. Nonequilibrium Energy Transfer in Nanostructures

247

- 7.1 Phenomenological Theories / 248
  - 7.1.1 Hyperbolic Heat Equation / 250
  - 7.1.2 Dual-Phase-Lag Model / 254
  - 7.1.3 Two-Temperature Model / 258

- 7.2 Heat Conduction Across Layered Structures / 262
  - 7.2.1 Equation of Phonon Radiative Transfer (EPRT) / 263
  - 7.2.2 Solution of the EPRT / 266
  - 7.2.3 Thermal Boundary Resistance (TBR) / 271
- 7.3 Heat Conduction Regimes / 275
- 7.4 Summary / 278
  - References / 278
  - Problems / 281

## Chapter 8. Fundamentals of Thermal Radiation

283

- 8.1 Electromagnetic Waves / 285
  - 8.1.1 Maxwell's Equations / 285
  - 8.1.2 The Wave Equation / 286
  - 8.1.3 Polarization / 288
  - 8.1.4 Energy Flux and Density / 290
  - 8.1.5 Dielectric Function / 291
  - 8.1.6 Propagating and Evanescent Waves / 293
- 8.2 Blackbody Radiation: The Photon Gas / 294
  - 8.2.1 Planck's Law / 294
  - 8.2.2 Radiation Thermometry / 298
  - 8.2.3 Entropy and Radiation Pressure / 301
  - 8.2.4 Limitations of Planck's Law / 305
- 8.3 Radiative Properties of Semi-Infinite Media / 306
  - 8.3.1 Reflection and Refraction of a Plane Wave / 306
  - 8.3.2 Emissivity / 311
  - 8.3.3 Bidirectional Reflectance / 312
- 8.4 Dielectric Function Models / 314
  - 8.4.1 Kramers-Kronig Dispersion Relations / 314
  - 8.4.2 The Drude Model for Free Carriers / 315
  - 8.4.3 The Lorentz Oscillator Model for Lattice Absorption / 318
  - 8.4.4 Semiconductors / 321
  - 8.4.5 Superconductors / 325
  - 8.4.6 Metamaterials with a Magnetic Response / 326
- 8.5 Summary / 329
  - References / 329
  - Problems / 330

## Chapter 9. Radiative Properties of Nanomaterials

333

- 9.1 Radiative Properties of a Single Layer / 333
  - 9.1.1 The Ray Tracing Method for a Thick Layer / 334
  - 9.1.2 Thin Films / 335
  - 9.1.3 Partial Coherence / 340
  - 9.1.4 Effect of Surface Scattering / 344
- 9.2 Radiative Properties of Multilayer Structures / 346
  - 9.2.1 Thin Films with Two or Three Layers / 347
  - 9.2.2 The Matrix Formulation / 348
  - 9.2.3 Radiative Properties of Thin Films on a Thick Substrate / 350
  - 9.2.4 Local Energy Density and Absorption Distribution / 352
- 9.3 Photonic Crystals / 352
- 9.4 Periodic Gratings / 356
  - 9.4.1 Rigorous Coupled-Wave Analysis (RCWA) / 358
  - 9.4.2 Effective Medium Formulations / 360
- 9.5 Bidirectional Reflectance Distribution Function (BRDF) / 362
  - 9.5.1 The Analytical Model / 363
  - 9.5.2 The Monte Carlo Method / 364

- 9.5.3 Surface Characterization / 367
- 9.5.4 BRDF Measurements / 368
- 9.5.5 Comparison of Modeling with Measurements / 370
- 9.6 Summary / 372
- References / 373
- Problems / 374

## Chapter 10. Near-Field Energy Transfer

377

- 10.1 Total Internal Reflection, Guided Waves, and Photon Tunneling / 378
  - 10.1.1 The Goos-Hänchen Shift / 379
  - 10.1.2 Waveguides and Optical Fibers / 382
  - 10.1.3 Photon Tunneling by Coupled Evanescent Waves / 386
  - 10.1.4 Thermal Energy Transfer between Closely Spaced Dielectrics / 389
  - 10.1.5 Resonance Tunneling through Periodic Dielectric Layers / 391
  - 10.1.6 Photon Tunneling with Negative Index Materials / 393
- 10.2 Polaritons or Electromagnetic Surface Waves / 395
  - 10.2.1 Surface Plasmon and Phonon Polaritons / 396
  - 10.2.2 Coupled Surface Polaritons and Bulk Polaritons / 401
  - 10.2.3 Polariton-Enhanced Transmission of Layered Structures / 405
  - 10.2.4 Radiation Transmission through Nanostructures / 408
  - 10.2.5 Superlens for Perfect Imaging and the Energy Streamlines / 410
- 10.3 Spectral and Directional Control of Thermal Radiation / 414
  - 10.3.1 Gratings and Microcavities / 417
  - 10.3.2 Metamaterials / 421
  - 10.3.3 Modified Photonic Crystals for Coherent Thermal Emission / 422
- 10.4 Radiation Heat Transfer at Nanometer Distances / 425
  - 10.4.1 The Fluctuational Electrodynamics / 426
  - 10.4.2 Heat Transfer between Parallel Plates / 428
  - 10.4.3 Asymptotic Formulation / 430
  - 10.4.4 Nanoscale Radiation Heat Transfer between Doped Silicon / 431
- 10.5 Summary / 436
- References / 437
- Problems / 440

## Appendix A. Physical Constants, Conversion Factors, and SI Prefixes

443

- Physical Constants / 443
- Conversion Factors / 443
- SI Prefixes / 443

## Appendix B. Mathematical Background

445

- B.1 Some Useful Formulae / 445
  - B.1.1 Series and Integrals / 445
  - B.1.2 The Error Function / 446
  - B.1.3 Stirling's Formula / 447
- B.2 The Method of Lagrange Multipliers / 447
- B.3 Permutation and Combination / 448
- B.4 Events and Probabilities / 450
- B.5 Distribution Functions and the Probability Density Function / 451
- B.6 Complex Variables / 454
- B.7 The Plane Wave Solution / 455
- B.8 The Sommerfeld Expansion / 459

*This page intentionally left blank*

---

# PREFACE

---

Over the past 20 years, there have been tremendous developments in microelectronics, microfabrication technology, MEMS and NEMS, quantum structures (e.g., superlattices, nanowires, nanotubes, and nanoparticles), optoelectronics and lasers, including ultrafast lasers, and molecular- to atomic-level imaging techniques (such as high-resolution electron microscopy, scanning tunneling microscopy, atomic force microscopy, near-field optical microscopy, and scanning thermal microscopy). The field is fast moving into scaling up and systems engineering to explore the unlimited potential that nanoscience and nanoengineering may offer to restructure the technologies in the new millennia. When the characteristic length becomes comparable to the mechanistic length scale, continuum assumptions that are often made in conventional thermal analyses may break down. Similarly, when the characteristic time becomes comparable to the mechanistic timescale, traditional equilibrium approaches may not be appropriate. Understanding the energy transport mechanisms in small dimensions and short timescales is crucial for future advances of nanotechnology. In recent years, a growing number of research publications have been in nano/microscale thermophysical engineering. Timely dissemination of the knowledge gained from contemporary research to educate future scientists and engineers is of emerging significance. For this reason, more and more universities have started to offer courses in microscale areas. A self-contained textbook suitable for engineering students is much needed. Many practicing engineers who have graduated earlier wish to learn what is going on in this fascinating area, but are often frustrated due to the lack of a solid background to comprehend the contemporary literature. A book that does not require prior knowledge in statistical mechanics, quantum mechanics, solid state physics, and electro-dynamics is extremely helpful. On the other hand, such a book should cover all these subjects in some depth without significant prerequisites.

This book is written for engineering senior undergraduate and graduate students, practicing engineers, and academic researchers who have not been extensively exposed to nanoscale sciences but wish to gain a solid background in the thermal phenomena occurring at small length scales and short timescales. The basic philosophy behind this book is to logically integrate the traditional knowledge in thermal engineering and physics with newly developed theories in an easy-to-understand approach, with ample examples and homework problems. The materials have been used in the graduate course and undergraduate elective that I have taught a number of times at two universities since 1999. While this book can be used as a text for a senior elective or an entry-level graduate course, it is not expected that all the materials will be covered in a one-semester course. The instructors should have the freedom to select materials from the book according to students' backgrounds and interests. Some chapters and sections can also be used to integrate with traditional thermal science courses in order to update the current undergraduate and graduate curricula with nanotechnology contents.

The content of this book includes microscopic descriptions and approaches, as well as their applications in thermal science and engineering, with an emphasis on energy transport in gases and solids by conduction (diffusion) and radiation (with or without a medium), as well as convection in micro/nanofluidics. Following the introduction in

Chapter 1, an in-depth overview on the foundation of macroscopic thermodynamics, heat transfer, and fluid mechanics is given in Chapter 2. Chapter 3 summarizes the well-established theories in statistical mechanics, including classical and quantum statistics; thermal properties of ideal gases are described in the context of statistical thermodynamics, followed by a concise presentation of quantum mechanics. Chapter 4 focuses on microfluidics and introduces the Boltzmann transport equation. The heat transfer and microflow regimes from continuous flow to free molecule flow are described. In Chapters 5, 6, and 7, heat transfer in solid nanostructures is extensively discussed. Chapter 5 presents the classical and quantum size effects on specific heat and thermal conductivity without involving detailed solid state theories, which are introduced in Chapter 6. This arrangement allows a more intuitive learning experience. Chapter 7 focuses on transient as well as nonequilibrium energy transport processes in nanostructures. The next three chapters deal with thermal radiation at nanoscales. Chapter 8 provides the fundamental understanding of electromagnetic waves and the dielectric properties of various materials. The concept of radiation entropy is also introduced, along with the recently demonstrated metamaterials with exotic properties. Chapter 9 describes interference effects of thin films and multilayers, the band structure of photonic crystals, diffraction from surface-relief gratings, and scattering from rough surfaces. Chapter 10 explores the evanescent wave and the coupling phenomena in the near field for energy transfer. Recent advances in nanophotonics and nanoscale radiation heat transfer are also summarized. The dual nature of particles and waves are emphasized throughout the book in explaining the energy carriers, such as molecules in ideal gases, electrons in metals, phonons in dielectric crystalline materials, and photons for radiative transfer.

In the early 1990s, I was fortunate to work with Professor Markus Flik for my Ph.D. dissertation on the infrared spectroscopy of thin (down to 10 nanometers) high-temperature superconducting films for microfabricated, highly sensitive radiation detectors, as well as to assist him in the summer short course on microscale heat transfer at the Massachusetts Institute of Technology. While I was still a postdoctoral researcher, late Professor Chang-Lin Tien, then Chancellor of the University of California at Berkeley, wrote an invitation letter to me to give a seminar in the Department of Mechanical Engineering of Berkeley in January 1994; he continuously supported me, including the development of the concept of this book. The last time I heard from him was just a few weeks prior to the 2000 National Heat Transfer Conference in Pittsburgh, where he delivered a plenary speech before he fell ill. In his letter dated August 10, 2000, Professor Tien enthusiastically endorsed my plan to write a microscale textbook and encouraged me to include nano aspects. He wrote “I would like to express to you my strongest support for your project; however, I would suggest that you broaden the content somewhat beginning with the title to ‘Micro/Nanoscale Heat ...,’ and to talk about some coverage on nano aspects.” Professor Tien opened my eyes, and it took me several years afterward to complete this book, which now has more emphasis on nanoscale thermal sciences and engineering.

I also benefited greatly from the encouragements and comments received through discussions with a large number of people in the heat transfer and thermophysics community, too many to be listed here. I am grateful to my colleagues and friends at both University of Florida (UF) and Georgia Tech for their help whenever needed. I especially want to thank Professor William Tiederman, who was Chair of the Department of Mechanical Engineering during my stay at UF, for his support and mentorship at the early stage of my independent research and teaching career. Professor David Tanner in the Department of Physics of UF helped me understand solid state physics; I have enjoyed collaboration with him since 1995. Through the years, Dr. Jack Hsia, former Chief of Academic Affairs at the National Institute of Standards and Technology (NIST), offered me much personal and professional advice. He is one of the many outstanding mentors I have had from NIST, where I gained my postdoctoral experience and worked for a number of summers afterward. This book

would not have been possible without my graduate students' hard work and dedication. Most of them have taken my classes and proofread different versions of the manuscript. Some materials in the last few chapters of the book were generated based on their thesis research. Many graduate and undergraduate students who have taken my classes or worked in the Nanoscale Thermal Radiation Lab also provided constructive suggestions. I enjoyed working with all of them. I must thank the Sponsoring Editor, Ken McCombs, for his endurance and persistence that kept me on the writing track over the past few years, and the whole production team, for carefully editing the manuscript and setting the final pages. While this project was partially supported by the National Science Foundation as part of my educational plan in the CAREER/PECASE grant, I take full responsibility for any inadvertent errors or mistakes.

Finally, I thank my family for their understanding and support throughout the writing journey. My three children, Emmy, Angie, and Bryan, have given me great happiness and made my life meaningful. This book is dedicated to my wife Lingyun for the unconditional love and selfless care she has provided to me and to our children.

ZHUOMIN M. ZHANG



*This page intentionally left blank*

---

# LIST OF SYMBOLS

---

<b>A</b>	area, m <sup>2</sup> ; Helmholtz free energy, J
$A_c$	cross-sectional area, m <sup>2</sup>
$A'_A$	directional-spectral absorptance of a semitransparent material
<b>a</b>	acceleration, m/s <sup>2</sup>
$a$	lattice constant, m; magnitude of acceleration, m/s <sup>2</sup>
$a_0$	Bohr radius, 0.0529 nm
$a_A$	absorption coefficient, m <sup>-1</sup>
<b>B</b>	magnetic induction or magnetic flux density, T (tesla) or Wb/m <sup>2</sup>
<b>C</b>	volumetric heat capacity ( $\rho c_p$ ), J/(m <sup>3</sup> · K)
$c$	phase velocity of electromagnetic wave, m/s
$c_0$	speed of light in vacuum, $2.998 \times 10^8$ m/s
$c_v$ or $\bar{c}_v$	mass or molar specific heat for constant volume, J/(kg · K) or J/(kmol · K)
$c_p$ or $\bar{c}_p$	mass or molar specific heat for constant pressure, J/(kg · K) or J/(kmol · K)
<b>D</b>	dynamical matrix; electric displacement, C/m <sup>2</sup>
<b>D</b>	density of states, m <sup>-3</sup> ; diameter, m
$D_{AB}$	binary diffusion coefficient, m <sup>2</sup> /s
$d$	diameter or film thickness, m
<b>E</b>	electric field vector, N/C or V/m
<b>E</b>	energy, J; magnitude of electric field, V/m
$E_F$	Fermi energy, J
$E_g$	bandgap energy, J
$e$	electron charge (absolute value), $1.602 \times 10^{-19}$ C
$e_b$	blackbody emissive power, W/m <sup>2</sup>
<b>F, F</b>	force, N
<b>F</b>	normalized distribution function
$f$	distribution function (sometimes normalized)
<b>G</b>	reciprocal lattice vector, m <sup>-1</sup> ; dyadic Green function
<b>G</b>	Gibbs free energy, J; electron-phonon coupling constant, W/(m <sup>3</sup> · K)
$\bar{g}$	molar specific Gibbs free energy, J/kmol
$g$	degeneracy
<b>H</b>	magnetic field vector, A/m or C/(m · s)
<b>H</b>	enthalpy, J; magnetic field strength, A/m or C/(m · s)
$h$	mass specific enthalpy, J/kg; convection heat transfer coefficient, W/(m <sup>2</sup> · K); Planck's constant, $6.626 \times 10^{-34}$ J · s
$h_m$	convection mass transfer coefficient, m/s
$\hbar$	Planck's constant divided by $2\pi$ , $h/2\pi$
$\bar{h}$	molar specific enthalpy, J/kmol
<b>I</b>	unit matrix; unit dyadic

$I$	moment of inertia, $\text{kg} \cdot \text{m}^2$ ; intensity or radiance, $\text{W}/(\text{m}^2 \cdot \mu\text{m} \cdot \text{sr})$ ; electric current, A
$i$	$\sqrt{-1}$
$i, j, k$	indices used in series
$\mathbf{J}$ or $J$	flux vector or magnitude (quantity transferred per unit area per unit time)
$\mathbf{J}, \mathbf{J}_e$	current density (also called electric charge flux), $\text{A}/\text{m}^2$
$J_E$	energy flux, $\text{W}/\text{m}^2$
$J_m$	mass flux, $\text{kg}/(\text{s} \cdot \text{m}^2)$
$J_N$	particle flux, $\text{m}^{-2}$
$J_p$	momentum flux, Pa ( $\text{N}/\text{m}^2$ )
$K$	spring constant, N/m; Thomson's coefficient, V/K; Bloch wavevector, $\text{m}^{-1}$
$\mathbf{k}$	wavevector, $\text{m}^{-1}$
$k$	magnitude of the wavevector, $\text{m}^{-1}$
$k_B$	Boltzmann's constant, $1.381 \times 10^{-23}$ J/K
$L$	characteristic length, m
$L_0$	average distance between molecules or atoms, m
$L_\lambda$	radiation entropy intensity, $\text{W}/(\text{K} \cdot \text{m}^2 \cdot \mu\text{m} \cdot \text{sr})$
$l$	length, m
$l, m, n$	index numbers
$M$	molecular weight, kg/kmol
$m$	mass of a system or a single particle, kg
$m_r$	reduced mass, kg
$m^*$	effective mass, kg
$\dot{m}$	mass flow rate or mass transfer rate, kg/s
$N$	number of particles; number of phonon oscillators
$N_A$	Avegado's constant, $6.022 \times 10^{26}$ kmol $^{-1}$ ; acceptor concentration, $\text{m}^{-3}$
$N_D$	donor concentration, $\text{m}^{-3}$
$\dot{N}$	particle flow rate, $\text{s}^{-1}$
$n$	number density, $\text{m}^{-3}$ ; quantum number; real part of refractive index or refractive index
$\bar{n}$	amount of substance, kmol
$\tilde{n}$	complex refractive index
$\mathbf{P}$	propagation matrix; polarization vector or dipole moment per unit volume, $\text{C}/\text{m}^2$
$P$	pressure, Pa ( $\text{N}/\text{m}^2$ )
$P_{ij}$	momentum flux component, Pa
$\mathbf{p}$	momentum vector ( $m\mathbf{v}$ or $\hbar\mathbf{k}$ ), $\text{kg} \cdot \text{m}/\text{s}$
$p$	momentum ( $m\mathbf{v}$ or $\hbar\mathbf{k}$ ), $\text{kg} \cdot \text{m}/\text{s}$ ; probability; specularly
$p, q$	index numbers
$Q$	heat, J, quality factor
$\dot{Q}$	heat transfer rate, W
$q$	number of coexisting phases; number of atoms per molecule
$\dot{q}$	thermal energy generation rate, $\text{W}/\text{m}^3$
$\mathbf{q}''$	heat flux vector, $\text{W}/\text{m}^2$
$q''$	heat flux, $\text{W}/\text{m}^2$
$R$	gas constant, $\text{J}/(\text{kg} \cdot \text{K})$ ; electrical resistance, $\Omega$ or V/A
$R'$	directional-hemispherical reflectance
$R_b''$	thermal boundary resistance, $\text{m}^2 \cdot \text{K}/\text{W}$
$R_1''$	thermal resistance, $\text{m}^2 \cdot \text{K}/\text{W}$
$\bar{R}$	universal gas constant, 8314.5 J/(kmol $\cdot$ K)

$r$	distance or radius, m; Fresnel reflection coefficient
$r_e$	electrical resistivity, $\Omega \cdot \text{m}$
$\tilde{r}$	complex Fresnel reflection coefficient
$\mathbf{S}$	Poynting vector, $\text{W}/\text{m}^2$
$S$	entropy, J/K
$S_j$	strength of the $j$ th phonon oscillator
$\dot{S}$	entropy transfer rate, W/K
$\dot{S}_{\text{gen}}$	entropy generation rate, W/K
$s$ or $\bar{s}$	specific entropy, $\text{J}/(\text{kg} \cdot \text{K})$ , $\text{J}/(\text{m}^3 \cdot \text{K})$ or $\text{J}/(\text{kmol} \cdot \text{K})$
$\dot{s}_{\text{gen}}$	volumetric entropy generation rate, $\text{W}/(\text{m}^3 \cdot \text{K})$
$\dot{s}''$	entropy flux, $\text{J}/(\text{m}^2 \cdot \text{K})$
$T$	temperature, K
$T'$	directional-hemispherical transmittance
$t$	time, s; Fresnel transmission coefficient
$\tilde{t}$	complex Fresnel transmission coefficient
$U$	internal energy, J; periodic potential; J
$\mathbf{u}_d$	drift velocity, m/s
$u$ or $\bar{u}$	specific internal energy: mass specific, J/kg, and volume specific (i.e., energy density), $\text{J}/\text{m}^3$ , or molar specific, J/kmol
$V$	volume, $\text{m}^3$ ; voltage, V
$\mathbf{v}$	velocity, m/s
$\mathbf{v}_B$	bulk or mean velocity, m/s
$\mathbf{v}_R$	random or thermal velocity, m/s
$v$	specific volume, $\text{m}^3/\text{kg}$ ; speed, m/s
$v_a$	speed of sound or average speed of phonons, m/s
$v_F$	Fermi velocity, m/s
$v_g$	magnitude of group velocity ( $d\omega/dk$ ), m/s
$v_P, v_t$	longitudinal, transverse phonon speed, m/s
$v_p$	phase speed ( $\omega/k$ ), m/s
$v_x, v_y, v_z$	velocity components, m/s
$\bar{v}$	molar specific volume, $\text{m}^3/\text{kmol}$ ; average speed, m/s
$W$	work, J; width, m
$x, y, z$	coordinates, m
$Z$	partition function

## ***DIMENSIONLESS PARAMETERS***

---

$Kn$	Knudsen number, $\Lambda/L$
$Le$	Lewis number, $D_{AB}/\alpha = Pr/Sc$
$Lz$	Lorentz number, $\kappa/\sigma T$
$Ma$	Mach number, $v/v_a$
$Nu$	Nusselt number, $hL/\kappa$
$Pe$	Peclet number, $RePr = v_\infty L/\alpha$
$Pr$	Prandtl number, $\nu/\alpha$
$Re$	Reynolds number, $\rho v_\infty L/\mu$
$Sc$	Schmidt number, $\nu/D_{AB}$
$ZT$	dimensionless figure of merit for thermoelectricity

## GREEK SYMBOLS

---

$\alpha$	thermal diffusivity, $\text{m}^2/\text{s}$ ; other constant
$\alpha$ and $\beta$	Lagrangian multipliers
$\alpha_T$	thermal accommodation coefficient
$\alpha_v$	(tangential) momentum accommodation coefficient
$\alpha_v'$	normal momentum accommodation coefficient
$\alpha'_\lambda$	directional-spectral absorptivity
$\beta$	phase shift, rad; various coefficients
$\beta_P$	isobaric thermal expansion coefficient, $\text{K}^{-1}$
$\beta_T$	$2\gamma(2 - \alpha_T)Kn/[\alpha_T(\gamma + 1)Pr]$
$\beta_v$	$(2 - \alpha_v)Kn/\alpha_v$
$\Gamma_{ij}$	hemispherical transmissivity for phonons from medium $i$ to $j$
$\Gamma_S$	Seebeck's coefficient, $\text{V}/\text{K}$
$\gamma$	specific heat ratio ( $c_p/c_v$ ); scattering rate ( $1/\tau$ ), $\text{rad}/\text{s}$
$\gamma_s$	Sommerfeld constant, $\text{J}/(\text{kg} \cdot \text{K}^2)$
$\delta$	differential small quantity; boundary layer thickness, $\text{m}$
$\delta_\lambda$	radiation penetration depth, $\text{m}$
$\varepsilon$	particle energy, $\text{J}$ ; electric permittivity, $\text{C}^2/(\text{N} \cdot \text{m}^2)$ ; ratio of permittivity to that of vacuum; emissivity
$\tilde{\varepsilon}$	complex dielectric function, i.e., ratio of permittivity to that of vacuum
$\varepsilon'_\lambda$	directional-spectral emissivity
$\eta_H$	Hall coefficient $E_y/J_x B$ , $\text{m}^3/\text{C}$
$\Theta$	characteristic temperature, $\text{K}$
$\Theta_D$	Debye temperature, $\text{K}$
$\theta$	zenith angle, $\text{rad}$
$\theta_B$	Brewster's angle, $\text{rad}$
$\theta_c$	critical angle, $\text{rad}$
$\kappa$	thermal conductivity, $\text{W}/(\text{m} \cdot \text{K})$ ; extinction coefficient (i.e., imaginary part of the refractive index)
$\kappa_T$	isothermal compressibility, $\text{Pa}^{-1}$
$\Lambda$	mean free path, $\text{m}$ ; period of a grating or photonic crystal, $\text{m}$
$\Lambda_a$	average distance between collisions, $\text{m}$
$\lambda$	wavelength, $\text{m}$
$\mu$	viscosity, $\text{N} \cdot \text{s}/\text{m}^2$ ; chemical potential, $\text{J}$ ; electron or hole mobility, $\text{m}^2/(\text{V} \cdot \text{s})$ , magnetic permeability, $\text{N}/\text{A}^2$ ; ratio of the permeability to that of vacuum
$\mu_F$	Fermi energy, $\text{J}$
$\nu$	kinematic viscosity, $\text{m}^2/\text{s}$ ; frequency, $\text{Hz}$
$\bar{\nu}$	wavenumber, $\text{cm}^{-1}$
$\Pi$	Peltier's coefficient, $\text{V}$
$\rho$	density, $\text{kg}/\text{m}^3$
$\rho_c$	charge density, $\text{C}/\text{m}^3$
$\rho'$	directional-hemispherical reflectivity
$\sigma$	electrical conductivity, $(\Omega \cdot \text{m})^{-1}$ ; standard deviation
$\sigma_{\text{rms}}$	root-mean-square surface roughness, $\text{m}$
$\sigma_{\text{SB}}$	Stefan-Boltzmann constant, $5.67 \times 10^{-8} \text{W}/(\text{m}^2 \cdot \text{K}^4)$
$\sigma'_{\text{SB}}$	phonon Stefan-Boltzmann constant, $\text{W}/(\text{m}^2 \cdot \text{K}^4)$

$\tau$	relaxation time, s; shear stress, Pa
$\tau'$	directional-hemispherical transmissivity
$\tau_{12}$	transmission coefficient
$\Phi$	scattering phase function; viscous dissipation function; potential function
$\phi$	number of degrees of freedom; azimuthal angle, rad; intermolecular potential
$\Psi$	Schrödinger's wavefunction; various functions
$\psi$	molecular quantity; wavefunction; work function, J
$\Omega$	solid angle, sr; thermodynamic probability
$\omega$	angular frequency, rad/s
$\omega_p$	plasma frequency, rad/s
$\varpi$	velocity space, $d\varpi = dv_x dv_y dv_z$

## SUBSCRIPTS

---

0	vacuum
1, 2, 3	medium 1, 2, 3
b	blackbody; boundary
d	defect or impurity
e	electron
h	hole
i	incident
$i, j, k, l, m, n$	indices
m	bulk or mean; maximum; medium
mp	most probable
$n$ or $p$	$n$ -type or $p$ -type semiconductor
p	TM wave or $p$ (parallel) polarization
r	reflected; rotational
s	TE wave or $s$ (perpendicular) polarization; scattered; surface; solid; lattice
t	transmitted; translational
th	thermal
v	vibrational
w	wall
$\infty$	free steam
$\lambda, \nu, \text{ or } \omega$	spectral property in terms of wavelength, frequency, or angular frequency

*This page intentionally left blank*

---

# **NANO/MICROSCALE HEAT TRANSFER**

---



*This page intentionally left blank*

---

# CHAPTER 1

---

# INTRODUCTION

---

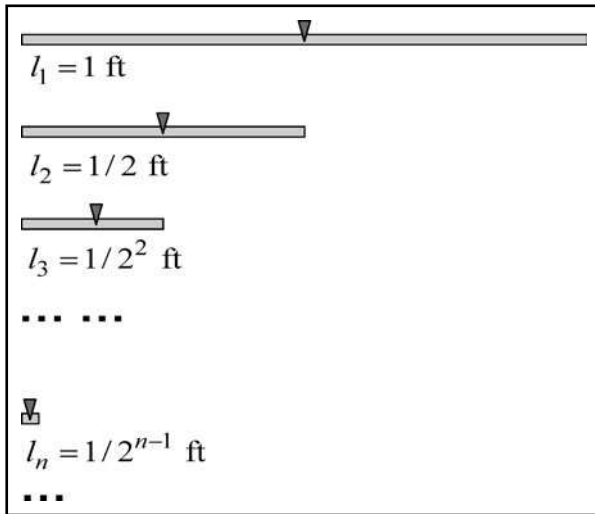
Improvement in performance and shrinkage of device sizes in microelectronics have been major driving forces for scientific and economic progress over the past 30 years. Developments in semiconductor processing and surface sciences have allowed precise control over critical dimensions with desirable properties for solid state devices. In the past 20 years, there have been tremendous developments in micro- and nanoelectromechanical systems (MEMS and NEMS), microfluidics and nanofluidics, quantum structures and devices, photonics and optoelectronics, nanomaterials for molecular sensing and biomedical diagnosis, and scanning probe microscopy for measurement and manipulation at the molecular and atomic levels.

Nanotechnology research has not only emerged as a new area of science and engineering, but it has also become an integral part of almost all natural science and engineering disciplines. According to the Web site of Georgia Institute of Technology ([www.gatech.edu](http://www.gatech.edu)), more than 10% of the faculty members at the university have been involved with research projects related to nanoscience and nanoengineering. The same can be said for most major research universities in the United States and in many other countries. Furthermore, the study of nanoscience and nanoengineering requires and has resulted in close interactions across the boundaries of many traditional disciplines. Knowledge of physical behavior at the molecular and atomic levels has played and will continue to play an important role in our understanding of the fundamental processes occurring in the macro world. This will enable us to design and develop novel devices and machines, ranging from a few nanometers all the way to the size of automobiles and airplanes. We have already enjoyed camera phones and the iPod that can store thousands of pictures and songs. In the next few decades, the advancement of nano/microscale science and engineering will fundamentally restructure the technologies currently used in manufacturing, energy production and utilization, communication, transportation, space exploration, and medicine.<sup>1,2</sup>

A key issue associated with miniaturization is the tremendous increase in the heat dissipation per unit volume. Micro/nanostructures may enable engineered materials with unique thermal properties to allow significant enhancement or reduction of the heat flow rate. Therefore, knowledge of thermal transport from the micrometer scale down to the nanometer scale and thermal properties of micro/nanostructures is of critical importance to future technological growth. Solutions to more and more problems in small devices and systems require a solid understanding of the heat (or more generally, energy) transfer mechanisms in reduced dimensions and/or short time scales, because classical equilibrium and continuum assumptions are not valid anymore. Examples are the thermal analysis and modeling of micro/nanodevices, ultrafast laser interaction with materials, micromachined thermal sensors and actuators, thermoelectricity in nanostructures, photonic crystals, microscale thermophotovoltaic devices, and so forth.<sup>3,4</sup>

## 1.1 LIMITATIONS OF THE MACROSCOPIC FORMULATION

As an ancient Chinese philosopher put it, suppose you take a foot-long wood stick and cut off half of it each day; you will never reach an end even after thousands of years, as shown in Fig. 1.1. Modern science has taught us that, at some stage, one would reach the molecular



**FIGURE 1.1** The length of the wood stick:  $l_1 = 1 \text{ ft}$  in day 1,  $l_2 = 1/2 \text{ ft}$  in day 2, and  $l_n = 1/2^{n-1} \text{ ft}$  in day  $n$ .

level and even the atomic level, below which the physical and chemical properties are completely different from those of the original material. The wooden stick or slice would eventually become something else that is not distinguishable from the other constituents in the atmosphere. Basically, properties of materials at very small scales may be quite different from those of the corresponding bulk materials. Note that 1 nm (nanometer) is one-billionth of a meter. The diameter of a hydrogen atom H is on the order of 0.1 nm, and that of a hydrogen molecule  $\text{H}_2$  is approximately 0.3 nm. Using the formula  $l_n = 0.3048/2^{n-1} \text{ m}$ , where  $n$  is number of days, we find  $l_{30} = 5.7 \times 10^{-10} \text{ m}$  (or 0.57 nm) after just a month, which is already near the diameter of a hydrogen atom.

While atoms can still be divided with large and sophisticated facilities, our ability to observe, manipulate, and utilize them is very limited. On the other hand, most biological processes occur at the molecular level. Many novel physical phenomena happen at the length scale of a few nanometers and can be integrated into large systems. This is why the nanometer is a critical length scale for the realization of practically important new materials, structures, and phenomena. For example, carbon nanotubes with diameters ranging from 0.4 to 50 nm or so have dramatically different properties. Some researchers have shown that these nanotubes hold promise as the building block of nanoelectronics. Others have found that the thermal conductivity of single-walled carbon nanotubes at room temperature could be an order of magnitude higher than that of copper. Therefore, carbon nanotubes have been considered as a candidate material for applications that require a high heat flux.

In conventional fluid mechanics and heat transfer, we treat the medium as a *continuum*, i.e., indefinitely divisible without changing its physical nature. All the intensive

properties can be defined locally and continuously. For example, the local density is defined as

$$\rho = \lim_{\delta V \rightarrow 0} \frac{\delta m}{\delta V} \quad (1.1)$$

where  $\delta m$  is the mass enclosed within a volume element  $\delta V$ . When the characteristic dimension is comparable with or smaller than that of the *mechanistic* length—for example, the molecular *mean free path*, which is the average distance that a molecule travels between two collisions—the continuum assumption will break down. The density defined in Eq. (1.1) will depend on the size of the volume,  $\delta V$ , and will fluctuate with time even at macroscopic equilibrium. Noting that the mean free path of air at standard atmospheric conditions is about 70 nm, the continuum assumption is well justified for many engineering applications until the submicrometer regime or the nanoscale is reached. Nevertheless, if the pressure is very low, as in an evacuated chamber or at a high elevation, the mean free path can be very large; and thus, the continuum assumption may break down even at relatively large length scales.

Within the macroscopic framework, we calculate the temperature distribution in a fluid or solid by assuming that the medium under consideration is not only a continuum but also at thermodynamic equilibrium everywhere. The latter condition is called the *local-equilibrium* assumption, which is required because temperature can be defined only for stable-equilibrium states. With extremely high temperature gradients at sufficiently small length scales and/or during very short periods of time, the assumption of local equilibrium may be inappropriate. An example is the interaction between short laser pulses and a material. Depending on the type of laser, the pulse duration or width can vary from a few tens of nanoseconds down to several femtoseconds ( $1 \text{ fs} = 10^{-15} \text{ s}$ ). In the case of ultrafast laser interaction with metals, free electrons in the metal could gain energy quickly to arrive at an excited state corresponding to an effective temperature of several thousand kelvins, whereas the crystalline lattices remain near room temperature. After an elapse of time represented by the electron relaxation time, the excess energy of electrons will be transferred to phonons, which are energy quanta of lattice vibration, thereby causing a heating effect that raises the temperature or changes the phase of the material under irradiation.

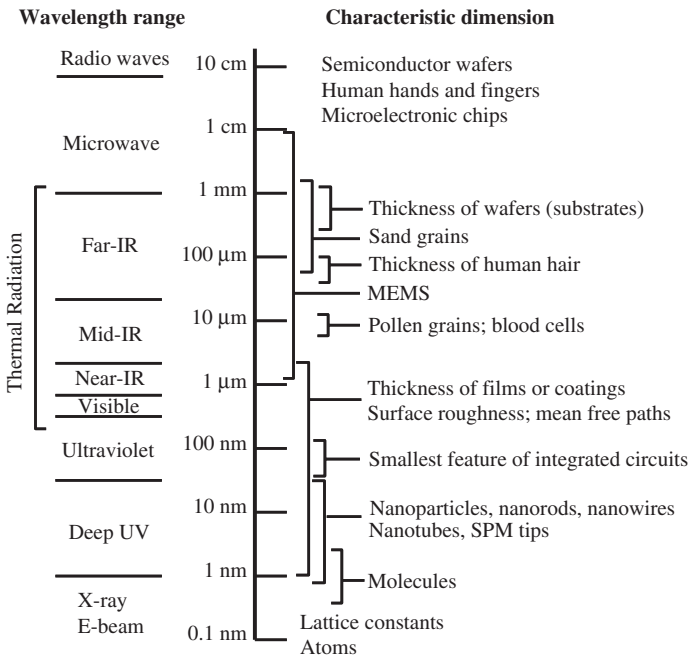
Additional mechanisms may affect the behavior of a system as the physical dimensions shrink or as the excitation and detection times are reduced. A scale-down of the theories developed from macroscopic observations often proves to be unsuitable for applications involving micro/nanoscale phenomena. Examples are reductions in the conductivity of thin films or thin wires due to boundary scattering (size effect), discontinuous velocity and temperature boundary conditions in microfluidics, wave interferences in thin films, and tunneling of electrons and photons through narrow gaps. In the quantum limit, the thermal conductance of a nanowire will reach a limiting value that is independent of the material that the nanowire is made of. At the nanoscale, the radiation heat transfer between two surfaces can exceed that calculated from the Stefan-Boltzmann law by several orders of magnitude. Another effect of miniaturization is that surface forces (such as shear forces) will scale down with  $L^2$ , where  $L$  is the characteristic length, while volume forces (such as buoyancy) will scale down with  $L^3$ . This will make surface forces predominant over volume forces at the microscale.

## 1.2 THE LENGTH SCALES

---

It is instructive to compare the length scales of different phenomena and structures, especially against the wavelength of the electromagnetic spectrum. Figure 1.2 compares the wavelength ranges with some characteristic dimensions. One can see that MEMS generally produce micromachining capabilities from several millimeters down to a few micrometers.

Currently, the smallest feature of integrated circuits is well below 100 nm. The layer thickness of thin films ranges from a few nanometers up to several micrometers. The wavelengths of the visible light are in the range from approximately 380 to 760 nm. On the other hand, thermal radiation covers a part of the ultraviolet, the entire visible and infrared, and a portion of the microwave region. The thickness of human hair is between 50 and 100  $\mu\text{m}$ , while the diameter of red blood cells is about 6 to 8  $\mu\text{m}$ . A typical optical microscope can magnify 100 times with a resolution of 200 to 300 nm, which is about half the wavelength and is limited due to the diffraction of light. Therefore, optical microscopy is commonly used to study micrometer-sized objects. On the other hand, atoms and molecules are on the order of 1 nm, which falls in the x-ray and electron-beam wavelength region. Therefore, x-ray and electron microscopes are typically used for determining crystal structures and defects, as well as for imaging nanostructures. The development of scanning probe microscopes (SPMs) and near-field scanning optical microscopes (NSOMs) in the 1980s enabled unprecedented capabilities for the visualization and manipulation of nanostructures, such as nanowires, nanotubes, nanocrystals, single molecules, individual atoms, and so forth, as will be discussed in Sec. 1.3.4. Figure 1.2 also shows that the mean free path of heat



**FIGURE 1.2** Characteristic length scales as compared with the wavelength of electromagnetic spectrum.

carriers (e.g., molecules in gases, electrons in metals, and phonons or lattice vibration in dielectric solids) often falls in the micrometer to nanometer scales, depending on the material, temperature, and type of carrier.

This book is motivated by the need to understand the thermal phenomena and heat transfer processes in micro/nanosystems and at very short time scales for solving problems occurring in contemporary and future technologies. A brief historical retrospective is given in the next section on the development of modern science and technologies, with a focus

on the recent technological advances leading to nanotechnology. The role of thermal engineering throughout this technological advancement is outlined.

### **1.3 FROM ANCIENT PHILOSOPHY TO CONTEMPORARY TECHNOLOGIES**

---

Understanding the fundamentals of the composition of all things in the universe, their movement in space and with time, and the interactions between one and another is a human curiosity and the inner drive that makes us different from other living beings on the earth. The ancient Chinese believed that everything was composed of the five elements: metal, wood, water, fire, and earth (or soil) that generate and overcome one another in certain order and time sequence. These simple beliefs were not merely used for fortune-telling but have helped the development of traditional Chinese medicine, music, military strategy, astronomy, and calendar. In ancient Greece, the four elements (fire, earth, air, and water) were considered as the realm wherein all things existed and whereof all things consisted. These classical element theories prevailed in several other countries in somewhat different versions for over 2000 years, until the establishment of modern atomic theory that began with John Dalton's experiment on gases some 200 years ago. In 1811, Italian chemist Amedeo Avogadro introduced the concept of the molecule, which consists of stable systems or bound state of atoms. A molecule is the smallest particle that retains the chemical properties and composition of a pure substance. The first periodic table was developed by Russian chemist Dmitri Mendeleev in 1869. Although the original meaning of atom in Greek is "indivisible," subatomic particles have since been discovered. For example, electrons as a subatomic particle were discovered in 1897 by J. J. Thomson, who won the 1906 Nobel Prize in Physics. An atom is known as the smallest unit of one of the 116 confirmed elements so far.

The first industrial revolution began in the late eighteenth century and boosted the economy of western countries from manual labor to the machine age by the introduction of machine tools and textile manufacturing. Following the invention of the steam engine in the mid-nineteenth century, the second industrial revolution had an even bigger impact on human life through the development of steam-powered ships and trains, along with the internal combustion engines, and the generation of electrical power. Newtonian mechanics and classical thermodynamics have played an indispensable role in the industrial revolutions. The development of machinery and the understanding of the composition of matter have allowed unprecedented precision of experimental investigation of physical phenomena, leading to the establishment of modern physics in the early twentieth century.

The nature of light has long been debated. At the turn of the eighteenth century, Isaac Newton formulated the corpuscular theory of light and observed with his prism experiment that sunlight is composed of different colors. In the early nineteenth century, the discovery of infrared and ultraviolet radiation and Young's double-slit experiment confirmed Huygens' wave theory, which was overshadowed by Newton's corpuscular theory for some 100 years. With the establishment of Maxwell's equations that fully describe the electromagnetic waves and Michelson's interferometric experiment, the wave theory of radiation had been largely accepted by the end of the nineteenth century. While the wave theory was able to explain most of the observed phenomena, it could not explain thermal emission over a wide spectrum, nor was it able to explain the photoelectric effect. Max Planck in 1901 used the hypothesis of light, or radiation quanta, or oscillators, to successfully derive the blackbody spectral distribution function. In 1905, Albert Einstein explained the photoelectric effect based on the concept of radiation quanta. To knock out an electron from the metal surface, the energy of each incoming radiation quantum ( $h\nu$ ) must be sufficiently large because one electron can absorb only one quantum. This explained why photoemission could not occur

at frequencies below the threshold value, no matter how intense the incoming radiation might be. It appears that light is not indefinitely divisible but must exist in multiples of the smallest massless quanta, which are known as photons. In 1924, Louis de Broglie hypothesized that particles should also exhibit wavelike characteristics. With the electron diffraction experiment, it was found that electrons indeed can behave like waves with a wavelength inversely proportional to the momentum. Electron microscopy was based on the principle of electron diffraction. The wave-particle duality was essential to the establishment of quantum mechanics in the early twentieth century. Quantum mechanics describes the phenomena occurring in minute particles, structures, and their interaction with radiation, for which classical mechanics and electrodynamics are not applicable. The fundamental scientific understanding gained during the first half of the twentieth century has facilitated the development of contemporary technologies that have transformed from the industrial economy to the knowledge-based economy and from the machine age to the information age. The major technological advancements in the last half of the century are highlighted in the following sections.

### 1.3.1 Microelectronics and Information Technology

In his master's thesis at MIT published in 1940, Claude Shannon (1916–2001) used the Boolean algebra and showed how to use TRUE and FALSE to represent function of switches in electronic circuits. Digital computers were invented during the 1940s in several countries, including the IBM Mark I which is 2.4 m high and 16 m long. In 1948, while working at Bell Labs, Shannon published an article, "A Mathematical Theory of Communication," which marked the beginning of modern communication and information technology.<sup>5</sup> In that paper, he laid out the basic principles of underlying communication of information with two symbols, 1 and 0, and coined the term "bit" for a binary digit. His theory made it possible for digital storage and transmission of pictures, sounds, and so forth.

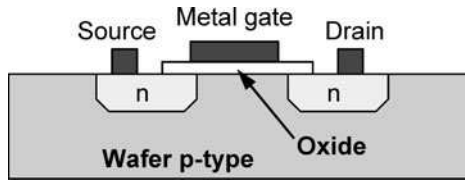
In December 1947, three scientists invented the semiconductor point contact transistor at Bell Labs using germanium. The earlier computers and radios were based on bulky vacuum tubes that generated a huge amount of heat. The invention of the transistor by William Shockley, John Bardeen, and Walter Brattain was recognized through the Nobel Prize in Physics conferred on them in 1956. There had been intense research on semiconductor physics using the atomic theory and the mechanism of point contact for the fabrication of transistors to become possible. The invention of transistors ushered the information age with a whole new industry.

In 1954, Gordon Teal at Texas Instruments built the first silicon transistor. The native oxide of silicon appeared to be particularly suitable as the electric insulator. In 1958, Jack Kilby (1923–2005) at Texas Instruments was able to cram all the discrete components onto a silicon base and later onto one piece of germanium. He filed a patent application the next year on "Miniaturized Electronic Circuits," where he described how to make integrated circuits and connect the passive components via gold wires. Working independently, Robert Noyce at Fairchild Electronics in California found aluminum to adhere well to both silicon and silicon oxide and filed a patent application in 1959 on "Semiconductor Device-and-Lead Structure." Kilby and Noyce are considered the coinventors of integrated circuits. Noyce was one of the founders of Intel and died in 1990. Kilby was awarded half of the Nobel Prize in Physics in 2000 "for his part in the invention of the integrated circuit." (See [http://nobelprize.org/nobel\\_prizes/](http://nobelprize.org/nobel_prizes/)). The other half was shared by Zhores Alferov and Herbert Kroemer for developing semiconductor heterogeneous structures used in optoelectronics, to be discussed in the next section.

In 1965, around 60 transistors could be packed on a single silicon chip. Seeing the fast development and future potential of integrated circuits, Gordon Moore, a cofounder of

Intel, made a famous prediction that the number and complexity of semiconductor devices would double every year.<sup>6</sup> This is Moore's law, well-known in the microelectronics industry. In the mid-1970s, the number of transistors on a chip increased from 60 to 5000. By 1985, the Intel 386 processor contained a quarter million transistors on a chip. In 2001, the Pentium 4 processor reached 42 million transistors. The number has now exceeded 1 billion per chip in 2006. When the device density is plotted against time in a log scale, the growth almost follows a straight line, suggesting that the packaging density has doubled approximately every 18 months.<sup>6</sup> Reducing the device size and increasing the packaging density have several advantages. For example, the processor speed increases by reducing the distance between transistors. Furthermore, new performance features can be added into the chip to enhance the performance. The cost for the same performance also reduces. Advanced supercomputer systems have played a critical role in enabling modeling and understanding micro/nanoscale phenomena.

The process is first to grow high-quality silicon crystals and then dice and polish into wafers. Devices are usually made on  $\text{SiO}_2$  layer that can be grown by heating the wafer to sufficiently high temperatures in a furnace with controlled oxygen partial pressure. The wafers are then patterned using photolithographic techniques combined with etching processes. Donors and acceptors are added to the wafer to form *n*- and *p*-type regions by ion implantation and then annealed in a thermal environment. Metals or heavily doped polycrystalline silicon are used as gates with proper coverage and patterns through lithography. A schematic of metal-oxide-semiconductor field-effect transistor (MOSFET) is shown in Fig. 1.3. Millions of transistors can be packed in 1-mm<sup>2</sup> area with several layers



**FIGURE 1.3** Schematic of a metal-oxide-semiconductor field-effect transistor (MOSFET).

through very-large-scale integration (VLSI) with the smallest features smaller than 100 nm. As mentioned earlier, managing heat dissipation is a challenge especially as the device dimension continues to shrink. Local heating or hot spots on the size of 10 nm can cause device failure. The principles governing the heat transfer at the nanoscale are very different from those at large scales. A fundamental understanding of the phonon transport is required for device-level thermal analysis. Furthermore, understanding heat transfer in microfluidics is necessary to enable reliable device cooling at the micro- and nanoscales. Additional discussions will be given in subsequent chapters of the book.

The progress in microelectronics is not possible without the advances in materials such as crystal growth and thermal processing during semiconductor manufacturing, as well as the deposition and photolithographic technologies. Rapid thermal processing (RTP) is necessary during annealing and oxidation to prevent ions from deep diffusion into the wafer. Thermal modeling of RTP must consider the combined conduction, convection, and radiation modes. A lightpipe thermometer is commonly used to monitor the temperature of the wafer. In an RTP furnace, the thermal radiation emitted by the wafer is collected by the light pipe and then transmitted to the radiometer for inferring the surface temperature.<sup>7</sup> In some cases,



the wafer surface is rough with anisotropic features. A better understanding of light scattering by anisotropic rough surfaces is also necessary.

According to the International Technology Roadmap for Semiconductors, the gate length and the junction depth will be 25 and 13.8 nm, respectively, for the 65-nm devices used in high-performance complementary-metal-oxide-semiconductor (CMOS) technology.<sup>8</sup> High-intensity Ar or Xe arc lamps with millisecond optical pulses are considered as a suitable annealing tool following ion implantation in ultra-shallow junction fabrication. Because the optical energy is absorbed within milliseconds, thermal diffusion cannot distribute heat uniformly across the wafer surface. Therefore, temperature uniformity across the nanometer-patterned wafer is expected to be a critical issue. To reduce the feature size further, deep-UV lithography and x-ray lithography have also been developed. It is predicted that Moore's law will reach its limit in 2017, when the critical dimensions would be less than 10 nm. Further reduction will be subjected to serious barriers due to problems associated with gate dielectrics and fabrication difficulties. Molecular nanoelectronics using self-assembly is sought as an alternative, along with quantum computing. Therefore, nanoelectronics and quantum computing are anticipated to brighten the electronics and computer future.

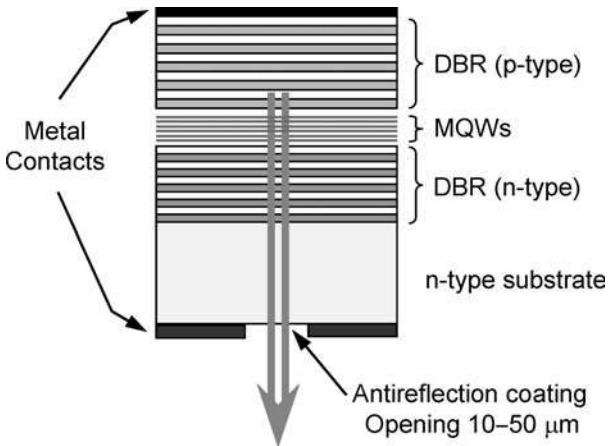
### 1.3.2 Lasers, Optoelectronics, and Nanophotonics

It is hard to imagine what the current technology would look like without lasers. Lasers of different types have tremendous applications in metrology, microelectronics fabrication, manufacturing, medicine, and communication. Examples are laser printers, laser bar code readers, laser Doppler velocimetry, laser machining, and laser corneal surgery for vision correction. The concept of laser was demonstrated in late 1950s independently in the United States and the Soviet Union during the cold war. The Nobel Prize in Physics of 1964 recognized the fundamental contributions in the field of quantum electronics by Charles Townes, Nicolay Basov, and Aleksandr Prokhorov. The first working laser was a Ruby laser built by Theodore Maiman at Hughes Aircraft Company in 1960. The principle of lasers dates back to 1917, when Einstein elegantly depicted his conception of stimulated emission of radiation by atoms. Unlike thermal emission and plasma emission, lasers are coherent light sources and, with the assistance of optical cavity, lasers can emit nearly monochromatic light and point to the same direction with little divergence. Lasers enabled a branch of nonlinear optics, which is important to understand the fundamentals of light-matter interactions, communication, as well as optical computing. In 1981, Nicolaas Bloembergen and Arthur Schawlow received the Nobel Prize in Physics for their contributions in laser spectroscopy. There are a variety of nonlinear spectroscopic techniques, including Raman spectroscopy, as reviewed by Fan and Longtin.<sup>9</sup> Two-photon spectroscopy has become an important tool for molecular detection.<sup>10</sup> Furthermore, two-photon 3-D lithography has also been developed for microfabrication.<sup>11</sup>

Gas lasers such as He-Ne (red) and Ar (green) have been extensively used for precision alignment, dimension measurements, and laser Doppler velocimetry due to their narrow linewidth. On the other hand, powerful Nd:YAG and CO<sub>2</sub> lasers are used in thermal manufacturing, where the heat transfer processes include radiation, phase change, and conduction.<sup>12</sup> Excimer lasers create nanosecond pulses in ultraviolet and have been extensively used in materials processing, ablation, eye surgery, dermatology, as well as photolithography in microelectronics and microfabrication. High-energy nanosecond pulses can also be produced by *Q*-switching, typically with a solid state laser such as Nd:YAG laser at a wavelength near 1  $\mu\text{m}$ . On the other hand, mode-locking technique allows pulse widths from picoseconds down to a few femtoseconds. Pulse durations less than 10 fs have been achieved since 1985. Ultrafast lasers have enabled the study of reaction dynamics and formed a branch in chemistry called *femtochemistry*. Ahmed Zewail of Caltech received the

1999 Nobel Prize in Chemistry for his pioneering research in this field. In 2005, John Hall and Theodor Hänsch received the Nobel Prize in Physics for developing laser-based precision spectroscopy, in particular, the frequency comb technique. Short-pulse lasers can facilitate fabrication, the study of electron-phonon interaction in the nonequilibrium process, measurement of thermal properties including interface resistance, nondestructive evaluation of materials, and so forth.<sup>13–16</sup>

Room-temperature continuous-operation semiconductor lasers were realized in May 1970 by Zhores Alferov and coworkers at the Ioffe Physical Institute in Russia, and independently by Morton Panish and Izuo Hayashi at Bell Labs a month later. Alferov received the Nobel Prize in Physics in 2000, together with Herbert Kroemer who conceived the idea of double-heterojunction laser in 1963 and was also an earlier pioneer of molecular beam epitaxy (MBE). Invented in 1968 by Alfred Cho and John Arthur at Bell Labs and developed in the 1970s, MBE is a high-vacuum deposition technique that enables the growth of highly pure semiconductor thin films with atomic precision. The name heterojunction refers to two layers of semiconductor materials with different bandgaps, such as GaAs/Al<sub>x</sub>Ga<sub>1-x</sub>As pair. In a double-heterojunction structure, a lower-bandgap layer is sandwiched between two higher-bandgap layers.<sup>17</sup> When the middle layer is made thin enough, on the order of a few nanometers, the structure is called a quantum well because of the discrete energy levels and enhanced density of states. Quantum well lasers can have better performance with a smaller driving current. Multiple quantum wells (MQWs), also called superlattices, that consist of periodic structures can also be used to further improve the performance. In a laser setting, an optical cavity is needed to confine the laser bandwidth as well as enhance the intensity at a desired wavelength with narrow linewidth. Distributed Bragg reflectors (DBRs) are used on both ends of the quantum well (active region). DBRs are the simplest photonic crystals made of periodic dielectric layers of different refractive indices; each layer thickness is equal to a quarter of the wavelength in that medium ( $\lambda/n$ ). DBRs are dielectric mirrors with nearly 100% reflectance, except at the resonance wavelength  $\lambda$ , where light will eventually escape from the cavity. Figure 1.4 illustrates a vertical cavity surface emitting laser (VCSEL), where light is emitted through the substrate (bottom of the structure). The energy transfer mechanisms through phonon waves and electron waves have been extensively investigated.<sup>18</sup>



**FIGURE 1.4** Schematic of a VCSEL laser made of heterogeneous quantum well structure. The smaller layer thickness can be 3 nm, and there can be as many as several hundred layers.

Further improvement in the laser efficiency and control of the wavelength has been made using quantum wires and quantum dots (QDs).<sup>17</sup>

Semiconductor lasers are the most popular lasers (in quantity), and several hundred-million units are sold each year. Their applications include CD/DVD reading/writing, optical communication, laser pointers, laser printers, bar code readers, and so forth. A simpler device is the light-emitting diode (LED), which emits incoherent light with a two-layer *p-n* junction without DBRs. LEDs have been used for lighting, including traffic lights with improved efficiency and decorating lights. The development of wide-bandgap materials, such as GaN and AlN epitaxially grown through metal-organic chemical vapor deposition (MOCVD), allows the LED and semiconductor laser wavelength to be pushed to the blue and ultraviolet. Organic light-emitting diodes (OLEDs) based on electroluminescence are being developed as a promising candidate for the next-generation computer and TV displays.

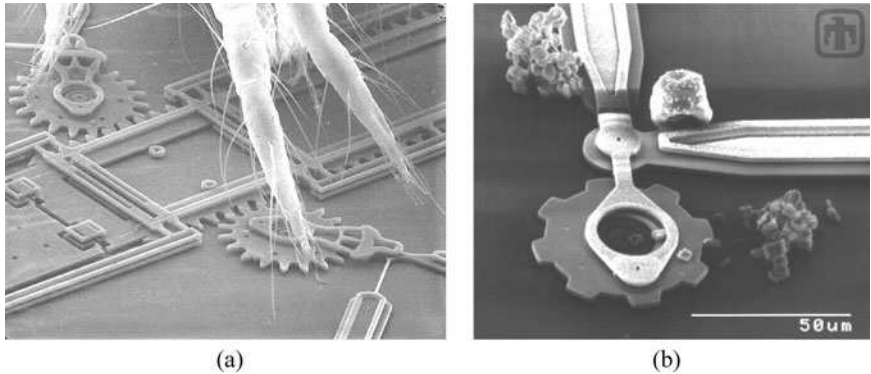
Alongside the development of light sources, there have been continuous development and improvement in photodetectors, mainly in focal plane arrays, charge-coupled devices (CCDs), quantum well detectors, readout electronics, data transfer and processing, compact refrigeration and temperature control, and so forth. On the other hand, optical fibers have become an essential and rapidly growing technology in telecommunication and computer networks. The optical fiber technology for communication was developed in the 1970s along with the development of semiconductor lasers. In 1978, Nippon Telegraph and Telephone (NTT) demonstrated the transmission of 32 Mbps (million-bits-per-second) through 53 km of graded-index fiber at 1.3- $\mu\text{m}$  wavelength. By 2001,  $3 \times 10^{11}$  m of fiber-optic wires had been installed worldwide; this is a round-trip from the earth to the sun. In March 2006, NEC Corporation announced a 40-Gbps optical-fiber transmission system. Optical fibers have also been widely applied as sensors for biochemical detection as well as temperature and pressure measurements. Fiber drawing process involves complicated heat transfer and fluid dynamics at different length scales and temperatures.<sup>19</sup>

Nanophotonics is an emerging frontier that integrates photonics with physics, chemistry, biology, materials science, manufacture, and nanotechnology. The foundation of nanophotonics is to study interactions between light and matter, to explore the unique characteristics of nanostructures for utilizing light energy, and to develop novel nanofabrication and sensing techniques. Recent studies have focused on photonic crystals, nanocrystals, plasmonic waveguides, nanofabrication and nanolithography, light interaction with organic materials, biophotonics, biosensors, quantum electrodynamic, nanocavities, quantum dot and quantum wire lasers, solar cells, and so forth. Readers are referred to Prasad<sup>20</sup> for an extensive discussion of the recent developments. In the field of thermal radiation, nanoscale radiative transfer and properties have become an active research area, and a special issue of the *Journal of Heat Transfer* is devoted to this exciting area.<sup>21</sup>

### 1.3.3 Microfabrication and Nanofabrication

Richard Feynman, one of the best theoretical physicists of his time and a Nobel Laureate in Physics, delivered a visionary speech at Caltech in December 1959, entitled "There's plenty of room at the bottom." At that time, lasers had never existed and integrated circuits had just been invented and were not practically useful, and a single computer that is not as fast as a present-day handheld calculator would occupy a whole classroom with enormous heat generation. Feynman envisioned the future of controlling and manipulating things on very small scales, such as writing (with an electron beam) the whole 24 volumes of *Encyclopedia Britannica* on the head of a pin and rearranging atoms one at a time.<sup>22</sup> Many of the things Feynman predicted were once considered scientific fictions or jokes but have been realized in practice by now, especially since the 1980s. In 1983, Feynman gave a second talk about the use of swimming machine as a medical device: the surgeon that you could swallow, as well as quantum computing.<sup>22</sup> In the 1990s, micromachining and MEMS

emerged as an active research area, with a great success by the commercialization of the micromachined accelerometers in the automobile airbag. Using the etching and lithographic techniques, engineers were able to manufacture microscopic machines with moving parts, as shown in Fig. 1.5, such as gears with a size less than the cross-section of human hair.



**FIGURE 1.5** MEMS structures. (a) A dust mite on a microfabricated mirror assembly, where the gears are smaller than the thickness of human hair. (b) Drive gear chain with linkages, where coagulated red blood cells are on the upper left and a grain of pollen is on the upper right. (Courtesy of Sandia National Laboratories, SUMMIT Technologies, [www.mems.sandia.gov](http://www.mems.sandia.gov).)

The technologies used in microfabrication have been extensively discussed in the text of Madou.<sup>23</sup> These MEMS devices were later developed as tools for biological and medical diagnostics, such as the so-called lab-on-a-chip, with pump, valve, and analysis sections on the 10 to 100  $\mu\text{m}$  scale. In aerospace engineering, an application is to build micro-air vehicles or microflyers, with sizes ranging from a human hand down to a bumblebee that could be used for surveillance and reconnaissance under extreme conditions. Microchannels and microscale heat pipes have also been developed and tested for electronic cooling applications. The study of microfluidics has naturally become an active research area in mechanical engineering. The development of SPM and MEMS technologies, together with materials development through self-assembly and other technologies, lead to further development of even smaller structures and the bottom-up approach of nanotechnology. Laser-based manufacturing, focused ion beam (FIB), and electron-beam lithography have also been developed to facilitate nanomanufacturing. In NEMS, quantum behavior becomes important and quantum mechanics is inevitable in understanding the behavior.

Robert Curl, Harold Kroto, and Richard Smalley were winners of the Nobel Prize in Chemistry in 1996 for their discovery of fullerenes in 1985 at Rice University, during a period Kroto visited from University of Sussex. The group used pulsed laser irradiation to vaporize graphite and form carbon plasma in a pressurized helium gas stream. The result as diagnosed by time-of-flight mass spectroscopy suggested that self-assembled  $\text{C}_{60}$  molecules were formed and would be shaped like a soccer ball with 60 vertices made of the 60 carbon atoms; see Kroto et al., *Nature*, **318**, 162 (1985). The results were confirmed later to be  $\text{C}_{60}$  molecules indeed with a diameter on the order of 1 nm with wave-particle duality. This type of carbon allotrope is called a buckminsterfullerene, or fullerene, or buckyball, after the famous architect Buckminster Fuller (1895–1983) who designed geodesic domes. In 1991, Sumio Iijima of NEC Corporation synthesized carbon nanotubes (CNTs) using

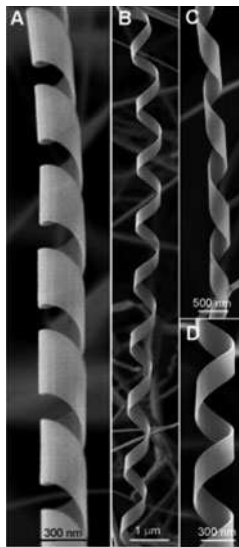
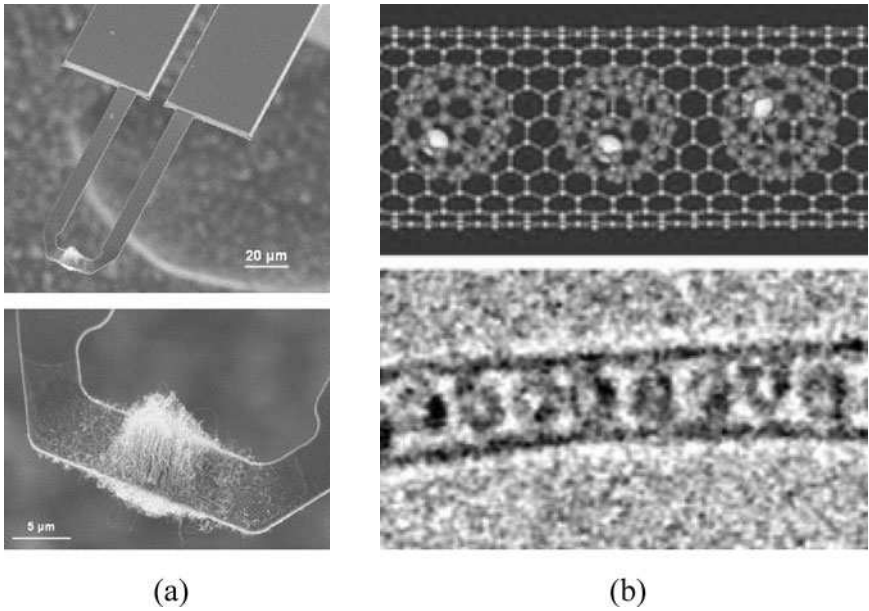
arc discharge. Soon his group and an IBM group were able to produce single-walled carbon nanotubes (SWNTs) with a diameter on the order of 1 nm. There have been intensive studies of CNTs for hydrogen storage, nanotransistors, field emission, light emission and absorption, quantum conductance, nanocomposites, and high thermal conductivity. Figure 1.6*a* shows CNTs growth at a room-temperature environment by chemical vapor deposition on a heated cantilever tip with a size around 5  $\mu\text{m}$ .<sup>24</sup> Figure 1.6*b* shows the synthesized SWNTs with encapsulated metallofullerenes of Gd:C<sub>82</sub> (i.e., a gadolinium inside a fullerene molecule). The high-resolution transmission electron microscope (TEM) image suggests that the diameter of the SWNT is from 1.4 to 1.5 nm.<sup>25</sup> It should be noted that electron microscopes, including SEM and TEM, have become a powerful tool for imaging micro/nanoscale objects with a magnification up to 2 million. The first electron microscope was built by Ernst Ruska and Max Knoll in Germany during the early 1930s, and Ruska shared the Nobel Prize in Physics in 1986 for his contributions to electron optics and microscopy.

Various nanostructured materials have been synthesized, such as silicon nanowires, InAs/GaAs QDs, and Ag nanorods. Figure 1.6*c* shows some images for nanohelices or nanosprings made of ZnO nanobelts or nanoribbons using a solid-vapor process.<sup>26</sup> These self-assembled structures under controlled conditions could be fundamental to the study of electromagnetic coupled nanodevices for use as sensors and actuators, as well as the growth dynamics at the nanoscale.

One of the successful technologies that operate in the regime of quantum mechanical domain is the giant magnetoresistive (GMR) head and hard drive. The GMR head is based on ferromagnetic layers separated by an extremely thin (about 1 nm) nonferromagnetic spacer, such as Fe/Cr/Fe and Co/Cu/Co. MBE enabled the metallic film growth with required precision and quality. The electrical resistance of GMR materials depends strongly on the applied magnetic field, which affects the spin states of electrons. IBM first introduced this technology in 1996, which was only about 10 years after the publication of the original research results; see Grünberg et al., *Phys. Rev. Lett.*, **57**, 2442 (1986); Baibich et al., *Phys. Rev. Lett.*, **61**, 2472 (1988). GMR materials have been extensively used in computer hard drive and read/write head. Overheating, due to friction with the disk surface, can render the data unreadable for a short period until the head temperature stabilizes; such an effect is called *thermal asperity*. Yang et al. performed a detailed thermal characterization of Cu/CoFe superlattices for GMR head applications using MEMS-based thermal metrology tools.<sup>27</sup>

### 1.3.4 Probing and Manipulation of Small Structures

Tunneling by elementary particles is a quantum mechanical phenomenon or wavelike behavior. It refers to a potential barrier of the particles that normally will confine the particles to either side of the barrier, like a mountain that is so high as to separate people on one side from those on the other. When the barrier thickness is thin enough, quantum tunneling can occur and particles can transmit through the barrier, as if a tunnel is dug through a mountain. An example is an insulator between two metal strips. Trained in mechanical engineering, Ivar Giaever performed the first tunneling experiment with superconductors in 1960 at the General Electric Research Laboratory and received the 1973 Nobel Prize in Physics, together with Leo Esaki of IBM and Brian Josephson. Esaki made significant contributions in semiconductor tunneling, superlattices, and the development of MBE technology. He invented a tunneling diode, called the Esaki diode, which is capable of very fast operation in the microwave region. Josephson further developed the tunneling theory and a device, called a Josephson junction, which is used in the superconducting quantum interface devices (SQUIDS), for measuring extremely small magnetic fields. SQUIDS are used in magnetic resonance imaging (MRI) for medical diagnostics.



(c)

**FIGURE 1.6** Examples of nanostructures. (a) SEM image of CNTs grown on heated cantilever tip. (Reprinted with permission from Sunden *et al.*,<sup>24</sup> copyright 2006, American Institute of Physics.) (b) Buckyballs inside a SWNT (the lower is a TEM image in which the nanotube diameter is 1.4 to 1.5 nm). (Reprinted with permission from Hirahara *et al.*,<sup>25</sup> copyright 2000, American Physical Society.) (c) TEM images of ZnO nanobelts that are coiled into nanohelices or nanosprings. [Reprinted with permission from Gao *et al.*,<sup>26</sup> copyright 2005, AAAS (image courtesy of Prof. Z. L. Wang, Georgia Tech).]

In 1981, Gerd Binnig and Heinrich Rohrer of IBM Zurich Research Laboratory developed the first scanning tunneling microscope (STM) based on electron tunneling through vacuum. This invention has enabled the detection and manipulation of surface phenomena at the atomic level and, thus, has largely shaped the nanoscale science and technology through further development of similar instrumentation. Binnig and Rohrer shared the Nobel Prize in Physics in 1986, along with Ruska who developed the first electron microscope as mentioned earlier. STM uses a sharp-stylus-probe tip and piezoelectricity for motion control. When the tip is near 1 nm from the surface, an electron can tunnel through the tip to the conductive substrate. The tunneling current is very sensitive to the gap. Therefore, by maintaining the tip in position and scanning the substrate in the  $x$ - $y$  direction with a constant current (or distance), the height variation can be obtained with extremely good resolution (0.02 nm). Using STM, Binnig et al. soon obtained the real-space reconstruction of the  $7 \times 7$  unit cells of Si(111).<sup>28</sup> In 1993, another group at IBM Almaden Research Center was able to manipulate iron atoms to create a 48-atom quantum corral on a copper substrate.<sup>29</sup> The images have appeared in the front cover of many magazines, including *Science* and *Physics Today*. STM can also be used to assemble organic molecules and to study DNA molecules.<sup>2</sup>

In 1996, Gerd Binnig, Calvin Quate, and Christoph Gerber developed another type of SPM, i.e., the atomic force microscope (AFM) that can operate without a vacuum environment and for electrical insulators.<sup>30</sup> AFM uses a tapered tip at the end of a cantilever and an optical position sensor, as shown in Fig. 1.7. The position sensor is very sensitive to the

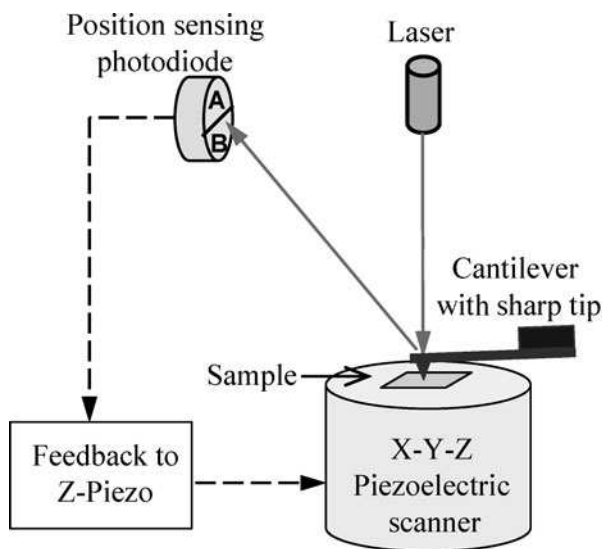


FIGURE 1.7 Schematic of an atomic force microscope (AFM).

bending of the cantilever (with a 0.1-nm vertical resolution). When the tip is brought close to the surface, there exist intermolecular forces (repulsive or attractive) between the tip and the atoms on the underneath surface. In the contact mode, the cantilever is maintained in position using the servo signal from the position-sensing diode to adjust the height of the sample, while it scans in the lateral direction. Surface topographic data can be obtained in

an ambient environment for nonconductive materials. Other SPMs have also been developed and the family of SPMs is quite large today. Wickramasinghe and coworkers first investigated thermal probing by attaching a thermocouple to the cantilever tip.<sup>31</sup> Later, Arun Majumdar's group developed several types of scanning thermal microscope (SThM) for nanoscale thermal imaging of heated samples, including microelectronic devices and nanotubes.<sup>32</sup> Recently, researchers have modified SThM for measuring and mapping thermoelectric power at nanoscales.<sup>33</sup>

Because of its simplicity, AFM has become one of the most versatile tools in nanoscale research, including friction measurements, nanoscale indentation, dip-pen nanolithography, and so forth. Heated cantilever tips were proposed for nanoscale indentation or writing on the polymethyl methacrylate (PMMA) surface, either using a laser or by heating the cantilever legs.<sup>34</sup> The method was further developed to concentrate the heat dissipation to the tip by using heavily doped legs as electrical leads, resulting in writing (with a density near 500 Gb/in<sup>2</sup>) and erasing (with a density near 400 Gb/in<sup>2</sup>) capabilities. The temperature signal measured by the tip resistance can also be used to read the stored data due to the difference in heat loss as the tip scans the area.<sup>35</sup> In an effort to improve the data-writing speed, IBM initiated the "millipede" project in 2000 and succeeded in making  $32 \times 32$  heated-cantilever array for which each cantilever was separately controlled.<sup>36</sup> Obviously, heat transfer and mechanical characteristics are at the center of these systems. The heated AFM cantilever tips have been used as a local heating source for a number of applications, including the above-mentioned CVD growth of CNTs locally and thermal dip-pen nanolithography.<sup>37</sup>

### 1.3.5 Energy Conversion Devices

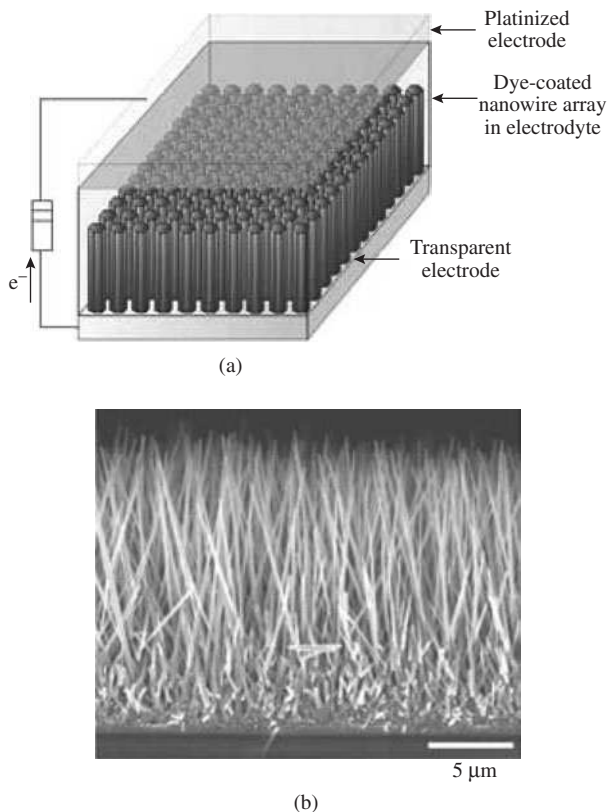
Nanostructures may have unique thermal properties that can be used to facilitate heat transfer for heat removal and thermal management applications. An example was mentioned earlier to utilize nanotubes with high thermal conductivity, although nanotube bundles often suffer from interface resistance and phonon scattering by defects and boundaries. Recently, there have been a number of studies on nanofluids, which are liquids with suspensions of nanostructured solid materials, such as nanoparticles, nanofibers, and nanotubes.<sup>38</sup> Enhanced thermal conductivity and increased heat flux have been demonstrated; however, the mechanisms that contributed to the enhancement and temperature dependence are still being debated.<sup>39</sup>

Thermoelectricity utilizes the irreversible thermodynamics principle for thermal-electrical conversion and can be used for cooling in microelectronics as well as miniaturized power generation. A critical issue is to enhance the figure of merit of performance, with a reduced thermal conductivity. Multilayer heterogeneous structures create heat barriers due to size effects and the boundary resistance. These structures have been extensively studied in the literature and demonstrate enhanced performances. Understanding the thermal and electrical properties of heterogeneous structures is critically important for future design and advancement.<sup>40</sup>

Fast-depleting reserves of conventional energy sources have resulted in an urgent need for increasing energy conversion efficiencies and recycling of waste heat. One of the potential candidates for fulfilling these requirements is thermophotovoltaic devices, which generate electricity from either the complete combustion of different fuels or the waste heat of other energy sources, thereby saving energy. The thermal radiation from the emitter is incident on a photovoltaic cell, which generates electrical currents. Applications of such devices range from hybrid electric vehicles to power sources for microelectronic systems. At present, thermophotovoltaic systems suffer from low conversion efficiency. Nanostructures have been extensively used to engineer surfaces with designed absorption, reflection, and emission characteristics. Moreover, at the nanoscale, the radiative energy transfer can be greatly enhanced due to tunneling and enhanced local density of states. A viable solution to



increase the thermophotovoltaic efficiency is to apply microscale radiation principles in the design of different components to utilize the characteristics of thermal radiation at small distances and in microstructures.<sup>41</sup> Nanostructures can also help increase the energy conversion efficiency and reduce the cost of solar cells. Figure 1.8 shows the device structure of a



**FIGURE 1.8** ZnO nanowires for dye-sensitized solar cells, from Law et al.<sup>42</sup> (Reprinted by permission from Macmillan Publishers Ltd.: *Nature Materials*, copyright 2005.) The height of the wires is near 16  $\mu\text{m}$  and their diameters vary between 130 and 200 nm. (a) Schematic of the cell with light incident through the bottom electrode. (b) SEM image of a cleaved nanowire array.

ZnO-nanowire array for dye-sensitized solar cells.<sup>42</sup> This structure can greatly enhance the absorption or quantum efficiency over nanoparticle-based films. Knowledge of the spectral and directional absorbance of nanostructures and heat dissipation mechanisms is critically important for further advancement of this type of device.<sup>43</sup>

Hydrogen technologies are being considered and actively pursued as the energy source of the future. There are two ways in which hydrogen  $\text{H}_2$  may be used: one is in a combustion heat engine where hydrogen reacts with oxygen intensively while releasing heat; the other is in a fuel cell where electrochemical reaction occurs quietly to generate electricity just like a battery. Because the only reaction product is water, hydrogen-powered automobiles can be made pollution free in principle. Grand challenges exist in generation, storage,

and transport of hydrogen. If all hydrogen is obtained from fossil fuels, there will be no reduction in either the fossil fuel consumption or the carbon dioxide emission, except that the emission is centralized in the hydrogen production plant. Alternatively, hydrogen may be produced from water with other energy sources, such as renewable energy sources. Nanomaterials are being developed for several key issues related to hydrogen technologies, such as hydrogen storage using nanoporous materials, effective hydrogen generation by harvesting solar energy with inexpensive photovoltaic materials, and fuel cells based on nanostructure catalysts.<sup>44</sup>

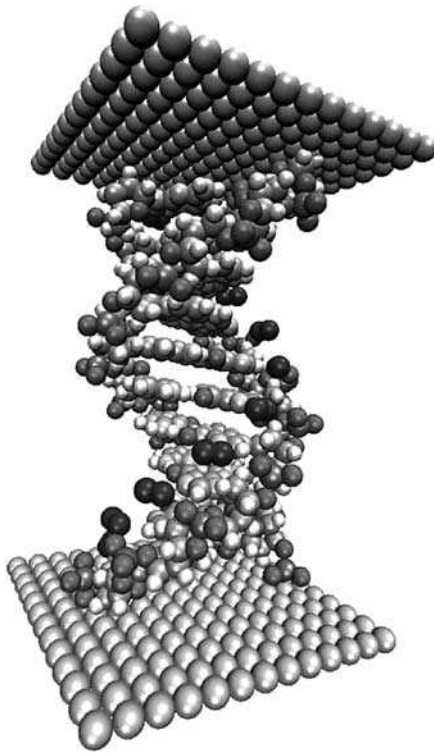
### 1.3.6 Biomolecule Imaging and Molecular Electronics

Optical microscopy has played an instrumental role in medical diagnoses because it allows us to see bacteria and blood cells. Optical wavelength is more desirable than x-ray or electron beam because of the less invasiveness and the more convenience. However, the resolution of a traditional microscope is on the order of half the wavelength due to the diffraction limit. While the concept of near-field imaging existed in the literature before 1930, it has been largely forgotten because of the inability in building the structures and controlling their motion. With the microfabrication and precision-positioning capabilities, near-field scanning optical microscopes (NSOMs, also called SNOMs) were realized in the early 1980s by different groups and extensively used for biomolecule imaging with a resolution of 20 to 50 nm.<sup>45</sup> The principle is to bring the light through an aperture of a tapered fiber of very small diameter at the end or to bring the light through an aperture of very small diameter. The beam out from the fiber tip or aperture will diverge quickly if the sample is placed in the far field, i.e., away from the aperture. However, high resolution can be achieved by placing the sample in close proximity to the aperture within a distance much less than the wavelength, i.e., in the near field, such that the beam size is almost the same as the aperture. An apertureless metallic tip can be integrated with an SPM to guide the electromagnetic wave via surface plasmon resonance with a spatial resolution as high as 10 nm, for high-resolution imaging and processing. There have since been extensive studies on near-field interactions between electromagnetic waves and nanostructured materials, from semiconductor QDs, metallic nanoaperture and nanohole arrays, to DNA and RNA structures.

Nanoparticles are among the earliest known nanostructures that have been used for centuries in making stained glass with gold or other metallic nanoparticles as well as photographic films with silver nanoparticles. A QD has a spherical core encapsulated in a shell made of another semiconductor material, such as a CdSe core in a ZnS shell. The outer shell is only several monolayers thick, and the diameters of QDs range from 2 to 10 nm. The material for the inner core has a smaller bandgap. Quantum confinement in the core results in size-dependent fluorescent properties. Compared with molecular dyes conventionally used for fluorescent labeling in cellular imaging, the emission from QD fluorophores is brighter with a narrower spectral width. QDs also allow excitation at shorter wavelengths, making it easier to separate the fluorescent signal from the scattered one, and are resistive to photobleaching that causes dyes to lose fluorescence. Furthermore, the emission wavelength can be selected by varying the core size of QDs to provide multicolor labeling. It was first demonstrated in 1998 that QDs could be conjugated to biomolecules such as antibodies, peptides, and DNAs, enabling surface passivation and water solubility. In recent years, significant development has been made to employ QDs for *in vivo* and *in vitro* imaging, labeling, and sensing.<sup>46,47</sup>

CMOS technology is a top-down semiconductor fabrication process, in which patterns are created by first making a mask and then printing the desired features onto the surface of the wafer via lithography. Integrated circuits have dominated the technological and economic progress in the past 30 years, and complex and high-density devices have been manufactured on silicon wafers. However, this technology is going to reach a limit in 10 to 15 years, when the smallest feature size is less than 10 nm. Molecular electronics is considered

as a promising alternative.<sup>48</sup> A 3-D assembly with short interconnect distances would greatly increase the information storage density and transfer speed with reduced power consumption and amount of heat dissipated. Self-assembly means naturally occurring processes, from biological growth to the galaxy formation. In materials synthesis, self-assembly implies that the end products or structures are formed under favorable conditions and environments. An example is the growth of bulk crystals from a seed. Fullerenes and nanotubes are formed by self-assembling, not by slicing a graphite piece and then rolling and bending it to the shape of a tube or a shell. Self-assembly is referred to as a bottom-up process, like constructing an airplane model with LEGO pieces. Biological systems rely on self-assembly and self-replication to develop. Since 2000, CNT-based transistors have been built by several groups and found to be able to outperform Si-based ones. Transistors have also been created using a single molecule of a transition-metal organic complex nanobridge between two electrodes.<sup>49</sup> Because of the small dimensions, quantum mechanics should govern the electrical and mechanical behaviors. Figure 1.9 illustrates an engineered



**FIGURE 1.9** An engineered DNA strand between metal-atom contacts that could function as a molecular electronics device. (Courtesy of NASA Ames Center of Nanotechnology, <http://ipt.arc.nasa.gov>.)

DNA strand between metallic atoms, noting that the width of a DNA strand is around 2 nm. Such a structure could function as a sensor and other electronic components. Molecular electronics, while at its infancy, is expected to revolutionize electronics industry and to enable continuous technological progress through the twenty-first century.

Nano/microscale research and discoveries have been instrumental to the development of technologies used today in microelectronics, photonics, communication, manufacture, and biomedicine. However, systematic and large-scale government investment toward nanoscience and engineering did not start until late 1990s, when the Interagency Working Group on Nanoscience, Engineering, and Technology (IWGN) was formed under the National Science and Technology Council (NSTC). The first report was released in fall 1999, entitled “Nanostructure Science and Technology,” followed by the report, “Nanotechnology Research Directions.” In July 2000, NSTC published the “National Nanotechnology Initiative (NNI).” A large number of nanotechnology centers and nanofabrication facilities have been established since then; see [www.nano.gov](http://www.nano.gov). In the United States, the government spending on nanotechnology R&D exceeded \$1 billion in 2005, as compared to \$464 million in 2001 and approximately \$116 million in 1997. The total government investment worldwide was over \$4 billion in 2005, and Japan and European countries invested similar amount of money as the United States did. Recognizing the increasing impact on engineering and science, the American Society of Mechanical Engineers established the ASME Nanotechnology Institute in mid-2001 and sponsored a large number of international conferences and workshops; see <http://nano.asme.org/>. Understanding the thermal transport and properties at the nanoscale is extremely important as mentioned earlier.

Engineers have the responsibility to transfer the basic science findings into technological advances, to design and develop better materials with desired functions, to build systems that integrate from small to large scales, to perform realistic modeling and simulation that facilitate practical realization of improved performance and continuously reduced cost, and to conduct quantitative measurements and tests that determine the materials properties and system performance. Like any other technology, nanotechnology may also have some adverse effects, such as toxic products and biochemical hazards, which are harmful to human health and the environment. There are also issues and debates concerning security, ethics, and religion. Government and industry standard organizations, as well as universities, have paid great attention to the societal implications and education issues in recent years. Optimists believe that we can harness nanobiotechnology to improve the quality of human life and benefit social progress, while overcoming the adverse effects, like we have done with electricity, chemical plants, and space technology.

## **1.4 OBJECTIVES AND ORGANIZATION OF THIS BOOK**

---

Scientists, engineers, entrepreneurs, and lawmakers must work together for the research outcomes to be transferred into practical products that will advance the technology and benefit society. Nanotechnology is still in the early stage and holds tremendous potential; therefore, it is important to educate a large number of engineers with a solid background in nanoscale analysis and design so that they will become tomorrow’s leaders and inventors. There is a growing demand of educating mechanical engineering students at both the graduate and undergraduate levels with a background in thermal transport at micro/nanoscales. Micro/nanoscale heat transfer courses have been introduced in a number of universities; however, most of these courses are limited at the graduate level. While an edited book on *Microscale Energy Transport* has been available since 1998,<sup>3</sup> it is difficult to use as a textbook due to the lack of examples, homework problems, and sufficient details on each subject. Some universities have introduced nanotechnology-related courses to the freshmen and sophomores, with no in-depth coverage on the fundamentals of physics. A large number of institutions have introduced joint mechanical-electrical engineering courses on MEMS/NEMS, with a focus on device-level manufacturing and processing technology. To understand the thermal transport phenomena and thermophysical properties at small length

scales, the concepts of quantum mechanics, solid state physics, and electrodynamics are inevitable. These concepts, however, are difficult to comprehend by engineering students.

The aim of this book is to introduce the much needed physics knowledge without overwhelming mathematical operators or notions that are unfamiliar to engineering students. Therefore, this book can be used as the textbook not only in a graduate-level course but also in a tech elective for senior engineering undergraduates. While the book contains numerous equations, the math requirement mostly does not exceed engineering calculus including series, differential and integral equations, and some vector and matrix algebra. The reason to include such a large number of equations is to provide necessary derivation steps, so that readers can follow and understand clearly. This is particularly helpful for practicing engineers who do not have a large number of references at hand. The emphasis of this book is placed on the fundamental understanding of the phenomena and properties: that is, why do we need particular equations and how can we apply them to solve thermal transport problems at the prescribed length and time scales? Selected and refined examples are provided that are both practical and illustrative. At the end of each of the remaining nine chapters, a large number of exercises are given at various levels of complexity and difficulty. Numerical methods are not presented in this book. Most of the problems can be solved with a personal computer using a typical software program or spreadsheet. For course instructors, the solutions of many homework problems can be obtained from the author.

The field of micro/nanoscale heat transfer was cultivated and fostered by Professor Chang-Lin Tien beginning in the late 1980s, along with the rapid development in micro-electronics, MEMS, and nanotechnology. His long-lasting and legendary contributions to the thermal science research have been summarized in the recent volume of *Annual Review of Heat Transfer*.<sup>50</sup> As early as in the 1960s, Professor Tien investigated the fundamentals of the radiative properties of gas molecules, the size effect on the thermal conductivity of thin films and wires, and radiation tunneling between closely spaced surfaces. He published (with John H. Lienhard) a book in 1971, titled *Statistical Thermodynamics*, which provided inspiring discussions on early quantum mechanics and models of thermal properties of gases, liquids, and crystalline solids. While thermodynamics is a required course for mechanical engineering students, the principles of thermodynamics cannot be understood without a detailed background in statistical thermodynamics. Statistical mechanics and kinetic theory are also critical for understanding thermal properties and transport phenomena.

Chapter 2 provides an overview of equilibrium thermodynamics, heat transfer, and fluid mechanics. Built up from the undergraduate mechanical engineering curricula, the materials are introduced in a quite different sequence to emphasize thermal equilibrium, the second law of thermodynamics, and thermodynamic relations. The concept of entropy is rigorously defined and applied to analyze conduction and convection heat transfer problems in this chapter. It should be noted that, in Chap. 8, an extensive discussion is given on the entropy of radiation.

Chapter 3 introduces statistical mechanics and derives the classical (Maxwell-Boltzmann) statistics and quantum (i.e., Bose-Einstein and Fermi-Dirac) statistics. The first, second, and third laws of thermodynamics are presented with a microscopic interpretation, leading to the discussion of Bose-Einstein condensate and laser cooling of atoms. The classical statistics are extensively used to obtain the ideal gas equation, the velocity distribution, and the specific heat. A concise presentation of elementary quantum mechanics is then provided. This will help students gain a deep understanding of the earlier parts of this chapter. For example, the quantization of energy levels and the energy storage mechanisms by translation, rotation, and vibration for modeling the specific heat of ideal polyatomic gases. The combined knowledge of quantum mechanics and statistical thermodynamics is important for subsequent studies. The concept of photon as an elementary particle and how it interacts with an atom are discussed according to Einstein's 1917 paper on the atomic absorption and emission mechanisms. Finally, the special theory of relativity is briefly introduced to help understand the limitation of mass conservation and the generality of the law of energy conservation.

Chapter 4 begins with a very basic kinetic theory of dilute gases and provides a microscopic understanding of pressure and shear. With the help of mean free path and average collision distance, the transport coefficients such as viscosity, thermal conductivity, and mass diffusion coefficient are described. Following a discussion of intermolecular forces, the detailed Boltzmann transport equation (BTE) is presented to fully describe hydrodynamic equations as well as Fourier's law of heat conduction, under appropriate approximations. In the next section, the regimes of microflow are described based on the Knudsen number, and the current methods to deal with microfluidics are summarized. The heat transfer associated with slip flow and temperature jump is presented in more detail with a simple planar geometry. Then, gas conduction between two surfaces under free molecular flow is derived. These examples, while simple, capture some of the basics of microfluidics. No further discussion is given on properties of liquids or multiphase fluids. It should be noted that several books on microflow already exist in the literature.

The next three chapters provide a comprehensive treatment of nano/microscale heat transfer in solids, with an emphasis on the physical phenomena as well as material properties. The materials covered in Chap. 5 are based on simple free-electron model, kinetic theory, and BTE without a detailed background of solid state physics, which is discussed afterward in Chap. 6. This not only helps students comprehend the basic, underlying physical mechanisms but also allows the instructor to integrate Chap. 5 into a graduate heat conduction course. For an undergraduate elective, Chap. 6 can be considered as reading material or reference without spending too much time going through the details in class. In Chap. 5, the theory of specific heat is presented with a detailed treatment on the quantum size effect. Similarly, the theory of thermal conductivity of metals and dielectric solids is introduced. Because of the direct relation between electrical and thermal conductivities and the importance of thermoelectric effects, irreversible thermodynamics and thermoelectricity are also introduced. The classical size effect on thermal conductivity due to boundary scattering is elaborated. Finally, the concept of quantum conductance (both electric and thermal) is introduced.

Chapter 6 introduces the electronic band structures and phonon dispersion relations in solids. It helps understand semiconductor physics and some of the difficulties of free-electron model for metals. Photoemission, thermionic emission, and electron tunneling phenomena are introduced. The electrical transport in semiconductors is described with applications in energy conversion and optoelectronic devices. Chapter 7 focuses on nonequilibrium energy transport in nanostructures, including non-Fourier equations for transient heat conduction. The equation of phonon radiative transfer is presented and solved for thin-film and multilayer structures. The phenomenon of thermal boundary resistance is studied microscopically. A regime map is developed in terms of the length scale and the time scale from macroscale to microscale to nanoscale heat conduction. Additional reading materials regarding multiscale modeling, atomistic modeling, and thermal metrology are provided as references.

The last three chapters give comprehensive discussion on nano/microscale radiation with extensive background on the fundamentals of electromagnetic waves, the optical and thermal radiative properties of materials and surfaces, and the recent advancement in nanophotonics and nanoscale radiative transfer. Chapter 8 presents the Maxwell equations of electromagnetic waves and the derivation of Planck's law and radiation entropy. The electric and magnetic properties of the newly developed class of materials, i.e., negative-refractive-index materials are also discussed. More extensive discussion of the radiative properties of thin films, gratings, and rough surfaces is given in Chap. 9. The wave interference, partial coherence, and diffraction phenomena are introduced with detailed formulations. In Chap. 10, attention is given to the evanescent wave, coupling and localization, surface plasmon polaritons, surface phonon polaritons, and near-field energy transfer. This chapter contains the most recent developments in near-field optics, nanophotonics, and nanoscale radiative transfer. These advancements will continue to impact on the energy conversion devices, sensors, and nanoscale photothermal manufacturing.

It is noteworthy that the book *Nanoscale Energy Transfer and Conversion*, by Professor G. Chen, has recently been published.<sup>4</sup> In his book, a parallel treatment is presented to deal with electron, molecule, phonon, and photon transport processes. Such a parallel treatment places emphasis on the similarity and analogy between different energy carriers and transport mechanisms. While the approaches are unique and interesting, it is difficult for use as a textbook at the entry level without some preliminary solid state physics and statistical thermodynamics background. The present book places materials within the context of each topic by presenting statistical thermodynamics, kinetic theory of ideal gases and microfluidics, electrons and phonons in solids, and electromagnetic waves and their interactions with nanomaterials in separate chapters. In addition to the differences in the organization and presentation, the coverage of the present text differs to some extent from Chen's book. The present book contains much more extensive discussion on statistical thermodynamics and nanoscale thermal radiation, while Chen's book includes additional chapters on liquids and their interfaces as well as molecular dynamics simulation. As a result, the two books complement each other in terms of the coverage and organization. It is hoped that the present text can be used either as a whole in a one-semester course, or in part for integration into an existing thermal science course for several weeks on a particular topic. Examples are graduate-level thermodynamics (Chaps. 2 and 3), convection heat transfer (Chap. 4), conduction heat transfer (Chaps. 5 and 7), and radiation heat transfer (Chaps. 8 and 9). Selected materials may also be used to introduce nanoscale thermal sciences in undergraduate heat transfer and fluid mechanics courses. Some universities offer a second course on thermodynamics at the undergraduate level for which statistical thermodynamics and quantum theory can also be introduced. This text can also be self-studied by researchers or practicing engineers, graduated from a traditional engineering discipline. A large effort is given to balance the depth with the breadth so that it is easy to understand and contains sufficient coverage of both the fundamentals and advanced developments in the field. Readers will gain the background necessary to understand the contemporary research in nano/microscale thermal engineering and to solve a variety of practical problems using the approaches presented in the text.

## REFERENCES

---

1. C. P. Poole, Jr. and F. J. Owens, *Introduction to Nanotechnology*, Wiley, New York, 2003.
2. E. L. Wolf, *Nanophysics and Nanotechnology—An Introduction to Modern Concepts in Nanoscience*, Wiley-VCH, Weinheim, Germany, 2004.
3. C. L. Tien, A. Majumdar, and F. M. Gerner (eds.), *Microscale Energy Transport*, Taylor & Francis, Washington, DC, 1998.
4. G. Chen, *Nanoscale Energy Transport and Conversion*, Oxford University Press, New York, 2005.
5. C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, **27**, 379–423, 623–656, July & October 1948. See <http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html>.
6. G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, **38**(8), 114–117, April 1965; G. E. Moore, "Progress in digital integrated electronics," *IEEE Tech. Digest* (International Electron Devices Meeting), 11–13, 1975. [www.intel.com/technology/mooreslaw](http://www.intel.com/technology/mooreslaw).
7. B. K. Tsai, "A summary of lightpipe radiation thermometry research at NIST," *J. Res. Natl. Inst. Stand. Technol.*, **111**, 9–30, 2006.
8. SIA, International Technology Roadmap for Semiconductors. [www.itrs.net](http://www.itrs.net).
9. C. H. Fan and J. P. Longtin, "Radiative energy transport at the spatial and temporal micro/nano scales," in *Heat Transfer and Fluid Flow in Microscale and Nanoscale Structures*, M. Faghri and B. Sunden (eds.), WIT Press, Southampton, UK, pp. 225–275, 2003.
10. W. Denk, J. H. Stricker, and W. W. Webb, "Two-photon laser scanning fluorescence microscopy," *Science*, **248**, 73–76, 1990.
11. T. Yu, C. K. Ober, S. M. Kuebler, W. Zhou, S. R. Marder, and J. W. Perry, "Chemically-amplified positive resist system for two-photon three-dimensional lithography," *Adv. Mater.*, **15**,

- 517–521, 2003; S. M. Kuebler, K. L. Braun, W. Zhou, et al., “Design and application of high-sensitivity two-photon initiators for three-dimensional microfabrication,” *J. Photochem. Photobiol. A: Chemistry*, **158**, 163–170, 2003.
12. M. F. Modest and H. Abakians, “Heat-conduction in a moving semi-infinite solid subject to pulsed laser irradiation,” *J. Heat Transfer*, **108**, 597–601, 1986; M. F. Modest and H. Abakians, “Evaporative cutting of a semi-infinite body with a moving cw laser,” *J. Heat Transfer*, **108**, 602–607, 1986.
  13. C. L. Tien, T. Q. Qiu, and P. M. Norris, “Microscale thermal phenomena in contemporary technology,” *Thermal Sci. Eng.*, **2**, 1–11, 1994.
  14. R. J. Stoner and H. J. Maris, “Kapitza conductance and heat flow between solids at temperatures from 50 to 300 K,” *Phys. Rev. B*, **48**, 16373–16387, 1993.
  15. W. S. Capinski, H. J. Maris, T. Ruf, M. Cardona, K. Ploog, and D. S. Katzer, “Thermal-conductivity measurements of GaAs/AlAs superlattices using a picosecond optical pump-and-probe technique,” *Phys. Rev. B*, **59**, 8105–8113, 1999.
  16. P. M. Norris, A. P. Caffrey, R. Stevens, J. M. Klopff, J. T. McLeskey, and A. N. Smith, “Femtosecond pump-probe nondestructive evaluation of materials,” *Rev. Sci. Instrum.*, **74**, 400–406, 2003; R. J. Stevens, A. N. Smith, and P. M. Norris, “Measurement of thermal boundary conductance of a series of metal-dielectric interfaces by the transient thermoreflectance techniques,” *J. Heat Transfer*, **127**, 315–322, 2005.
  17. O. Manasreh, *Semiconductor Heterojunctions and Nanostructures*, McGraw-Hill, New York, 2005.
  18. G. Chen, “Heat transfer in micro- and nanoscale photonic devices,” *Annu. Rev. Heat Transfer*, **7**, 1–18, 1996.
  19. Y. Jaluria, “Thermal processing of materials: from basic research to engineering,” *J. Heat Transfer*, **125**, 957–979, 2003; X. Cheng and Y. Jaluria, “Optimization of a thermal manufacturing process: drawing of optical fiber,” *Int J. Heat Mass Transfer*, **48**, 3560–3573, 2005; C. Chen and Y. Jaluria, “Modeling of radiation heat transfer in the drawing of an optical fiber with multi-layer structure,” *J. Heat Transfer*, **129**, 342–352, 2007.
  20. P. N. Prasad, *Nanophotonics*, Wiley, New York, 2004.
  21. Z. M. Zhang and M. P. Mengüç, “Guest editorial: special issue on nano/microscale radiative transfer,” *J. Heat Transfer*, **129**, 1–2, 2007.
  22. R. P. Feynman, “There’s plenty of room at the bottom,” *J. Microelectromechanical Systems*, **1**, 60–66, 1992; R.P. Feynman, “Infinitesimal Machinery,” *J. Microelectromechanical Systems*, **2**, 4–14, 1993. [www.zyvex.com/nanotech/feynman.html](http://www.zyvex.com/nanotech/feynman.html).
  23. M. J. Madou, *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd ed., CRC Press, Boca Raton, FL, 2002.
  24. E. O. Sunden, T. L. Wright, J. Lee, W. P. King, and S. Graham, “Room-temperature chemical vapor deposition and mass detection on a heated atomic force microscope cantilever,” *Appl. Phys. Lett.*, **88**, 033107, 2006.
  25. K. Hirahara, K. Suenaga, S. Bandow, et al., “One-dimensional metallofullerene crystal generated inside single-walled carbon nanotubes,” *Phys. Rev. Lett.*, **85**, 5384, 2000. Also see *Phys. Rev. Focus*, 19 December 2000 at <http://focus.aps.org/story/v6/st27>.
  26. P. X. Gao, Y. Ding, W. J. Mai, W. L. Hughes, C. S. Lao, and Z. L. Wang, “Conversion of zinc oxide nanobelt into superlattice-structured nanohelices,” *Science*, **309**, 1700–1704, 2005; X. Y. Kong, Y. Ding, R. Yang, and Z. L. Wang, “Single-crystal nanorings formed by epitaxial self-coiling of polar nanobelts,” *Science*, **309**, 1348–1351, 2004.
  27. Y. Yang, W. Liu, and M. Asheghi, “Thermal and electrical characterization of Cu/CoFe superlattices,” *Appl. Phys. Lett.*, **84**, 3121–3123, 2004; Y. Yang, R. M. White, and M. Asheghi, “Thermal characterization of Cu/CoFe multilayer for giant magnetoresistive (GMR) head applications,” *J. Heat Transfer*, **128**, 113–120, 2006.
  28. G. Binnig and H. Rohrer, “Scanning tunneling microscopy,” *Helv. Phys. Acta*, **55**, 726–735, 1982; G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, “Surface studies by scanning tunneling microscopy,” *Phys. Rev. Lett.*, **49**, 57–61, 1982; G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, “ $7 \times 7$  reconstruction on Si(111) resolved in real space,” *Phys. Rev. Lett.*, **50**, 120–123, 1983.
  29. M. F. Crommie, C. P. Lutz, and D. M. Eigler, “Confinement of electrons to quantum corrals on a metal surface,” *Science*, **262**, 218–220, 1993.



30. G. Binnig, C. F. Quate, and Ch. Gerber, "Atomic force microscope," *Phys. Rev. Lett.*, **56**, 930–933, 1986.
31. C. C. Williams and H. K. Wickramasinghe, "Scanning thermal profiler," *Appl. Phys. Lett.*, **49**, 1587–89, 1986; J. M. R. Weaver, L. M. Walpita, and H. K. Wickramasinghe, "Optical absorption microscopy with nanometer resolution," *Nature*, **342**, 783–85, 1989; M. Nonnenmacher and H. K. Wickramasinghe, "Optical absorption spectroscopy by scanning force microscopy," *Ultramicroscopy*, **42–44**, 351–354, 1992.
32. A. Majumdar, "Scanning thermal microscopy," *Annu. Rev. Mater. Sci.*, **29**, 505–585, 1999.
33. H.-K. Lyeo, A. A. Khajetoorians, L. Shi, et al., "Profiling the thermoelectric power of semiconductor junctions with nanometer resolution," *Science*, **303**, 818–820, 2004; Z. Bian, A. Shakouri, L. Shi, H.-K. Lyeo, and C. K. Shih, "Three-dimensional modeling of nanoscale Seebeck measurement by scanning thermoelectric microscopy," *Appl. Phys. Lett.*, **87**, 053115, 2005.
34. H. J. Mamin and D. Rugar, "Thermomechanical writing with an atomic force microscope tip," *Appl. Phys. Lett.*, **61**, 1003–1005, 1992; H. J. Mamin, "Thermal writing using a heated atomic force microscope tip," *Appl. Phys. Lett.*, **69**, 433–435, 1996.
35. G. Binnig, M. Despont, U. Drechsler, et al., "Ultrahigh-density atomic force microscopy data storage with erase capability," *Appl. Phys. Lett.*, **74**, 1329–1331, 1999; W. P. King, T. W. Kenny, K. E. Goodson, et al., "Atomic force microscope cantilevers for combined thermomechanical data writing and reading," *Appl. Phys. Lett.*, **78**, 1300–1302, 2001.
36. U. Dürig, G. Cross, M. Despont, et al., "'Millipede'—an AFM data storage system at the frontier of nanotechnology," *Tribology Lett.*, **9**, 25–32, 2000; P. Vettiger, G. Cross, M. Despont, et al., "The 'millipede'—nanotechnology entering data storage," *IEEE Trans. Nanotechnol.*, **1**, 39–55, 2002.
37. P. E. Sheehan, L. J. Whitman, W. P. King, and B. A. Nelson, "Nanoscale deposition of solid inks via thermal dip pen nanolithography," *Appl. Phys. Lett.*, **85**, 1589–1591, 2004.
38. J. A. Eastman, S. R. Phillpot, S. U. S. Choi, and P. Kablinski, "Thermal transport in nanofluids," *Annu. Rev. Mater. Res.*, **34**, 219–246, 2004.
39. R. S. Prasher, P. Bhattacharya, and P. E. Phelan, "Thermal conductivity of nanoscale colloidal solutions (nanofluids)," *Phys. Rev. Lett.*, **94**, 025901, 2005; R. Prasher, P. Bhattacharya, and P. E. Phelan, "Brownian-motion-based convective-conductive model for the effective thermal conductivity of nanofluids," *J. Heat Transfer*, **128**, 588–595, 2006.
40. G. Chen and A. Shakouri, "Heat transfer in nanostructures for solid-state energy conversion," *J. Heat Transfer*, **124**, 242–252, 2002; H. Böttner, G. Chen, and R. Venkatasubramanian, "Aspects of thin-film superlattice thermoelectric materials, devices and applications," *MRS Bulletin*, **31**, 211–217, March 2006.
41. S. Basu, Y.-B. Chen, and Z. M. Zhang, "Microscale radiation in thermophotovoltaic devices—A review," *Int. J. Ener. Res.*, **31**, in press, 2007. (Published online 6 Dec. 2006.)
42. M. Law, L. E. Greene, J. C. Johnson, R. Saykally, and P. Yang, "Nanowire dye-sensitized solar cells," *Nature Mater.*, **4**, 455–459, 2005.
43. A. Mihi and H. Miguez, "Origin of light-harvesting enhancement in colloidal-photon-crystal-based dye-sensitized solar cells," *J. Phys. Chem. B*, **109**, 15968–15976, 2005.
44. G. Crabtree, M. Dresselhaus, and M. Buchanan, "The hydrogen economy," *Physics Today*, 39–44, December 2004.
45. A. Lewis, H. Taha, A. Strinkovski, et al., "Near-field optics: from subwavelength illumination to nanometric shadowing," *Nature Biotechnol.*, **21**, 1378–1386, 2003.
46. X. Michalet, F. F. Pinaud, L. A. Bentolila, et al., "Quantum dots for live cells, in vivo imaging, and diagnostics," *Science*, **307**, 538–544, 2005.
47. I. L. Medintz, H. T. Uyeda, E. R. Goldman, and H. Mattoussi, "Quantum dot bioconjugates for imaging, labelling and sensing," *Nature Mater.*, **4**, 435–446, 2005.
48. B. Yu and M. Meyyappan, "Nanotechnology: role in emerging nanoelectronics," *Solid-State Electronics*, **50**, 536–544, 2006.
49. S. De Franceschi and L. Kouwenhoven, "Electronics and the single atom," *Nature*, **417**, 701–702, 2002.
50. V. Prasad, Y. Jaluria, and G. Chen (eds.), *Annual Review of Heat Transfer*, Vol. 14, Begell House, New York, 2005.

---

# CHAPTER 2

---

# OVERVIEW OF MACROSCOPIC THERMAL SCIENCES

---

This chapter provides a concise description of the basic concepts and theories underlying classical thermodynamics and heat transfer. Different approaches exist in presenting the subject of thermodynamics. Most engineering textbooks first introduce temperature, then discuss energy, work, and heat, and define entropy afterward. Callen developed an axiomatic structure using a simple set of abstract postulates to combine the physical information that is included in the laws of thermodynamics.<sup>1</sup> Continuing the effort pioneered by Keenan and Hatsopoulos,<sup>2</sup> Gyftopoulos and Beretta<sup>3</sup> developed a logical sequence to introduce the basic concepts with a rigorous definition of each thermodynamic term. Their book has been a great inspiration to the present author in comprehending and teaching thermodynamics. Here, an overview of classical thermodynamics is provided that is somewhat beyond typical undergraduate textbooks.<sup>4,5</sup> Details on the historic development of classical thermodynamics can be found from Bejan<sup>6</sup> and Kestin<sup>7</sup>, and references therein. The basic phenomena and governing equations in energy, mass, and momentum transfer will be presented subsequently in a self-consistent manner without invoking microscopic theories.

---

## 2.1 FUNDAMENTALS OF THERMODYNAMICS

---

A *system* is a collection of constituents (whose amounts may be fixed or varied within a specified range) in a defined space (e.g., a container whose volume may be fixed or varied within a specified range), subject to other external forces (such as gravitational and magnetic forces) and constraints. External forces are characterized by *parameters*. An example is the volume of a container, which is a parameter associated with the forces that confine the constituents within a specified space. Everything that is not included in the system is called the *environment* or *surroundings* of the system.

Quantities that characterize the behavior of a system at any instant of time are called *properties* of the system. Properties must be measurable and their values are independent of the measuring devices. Properties supplement constituents and parameters to fully characterize a system. At any given time, the system is said to be in a *state*, which is fully characterized by the types and amount of constituents, a set of parameters associated with various types of external forces, and a set of properties. Two states are identical if the amount of each type of constituents and values of all the parameters and properties are the same. A system may experience a *spontaneous change of state*, when the change of state does not involve any interaction between the system and its environment. If the

system changes its state through interactions with other systems in the environment, it is said to experience an *induced change of state*. If a system can experience only spontaneous changes of state, it is said to be an *isolated system*, that is, the change of state of the system does not affect the environment of the system. The study of the possible and allowed states of a system is called *kinematics*, and the study of the time evolution of the state is called *dynamics*.

The relation that describes the change of state of a system as a function of time is the *equation of motion*. In practice, the complete equations of motion are often not known. Therefore, in thermodynamics the description of the change of state is usually given in terms of the end states (i.e., the initial and final states) and the *modes of interaction* (for example, work and heat, which are discussed later). The end states and the modes of interaction specify a *process*. A spontaneous change of state is also called a *spontaneous process*. A process is *reversible* if there is at least one way to restore both the system and its environment to their initial states. Otherwise, the process is *irreversible*, i.e., it is not possible to restore both the system and its environment to their initial states. A *steady state* is one that does not change as a function of time despite interactions between the system and other systems in the environment.

### 2.1.1 The First Law of Thermodynamics

*Energy* is a property of every system in any state. The first law of thermodynamics states that *energy can be transferred to or from a system but can be neither created nor destroyed*. The energy balance for a system can be expressed as

$$\Delta E = E_2 - E_1 = E_{\text{net,in}} \quad (2.1a)$$

where  $\Delta$  denotes a finite change, subscripts 1 and 2 refer to the initial and final states, respectively, and  $E_{\text{net,in}} = E_{\text{in}} - E_{\text{out}}$  is the net amount of energy transferred into the system. For an infinitesimal change, the differential form of the energy balance is

$$dE = \delta E_{\text{net,in}} \quad (2.1b)$$

Here,  $d$  is used to signify a differential change of the property of a system, and  $\delta$  is used to specify a differentially small quantity that is not a property of any system. Clearly, the energy of an isolated system is conserved. Energy is an additive property, i.e., the energy of a composite system is the sum of the energies of all individual subsystems. Examples are kinetic energy and potential energy, as defined in classical mechanics, and internal energy, which will be discussed later. A similar expression for mass balance can also be written.

The term *mechanical effect* is used for the kind of processes described in mechanics, such as the change of the height of a weight in a gravitational field, the change of the relative positions of two charged particles, the change of the velocity of a point mass, the change of the length of a spring, or a combination of such changes. All mechanical effects are equivalent in the sense that it is always possible to arrange forces and processes that annul all the mechanical effects except one that we choose. It is common to choose the rise and fall of a weight in a gravity field to represent this kind of processes.

A *cyclic process* (also called a *cycle*) is one with identical initial and final states. A *perpetual-motion machine of the first kind (PMM1)* is any device (or system) undergoing a cyclic process that produces no external effects but the rise or fall of a weight in a gravity field. A PMM1 violates the first law of thermodynamics, and hence, it is impossible to build a PMM1. Perpetual motion, however, may exist as long as it produces zero net external effect. Examples of perpetual motion are a lossless oscillating pendulum, an electric current through a superconducting coil, and so forth.

## 2.1.2 Thermodynamic Equilibrium and the Second Law

An *equilibrium* state is a state that cannot change spontaneously with time. There are different types of equilibrium: unstable, stable, and metastable. A *stable-equilibrium state* is a state that cannot be altered to a different state without leaving any net effect on the environment. In the following, a stable-equilibrium state is frequently referred to as a state at *thermodynamic equilibrium*.

The *stable-equilibrium-state principle*, or *state principle*, can be phrased as follows: *Among all states of a system with a given set of values of energy, parameters, and constituents, there exists one and only one stable-equilibrium state.* That is to say that, in a stable-equilibrium state, all properties are uniquely determined by the amount of energy, the value of each parameter, and the amount of each type of constituents. This principle is an integral part of the second law of thermodynamics.<sup>2,3,7</sup> It is important for the thermodynamic definition of temperature and the derivation of thermodynamic relations in stable-equilibrium states. Another aspect of the second law of thermodynamics is the definition of an important property, called *entropy*, as discussed next.

Entropy is an additive property of every system in any state. The second law of thermodynamics asserts that, *in an isolated system, entropy cannot be destroyed* but can either be created (in an irreversible process) or remain the same (in a reversible process). The entropy produced as time evolves during an irreversible process is called the *entropy generation* ( $S_{\text{gen}}$ ) of the process due to *irreversibility*. Like energy, entropy can be transferred from one system to another. One can write the entropy balance as follows (keeping in mind that entropy generation must not be negative):

$$\Delta S = S_2 - S_1 = S_{\text{net,in}} + S_{\text{gen}}$$

with 
$$S_{\text{gen}} \geq 0 \quad (2.2a)$$

or 
$$dS = \delta S_{\text{net,in}} + \delta S_{\text{gen}}$$

with 
$$\delta S_{\text{gen}} \geq 0 \quad (2.2b)$$

Here again,  $\delta$  is used to indicate an infinitesimal quantity that is *not* a property of any system. For a system with fixed values of energy ( $E$ ), parameters, and constituents, the entropy of the system is the largest in the stable-equilibrium state. This is *the highest entropy principle*. Applying this principle to an isolated system for which the energy is conserved, the entropy of the system will increase until a thermodynamic equilibrium is reached. Spontaneous changes of state are usually irreversible and accompanied by entropy generation.

The second law of thermodynamics can be summarized with the following three statements: (1) There exists a unique stable-equilibrium state for any system with given values of energy, parameters, and constituents. (2) Entropy is an additive property, and for an isolated system, the entropy change must be nonnegative. (3) Among all states with the same values of energy, parameters, and constituents, the entropy of the stable-equilibrium state is the maximum.

The energy of a system with volume ( $V$ ) as its only parameter (neglecting other external forces) is called the *internal energy* ( $U$ ). The state principle implies that there are  $r + 2$  (where  $r$  is the number of different constituents) independent variables that fully characterize a stable-equilibrium state of such a system. Therefore in a stable-equilibrium state, all properties are functions of  $r + 2$  independent variables. Since entropy is a property of the system, we have

$$S = S(U, V, N_1, N_2, \dots, N_r) \quad (2.3)$$

where  $N_i$  is the number of particles of the  $i$ th species (or type of constituents). This function is continuous and differentiable, and furthermore, it is a monotonically increasing function of energy for fixed values of  $V$  and  $N_{j,s}$ .<sup>1,3,6</sup> Equation (2.3) can be uniquely solved for  $U$  so that

$$U = U(S, V, N_1, N_2, \dots, N_r) \quad (2.4)$$

which is also continuous and admits partial derivatives of all orders. Each first order partial derivative of Eq. (2.3) or (2.4) represents a property of the stable-equilibrium state. For example, *temperature* and *pressure* are properties of a system at thermodynamic equilibrium. The (absolute) temperature is defined by

$$T = \left( \frac{\partial U}{\partial S} \right)_{V, N_{j,s}} \quad (2.5a)$$

and the pressure is defined by

$$P = - \left( \frac{\partial U}{\partial V} \right)_{S, N_{j,s}} \quad (2.5b)$$

The partial derivative with respect to the  $i$ th type of constituents defines its chemical potential of that species,

$$\mu_i = \left( \frac{\partial U}{\partial N_i} \right)_{S, V, N_{j,s} (j \neq i)} \quad (2.5c)$$

Equation (2.3) or (2.4) is called the *fundamental relation* for states at thermodynamic equilibrium. The differential form of Eq. (2.4) is the Gibbs relation:

$$dU = TdS - PdV + \sum_{i=1}^r \mu_i dN_i \quad (2.6)$$

where Eq. (2.5) has been used. The above equation may be rearranged into the form

$$dS = \frac{1}{T}dU + \frac{P}{T}dV - \sum_{i=1}^r \frac{\mu_i}{T}dN_i \quad (2.7)$$

Therefore,

$$\frac{1}{T} = \left( \frac{\partial S}{\partial U} \right)_{V, N_{j,s}}, \quad \frac{P}{T} = \left( \frac{\partial S}{\partial V} \right)_{U, N_{j,s}}, \quad \text{and} \quad \frac{\mu_i}{T} = - \left( \frac{\partial S}{\partial N_i} \right)_{U, V, N_{j,s} (j \neq i)} \quad (2.8)$$

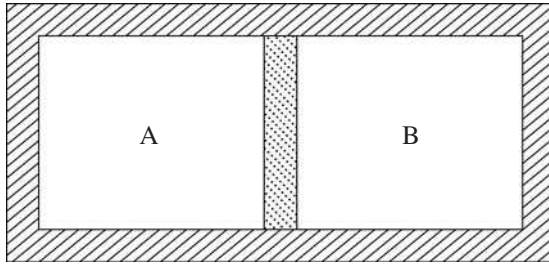
An interaction between two systems that results in a transfer of energy without net exchanges of entropy and constituents is called a *work interaction*. The amount of energy transferred in such an interaction is called *work* ( $W$ ). An interaction that has only mechanical effects is a work interaction, but a work interaction may involve nonmechanical effects. A process that involves only work interaction is called an *adiabatic* process. Another kind of a typical interaction is *heat interaction*, in which both energy and entropy are transferred without net exchanges of constituents and parameters between two systems. The amount of energy transferred in a heat interaction is called *heat* ( $Q$ ). Furthermore, the amount of entropy transferred  $\delta S$  is equal to the amount of energy transferred  $\delta Q$  divided by the temperature  $T_Q$  at which the heat interaction happens, i.e.,  $\delta S = \delta Q/T_Q$ . If a system cannot exchange constituents with other systems, it is said to be a *closed* system; otherwise, it is an *open* system.

Reversible processes are considered as the limiting cases of real processes, which are always accompanied by a certain amount of irreversibility. Such an ideal process is called

a *quasi-equilibrium* (or *quasi-static*) process, in which each stage can be made as close to thermodynamic equilibrium as possible if the movement is frictionless and very slow. In an ideal process, a finite amount of heat can be transferred reversibly from one system to another at a constant temperature. In practice, heat transfer can only happen when there is a temperature difference, and the process is always irreversible.

A *perpetual-motion machine of the second kind (PMM2)* is a cyclic device that interacts with a system at thermodynamic equilibrium and produces no external effect other than the rise of a weight in a gravity field, without changing the values of parameters and the amounts of constituents of the system. Historically, there exist different statements of the second law of thermodynamics: The Kelvin-Planck statement of the second law is that *it is impossible to build a PMM2*. The Clausius statement of the second law is that *it is not possible to construct a cyclic machine that will produce no effect other than the transfer of heat from a system at lower temperature to a system at higher temperature*. These statements can be proved using the three statements of the second law of thermodynamics given earlier in this chapter.

**Example 2-1. Criteria for thermodynamic equilibrium.** Consider a moveable piston (adiabatic and impermeable to matter) that separates a cylinder into two compartments (systems A and B), as shown in Fig. 2.1. We learned from mechanics that a mechanical equilibrium requires a balance of



**FIGURE 2.1** Illustration of two systems that may exchange work, heat, and species.

forces on both sides of the piston, that is to say the pressure of system A must be the same as that of system B (i.e.,  $P_A = P_B$ ). If the piston wall is made of materials that are diathermal (allowing heat transfer) and permeable to all species, under what conditions will the composite system C consisting of systems A and B be at stable equilibrium?

**Solution.** Assume system C is isolated from other systems, and each of the subsystems A and B is at a thermodynamic equilibrium state, whose properties are solely determined by its internal energy, volume, and amount of constituents:  $U_A, V_A, N_{j,s,A}$  and  $U_B, V_B, N_{j,s,B}$ , respectively. There exist neighboring states for both subsystems with small differences in  $U, V$ , and  $N_{j,s}$ , but the values of the composite system must be conserved, i.e.,  $dU_A = -dU_B, dV_A = -dV_B$ , and  $dN_{i,A} = -dN_{i,B}$  ( $i = 1, 2, \dots, r$ ). The differential entropy of system C can be expressed as:

$$\begin{aligned}
 dS_C &= dS_A + dS_B \\
 &= \frac{1}{T_A} dU_A + \frac{P_A}{T_A} dV_A - \sum_{i=1}^r \frac{\mu_{i,A}}{T_A} dN_{i,A} + \frac{1}{T_B} dU_B + \frac{P_B}{T_B} dV_B - \sum_{i=1}^r \frac{\mu_{i,B}}{T_B} dN_{i,B} \\
 &= \left( \frac{1}{T_A} - \frac{1}{T_B} \right) dU_A + \left( \frac{P_A}{T_A} - \frac{P_B}{T_B} \right) dV_A - \sum_{i=1}^r \left( \frac{\mu_{i,A}}{T_A} - \frac{\mu_{i,B}}{T_B} \right) dN_{i,A}
 \end{aligned} \tag{2.9}$$

If system C is in a stable-equilibrium state, its entropy is maximum and  $dS_C = 0$ . Since the values of  $dU_A$ ,  $dV_A$ , and  $dN_{i,A}$  are arbitrary, we must have

$$\frac{1}{T_A} = \frac{1}{T_B}, \quad \frac{P_A}{T_A} = \frac{P_B}{T_B}, \quad \text{and} \quad \frac{\mu_{i,A}}{T_A} = \frac{\mu_{i,B}}{T_B} \quad (i = 1, 2, \dots, r)$$

or 
$$T_A = T_B, \quad P_A = P_B, \quad \text{and} \quad \mu_{i,A} = \mu_{i,B} \quad (i = 1, 2, \dots, r) \quad (2.10)$$

These conditions correspond to thermal equilibrium, mechanical equilibrium, and chemical equilibrium, respectively. The combination forms the criteria for thermodynamic equilibrium.

**Discussion.** In the case when the piston is diathermal but rigid and impermeable to matter, the entropy change of system C must be nonnegative, i.e.,

$$dS_C = dS_A + dS_B = \left( \frac{1}{T_A} - \frac{1}{T_B} \right) dU_A \geq 0 \quad (2.11)$$

The above expression implies that  $dU_A \leq 0$  for  $T_A > T_B$ , and  $dU_A \geq 0$  for  $T_A < T_B$ . Spontaneous heat transfer can occur only from regions of higher temperature to regions of lower temperature. This essentially proves the Clausius statement of the second law of thermodynamics.

The concept of thermal equilibrium provides the physical foundation for *thermometry*, which is the science of temperature measurement. The temperature of a system at a thermodynamic equilibrium state is measured through changes in resistance, length, volume, or other physical parameters of the sensing element used in the thermometer, which is brought to thermal equilibrium with the system. Based on the inclusive statement of the second law of thermodynamics given previously, it can be inferred that two systems are in thermal equilibrium with each other if they are separately in thermal equilibrium with a third system. This is sometimes referred to as the *zeroth law of thermodynamics*.<sup>6</sup>

The International Temperature Scale of 1990 (ITS-90) was adopted by the International Committee of Weights and Measures in 1989.<sup>8</sup> The unit of thermodynamic temperature is kelvin (K), which is defined as  $1/273.16$  of the thermodynamic temperature of the triple point of water. The Celsius temperature is defined as the difference of the thermodynamic temperature and 273.15 K (the ice point). A difference of temperature may be expressed in either kelvins or degrees Celsius ( $^{\circ}\text{C}$ ). Although earlier attempts were made to define a temperature scale consistent with the original Celsius temperature scale (i.e.,  $0^{\circ}\text{C}$  for the ice point and  $100^{\circ}\text{C}$  for the steam point), a  $0.026^{\circ}\text{C}$  departure arose from more accurate measurements of the steam point, as shown in Table 2.1.<sup>9</sup> The steam point is therefore no longer

**TABLE 2.1** Two-Phase Points and the Triple Point of Water

	Temperature	
	(K)	( $^{\circ}\text{C}$ )
Ice point*	273.15	0
Triple point†	273.16	0.01
Steam point‡	373.124	99.974

\* Solid and liquid phases are in equilibrium at a pressure of 1 atm (101.325 kPa).

† Solid, liquid, and vapor phases are in equilibrium.

‡ Liquid and vapor phases are in equilibrium at 1 atm.

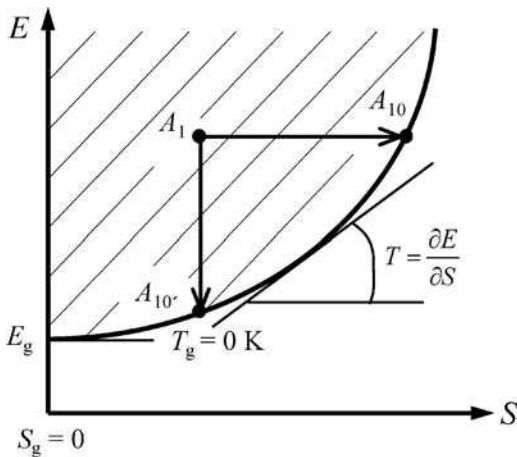
used as a defining fixed point in the ITS-90. More accurate steam tables were developed in the 1990s.

The ITS-90 defines 17 fixed points, which are determined by primary thermometry with standard uncertainties less than 0.002 K below 303 K and up to 0.05 K at the freezing point of copper ( $\approx 1358$  K). Cryogenic thermometry is essentially based on ideal gas thermometers (up to about 20 K). Platinum resistance thermometers, calibrated at specified sets of fixed points, are used to define the temperature scale from the triple point of hydrogen ( $\approx 13.8$  K) to the freezing point of silver ( $\approx 1235$  K). Platinum resistance thermometers have been chosen because of their excellent reproducibility, even though they are not primary thermometers. Radiation thermometers based on Planck's law of thermal radiation are used to define the temperature scale above 1235 K.

### 2.1.3 The Third Law of Thermodynamics

For each given set of values of constituents and parameters, there exists a unique stable-equilibrium state with *zero absolute temperature* (though not physically attainable). Furthermore, the entropy of any pure substance (in the form of a crystalline solid) vanishes at this state (zero absolute entropy). This is the third law of thermodynamics, also called *the Nernst theorem* after Walther Nernst who received the Nobel Prize in chemistry in 1920.<sup>3,6</sup> The energy is the lowest at this state, which is called the *ground-state energy* ( $E_g > 0$ ). The ground-state energy of a system consisting of independent particles may be related to its mass  $m$  using the relativistic theory, i.e.,  $E_g = mc^2$ , where  $c$  is the speed of light. Although absolute energy and entropy can be defined according to the third law of thermodynamics, in practice, reference states are often chosen so that the relative values of energy and entropy can be tabulated with respect to those of the reference states.

After reviewing the laws of thermodynamics, it is instructive to give a pictorial presentation to illustrate some of the fundamental concepts in thermodynamics, as done by Gyftopoulos and Beretta.<sup>3</sup> For a system that contains a single type of constituents (i.e., pure substance) with fixed values of parameters and amount of constituents, the stable-equilibrium states can be represented as a *convex E-S curve*, whose slope  $T = \partial E/\partial S$  defines the temperature of each state on the curve, as shown in Fig. 2.2. The stable-equilibrium-state curve



**FIGURE 2.2** The  $E$ - $S$  graph for a pure substance with fixed values of parameters and amount of constituents.



intersects the vertical axis at the ground state, whose energy is the ground-state energy  $E_g$  and whose absolute entropy is zero. Furthermore, the temperature at the ground state is 0 K. This provides a graphical illustration of the third law of thermodynamics. Along the stable-equilibrium-state curve, temperature increases with increasing energy or entropy. The vertical axis above  $E_g$  represents *zero-entropy states*, which are not at stable equilibrium (except when  $E = E_g$ ). These are states defined in mechanics, where entropy is not a concern. A spontaneous change of state can be illustrated with this graph as a horizontal line, e.g., from  $A_1$  to  $A_{10}$ , where  $A_{10}$  corresponds to the stable-equilibrium state that has the same values of energy, parameters, and constituents as those of  $A_1$ . No states exist below the stable-equilibrium-state curve because this would violate the highest entropy principle. Each point in the shaded area corresponds to one or more state(s) that is or are not at thermodynamic equilibrium, for which temperature may not be defined. Such a state in general cannot be uniquely determined by the values of its energy (or entropy) and parameters and the amount of constituents. The *lowest energy principle* is expressed as follows: Among all states with the same values of entropy and parameters and the amount of constituents, there exists a stable-equilibrium state whose energy is the lowest. Starting with any state that is not at stable equilibrium, there exists a reversible adiabatic process, in which work can be done by the system until it reaches a stable-equilibrium state. This process is illustrated in the  $E$ - $S$  graph by a vertical line from  $A_1$  to  $A_{10}$ . The corresponding work, which is equal to the energy difference between  $A_1$  and  $A_{10}$ , is called the *adiabatic availability*.<sup>3</sup> It defines the largest amount of work that can be extracted from a system without any other net effect on the environment of the system.

## 2.2 THERMODYNAMIC FUNCTIONS AND PROPERTIES

---

Several additional properties defined in this section are important in the study of states at thermodynamic equilibrium. The functional relations are derived based on the fundamental relation and are useful under specific circumstances. The phase equilibrium is summarized with an emphasis on pure substances. The concepts of specific heat and latent heat are then introduced. Combining the specific heat and the equation of state, we can evaluate the internal energy and entropy for ideal gases and incompressible solids and liquids.

### 2.2.1 Thermodynamic Relations

When dealing with substances within the container, the volume is a parameter that characterizes external forces, i.e., the interaction between the system and the wall of the container. If the constituents are confined within a surface, then the surface area will be a parameter instead of the volume. Parameters associated with other external forces (such as gravitational and magnetic forces) can also be included, if necessary. For simplicity, we assume that volume is the only parameter of the systems under investigation, unless otherwise specified.

Enthalpy is defined as  $H = U + PV$ , thus we have  $dH = dU + PdV + VdP$ . From Eq. (2.6), we obtain

$$dH = TdS + VdP + \sum_{i=1}^r \mu_i dN_i \quad (2.12a)$$

The significance of Eq. (2.12a) is that enthalpy can be expressed as a function of  $S$ ,  $P$ , and  $N_{j,s}$ ,

$$H = H(S, P, N_1, N_2, \dots, N_r) \quad (2.12b)$$

Furthermore,

$$T = \left( \frac{\partial H}{\partial S} \right)_{P, N_{j_s}}, \quad V = \left( \frac{\partial H}{\partial P} \right)_{S, N_{j_s}}, \quad \text{and} \quad \mu_i = \left( \frac{\partial H}{\partial N_i} \right)_{S, P, N_{j_s} (j \neq i)} \quad (2.12c)$$

Note that the subscripts in Eq. (2.12c) are different from those in Eq. (2.5). Enthalpy  $H(S, P, N_{j_s})$  is said to be a *characteristic function*, since it allows us to find out all the information about a stable-equilibrium state. There are a large number of characteristic functions. Depending on the particular situation and measurements available, it is advantageous to choose the most convenient one. Two other characteristic functions are now introduced. The first one is called *Helmholtz free energy*  $A(T, V, N_{j_s})$ , defined as  $A = U - TS$ . It follows that

$$dA = -SdT - PdV + \sum_{i=1}^r \mu_i dN_i \quad (2.13a)$$

and 
$$S = -\left( \frac{\partial A}{\partial T} \right)_{V, N_{j_s}}, \quad P = -\left( \frac{\partial A}{\partial V} \right)_{T, N_{j_s}}, \quad \text{and} \quad \mu_i = \left( \frac{\partial A}{\partial N_i} \right)_{T, V, N_{j_s} (j \neq i)} \quad (2.13b)$$

The second is *Gibbs free energy*  $G(T, P, N_{j_s})$ :  $G = U + PV - TS = H - TS = A + PV$ . It follows that

$$dG = -SdT + VdP + \sum_{i=1}^r \mu_i dN_i \quad (2.14a)$$

and 
$$S = -\left( \frac{\partial G}{\partial T} \right)_{P, N_{j_s}}, \quad V = \left( \frac{\partial G}{\partial P} \right)_{T, N_{j_s}}, \quad \text{and} \quad \mu_i = \left( \frac{\partial G}{\partial N_i} \right)_{T, P, N_{j_s} (j \neq i)} \quad (2.14b)$$

Characteristic functions supplement the fundamental relation and are very useful in evaluation of the properties of systems under thermodynamic equilibrium.

In a stable-equilibrium state,  $T$ ,  $P$ , and  $\mu_i$  ( $i = 1, 2, \dots, r$ ) must be uniform everywhere in the system. If the system is divided into  $k$  equal-volume subsystems, the energy, entropy, and the amount of each type of constituents of the system are the sums of these quantities in all subsystems. If the energy and the amount of each type of constituents in every subsystem are the same, then all subsystems are exactly identical to each other. If this is the case, the system is said to be in a *homogeneous* state; otherwise, it is *heterogeneous*. Examples of homogeneous states are air (which is a mixture of many different kinds of gases) and a well-mixed solution. Examples of heterogeneous states are ice water, and water and steam mixture in a boiler.

A system that experiences only homogeneous states is called a *simple system*. In a simple system,  $T$ ,  $P$ , and  $\mu_{j_s}$  of each subsystem are the same as the system itself and independent of  $k$ ; hence, they are called *intensive properties*. Taking  $T$  as an example, we have

$$T \left( \frac{U}{k}, \frac{V}{k}, \frac{N_1}{k}, \frac{N_2}{k}, \dots, \frac{N_r}{k} \right) = T(U, V, N_1, N_2, \dots, N_r) \quad (2.15)$$

In the above equation, the left-hand side is the temperature of the subsystem, and the right-hand side is the temperature of the whole system. Unlike temperature and pressure, the properties such as  $U$ ,  $S$ ,  $V$ , and  $N$  of each subsystem are inversely proportional to  $k$ , e.g.,

$$S \left( \frac{U}{k}, \frac{V}{k}, \frac{N_1}{k}, \frac{N_2}{k}, \dots, \frac{N_r}{k} \right) = \frac{1}{k} S(U, V, N_1, N_2, \dots, N_r) \quad (2.16)$$

Properties whose values are proportional to the total amount of constituents are called *extensive properties*. Therefore,  $U$ ,  $V$ ,  $S$ , and  $H$  are extensive properties. Notice that  $k$  cannot be arbitrarily large because of the continuum requirement.

The ratio or derivative of two extensive properties is an intensive property, e.g., the density (the ratio of mass to volume) is an intensive property and uniform in a simple system. Note that temperature, pressure, and chemical potentials are derivatives of two extensive properties. The properties  $T$ ,  $P$ , and  $\mu_{j_s}$  distinguish themselves from other intensive properties in that they are uniform in both homogeneous and heterogeneous states, whereas others may or may not be uniform in a heterogeneous state. A *specific property* is the ratio of an extensive property to the total amount of constituents (expressed as mass, mole, or number). For example, the mass-specific enthalpy is the enthalpy per kilogram of the substance. Specific properties are intensive properties.

For simple systems, the Gibbs relation given in Eq. (2.6) can be integrated to obtain

$$U = TS - PV + \sum_{i=1}^r \mu_i N_i \quad (2.17)$$

which is the *Euler relation*. By differentiating Eq. (2.17) and then subtracting Eq. (2.6) from it, we obtain the *Gibbs-Duhem relation*:

$$SdT - VdP + \sum_{i=1}^r N_i d\mu_i = 0 \quad (2.18)$$

The Euler relation for a system containing only one type of constituents ( $r = 1$ ) is

$$G = U + PV - TS = \mu N$$

or

$$\mu(T,P) = \frac{G}{N} = g(T,P) \quad (2.19)$$

Hence, the chemical potential of a pure substance is nothing but the specific Gibbs free energy. For a system containing two or more types of constituents, Eq. (2.14b) relates the chemical potential to the partial derivative of the Gibbs free energy with respect to  $N_i$  for fixed  $T$  and  $P$ , which is called the *partial* Gibbs free energy of the  $i$ th type of constituents.

## 2.2.2 The Gibbs Phase Rule

In a heterogeneous state, we consider a subdivision of the system into subsystems, each being a simple system. The collection of all subsystems that have the same values of all intensive properties is called a *phase*. Solid, liquid, and gas (or vapor) are the three distinct phases. The boundary between subsystems of different phases is called an *interface*. Different phases may appear to be clearly separated or well mixed. In space, liquid water droplets could be dispersed throughout water vapor, whereas on the earth, the liquid would occupy the lower part of the container due to gravity.

Assume that there are  $q$  coexisting phases, called a  $q$ -phase heterogeneous state. We can write the Gibbs-Duhem relation for each phase, and thus reduce the independent variables for  $T$ ,  $P$ ,  $\mu_i$  ( $i = 1, 2, \dots, r$ ) by  $q$ . The number of independent variables among  $T$ ,  $P$ ,  $\mu_{j_s}$  is determined by the Gibbs phase rule:

$$\phi = r + 2 - q \quad (2.20)$$

For a pure substance, Eq. (2.20) implies that, for a single-phase state, there are only two independent variables among the three intensive properties  $T$ ,  $P$ , and  $\mu$ . If  $T$  and  $P$  are chosen

as the independent variables, then all other intensive properties are functions of  $T$  and  $P$ , e.g., specific internal energy  $u = u(T,P)$ , specific enthalpy  $h = h(T,P)$ , and specific entropy  $s = s(T,P)$ . Extensive properties can be determined from the specific properties if the total mass or volume is specified. For a two-phase mixture, such as ice and water or water and steam, only one of  $T, P$ , and  $\mu$  is independent. If  $T$  is chosen as the variable, then  $P$  and  $\mu$  can be expressed as functions of  $T$ , i.e.,  $P = P(T)$  and  $\mu = \mu(T)$ . In order to completely describe the state, however, we will also need to know the amount of constituents in each phase (which may be expressed by the total mass and a mass fraction  $x$  of one phase). For example, the specific entropy of a mixture can be expressed as  $s = s(T,x)$  or  $s = s(P,x)$ . In a three-phase mixture,  $T, P$ , and  $\mu$  are all fixed. For a pure substance, the solid, liquid, and vapor phases can only coexist at fixed temperature and pressure, which are called *triple point* properties. Taking water as an example, we have  $T_{\text{tp.}} = 0.01^\circ\text{C}$  and  $P_{\text{tp.}} = 0.61 \text{ kPa}$ . One needs to know the amount of constituents in each phase to completely characterize the state. No more than three phases can coexist for any pure substance. It should be noticed that a substance can have different solid phases, e.g., diamond and graphite are allotropes of carbon but with distinct differences in their physical and chemical properties; silicon dioxide can exist in the forms of crystalline quartz or fused silica (glass).

Figure 2.3 shows regions of solid, liquid, and vapor in a  $P$ - $T$  diagram. The S-L, S-V, and L-V lines indicate the coexistence of solid-liquid, solid-vapor, and liquid-vapor phases in

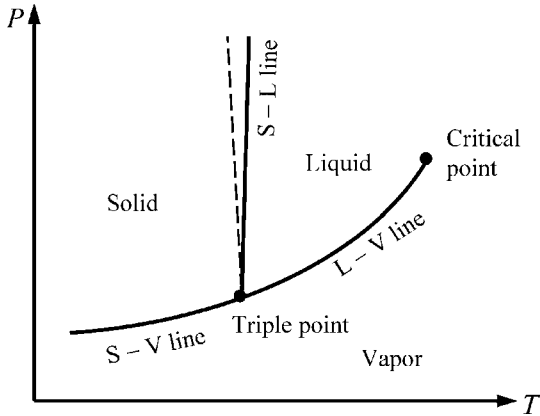
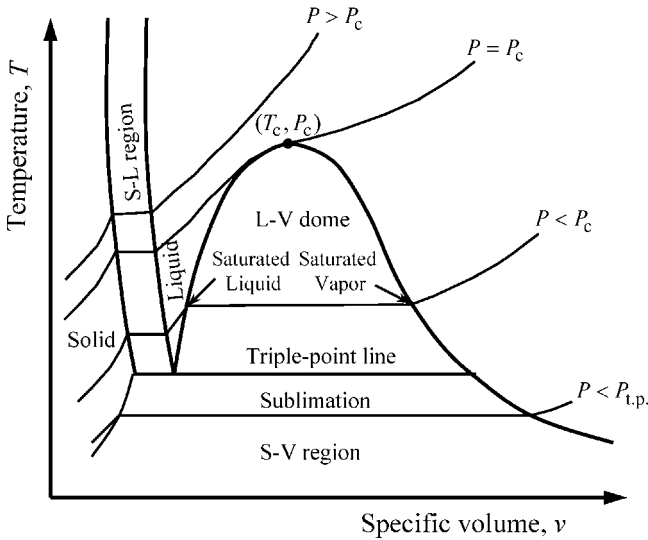


FIGURE 2.3 Schematic of a  $P$ - $T$  diagram for a pure substance.

thermodynamic equilibrium. The three lines merge to the triple point where all three phases can coexist in thermodynamic equilibrium. There are two S-L lines: the solid line represents a material that expands upon melting, and the dashed line represents a material that contracts upon melting (such as water). There exists a *critical point* or a critical state; the temperature and the pressure at the critical state are called *critical temperature* ( $T_c$ ) and *critical pressure* ( $P_c$ ). The distinction between liquid and vapor phases disappears beyond the critical point. This can be seen clearer in the  $T$ - $v$  diagram shown in Fig. 2.4. The S-L line in Fig. 2.3 becomes an S-L region in Fig. 2.4; the L-V line becomes a dome, called the *saturation dome*. Starting from a solid state, in a constant-pressure (isobaric) heating process with  $P_{\text{tp.}} < P < P_c$ , the temperature increases until melting starts. As more energy is added to the system, the fraction of solid decreases whereas the fraction of liquid increases, at a constant temperature. The amount of heat needed to completely melt a unit mass of solid to liquid is called the *specific latent heat of melting*. Once all the substance is in the liquid phase, the temperature rises again with increasing energy until a saturated liquid state is reached. Hereafter, vaporization occurs at constant temperature (saturation



**FIGURE 2.4** Schematic of a  $T$ - $v$  diagram for a material that expands upon melting.

temperature) until reaching the right side of the saturation dome, which is a saturated vapor state. The amount of energy needed to vaporize a unit mass of a substance is called the *specific latent heat of vaporization*. When the pressure is higher than the critical pressure, however, no vaporization can happen. The liquid and gas forms of aggregation differ in degree rather than in kind. At a pressure lower than the triple point pressure, the change from solid to vapor can occur without passing through a liquid phase. Such a process is called sublimation. An example is the sublimation of dry ice into  $\text{CO}_2$  gas at room temperature and atmospheric pressure, creating cooling or some theatrical effects.

### 2.2.3 Specific Heats

Specific heats are properties of a system (at stable equilibrium). The *specific heat at constant volume* ( $c_v$ ) and the *specific heat at constant pressure* ( $c_p$ ) are defined as

$$c_v = \left( \frac{\partial u}{\partial T} \right)_v = T \left( \frac{\partial s}{\partial T} \right)_v \quad (2.21a)$$

and

$$c_p = \left( \frac{\partial h}{\partial T} \right)_p = T \left( \frac{\partial s}{\partial T} \right)_p \quad (2.21b)$$

where subscripts  $V$  and  $P$  signify fixed volume and fixed pressure, respectively. The *heat capacity* is the product of the corresponding specific heat and the mass of the system. Note that only in a reversible process, the amount of heat transferred to a system is  $\delta Q = Tds$ . The heat capacity at constant volume of a closed system can be measured in terms of the total amount of energy supplied to it divided by its temperature rise in a constant-volume process. On the other hand, the heat capacity at constant pressure of a

closed system (such as in a piston-cylinder arrangement) can be measured in terms of the amount of energy per unit mass supplied to the system, excluding the volume work done by the system ( $\delta W = pdV$ ), divided by the temperature rise in an isobaric process. For example, in a reversible isobaric process,  $dU = \delta Q - pdV$  and  $dH = \delta Q$ . Therefore,  $c_p = (1/m)(dH/dT) = (1/m)(\delta Q/\delta T)$ .

Specific heats are not defined for all equilibrium states. For example, in a two-phase state, temperature and pressure are not independent, and enthalpy can be varied without changing the temperature at a constant pressure. This means that the specific heat approaches infinity in these states. In fact, the discontinuity in  $c_p$ - $T$  curve suggests some kind of phase transformation.

A *heat reservoir* is an idealized system that experiences only reversible heat interactions. For any finite amount of energy transfer, its temperature remains unchanged. Therefore, the heat capacity of a reservoir is infinitely large. For a reservoir at temperature  $T_R$ , the energy-entropy relation is a straight line in the  $E$ - $S$  graph, i.e.,

$$E_{R,2} - E_{R,1} = T_R(S_{R,2} - S_{R,1}) \quad (2.22a)$$

Furthermore, the amount of heat transferred to the reservoir from state 1 to state 2 is given by

$$Q = E_{R,2} - E_{R,1} \quad (2.22b)$$

For a pure substance in a single phase, temperature and pressure are independent, and all other properties can be expressed as functions of  $T$  and  $P$ . The relation among temperature, pressure, and specific volume, i.e.,

$$f(T, P, v) = 0 \quad \text{or} \quad v = v(T, P) \quad (2.23)$$

is called the *equation of state*. This equation does not contain information about the internal energy or the entropy. However, we can use the function  $c_p = c_p(T, P)$ , in addition to the equation of state, to fully determine all intensive properties. For example,  $ds = (\partial s/\partial T)_P dT + (\partial s/\partial P)_T dP$ . Using  $(\partial s/\partial T)_P = c_p(T, P)/T$ , from the definition of specific heat, and  $(\partial s/\partial P)_T = -(\partial v/\partial T)_P$ , which is a Maxwell relation, see Problem 2.11, we obtain

$$ds = \frac{c_p(T, P)}{T} dT - \left( \frac{\partial v}{\partial T} \right)_P dP \quad (2.24)$$

Furthermore, 
$$dh = c_p(T, P) dT + \left[ v(T, P) - \left( \frac{\partial v}{\partial T} \right)_P \right] dP \quad (2.25)$$

Under certain circumstances, the equation of state is rather simple and the specific heats can be assumed as functions of the temperature only, i.e., independent of the pressure. These ideal behaviors will be discussed in the next section.

**Example 2-2.** *Specific heat and latent heat.* A system consists of 10 kg of  $H_2O$  in a closed container maintained at a constant pressure of 100 kPa. Initially, the temperature is at  $-40^\circ\text{C}$  (ice) and it is heated to  $130^\circ\text{C}$  (vapor). How much energy must be provided to the system? What is the entropy change of the system? The specific heats of  $H_2O$  in the solid, liquid, and vapor states are  $c_{p,s} = 2 \text{ kJ}/(\text{kg} \cdot \text{K})$ ,  $c_{p,l} = 4.2 \text{ kJ}/(\text{kg} \cdot \text{K})$ , and  $c_{p,g} = 2 \text{ kJ}/(\text{kg} \cdot \text{K})$ , respectively. The specific latent heats of melting and evaporation are  $h_{sf} = 334 \text{ kJ}/\text{kg}$  and  $h_{fg} = 2257 \text{ kJ}/\text{kg}$ .

**Solution.** From the first law of the closed system in an isobaric process,  $\Delta U = Q - W$ . Since  $\Delta P = 0$ ,  $W = P\Delta V$ . Hence,  $Q = \Delta H = H_2 - H_1$ . Let  $T_1 = 233 \text{ K}$  and  $T_2 = 403 \text{ K}$  be the initial and

final temperatures, respectively, and  $T_{\text{sat,m}} = 273 \text{ K}$  and  $T_{\text{sat}} = 373 \text{ K}$  be the saturation temperatures. Based on the definition of specific heats, we obtain

$$Q = H_2 - H_1 = m[c_{p,s}(T_{\text{sat,m}} - T_1) + h_{\text{sf}} + c_{p,f}(T_{\text{sat}} - T_{\text{sat,m}}) + h_{\text{fg}} + c_{p,g}(T_2 - T_{\text{sat}})] = 31.51 \text{ MJ}$$

In the single-phase regions, entropy difference can be evaluated by integrating Eq. (2.21b) or Eq. (2.24) since  $P$  is fixed. During the phase change,  $\Delta S = \Delta H/T$  since the temperature is a constant.

$$S_2 - S_1 = m \left[ c_{p,s} \ln \left( \frac{T_{\text{sat,m}}}{T_1} \right) + \frac{h_{\text{sf}}}{T_{\text{sat,m}}} + c_{p,f} \ln \left( \frac{T_{\text{sat}}}{T_{\text{sat,m}}} \right) + \frac{h_{\text{fg}}}{T_{\text{sat}}} + c_{p,g} \ln \left( \frac{T_2}{T_{\text{sat}}} \right) \right] = 90.6 \text{ kJ/K}$$

**Discussion.** From the Steam Table or software accompanied with common thermodynamics text,<sup>4,5</sup> we can find the specific properties of water as follows:  $h_1 = -411.7 \text{ kJ/kg}$ ;  $s_1 = -1.532 \text{ kJ/(kg} \cdot \text{K)}$ ;  $h_2 = 2737 \text{ kJ/kg}$ ;  $s_2 = 7.517 \text{ kJ/(kg} \cdot \text{K)}$ . Therefore,  $Q = \Delta H = m(h_2 - h_1) = 31.49 \text{ MJ}$ ;  $\Delta S = m(s_2 - s_1) = 90.5 \text{ kJ/K}$ . The negligibly small difference is caused by the assumption of constant specific heat in each phase.

## 2.3 IDEAL GAS AND IDEAL INCOMPRESSIBLE MODELS

The amount of constituents is commonly expressed in terms of the amount of matter in mole. The *mole* is the amount of substance of a system that contains as many elementary entities as there are atoms in 0.012 kg of carbon 12. One mole of substance contains  $6.022 \times 10^{23}$  molecules, atoms, or other particles. This value is called the Avogadro's constant, i.e.,  $N_A = 6.022 \times 10^{26} \text{ kmol}^{-1}$ . Quantities like molecules and particles do not appear in the units. The mass  $m = \bar{n}M$ , where  $\bar{n}$  is the amount of constituents in kmol and  $M$  is called the molecular weight. For example,  $M = 18.012 \text{ kg/kmol}$  for water.

### 2.3.1 The Ideal Gas

At relatively high temperature and sufficiently low pressure, most substances behave as a single-phase fluid, in which the interactions between its molecules are generally negligible. The equation of state can be expressed as

$$P\bar{v} = \bar{R}T \quad \text{or} \quad PV = \bar{n}\bar{R}T \quad (2.26a)$$

where  $\bar{v} = V/\bar{n}$  is the molar specific volume in  $\text{m}^3/\text{kmol}$ ,  $\bar{R} = 8314 \text{ J/(kmol} \cdot \text{K)}$  is the *universal gas constant*. Equation (2.26a) is called the ideal gas equation since it can be considered as the definition of an ideal gas. Under *standard conditions* (temperature of  $25^\circ\text{C}$  and pressure of 1 atm), 1 kmol of an ideal gas occupies a volume of  $22.5 \text{ m}^3$ . Dry air can be treated as an ideal gas with an average molecular weight of  $M = 29 \text{ kg/kmol}$ . The ideal gas equation of state can be written in terms of the mass quantities for a given substance, i.e.,

$$Pv = RT \quad \text{or} \quad PV = mRT \quad (2.26b)$$

In the above equation,  $v = V/m$  is the specific volume, and  $R = \bar{R}/M$  is called the gas constant of the particular substance. The Boltzmann constant is defined as  $k_B = \bar{R}/N_A = 1.381 \times 10^{-23}$  J/K. It can be considered as the universal gas constant in terms of particles. Furthermore, if we denote the number density (number of particles per unit volume) as  $n$ , then the ideal gas equation can be written as  $P = nk_B T$  since  $n = N_A \bar{n}/V$ .

For ideal gases, both  $c_p$  and  $c_v$  are independent of the pressure, as will be shown from statistical thermodynamics in Sec. 3.3, but are generally dependent on temperature. The specific internal energy and enthalpy are functions of temperature only, i.e.,

$$du = c_v(T)dT \quad \text{and} \quad dh = c_p(T)dT \quad (2.27)$$

The specific heats  $c_p$  and  $c_v$  are related by the Mayer relation as

$$\bar{c}_p - \bar{c}_v = \bar{R} \quad \text{or} \quad c_p - c_v = R \quad (2.28)$$

If  $c_v(T) = \text{const.}$ , which is sometimes referred to as *perfect gas* behavior, then Eq. (2.27) can be integrated to yield

$$u_2 - u_1 = c_v(T_2 - T_1) \quad (2.29a)$$

and

$$h_2 - h_1 = c_p(T_2 - T_1) \quad (2.29b)$$

where subscripts 1 and 2 can be any two (thermodynamic equilibrium) states. The specific entropy depends on both the temperature and the pressure, i.e.,

$$ds = c_p \frac{dT}{T} - R \frac{dP}{P} \quad (2.30a)$$

Integrating the above equation from state 1 to state 2 yields

$$s_2 - s_1 = \int_1^2 \frac{c_p(T)}{T} dT - R \ln \left( \frac{P_2}{P_1} \right) \quad (2.30b)$$

In an isentropic process ( $ds = 0$ ) of a perfect gas, it can be shown that  $Pv^\gamma = \text{const.}$ , where  $\gamma = c_p/c_v$  is the *specific heat ratio*. Note that  $Pv = \text{const.}$  in an isothermal process.

**Example 2-3.** A cylinder contains 0.01 kmol of  $N_2$  gas (0.28 kg), which may be modeled as an ideal diatomic gas with  $c_v = 2.5R$ . A piston maintains the gas at constant pressure,  $P_0 = 100$  kPa. The cylinder interacts with a cyclic machine, which in turn interacts with a reservoir at  $T_R = 1000$  K. The cylinder, the reservoir, and the machinery cannot interact with any other systems. The cyclic machine may produce work  $W$  (which cannot be negative). A process brings the volume of the cylinder from  $V_1 = 0.224$  m<sup>3</sup> to  $V_2 = 0.448$  m<sup>3</sup>.

- What is the least amount of energy that must be transferred out from the reservoir? In such a case, how much work does the cyclic machine produce? How much entropy is generated in the process?
- Find the maximum work that the cyclic machine can produce.

**Analysis.** A schematic drawing is made first as shown in Fig. 2.5. From the ideal gas equation,  $T_1 = P_1 V_1 / \bar{n} \bar{R} = 269.4$  K and  $T_2 = 538.8$  K. The initial and final states of the cylinder are fully prescribed. The work done by the cylinder is  $W_B = \int P dV = P(V_2 - V_1) = 22.4$  kJ, which is also fixed.



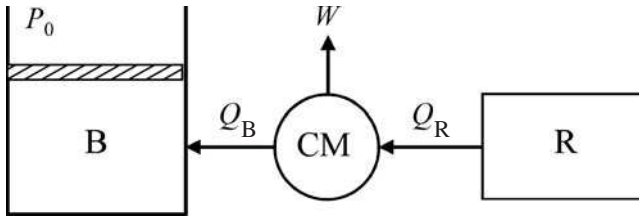


FIGURE 2.5 Schematic drawing for Example 2-3.

By applying the first law to the cylinder in an isobaric process,  $Q_B = m(h_2 - h_1) = mc_p(T_2 - T_1) = 3.5 \times P(V_2 - V_1) = 78.4$  kJ. The work done by the cyclic machine is  $W = Q_R - Q_B$ . Because  $Q_B$  is prescribed and  $W \geq 0$ , the least amount of energy that must be transferred from the reservoir is when  $W = 0$  and  $Q_R = Q_B$ .

**Solution.** (a)  $Q_R = Q_B = 78.4$  kJ and  $W = 0$ . We can evaluate the entropy change of the combined system by the following:

$$\begin{aligned} \Delta S &= m(s_2 - s_1) + \Delta S_{\text{CM}} + (-Q_R/T_R) = m[c_p \ln(T_2/T_1) - R \ln(P_2/P_1)] - 78.4/1000 \\ &= (0.2017 - 0.0784) \text{ kJ/K} = 123.3 \text{ J/K} \end{aligned}$$

Since the system does not have any interactions with any other systems, the entropy change is caused solely by entropy generation.

(b) The maximum work that can be produced is through a reversible process (*not a Carnot cycle since the temperature of the cylinder is not constant*). By setting  $\Delta S = m(s_2 - s_1) - Q_R/T_R = 0$ , we find  $Q_R = T_R m c_p \ln(T_2/T_1) = 201.7$  kJ. The maximum amount of work is therefore  $W_{\text{max}} = Q_R - Q_B = 123.3$  kJ.

## 2.3.2 Incompressible Solids and Liquids

The assumption for *ideal incompressible* behavior is  $v = \text{const.}$ , which is the equation of state for incompressible solids and liquids. It can be shown that in this case  $c_p = c_v$  and, to a good approximation, the specific heat depends on temperature only. It is common to use  $c_p$  for the specific heat of solids and liquids. Using Eq. (2.24) and Eq. (2.25), we obtain the specific internal energy, enthalpy, and entropy for an ideal incompressible solid or liquid as follows:

$$du = c_p(T)dT \quad (2.31)$$

$$ds = c_p(T) \frac{dT}{T} \quad (2.32)$$

and

$$dh = c_p(T)dT + vdP \quad (2.33)$$

Notice that the internal energy and the entropy are functions of temperature only, but the enthalpy depends also on the pressure, though the second term on the right-hand side of Eq. (2.33) is relatively small unless the pressure is high.

**Example 2-4.** In a Rankine cycle, water at  $15^\circ\text{C}$ , 100 kPa is compressed through a pump to 10 MPa before entering the boiler. Model the water as an incompressible liquid with a constant

specific heat  $c_p = 4.2 \text{ kJ}/(\text{kg} \cdot \text{K})$ . What is the least amount of work required to pump 1 kg of water? What is the exit temperature of the water? If the pump efficiency is  $\eta_p = 80\%$ , what is the actual specific work and exit temperature of the pump?

**Solution.** Take  $v = 0.001 \text{ m}^3/\text{kg}$  as an approximation. The least amount of work is needed in a reversible process. It has been shown that the reversible work *done by the system* between bulk flow states is  $\delta w = -v dP$ . Hence, the work needed in a reversible process is

$$w_{\text{rev}} = h_{2s} - h_1 = 0.001(10,000 - 100) \text{ kJ}/\text{kg} = 9.9 \text{ kJ}/\text{kg}$$

Because it is an adiabatic and reversible process, it must be isentropic, i.e.,  $s_{2s} - s_1 = c_p \ln(T_{2s}/T_1) = 0$ . Hence,  $T_{2s} = T_1 = 15^\circ\text{C}$ . Actual work  $w = w_{\text{rev}}/\eta_p = 12.375 \text{ kJ}/\text{kg}$ . Since  $w = h_2 - h_1 = c_p(T_2 - T_1) + v(P_2 - P_1)$ ,

$$T_2 = T_1 + \frac{h_2 - h_1}{c_p} - \frac{v}{c_p}(P_2 - P_1) = T_1 + \frac{w - w_{\text{rev}}}{c_p} = 15.59^\circ\text{C}$$

which is less than 1 K higher. The entropy generation is  $s_{\text{gen}} = c_p \ln(T_2/T_1) = 8.6 \text{ J}/(\text{kg} \cdot \text{K})$ .

**Discussion.** We can use the Steam Table and notice that all states are compressed liquid. The properties at state 1 can be evaluated at  $T_1 = 15^\circ\text{C}$  and  $P_1 = 100 \text{ kPa}$ , at state 2s (reversible) can be evaluated at  $P_{2s} = 10 \text{ MPa}$  and  $s_{2s} = s_1$ , and at state 2 can be evaluated at  $P_2 = 10 \text{ MPa}$  and  $h_2 = h_1 + w$ . Hence,

$$w_{\text{rev}} = 9.88 \text{ kJ}/\text{kg}; T_{2s} = 15.11^\circ\text{C}; w = 12.35 \text{ kJ}/\text{kg};$$

$$T_2 = 15.67^\circ\text{C}; s_{\text{gen}} = 8.2 \text{ J}/(\text{kg} \cdot \text{K})$$

The differences are negligibly small compared with those obtained from the incompressible assumption. Note that the temperature change in the pump is usually very small. On a  $T$ - $s$  diagram, it is difficult to distinguish states 1, 2s, and 2. In fact, state 2 crosses the saturated-liquid line to overlap with a two-phase-mixture state at  $T_2$  and  $s_2$ . This is because  $T$  and  $s$  together cannot uniquely determine a stable equilibrium state.

## 2.4 HEAT TRANSFER BASICS

---

Classical thermodynamics deals with the changes of mass, energy, and entropy of a system between equilibrium states, and establishes the required balance equations between end states during a given process. For example, we have learned that spontaneous transfer of energy can occur only from a higher temperature to a lower temperature. In thermodynamics, heat interaction is defined as the transfer of energy at the mutual (interface) temperature between two systems. Heat transfer is a subject that extends the thermodynamic principles to detailed energy transport processes that occur as a consequence of temperature differences. Heat transfer phenomena are abundant in our everyday life and play an important role in many industrial, environmental, and biological processes. Examples include energy conversion and storage, electrical power generation, combustion processes, heat exchangers, building-temperature regulation, thermal insulation, refrigeration, micro-electronic cooling, materials processing, manufacturing, global thermal budget, agriculture, food industry, and biological systems. Based on the local-equilibrium assumption, heat transfer analysis deals with the rate of heat transfer and/or the temperature distributions (steady state or transient) for given geometries, materials, and initial and boundary conditions. Thermal design, on the other hand, determines the necessary geometric structure and

materials for use to achieve optimum performance for a specific task, such as a heat exchanger.

Heat conduction refers to the transfer of heat in a stationary (from the macroscopic point of view) medium, which may be a solid, a liquid, or a gas. Energy can also be transferred between objects by the emission and absorption of electromagnetic waves without any intervening medium; this is called *thermal radiation*, such as the radiation from the sun. When the transfer of heat involves fluid motion, we call it *convection heat transfer*, or simply, *convection*. Examples of convection are the cooling of a cup of tea, hot water flowing in a pipe, and cold air blowing outside the wall of a building. The basic macroscopic formulations of conduction, convection, and radiation heat transfer are summarized in this section. The microscopic understanding of the underlying mechanisms, as well as the effect of small dimensions and short time duration on the transfer processes, will be the subject of the remaining chapters. Some historic aspects and an integrated approach of heat transfer processes can be found in Kaviany.<sup>10</sup>

### 2.4.1 Conduction

In a stationary medium, heat transfer occurs if the medium is not at thermal equilibrium. The assumption of local equilibrium allows us to define the temperature at each location. Fourier's law states that the heat flux (or heat transfer rate per unit area)  $\mathbf{q}''$  is proportional to the temperature gradient  $\nabla T$ , i.e.,

$$\mathbf{q}'' = -\kappa \nabla T \quad (2.34)$$

where  $\kappa$  is called *thermal conductivity*, which is a material property that may depend on temperature. Notice that  $\mathbf{q}''$  is a vector and its direction is always perpendicular to the isotherms and opposite to the temperature gradient. In an anisotropic medium, such as a thin film or a thin wire, the thermal conductivity depends on the direction along which it is measured.

By doing a control volume analysis using energy balance, a differential equation can be obtained for the transient temperature distribution  $T(t, \mathbf{r})$  in a homogeneous isotropic medium; that is<sup>10–12</sup>

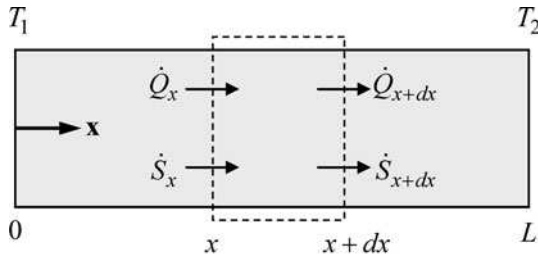
$$\nabla \cdot (\kappa \nabla T) + \dot{q} = \rho c_p \frac{\partial T}{\partial t} \quad (2.35)$$

where  $\nabla \cdot$  is the divergence operator,  $\dot{q}$  is the volumetric thermal energy generation rate, and  $\rho c_p$  can be considered as volumetric heat capacity. Equation (2.35) is called the heat diffusion equation or heat equation. Note that the concept of thermal energy generation is very different from the concept of entropy generation. Thermal energy generation refers to the conversion of other types of energy (such as electrical, chemical, or nuclear energies) to the internal energy of the system, while the total energy is always conserved. Entropy need not be conserved, and entropy generation refers to the creation of entropy by an irreversible process. If there is no thermal energy generation and the thermal conductivity can be assumed to be independent of temperature, Eq. (2.35) reduces to  $\nabla^2 T = 0$  at steady state, where  $\nabla^2 T = \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}$  in the Cartesian coordinates. With the prescribed initial temperature distribution and boundary conditions, the heat equation can be solved analytically for simple cases and numerically for more complex geometries as well as initial and boundary conditions. Typical boundary conditions include (a) constant temperature, (b) constant heat flux, (c) convection, and (d) radiation.

Generally speaking, metals with high electrical conductivities and some crystalline solids have very high thermal conductivities [100 to 1000 W/(m · K)]; alloys and metals with low electric conductivities have slightly lower thermal conductivities [10 to 100 W/(m · K)]; water, soil, glass, and rock have thermal conductivities from 0.5 to 5 W/(m · K); thermal insulation materials usually have a thermal conductivity on the order of 0.1 W/(m · K); and gases have the lowest thermal conductivity, e.g., the thermal conductivity of air at 300 K is 0.026 W/(m · K). Notice that thermal conductivity generally depends on temperature. A comprehensive collection of thermal-property data can be found in Touloukian and Ho.<sup>13</sup> At room temperature, Diamond IIa has the highest thermal conductivity,  $\kappa = 2300$  W/(m · K) among all natural materials. Researchers have shown that single-walled carbon nanotubes can have even higher thermal conductivity at room temperature. More detailed discussion about the mechanisms of thermal conduction and thermal properties of nanostructures will be provided in subsequent chapters.

**Example 2-5.** Consider the steady-state heat conduction through a solid rod, whose sides are insulated, between a constant-temperature source at  $T_1 = 600$  K and a constant-temperature sink at  $T_2 = 300$  K. Assume the thermal conductivity of the rod is independent of temperature,  $\kappa = 150$  W/(m · K). The rod has a length  $L = 0.2$  m and cross-sectional area  $A = 0.001$  m<sup>2</sup>. Show that the temperature distribution along the rod is linear. What is the heat transfer rate? What is the volumetric entropy generation rate? What is the total entropy generation rate?

**Solution.** This is a 1-D heat conduction problem with no thermal energy generation, as shown in Fig. 2.6. Fourier’s law can be written as  $\dot{Q}_x = -\kappa A(dT/dx)$ . Note that at steady state, the heat transfer



$$\dot{Q}_x = -\kappa A \frac{dT}{dx} \quad \text{and} \quad \dot{S}_x = \frac{\dot{Q}_x}{T(x)}$$

**FIGURE 2.6** Illustration of the control volume for energy and entropy balances in a solid rod with heat conduction.

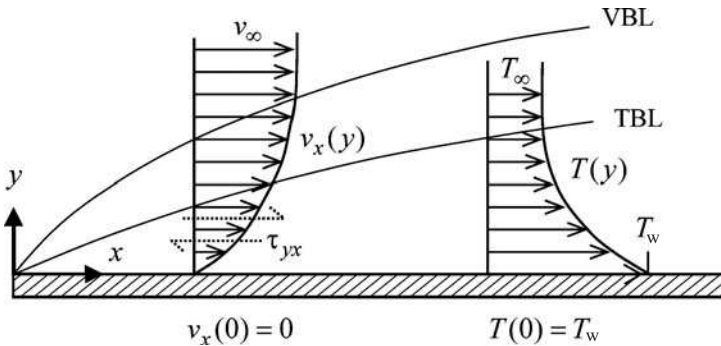
rate  $\dot{Q}_x$  is independent of  $x$  because there is no thermal energy generation. Because both  $\kappa$  and  $A$  are constant,  $dT/dx$  must not be a function of  $x$ , i.e., the temperature distribution is linear. From the boundary conditions  $T(0) = T_1$  and  $T(L) = T_2$ , we have  $T(x) = T_1 + (T_2 - T_1)(x/L)$ . Furthermore,  $\dot{Q}_x = \kappa A(T_1 - T_2)/L = 225$  W. To evaluate the entropy generation rate, we can apply Eq. (2.2b) to the control volume  $A dx$  to obtain  $\dot{s}_{\text{gen}}(x) A dx$ . The net entropy transferred to the control volume is  $\dot{S}_x - \dot{S}_{x+dx} = -d(\dot{Q}_x/T)$ . The sum of the entropy generation and entropy transferred is equal to the entropy change, which is zero at steady state. Therefore,  $\dot{s}_{\text{gen}}(x) = q''_x d(1/T)/dx = (\kappa/T^2)(dT/dx)^2$ , where  $q''_x = \dot{Q}_x/A$  is the heat flux. To calculate the total entropy generation rate, we can integrate  $\dot{s}_{\text{gen}}(x)$  over the whole rod. Alternatively, we can perform an entropy balance for the rod as a whole, which gives the rate of entropy generation for a heat transfer rate  $\dot{Q}_x$  from  $T_1$  to  $T_2$  as  $\dot{S}_{\text{gen}} = \dot{Q}_x(1/T_2 - 1/T_1) = 0.375$  W/K.

This example shows that the entropy generation occurs in a finite volume while the entropy flows through the interface. The amount of entropy flux increases with  $x$  as more and more entropy is generated through the irreversible process. More discussion on the entropy generation in heat transfer and fluid flow processes can be found in Bejan.<sup>14</sup>

*Contact resistance* is important in microelectronics thermal management and cryogenic heat transfer. A large thermal resistance may exist due to imperfect contact, such as surface roughness. The result is a large temperature difference across the interface. The value of contact resistance depends on the surface conditions, adjacent materials, and contact pressure. As an example, assume a contact resistance between two stainless steel plates to be  $R_c'' = 0.001 \text{ m}^2 \cdot \text{K/W}$  and the thermal conductivity of the stainless steel  $\kappa = 50 \text{ W/(m} \cdot \text{K)}$ . If the thickness of each plate is  $L = 5 \text{ mm}$  and the area of the plate is  $A = 0.01 \text{ m}^2$ , the total thermal resistance is then  $R_t = L/\kappa A + R_c''/A + L/\kappa A = (0.01 + 0.1 + 0.01) \text{ K/W} = 0.12 \text{ K/W}$ , which is mostly due to the contact resistance. Interfacial fluids and interstitial (filler) materials can be applied to reduce the contact resistance in some cases. Even with a perfect contact, thermal resistance exists between dissimilar materials due to acoustic mismatch, which is especially important at low temperatures.<sup>15</sup>

## 2.4.2 Convection

Convection heat transfer refers to the heat transfer from solid to fluid near the boundary when the fluid is in bulk motion relative to the solid. The combination of the bulk motion, known as *advection*, with the random motion of the fluid molecules (i.e., diffusion) is the key for convection heat transfer. Examples are flows over an object or inside a tube, a spray leaving a nozzle that is impinging on a microelectronic component for cooling purposes, and boiling in a pan. The velocity and temperature distributions for a fluid flowing over a heated flat plate are illustrated in Fig. 2.7. A *hydrodynamic* or *velocity boundary layer* is formed



**FIGURE 2.7** Illustration of the velocity boundary layer (VBL) and the thermal boundary layer (TBL).

near the surface, and the fluid moves at the free-stream velocity outside the boundary layer. Similarly, a *thermal boundary layer* is developed near the surface of the plate where a temperature gradient exists. When the flow speed is not very high and the density of the fluid not too low, the average velocity of the fluid is zero, and the fluid temperature equals the wall temperature in the vicinity of the wall, i.e.,  $v_x(y = 0) = 0$  and  $T(y = 0) = T_w$ . For Newtonian fluids, a linear relationship exists between the stress components and the

velocity gradients. Many common fluids like air, water, and oil belong to this catalog. The shear stress in the fluid is

$$\tau_{yx} = -\mu \frac{\partial v_x}{\partial y} \quad (2.36)$$

where  $\mu$  is the viscosity. Throughout this book, we will use  $v_x, v_y$ , and  $v_z$  (or  $v_i$  with  $i = 1, 2$ , and  $3$ ) for the velocity components in the  $x, y$  and  $z$  directions, respectively. When Eq. (2.36) is evaluated at the boundary  $y = 0$ , it gives the force per unit area exerted to the fluids by the wall and is used to calculate the *friction factor* in fluid mechanics.<sup>16</sup>

The heat flux between the solid and the fluid can be predicted by applying Fourier's law to the fluid at the boundary; thus,

$$q''_w = -\kappa \frac{\partial T}{\partial y} \Big|_{y=0} \quad (2.37)$$

where  $\kappa$  is the thermal conductivity of the fluid. Equation (2.37) shows that the basic heat transfer mechanism for convection is the same as that for conduction, i.e., both are caused by heat diffusion and governed by the same equation. Without bulk motion, however, the temperature gradient at the boundary would be smaller. Therefore, advection generally increases the heat transfer rate. Newton's law of cooling is a phenomenological equation for convection. It states that the convective heat flux is proportional to the temperature difference, therefore,

$$q''_w = h(T_w - T_\infty) \quad (2.38)$$

where  $h$  is called the *convection heat transfer coefficient* or *convection coefficient*.  $T_w$  is the surface temperature, and  $T_\infty$  is the fluid temperature. From Eq. (2.37) and Eq. (2.38), we have

$$h = \frac{-\kappa}{T_w - T_\infty} \frac{\partial T}{\partial y} \Big|_{y=0} \quad (2.39)$$

Although  $h$  depends on the location, the average convection coefficient is often used in heat transfer calculations. The convection coefficient depends on the fluid thermal conductivity, velocity, and flow conditions (laminar versus turbulent flow, internal versus external flow, and forced versus free convection). Convection can also happen with phase change, such as boiling, which usually causes vigorous fluid motion and enhanced heat transfer. Convection correlations are recommended in most heat transfer textbooks to determine the convection coefficient. For laminar flow over a flat plate of length  $L$  with a free-stream velocity  $v_\infty$ , the following equation correlates the average Nusselt number to the Reynolds number at  $x = L$  and the Prandtl number:<sup>11</sup>

$$\overline{Nu}_L = \frac{\overline{h}_L L}{\kappa} = 0.664 Re_L^{1/2} Pr^{1/3}, \quad \text{for } Pr > 0.6 \quad \text{and} \quad Re_L < 5 \times 10^5 \quad (2.40)$$

The Reynolds number, defined as  $Re_L = \rho v_\infty L / \mu$ , is key to the study of hydrodynamics. The Prandtl number  $Pr = \nu / \alpha$  is the ratio of *kinematic viscosity*  $\nu = \mu / \rho$ , which is also known as the *momentum diffusivity*, to the thermal diffusivity  $\alpha = \kappa / (\rho c_p)$  of the fluid. A detailed understanding of the fluid flow and convection heat transfer requires the solution of the conservation equations, as summarized in the following:

The differential form of the continuity equation or mass conservation is

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0 \quad (2.41)$$

where  $D/Dt = (\partial/\partial t + \mathbf{v} \cdot \nabla)$  is called the substantial derivative or material derivative. Notice that for an incompressible fluid, the continuity equation reduces to  $\nabla \cdot \mathbf{v} = 0$ .

Using Stokes' hypothesis that relates the second coefficient of viscosity to the viscosity for Newtonian fluids, the Navier-Stokes equation that describes the momentum conservation can be expressed as follows:<sup>16</sup>

$$\frac{D\mathbf{v}}{Dt} = -\frac{\nabla P}{\rho} + \mathbf{a} + \nu \nabla^2 \mathbf{v} + \frac{\nu}{3} \nabla(\nabla \cdot \mathbf{v}) \quad (2.42)$$

where  $\mathbf{a}$  is the body force per unit mass exerted on the fluid, i.e., the acceleration vector.

Energy equation for constant thermal conductivity without thermal energy generation for a moving fluid can be expressed as

$$\rho \frac{Du}{Dt} = \kappa \nabla^2 T - P \nabla \cdot \mathbf{v} + \mu \Phi \quad (2.43a)$$

where  $u$  is the specific internal energy ( $du = c_v dT$ ) and the last term accounts for the viscous dissipation, which is

$$\begin{aligned} \Phi = 2 & \left[ \left( \frac{\partial v_x}{\partial x} \right)^2 + \left( \frac{\partial v_y}{\partial y} \right)^2 + \left( \frac{\partial v_z}{\partial z} \right)^2 \right] + \left( \frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right)^2 + \left( \frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y} \right)^2 \\ & + \left( \frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z} \right)^2 - \frac{2}{3} (\nabla \cdot \mathbf{v})^2 \end{aligned} \quad (2.43b)$$

in the Cartesian coordinates. Equations (2.41) through (2.43) are usually simplified for specific conditions and solved analytically or numerically using computation fluid dynamics software. In Chap. 4, we will show that the conservation equations can also be derived from the microscopic theories, which are also applicable for rarefied flows and microfluidics.

### 2.4.3 Radiation

Thermal radiation refers to the electromagnetic radiation in a broad wavelength range from approximately 100 nm to 1000  $\mu\text{m}$ . It includes a portion of the ultraviolet region, the entire visible (400 to 760 nm) region, and the infrared region. Monochromatic radiation refers to radiation at a single wavelength (or a very narrow spectral band), such as lasers and some atomic emission lines. Radiation emitted from a thermal source, such as the sun, an oven, or a blackbody cavity, covers a broad spectral region and can be considered as the spectral integration of monochromatic radiation. In contrast to conduction or convection heat transfer, radiative energy propagates in the form of electromagnetic waves that do not require an intervening medium. Regardless of its wavelength, an electromagnetic wave travels in vacuum at the speed of light,  $c_0 = 2.998 \times 10^8$  m/s. Radiation can also be viewed as a collection of particles, called photons, whose energy is proportional to the frequency of radiation. Starting with the definition of intensity and its linkage to the radiative energy flux, radiative transfer between surfaces and in participating media will be briefly described later in this section. More detailed treatment of the mechanism of thermal radiation, radiative properties, and radiative transfer at small length scales will be given in Chaps. 8, 9, and 10.

The *spectral intensity* or *radiance* is defined as the radiative power received within a solid angle, a unit projected area, and a unit wavelength interval; hence,<sup>11</sup>

$$I_{\lambda}(\lambda, \theta, \phi) = \frac{d\dot{Q}}{dA \cos \theta d\Omega d\lambda} \quad (2.44)$$

where  $(\theta, \phi)$  is the direction of propagation, measured with respect to the surface normal,  $dA \cos \theta$  is therefore the projected area, and  $d\Omega$  is an element solid angle. It is convenient to describe the relationship between intensity and radiative power using the spherical coordinates, as shown in Fig. 2.8, where an element area  $dA$  whose surface normal is in

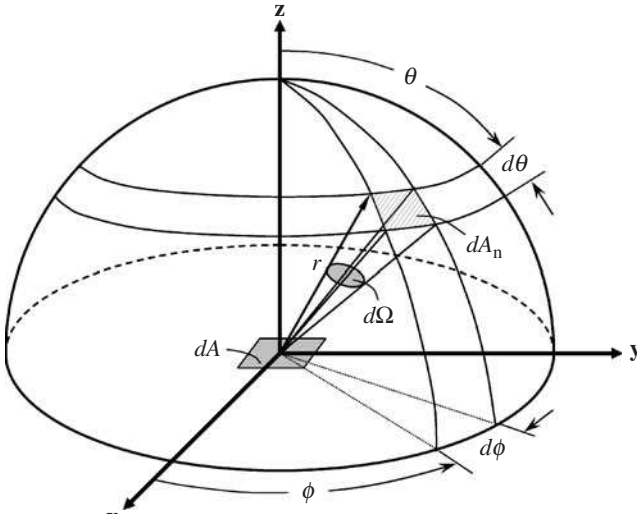


FIGURE 2.8 Illustration of the solid angle in spherical coordinates.

the  $z$  direction is placed at the origin. Note that  $r = (x^2 + y^2 + z^2)^{1/2}$ ,  $\theta = \cos^{-1}(z/r)$ , and  $\phi = \tan^{-1}(y/x)$ . The solid angle, defined as  $d\Omega = dA_n/r^2$ , can be expressed as  $d\Omega = (rd\theta)(r \sin \theta d\phi)/r^2 = \sin \theta d\theta d\phi$ .

The spectral heat flux from an element surface  $dA$  to the upper hemisphere can be obtained by integrating Eq. (2.44), i.e.,

$$q_{\lambda}''(\lambda) = \int_0^{2\pi} \int_0^{\pi/2} I_{\lambda}(\lambda, \theta, \phi) \cos \theta \sin \theta d\theta d\phi \quad (2.45)$$

The total heat flux is equal to the heat flux integrated over all wavelengths:

$$q_{\text{rad}}'' = \int_0^{\infty} q_{\lambda}''(\lambda) d\lambda \quad (2.46)$$

We can also define the total intensity as the integral of the spectral intensity over all wavelengths,  $I(\theta, \phi) = \int_0^{\infty} I_{\lambda}(\lambda, \theta, \phi) d\lambda$ . An equation similar to Eq. (2.45) holds between the total heat flux and the total intensity. If the radiation is emitted from a surface, the radiative heat flux  $q_{\text{rad}}''$  is termed as the (hemispherical) *emissive power*. When the intensity is same in all



directions, the surface is said to be diffuse, and Eq. (2.45) can be integrated to obtain the relation,  $q'' = \pi I_\lambda(\lambda, \theta, \phi)$ . Similarly, we can obtain  $q'' = \pi I$ .

The maximum power that can be emitted by a thermal source at a given temperature is from a blackbody. A blackbody is an ideal surface which absorbs all incoming radiation and gives out the maximum emissive power. Radiation inside an isothermal enclosure behaves like a blackbody. In practice, a blackbody cavity is made with a small aperture on an isothermal cavity. The emissive power of a blackbody is given by the Stefan-Boltzmann law, also proportional to the absolute temperature to the fourth power, viz.,

$$e_b(T) = \pi I_b(T) = \sigma_{\text{SB}} T^4 \quad (2.47)$$

where  $\sigma_{\text{SB}} = 5.67 \times 10^{-8} \text{ W}/(\text{m}^2 \cdot \text{K}^4)$  is the Stefan-Boltzmann constant. A blackbody is also a diffuse emitter, i.e., its intensity is independent of the direction. The spectral distribution of blackbody emission is described by Planck's law, which gives the spectral intensity as a function of temperature and wavelength as follows:

$$I_{b,\lambda}(\lambda, T) = \frac{e_{b,\lambda}(\lambda, T)}{\pi} = \frac{2hc^2}{\lambda^5 (e^{hc/k_B \lambda T} - 1)} \quad (2.48)$$

where  $h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$  is the Planck constant,  $c$  is the speed of light, and  $k_B$  is the Boltzmann constant. The derivation of Planck's law will be given in Chap. 8.

The ratio of the emissive power of a real material to that of the blackbody defines the (total-hemispherical) *emissivity*,  $\varepsilon(T) = e(T)/\sigma_{\text{SB}} T^4$ . The spectral-directional emissivity is defined as the spectral intensity emitted by the surface to  $I_{b,\lambda}$ , i.e.,  $\varepsilon'_\lambda(\lambda, \theta, \phi, T) = \pi I_\lambda(\lambda, \theta, \phi, T)/e_{b,\lambda}(\lambda, T)$ . Using  $e(T) = \int_0^\infty d\lambda \left[ \int_0^{2\pi} \int_0^{\pi/2} I_\lambda(\lambda, \theta, \phi, T) \cos \theta \sin \theta d\theta d\phi \right]$ , we have

$$\varepsilon(T) = \frac{\pi}{\sigma T^4} \int_0^\infty e_{b,\lambda}(\lambda, T) d\lambda \left[ \int_0^{2\pi} \int_0^{\pi/2} \varepsilon'_\lambda(\lambda, \theta, \phi, T) \cos \theta \sin \theta d\theta d\phi \right] \quad (2.49)$$

This equation suggests that the relationship between the total-hemispherical emissivity and the spectral-directional emissivity is rather complicated in general. For a gray surface, the spectral emissivity is not a function of the wavelength. For a diffuse surface, the intensity emitted by the surface is independent of the direction. For a diffuse-gray surface, Eq. (2.49) reduces to a simple form  $\varepsilon = \varepsilon'_\lambda$ , because the emissivity is independent of wavelength and the direction.

Real materials also reflect radiation in contrast to a blackbody. The reflection may be specular for mirrorlike surfaces and more diffuse for rough surfaces. Some window material and thin films are semitransparent. Generally speaking, reflection and transmission are highly dependent on the wavelength, angle of incidence, and polarization status of the incoming electromagnetic wave. The *absorptance*, *reflectance*, and *transmittance* of a material can be defined as the fraction of the absorbed, reflected, and transmitted radiation. The spectral-directional absorptance, spectral-directional-hemispherical reflectance, and spectral-directional-hemispherical transmittance are related by

$$A'_\lambda + R'_\lambda + T'_\lambda = 1 \quad (2.50)$$

For an opaque material, the transmittance  $T'_\lambda = 0$ . Very often, we use absorptivity  $\alpha'_\lambda$  and reflectivity  $\rho'_\lambda$  for opaque materials; hence,  $\alpha'_\lambda + \rho'_\lambda = 1$ . However, the distinction between words ending with “-tivity” and “-tance” is not always clear. Both endings are used interchangeably in the literature. The complete nomenclature of radiative quantities and properties can be found in Siegel and Howell.<sup>17</sup> Further discussion about the mechanisms and applications of radiation heat transfer will be provided in Chap. 8.

Kirchhoff's law states that the spectral-directional emissivity is always the same as the spectral-directional absorptivity, i.e.,  $\varepsilon'_\lambda \equiv \alpha'_\lambda$ . For diffuse-gray surfaces, it can also be shown that  $\varepsilon = \alpha$ , which may not be generally true for surfaces that are not diffuse-gray, unless they are in thermal equilibrium with the environment.

**Example 2-6.** Find the net radiative heat flux between two, large parallel surfaces. Surface 1 at  $T_1 = 600^\circ\text{C}$  has an emissivity  $\varepsilon_1 = 0.8$ , and surface 2 at  $T_2 = 27^\circ\text{C}$  has an emissivity  $\varepsilon_2 = 0.5$ .

**Solution.** Assume that the medium in between is transparent, and both surfaces are opaque and diffuse-gray. Note that radiation from one surface to another will be partially absorbed and partially reflected back. Furthermore, the reflected radiation will continue to experience the absorption/reflection processes between the two surfaces. Surface 1 emits  $\varepsilon_1\sigma_{\text{SB}}T_1^4$  radiation toward surface 2. The fraction of this emitted radiation that is absorbed by surface 2 can be calculated by tracing the rays between the two surfaces, which is  $\varepsilon_2 + (1 - \varepsilon_2)(1 - \varepsilon_1)\varepsilon_2 + (1 - \varepsilon_2)^2(1 - \varepsilon_1)^2\varepsilon_2 + \dots$  since the reflectivity is one *minus* the emissivity. The radiative heat flux from surface 1 to surface 2 is  $q''_{1\rightarrow 2} = \varepsilon_1\varepsilon_2\sigma_{\text{SB}}T_1^4/[1 - (1 - \varepsilon_1)(1 - \varepsilon_2)] = \sigma_{\text{SB}}T_1^4/[1/\varepsilon_1 + 1/\varepsilon_2 - 1]$ , and that from surface 2 to surface 1 is  $q''_{2\rightarrow 1} = \sigma_{\text{SB}}T_2^4/[1/\varepsilon_1 + 1/\varepsilon_2 - 1]$ . Subsequently, the net radiative flux from surface 1 to surface 2 is

$$q''_{12} = q''_{1\rightarrow 2} - q''_{2\rightarrow 1} = \frac{\sigma_{\text{SB}}(T_1^4 - T_2^4)}{1/\varepsilon_1 + 1/\varepsilon_2 - 1} \quad (2.51)$$

Plugging in  $T_1 = 873\text{ K}$ ,  $T_2 = 300\text{ K}$ , and other numerical values, we obtain  $q''_{12} = 14,433\text{ W/m}^2$ .

Gas emission, absorption, and scattering are important for atmospheric radiation and combustion. When radiation travels through a cloud of gas, some of the energy may be absorbed. The absorption of photons raises the energy levels of individual molecules. At sufficiently high temperatures, gas molecules may spontaneously lower their energy levels and emit photons. These changes in energy levels are called *radiative transitions*, which include bound-bound transitions (between nondissociated molecular states), bound-free transitions (between nondissociated and dissociated states), and free-free transitions (between dissociated states). Bound-free and free-free transitions usually occur at very high temperatures (greater than about 5000 K) and emit in the ultraviolet and visible regions. The most important transitions for radiative heat transfer are bound-bound transitions between vibrational energy levels coupled with rotational transitions. The photon energy (or frequency) must be exactly the same as the difference between two energy levels in order for the photon to be absorbed or emitted; therefore, the quantization of the energy levels results in discrete spectral lines for absorption and emission. The rotational lines superimposed on a vibrational line give a band of closely spaced spectral lines, called the vibration-rotation spectrum. Additional discussion will be given in Chap. 3 about quantized transitions in atoms and molecules.

Particles can also scatter electromagnetic waves or photons, causing a change in the direction of propagation. In the early twentieth century, Gustav Mie developed a solution of Maxwell's equations for scattering of electromagnetic waves by spherical particles, known as the Mie scattering theory which can be used to predict the scattering phase function. In the case when the particle sizes are small compared with the wavelength, the formulation reduces to the simple expression obtained earlier by Lord Rayleigh; and the phenomenon is called Rayleigh scattering, in which the scattering efficiency is inversely proportional to the wavelength to the fourth power. The wavelength-dependent characteristic of light scattering by small particles helps explain why the sky is blue and why the sun appears red at sunset. For spheres whose diameters are much greater than the wavelength, geometric optics can be applied by treating the surface as specular or diffuse.

The spectral intensity in a *participating medium*,  $I_\lambda = I_\lambda(\xi, \Omega, t)$ , depends on the location (the coordinate  $\xi$ ), its direction (the solid angle  $\Omega$ ), and time  $t$ . In a time interval  $dt$ , the beam travels from  $\xi$  to  $\xi + d\xi$  ( $d\xi = cdt$ ), and the intensity is attenuated by absorption and out-scattering, but enhanced by emission and in-scattering. The macroscopic description of the radiation intensity is known as the *equation of radiative transfer (ERT)*.<sup>17</sup>

$$\frac{1}{c} \frac{\partial I_\lambda}{\partial t} + \frac{\partial I_\lambda}{\partial \xi} = a_\lambda I_{b,\lambda}(T) - (a_\lambda + \sigma_\lambda) I_\lambda + \frac{\sigma_\lambda}{4\pi} \int_{4\pi} I_\lambda(\xi, \Omega', t) \Phi_\lambda(\Omega', \Omega) d\Omega' \quad (2.52)$$

where  $a_\lambda$  and  $\sigma_\lambda$  are the absorption and scattering coefficients, respectively,  $\Omega$  is the solid angle and direction of  $I_\lambda$ , and  $\Omega'$  is the in-scattering solid angle and direction of  $I_\lambda(\xi, \Omega', t)$ . Here,  $\Phi_\lambda(\Omega', \Omega)$  is the *scattering phase function* ( $\Phi_\lambda = 1$  for isotropic scattering), which satisfies the equation:  $\int_{4\pi} \Phi_\lambda(\Omega', \Omega) d\Omega' = 4\pi$ . The right-hand side of Eq. (2.52) is composed of three terms: the first accounts for the contribution of emission (which depends on the local gas temperature  $T$ ); the second is the attenuation by absorption and out-scattering; and the third is the contribution of in-scattering from all directions (solid angle  $4\pi$ ) to the direction  $\Omega$ .

Unless ultrafast laser pulses are involved, the transient term is negligible. The ERT for the steady state can be simplified as

$$\frac{\partial I_\lambda(\zeta_\lambda, \Omega)}{\partial \zeta_\lambda} + I_\lambda(\zeta_\lambda, \Omega) = (1 - \eta_\lambda) I_{b,\lambda} + \frac{\eta_\lambda}{4\pi} \int_{4\pi} I_\lambda(\zeta_\lambda, \Omega') \Phi_\lambda(\Omega', \Omega) d\Omega' \quad (2.53)$$

where  $\zeta_\lambda = \int_0^\xi (a_\lambda + \sigma_\lambda) d\xi$  is the *optical path length*, and  $\eta_\lambda = \sigma_\lambda / (a_\lambda + \sigma_\lambda)$  is called the *scattering albedo*. This is an integro-differential equation, and its right-hand side is called the source function. The integration of the spectral intensity over all wavelengths and all directions gives the radiative heat flux. Unless the temperature field is prescribed, Eq. (2.53) is coupled with the heat conduction equation in a macroscopically stationary medium and the energy conservation equation in a fluid with convection.

Analytical solutions of the ERT rarely exist for applications with multidimensional and nonhomogeneous media. Approximate models have been developed to deal with special types of problems, including Hotel's *zonal method*, the *differential and moment methods* (often using the spherical harmonics approximation), and the *discrete ordinates method*. The statistical model using the Monte Carlo method is often used for complicated geometries and radiative properties.<sup>17</sup> Analytical solutions can be obtained only for limited simple cases.

**Example 2-7.** A gray, isothermal gas at a temperature  $T_g = 3000$  K occupies the space between two, large parallel blackbody surfaces. Surface 1 is heated to a temperature  $T_1 = 1000$  K, while surface 2 is maintained at a relatively low temperature by water cooling. It is desired to know the amount of heat that must be removed from surface 2. If the scattering is negligible, calculate the heat flux at surface 2 for  $a_\lambda L = 0.01, 0.1, 1, \text{ and } 10$ , where  $L$  is the distance between the two surfaces.

**Solution.** For a gray medium without scattering, Eq. (2.53) becomes  $(1/a_\lambda) dI/d\xi + I(\xi, \theta) = I_b(T_g)$ , where  $\theta$  is the angle between  $\xi$  and  $x$ . With  $I_b(T_g) = \sigma_{SB} T_g^4 / \pi$  and  $I(0) = I_b(T_1) = \sigma_{SB} T_1^4 / \pi$ , the ERT can be integrated from  $x = 0$  to  $x = L$ . The result is  $I(L, \theta) = (\sigma_{SB} / \pi) T_1^4 \exp(-a_\lambda L / \cos \theta) + (\sigma_{SB} / \pi) T_g^4 [1 - \exp(-a_\lambda L / \cos \theta)]$ . The radiative flux at  $x = L$  can be obtained by integrating the intensity over the hemisphere, i.e.,

$$\begin{aligned} q''(a_\lambda L) &= \int_0^{2\pi} \int_0^{\pi/2} \frac{\sigma_{SB}}{\pi} [T_g^4 - (T_g^4 - T_1^4) e^{-a_\lambda L / \cos \theta}] \cos \theta \sin \theta d\theta d\phi \\ &= \sigma_{SB} T_g^4 - 2\sigma_{SB} (T_g^4 - T_1^4) E_3(a_\lambda L) \end{aligned}$$

where  $E_3(\zeta) = \int_0^1 e^{-\zeta/\mu} d\mu$  is called the *third exponential integral* and can be numerically evaluated. The final results are tabulated as follows:

$a_\lambda L$	0.01	0.1	1	10
$E_3(a_\lambda L)$	0.49	0.416	0.11	$3.48 \times 10^{-6}$
$q''$ (W/m <sup>2</sup> )	$1.474 \times 10^5$	$8.187 \times 10^5$	$3.595 \times 10^6$	$4.593 \times 10^6$

**Discussion.** In the optically thick limit ( $a_\lambda L \gg 1$ ),  $q'' \approx \sigma_{\text{SB}} T_g^4$ , and all radiation leaving surface 1 will be absorbed by the gas before reaching surface 2. On the other hand, the heat flux is much greater than  $\sigma_{\text{SB}} T_1^4 = 56.7 \text{ kW/m}^2$  at  $a_\lambda L = 0.01$ . The gas absorption can be neglected in the optically thin limit; however, its emission contributes significantly to the radiative flux at surface 2. This is because the gas temperature is much higher than that of surface 1 and  $L/\cos \theta$  can be much longer than  $L$  for large  $\theta$  values.

## 2.5 SUMMARY

---

This chapter provided an overview of classical or equilibrium thermodynamics, derived following logical steps and on a general basis, as well as the functional relations and thermodynamic properties of simple systems and ideal pure substances. Built upon the foundations of thermodynamics, the basic heat transfer modes were elaborated in a coherent way. Entropy generation is inevitably associated with any heat transfer process. The connection between heat transfer and entropy generation, which has been omitted by most heat transfer textbooks, was also discussed. The introduction of thermal radiation not only covered most of the undergraduate-level materials but also presented some basic graduate-level materials. This chapter should serve as a bridge or a reference to the rest of the book, dealing with energy transfer processes in micro/nanosystems and/or from a microscopic viewpoint of macroscopic phenomena.

## REFERENCES

---

1. H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed., Wiley, New York, 1985.
2. G. N. Hatsopoulos and J. H. Keenan, *Principles of General Thermodynamics*, Wiley, New York, 1965; J. H. Keenan, *Thermodynamics*, Wiley, New York, 1941.
3. E. P. Gyftopoulos and G. P. Beretta, *Thermodynamics: Foundations and Applications*, Macmillan, New York, 1991; Also see the augmented edition, Dover Publications, New York, 2005.
4. R. E. Sonntag, C. Borgnakke, and G. J. van Wylen, *Fundamentals of Thermodynamics*, 5th ed., Wiley, New York, 1998.
5. M. J. Moran and H. N. Shapiro, *Fundamentals of Engineering Thermodynamics*, 4th ed., Wiley, New York, 2000.
6. A. Bejan, *Advanced Engineering Thermodynamics*, 2nd ed., Wiley, New York, 1997.
7. J. Kestin (ed.), *The Second Law of Thermodynamics*, Dowden, Hutchinson & Ross, Inc., Stroudsburg, PA, 1976.
8. H. Preston-Thomas, "The International Temperature Scale of 1990 (ITS-90)," *Metrologia*, **27**, 3–10, 1990.
9. Z. M. Zhang, "Surface temperature measurement using optical techniques," *Annu. Rev. Heat Transfer*, **11**, 351–411, 2000.
10. M. Kaviany, *Principles of Heat Transfer*, Wiley, New York, 2002.
11. F. P. Incropera and D. P. DeWitt, *Fundamentals of Heat and Mass Transfer*, 5th ed., Wiley, New York, 2002.
12. M. N. Özışik, *Heat Conduction*, 2nd ed., Wiley, New York, 1993.

13. Y. S. Touloukian and C. Y. Ho (eds.), *Thermophysical Properties of Matter—The TPRC Data Series* (13 volumes compilation of data on thermal conductivity, specific heat, linear expansion coefficient, thermal diffusivity, and radiative properties), Plenum Press, New York, 1970–1977.
14. A. Bejan, *Entropy Generation Minimization*, CRC Press, Boca Raton, FL, 1996.
15. R. F. Barron, *Cryogenic Heat Transfer*, Taylor & Francis, Philadelphia, PA, 1999.
16. M. C. Potter and D. C. Wiggert, *Mechanics of Fluids*, Prentice Hall, New Jersey, 1991.
17. R. Siegel and J. R. Howell, *Thermal Radiation Heat Transfer*, 4th ed., Taylor & Francis, New York, 2002.

## PROBLEMS

---

- 2.1. Give examples of steady state. Give examples of thermodynamic equilibrium state. Give an example of spontaneous process. Is the growth of a plant a spontaneous process? Give an example of adiabatic process.
- 2.2. What is work? Describe an experiment that can measure the amount of work. What is heat? Describe an apparatus that can be used to measure heat. Are work and heat properties of a system?
- 2.3. Expand Eq. (2.1) and Eq. (2.2) in terms of the rate of energy and entropy change of an open system, which is subjected to work output, heat interactions, and multiple inlets and outlets of steady flow.
- 2.4. Discuss the remarks of Rudolf Clausius in 1867:
  - (a) The energy of the universe is constant.
  - (b) The entropy of the universe strives to attain a maximum value.
- 2.5. For a cyclic device experiencing heat interactions with reservoirs at  $T_1, T_2, \dots$ , the Clausius inequality can be expressed as  $\sum_i \delta Q_i/T_i \leq 0$  or  $\oint \delta Q/T \leq 0$ , regardless of whether the device produces or consumes work. Note that  $\delta Q$  is positive when heat is received by the device. Prove the Clausius inequality by applying the second law to a closed system.
- 2.6. In the stable-equilibrium states, the energy and the entropy of a solid are related by  $E = 3 \times 10^5 \exp[(S - S_0)/1000]$ , where  $E$  is in J,  $S$  is in J/K, and  $S_0$  is the entropy of the solid at a reference temperature of 300 K. Plot this relation in an  $E$ - $S$  graph. Find expressions for  $E$  and  $S$  in terms of its temperature  $T$  and  $S_0$ .
- 2.7. For an isolated system, give the mathematical expressions of the first and second laws of thermodynamics. Give graphic illustrations using  $E$ - $S$  graph.
- 2.8. Place two identical metal blocks A and B, initially at different temperatures, in contact with each other but without interactions with any other systems. A thermal equilibrium is reached quickly. System C represents the combined system of both A and B.
  - (a) Is the process reversible or not? Which system has experienced a spontaneous change of state? Which systems have experienced an induced change of state?
  - (b) Assume that the specific heat of the metal is independent of temperature,  $c_p = 240 \text{ J/(kg} \cdot \text{K)}$ , the initial temperatures are  $T_{A1} = 800 \text{ K}$  and  $T_{B1} = 200 \text{ K}$ , and the mass of each block is 5 kg. What is the final temperature? What is the total entropy generation in this process?
  - (c) Show the initial and final states of systems A, B, and C in a  $u$ - $s$  diagram, and indicate which state is not an equilibrium state. Determine the adiabatic availability of system C in the initial state.
- 2.9. Two blocks made of the same material with the same mass are allowed to interact with each other but isolated from the surroundings. Initially, block A is at 800 K and block B at 200 K. Assuming that the specific heat is independent of temperature, show that the final equilibrium temperature is 500 K. Determine the maximum and minimum entropies that may be transferred from block A to block B.
- 2.10. A cyclic machine receives 325 kJ heat from a 1000 K reservoir and rejects 125 kJ heat to a 400 K reservoir in a cycle that produces 200 kJ work. Is this cycle reversible, irreversible, or impossible?
- 2.11. If  $z = z(x, y)$ , then  $dz = f dx + g dy$ , where  $f(x, y) = \partial z/\partial x$ ,  $g(x, y) = \partial z/\partial y$ . Therefore,

$$\frac{\partial f}{\partial y} = \frac{\partial^2 z}{\partial y \partial x} = \frac{\partial^2 z}{\partial x \partial y} = \frac{\partial g}{\partial x}$$

The second-order derivatives of the fundamental equation and each of the characteristic function yield a Maxwell relation. Maxwell's relations are very useful for evaluating the properties of a system in the stable-equilibrium states. For a closed system without chemical reactions, we have  $dN_i = 0$ . Show that

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial P}{\partial S}\right)_V, \left(\frac{\partial T}{\partial P}\right)_S = \left(\frac{\partial V}{\partial S}\right)_P, \left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V, \text{ and } \left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P$$

**2.12.** The *isobaric volume expansion coefficient* is defined as  $\beta_p = (1/v)(\partial v/\partial T)_p$ , the isothermal compressibility is  $\kappa_T = -(1/v)(\partial v/\partial P)_T$ , and the *speed of sound* is  $v_a = \sqrt{(\partial P/\partial \rho)_s}$ . For an ideal gas, show that  $\beta_p = 1/T$ ,  $\kappa_T = 1/P$ , and  $v_a = \sqrt{\gamma RT}$ .

**2.13.** For a system with single type of constituents, the fundamental relation obtained by experiments gives  $S = \alpha(NVU)^{1/3}$ , where  $\alpha$  is a positive constant, and  $N$ ,  $V$ ,  $S$ , and  $U$  are the number of molecules, the volume, the entropy, and the internal energy of the system, respectively. Obtain expressions of the temperature and the pressure in terms of  $N$ ,  $V$ ,  $U$ , and  $\alpha$ . Show that  $S = 0$  at zero temperature for constant  $N$  and  $V$ .

**2.14.** For blackbody radiation in an evacuated enclosure of uniform wall temperature  $T$ , the energy density can be expressed as  $u_v = U/V = (4c)\sigma_{SB}T^4$ , where  $U$  is the internal energy,  $V$  the volume,  $c$  the speed of light, and  $\sigma_{SB}$  the Stefan-Boltzmann constant. Determine the entropy  $S(T, V)$  and the pressure  $P(T, V)$ , which is called the *radiation pressure*. Show that the radiation pressure is a function of temperature only and negligibly small at moderate temperatures. Hint:

$$S = \int_0^T \frac{1}{T} \left(\frac{\partial U}{\partial T}\right)_V dT \quad \text{and} \quad P = T \left(\frac{\partial S}{\partial V}\right)_T - \left(\frac{\partial U}{\partial V}\right)_T$$

**2.15.** A cyclic machine can only interact with two reservoirs at temperatures  $T_A = 298$  K and  $T_B = 77.3$  K, respectively.

- (a) If heat is extracted from reservoir A at a rate of  $\dot{Q} = 1000$  W, what is the maximum rate of work that can be generated ( $\dot{W}_{\max}$ )?
- (b) If no work is produced, what is the rate of entropy generation ( $\dot{S}_{\text{gen}}$ ) of the cyclic machine?
- (c) Plot  $\dot{S}_{\text{gen}}$  versus  $\dot{W}$  (the power produced).

**2.16.** An engineer claimed that it requires much more work to remove 0.1 J of heat from a cryogenic chamber at an absolute temperature of 0.1 K than to remove 270 J of heat from a refrigerator at 270 K. Assuming that the environment is at 300 K, justify this claim by calculating the minimum work required for each refrigeration task.

**2.17.** A solid block [ $m = 10$  kg and  $c_p = 0.5$  kJ/(kg · K)], initially at room temperature ( $T_{A,1} = 300$  K) is cooled with a large tank of liquid-gas mixture of nitrogen at  $T_B = 77.3$  K and atmospheric pressure.

- (a) After the block reaches the liquid-nitrogen temperature, what is the total entropy generation ( $S_{\text{gen}}$ )?
- (b) Given the specific enthalpy of evaporation of nitrogen,  $h_{\text{fg}} = 198.8$  kJ/kg, what must be its specific entropy of evaporation  $s_{\text{fg}}$  in kJ/(kg · K), in order for the nitrogen tank to be modeled as a reservoir? Does  $h_{\text{fg}} = T_{\text{sat}} \times s_{\text{fg}}$  always hold?

**2.18.** Two same-size solid blocks of the same material are isolated from other systems [specific heat  $c_p = 2$  kJ/(kg · K); mass  $m = 5$  kg]. Initially block A is at a temperature  $T_{A,1} = 300$  K and block B at  $T_{B,1} = 1000$  K.

- (a) If the two blocks are put together, what will be the equilibrium temperature ( $T_2$ ) and how much entropy will be generated ( $S_{\text{gen}}$ )?
- (b) If the two blocks are connected with a cyclic machine, what is the maximum work that can be obtained ( $W_{\max}$ )? What would be the final temperature of the blocks ( $T_3$ ) if the maximum work were obtained?

**2.19.** A rock [density  $\rho = 2800$  kg/m<sup>3</sup> and specific heat  $c_p = 900$  J/(kg · K)] of 0.8 m<sup>3</sup> is heated to 500 K using solar energy. A heat engine (cyclic machine) receives heat from the rock and rejects heat to the ambient at 290 K. The rock therefore cools down.

- (a) Find the maximum energy (heat) that the rock can give out.
- (b) Find the maximum work that can be done by the heat engine,  $W_{\max}$ .
- (c) In an actual process, the final temperature of the rock is 330 K and the work output from the engine is only half of  $W_{\max}$ . Determine the entropy generation of the actual process.

**2.20.** Consider three identical solid blocks with a mass of 5 kg each, initially at 300, 600, and 900 K, respectively. The specific heat of the material is  $c_p = 2000 \text{ J}/(\text{kg} \cdot \text{K})$ . A cyclic machine is available that can interact only with the three blocks.

- What is the maximum work that can be produced? What are the final temperatures of each block? Is the final state in equilibrium?
- If no work is produced, i.e., simply putting the three blocks together, what will be the maximum entropy generation? What will be the final temperature?
- If the three blocks are allowed to interact via cyclic machine but not with any other systems in the environment, what is the highest temperature that can be reached by one of the blocks?
- If the three blocks are allowed to interact via cyclic machine but not with any other systems in the environment, what is the lowest temperature that can be reached by one of the blocks?

**2.21.** Electrical power is used to raise the temperature of a 500 kg rock from 25 to 500°C. The specific heat of the rock material is  $c_p = 0.85 \text{ kJ}/(\text{kg} \cdot \text{K})$ .

- If the rock is heated directly through resistive (Joule) heating, how much electrical energy is needed? Is this process reversible? If not, how much entropy is generated in this process?
- By using cyclic devices that can interact with both the rock and the environment at 25°C, what is the minimum electrical energy required?

**2.22.** An insulated cylinder of  $2 \text{ m}^3$  is divided into two parts of equal volume by an initially locked piston. Side A contains air at 300 K and 200 kPa; side B contains air at 1500 K and 1 MPa. The piston is now unlocked so that it is free to move and it conducts heat. An equilibrium state is reached between the two sides after a while.

- Find the masses in both A and B.
- Find the final temperatures, pressures, and volumes for both A and B.
- Find the entropy generation in this process.

**2.23.** A piston-cylinder contains 0.56 kg of  $\text{N}_2$  gas, initially at 600 K. A cyclic machine receives heat from the cylinder and releases heat to the environment at 300 K. Assume that the specific heat of  $\text{N}_2$  is  $c_p = 1.06 \text{ kJ}/(\text{kg} \cdot \text{K})$  and the pressure inside the cylinder is maintained at 100 kPa by the environment. What is the maximum work that can be produced by the machine? What is the thermal efficiency (defined as the ratio of the work output to the heat received)? The thermodynamic efficiency can be defined as the ratio of the actual work produced to the maximum work. Plot the thermodynamic efficiency as a function of the entropy generation. What is the maximum entropy generation?

**2.24.** An air stream [ $c_p = 1 \text{ kJ}/(\text{kg} \cdot \text{K})$  and  $M = 29.1 \text{ kg}/\text{kmol}$ ] flows through a power plant. The stream enters a turbine at  $T_1 = 750 \text{ K}$  and  $P_1 = 6 \text{ MPa}$ , and exits at  $P_2 = 1.2 \text{ MPa}$  into a recovery unit, which can exchange heat with the environment at 25°C and 100 kPa. The stream then exits the recovery unit to the environment. The turbine is thermally insulated and has an efficiency  $\eta_t = 0.85$ .

- Find the power per unit mass flow rate produced by the turbine.
- Calculate the entropy generation rate in the turbine.
- Determine the largest power that can be produced by the recovery unit.

**2.25.** Water flows in a perfectly insulated, steady state, horizontal duct of variable cross-sectional area. Measurements were taken at two ports and the data were recorded in a notebook as follows. For port 1, speed  $\xi_1 = 3 \text{ m/s}$ , pressure  $P_1 = 50 \text{ kPa}$ , and temperature  $T_1 = 40^\circ\text{C}$ ; for port 2,  $\xi_2 = 5 \text{ m/s}$  and  $P_2 = 45 \text{ kPa}$ . Some information was accidentally left out by the student taking the notes. Can you determine  $T_2$  and the direction of the flow based on the available information? Hint: Model the water as an ideal incompressible liquid with  $c_p = 4.2 \text{ kJ}/(\text{kg} \cdot \text{K})$  and specific volume  $v = 10^{-3} \text{ m}^3/\text{kg}$ .

**2.26.** An insulated rigid vessel contains 0.4 kmol of oxygen at 200 kPa separated by a membrane from 0.6 kmol of carbon dioxide at 400 kPa; both sides are initially at 300 K. The membrane is suddenly broken and, after a while, the mixture comes to a uniform state (equilibrium).

- Find the final temperature and pressure of the mixture.
- Determine the entropy generation due to irreversibility.

**2.27.** Pure  $\text{N}_2$  and air (21%  $\text{O}_2$  and 79%  $\text{N}_2$  by volume), both at 298 K and 120 kPa, enter a chamber at a flow rate of 0.1 and 0.3 kmol/s, respectively. The new mixture leaves the chamber at the same temperature and pressure as the incoming streams.

- What are the mole fractions and the mass fractions of  $\text{N}_2$  and  $\text{O}_2$  at the exit?
- Find the enthalpy change in the mixing process. Find the entropy generation rate of the mixing process.
- Consider a process in which the flow directions are reversed. The chamber now contains necessary devices for the separation, and it may transfer heat to the environment at 298 K. What is the minimum amount of work per unit time needed to operate the separation devices?

**2.28.** A Carnot engine receives energy from a reservoir at  $T_H$  and rejects heat to the environment at  $T_0$  via a heat exchanger. The engine works reversibly between  $T_H$  and  $T_L$ , where  $T_L$  is the temperature of the higher-temperature side of the heat exchanger. The *product* of the area and the heat transfer coefficient of the heat exchanger is  $\alpha$ . Therefore, the heat that must be rejected to the environment through the heat exchanger is  $\dot{Q}_L = \alpha(T_L - T_0)$ . Given  $T_H = 800$  K,  $T_0 = 300$  K, and  $\alpha = 2300$  W/K. Determine the value of  $T_L$  so that the heat engine will produce maximum work, and calculate the power production and the entropy generation in such a case.

**2.29.** To measure the thermal conductivity, a thin-film electric heater is sandwiched between two plates whose sides are well insulated. Each plate has an area of  $0.1$  m<sup>2</sup> and a thickness of  $0.05$  m. The outside of the plates are exposed to air at  $T_\infty = 25^\circ\text{C}$  with a convection coefficient of  $h = 40$  W/(m<sup>2</sup> · K). The electric power of the heat is  $400$  W and a thermocouple inserted between the two plates measures a temperature of  $T_j = 175^\circ\text{C}$  at steady state. Determine the thermal conductivity of the plate material. Find the total entropy generation rate. Comment on the fraction of entropy generation due to conduction and convection.

**2.30.** An electric current,  $I = 2$  A, passes through a resistive wire of diameter  $D = 3$  mm with a resistivity  $r_e = 1.5 \times 10^{-4}$   $\Omega \cdot \text{m}$ . The cable is placed in ambient air at  $27^\circ\text{C}$  with a convection coefficient  $h = 20$  W/(m<sup>2</sup> · K). Assume a steady state has been reached and neglect radiation. Determine the radial temperature distribution inside the wire. Determine the volumetric entropy generation rate  $\dot{s}_{\text{gen}}$  as a function of radius. Determine the total entropy generation rate per unit length of the cable. Hint: For steady-state conduction,  $\dot{s}_{\text{gen}} = (1/T)\nabla \cdot \mathbf{q}'' - (1/T^2)(\mathbf{q}'' \cdot \nabla T)$ .

**2.31.** Find the thermal conductivity of intrinsic (undoped) silicon, heavily doped silicon, quartz, glass, diamond, graphite, and carbon from 100 to 1000 K from Touloukian and Ho.<sup>13</sup> Discuss the variations between different materials, crystalline structures, and doping concentrations.

**2.32.** Find the thermal conductivity of copper from 1 to 1000 K from Touloukian and Ho.<sup>13</sup> Discuss the general trend in terms of temperature dependence, and comment on the effect of impurities.

**2.33.** For laminar flow over a flat plate, the velocity and thermal boundary layer thicknesses can be calculated by  $\delta(x) = 5xRe_x^{-1/2}$  and  $\delta_t(x) = 5xRe_x^{-1/2}Pr^{-1/3}$ , respectively. Use room temperature data to calculate and plot the boundary layer thicknesses for air, water, engine oil, and mercury for different values of  $U_\infty$ . Discuss the main features. Hint: Property data can be found from Incropera and DeWitt.<sup>11</sup>

**2.34.** Air at  $14^\circ\text{C}$  and atmospheric pressure is in parallel flow over a flat plate of  $2 \times 2$  m<sup>2</sup>. The air velocity is  $3$  m/s and the surface is maintained at  $140^\circ\text{C}$ . Determine the average convection coefficient and the rate of heat transfer from the plate to air. (For air at  $350$  K, which is the average temperature between the surface and fluid,  $\kappa = 0.03$  W/(m · K),  $\nu = 20.9 \times 10^{-6}$  m<sup>2</sup>/s, and  $Pr = 0.7$ .)

**2.35.** Plot the blackbody intensity (Planck's law) as a function of wavelength for several temperatures. Discuss the main features of this function. Show that in the long-wavelength limit, the blackbody function can be approximated by  $e_{b,\lambda}(\lambda, T) \approx \pi ck_B T / \lambda^4$ , which is the Rayleigh-Jeans formula.

**2.36.** Calculate the net radiative heat flux from the human body at a surface temperature of  $T_s = 308$  K, with an emissivity  $\varepsilon = 0.9$ , to the room walls at  $298$  K. Assume air at  $298$  K has a natural convection coefficient of  $5$  W/(m<sup>2</sup> · K). Neglect evaporation, calculate the natural convection heat flux from the person to air. Comment on the significance of thermal radiation.

**2.37.** A combustion fired in a spherical enclosure of diameter  $D = 50$  cm with a constant wall temperature of  $600$  K. The temperature of the combustion gas may be approximated as uniform at  $2300$  K. The absorption coefficient of the gas  $a_\lambda = 0.01$  cm<sup>-1</sup>, which is independent of wavelength. Assuming that the wall is black and neglecting the scattering effect, determine the net heat transfer rate between the gas and the inner wall of the sphere.



*This page intentionally left blank*

---

# CHAPTER 3

---

# ELEMENTS OF STATISTICAL THERMODYNAMICS AND QUANTUM THEORY

---

Classical statistical mechanics is based on the assumption that all matters are composed of a myriad of small discrete particles, such as molecules and atoms, in any given macroscopic volume.<sup>1-5</sup> There are about  $N = 2.5 \times 10^{16}$  molecules per cubic millimeter of air at standard conditions (25°C and 1 atm). These particles are in continuous random motion, which generally obeys the laws of classical mechanics. A complete microscopic description of a system requires the identification of the position  $\mathbf{r}_i(t)$  and velocity  $\mathbf{v}_i(t)$  of each particle (here, subscript  $i$  indicates the  $i$ th particle) at any time. For a simple system of  $N$  molecules in a box of volume  $V$ , one can write Newton's law of motion for each molecule as

$$\sum_j \mathbf{F}_{ij}(\mathbf{r}_i, \mathbf{r}_j, t) = m_i \frac{d\mathbf{v}_i}{dt}, i = 1, 2, \dots, N \quad (3.1)$$

where  $\mathbf{F}_{ij}$  is the intermolecular force that the  $j$ th molecule exerts on the  $i$ th molecule, and  $m_i$  is the mass of the  $i$ th molecule. The initial position and velocity, as well as the nature of collisions among particles and that between particles and the walls of the box, must be specified in order to solve the  $N$  equations. Although this approach is straightforward, there are two major barriers. First, the intermolecular forces or potentials are often complicated and difficult to determine. Second, the solution of Eq. (3.1) requires significant computer resources even for rather simple problems. Statistical methods are often used instead to obtain microscopic descriptions that are related to macroscopic behaviors. *Statistical mechanics* aims at finding the equilibrium distribution of certain types of particles in the velocity space. It provides a linkage between macroscopic thermodynamic properties and the microscopic behavior and a means to evaluate some thermodynamic properties. *Kinetic theory*, on the other hand, deals with nonequilibrium processes. It gives a microscopic description of transport phenomena and helps predict some important transport properties, as will be seen in Chap. 4.

Along with the rapid development in computing speed and memory, *molecular dynamics* (MD) simulation has become a powerful tool for the investigation of phenomena occurring in nanostructures and/or at very short time scales. In the MD method, the location and the velocity of every particle are calculated at each time step by applying Eq. (3.1) with a suitable potential function.<sup>6,7</sup> Thermodynamic properties are then evaluated using statistical mechanics formulation. Further discussion about the application of MD simulation to predict the thermal properties of nanostructures will be given in Chap. 7.

This chapter starts with a statistical model of independent particles and a brief introduction to the basic principles of quantum mechanics. The necessary mathematical background is summarized in Appendix B. It is highly recommended that one review the materials covered in the appendix before studying this chapter. The three important distributions are derived

based on the statistics for different types of particles. The microscopic descriptions and results are then linked to macroscopic quantities and the laws of thermodynamics. The application to ideal gases is presented at the end of this chapter, while discussions of blackbody radiation, lattice vibrations, and free electron gas will be deferred to later chapters.

### 3.1 STATISTICAL MECHANICS OF INDEPENDENT PARTICLES

We say particles are independent when their energies are independent of each other and the total energy is the sum of the energies of individual particles. Consider a system that has  $N$  independent particles of the same type confined in a volume  $V$ . The total internal energy of the system is  $U$ , which is the sum of the energies of all particles. Particles may have different energies and can be grouped according to their energies. It is of interest to know how many particles are there within certain energy intervals. We can subdivide energy into a large number of discretized energy levels. As illustrated in Fig. 3.1, there are  $N_i$  particles on the  $i$ th energy level, each with energy exactly equal to  $\epsilon_i$ .

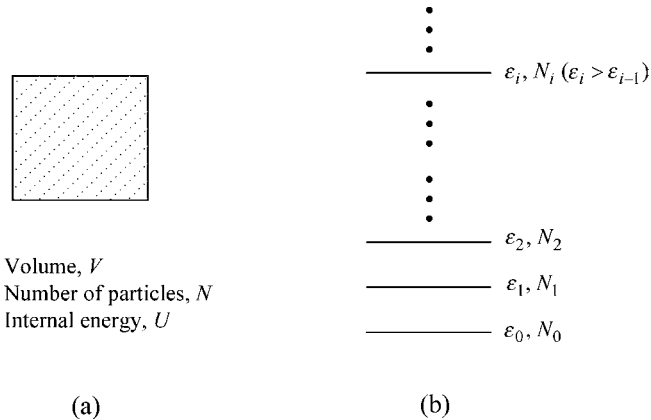


FIGURE 3.1 Illustration of (a) a simple system of independent particles and (b) energy levels.

From the classical mechanics point of view, it appears that the increment between adjacent energy levels can be indefinitely small. The particles are distinguishable, and there is no limit on the number of particles on each energy level. Quantum mechanics predicts that the energy levels are indeed discretized with finite increments between adjacent energy levels, and the particles are unidentifiable (*indistinguishable*). An introduction to the basic principles of quantum mechanics is given in Sec. 3.1.3 and a more detailed introduction of the quantum theory is given near the end of this chapter. The conservation equations for the system shown in Fig. 3.1 are

$$\sum_{i=0}^{\infty} N_i = N \quad (3.2)$$

and

$$\sum_{i=0}^{\infty} \epsilon_i N_i = U \quad (3.3)$$

### 3.1.1 Macrostates versus Microstates

The thermodynamic state may be viewed in terms of the gross behavior that ignores any differences at the molecular or atomic level, or in terms of the individual particles. A *macrostate* is determined by the values of  $N_0, N_1, N_2, \dots$  for a given volume (which somehow confines the quantized energy levels) though two different macrostates can have the same energy. Each macrostate may be made up of a number of microscopic arrangements; each microscopic arrangement is called a *microstate*. In statistical mechanics, all microstates are assumed *equally probable*. There may be a large number of microstates that correspond to the same macrostate. The number of microstates for each macrostate is termed the *thermodynamic probability*  $\Omega$  of that macrostate. Unlike the stochastic probability that lies between 0 and 1, the thermodynamic probability  $\Omega$  is usually a very large number. One of the principles underlying statistical mechanics is that the stable-equilibrium state corresponds to the *most probable macrostate*. Therefore, for given values of  $U, N$ , and  $V$ , the thermodynamic probability is the largest in the stable-equilibrium state. We will use the following example to illustrate the concepts of microstate and macrostate.

**Example 3-1.** There are four distinguishable particles in a confined space, and there are two energy levels. How many macrostates are there? How many microstates are there for the macrostate with two particles on each energy level?

**Solution.** There are five macrostates in total with  $(N_1, N_2) = (0, 4), (1, 3), (2, 2), (3, 1),$  and  $(4, 0)$ , respectively. Because the particles are distinguishable, the microstates will be different only if the particles from different energy levels are interchanged. Using the combination theory, we can figure out that  $\Omega(N_1, N_2) = N!/(N_1! N_2!) = 4!/(2!2!) = 6$ , i.e., there are six microstates for the macrostate with two particles on each energy level. It can be shown that this is also the most probable macrostate.

### 3.1.2 Phase Space

The *phase space* is a six-dimensional space formed by three coordinates for the position  $\mathbf{r}$  and three coordinates for the momentum  $\mathbf{p} = m\mathbf{v}$  or velocity  $\mathbf{v}$ . Each point in the phase space defines the exact location and momentum of an individual particle. If both the space and the momentum are described with the Cartesian system, then a volume element in the phase space is  $dx dy dz dp_x dp_y dp_z$ . Figure 3.2 shows a phase space projected to the  $x$ - $p_x$  plane.

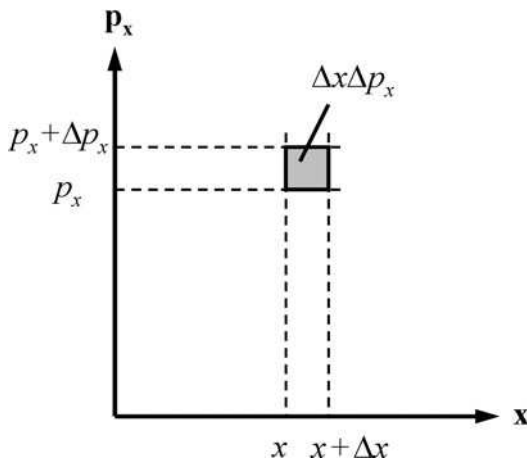


FIGURE 3.2 Phase space projected to the  $x$ - $p_x$  plane, where  $\Delta x \Delta p_x$  is an area element.

The three coordinates  $(p_x, p_y, p_z)$  form a *momentum space*. One may choose to use  $(v_x, v_y, v_z)$  to form a *velocity space*. If the momentum space is described in spherical coordinates, the volume element is  $dp_x dp_y dp_z = p^2 \sin \theta dp d\theta d\phi$ . The volume contained in a spherical shell from  $p$  to  $p + dp$  is  $4\pi p^2 dp$ . Figure 3.3 illustrates the momentum space projected to the  $p_x$ - $p_y$  plane, with a spherical shell.

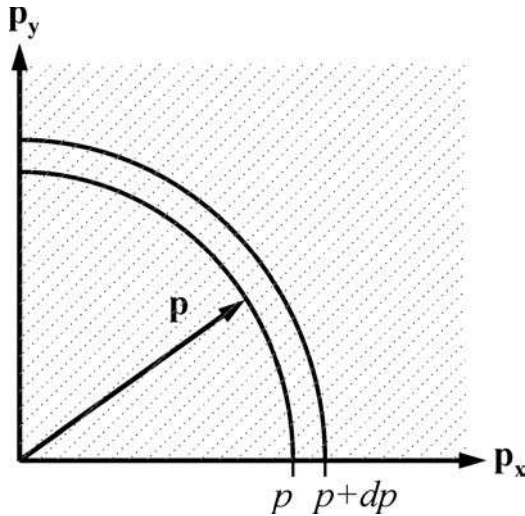


FIGURE 3.3 The  $p_x$ - $p_y$  plane of the momentum space, showing a spherical shell.

### 3.1.3 Quantum Mechanics Considerations

The origin of quantum theory can be traced back to about 100 years ago when Planck first used a discrete set of energies to describe the electromagnetic radiation, and thus obtained Planck's distribution (details to be presented in Sec. 8.1). For any given frequency of radiation  $\nu$ , the smallest energy increment is given by  $h\nu$ , where  $h = 6.626 \times 10^{-34} \text{ J} \cdot \text{s}$  is called Planck's constant. Radiation can be alternatively viewed as electromagnetic waves or traveling energy quanta. The corpuscular theory treats radiation as a collection of energy quanta, called *photons*. The energy of a photon is given by

$$\varepsilon = h\nu \quad (3.4)$$

From the wave theory, the speed of light  $c$  is related to the wavelength  $\lambda$  and the frequency by

$$c = \lambda\nu \quad (3.5)$$

In a medium with a refractive index of  $n$ ,  $c = c_0/n$  and  $\lambda = \lambda_0/n$ , where subscript 0 is used to indicate quantities in vacuum with  $n = 1$ . The speed of light in vacuum is  $c_0 = 299,792,458 \text{ m/s}$ , which is a defined quantity as given in Appendix A. Note that frequency does not change from one medium to another. Based on the relativistic theory, the rest energy  $E_0$  of a particle with mass  $m$  is

$$E_0 = mc^2 \quad (3.6)$$

The momentum of the particle traveling with speed  $v$  is  $p = mv$ . Since the energy of a photon is  $h\nu$  and its speed is  $c$ , the momentum of a (massless) photon is (see Sec. 3.7)

$$p = \frac{h\nu}{c} = \frac{h}{\lambda} \quad (3.7)$$

Another hypothesis of quantum theory is that the motion of matter may be wavelike, with characteristic wavelength and frequency. Therefore, for a particle moving with velocity  $v \ll c$ ,

$$\lambda_{\text{DB}} = \frac{h}{p} = \frac{h}{mv} \quad \text{and} \quad \nu_{\text{DB}} = \frac{mc^2}{h} \quad (3.8)$$

which are called *de Broglie wavelength* and *de Broglie frequency*, respectively. In 1923, Louis de Broglie postulated that matter may also possess wave characteristics and thereafter resolved the controversy as per the nature of radiation. Note that the phase speed of the wave defined by Eq. (3.8) is  $c^2/v$ , which is greater than the speed of light. The discovery of electron diffraction confirmed de Broglie's hypothesis. For this prediction, de Broglie received the Nobel Prize in physics in 1929. Seven years later, the 1937 Nobel Prize in physics was shared by Clinton J. Davisson and George P. Thomson for their independent experiments that demonstrated diffraction of electrons by crystals.

**Example 3-2.** Calculate the frequency in Hz and photon energy in eV of an ultraviolet (UV) laser beam at a wavelength of  $\lambda = 248$  nm and a microwave at  $\lambda = 10$  cm. Calculate the de Broglie wavelength of an He atom at  $200^\circ\text{C}$ , using the average speed of 1717 m/s, and an electron traveling with a speed of  $10^6$  m/s.

**Solution.** The equations are  $\nu = c/\lambda$  and  $\varepsilon = hc/\lambda$ . Assume the refractive index is 1. For the UV beam at  $\lambda = 248$  nm,  $\nu = 1.2 \times 10^{15}$  Hz and  $\varepsilon = 8.01 \times 10^{-19}$  J = 5 eV. For  $\lambda = 10$  cm,  $\nu = 3 \times 10^9$  Hz = 3 GHz and  $\varepsilon = 2 \times 10^{-24}$  J =  $1.24 \times 10^{-5}$  eV = 124 meV. The mass of an He atom is  $m = M/N_A = 6.64 \times 10^{-27}$  kg. Hence,  $\lambda_{\text{DB}} = h/mv = 5.8 \times 10^{-11}$  m = 58 pm. From Appendix A,  $m_e = 9.11 \times 10^{-31}$  kg, therefore,  $\lambda_{\text{DB}} = 7.3 \times 10^{-10}$  m = 0.73 nm, which is in the x-ray region.

The foundation of quantum mechanics is the Schrödinger equation, which is a partial-differential equation of the time-space dependent complex *probability density function*. More details can be found from Tien and Lienhard,<sup>1</sup> Carey,<sup>5</sup> and Griffiths.<sup>8</sup> The solutions of the Schrödinger equation support the dual nature of wave and matter, and result in discrete quantized energy levels. Furthermore, there are usually more than one distinguishable *quantum state* at each energy level, i.e., the energy levels may be degenerate. The number of quantum states for a given energy level is called the *degeneracy*, denoted by  $g_i$  for the  $i$ th energy level, as shown in Fig. 3.4.

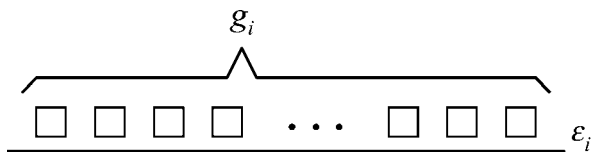


FIGURE 3.4 The degeneracy of the  $i$ th energy level.

The *uncertainty principle* states that the position and momentum of a given particle cannot be measured simultaneously with arbitrary precision. The limit is given by

$$\Delta x \Delta p_x \geq h/4\pi \quad (3.9)$$

This result implies that we cannot locate the exact position of a particle in the phase space; all we can say is that the particle is somewhere in a domain whose volume is around  $h^3$ . The uncertainty principle is one of the cornerstones of quantum mechanics and was formulated in 1927 by Werner Heisenberg, a Nobel Laureate in Physics.

In quantum theory, independent particles of the same type are indistinguishable. For certain particles, such as electrons, each quantum state cannot be occupied by more than one particle. This is the *Pauli exclusion principle*, discovered by Nobel Laureate Wolfgang Pauli in 1925. The result, as we will see, is the Fermi-Dirac statistics that can be used to describe the behavior of free electrons. The collection of free electrons in metals is sometimes called the free electron gas, which exhibits very different characteristics from ideal molecular gases.

### 3.1.4 Equilibrium Distributions for Different Statistics

The characteristics of various types of particles can be described by different statistics. In this section, we will first introduce three statistics and then apply them to obtain the distribution functions, i.e., the number of particles on each energy level. The application of the distribution functions to the study of thermodynamic properties of ideal molecular gases will be discussed later in this chapter. The applications of statistical thermodynamics to blackbody radiation, lattice vibration, free electrons in metals, and electrons and holes in semiconductors will be discussed in subsequent chapters.

- *The Maxwell-Boltzmann (MB) statistics:* Particles are distinguishable and there is no limit for the number of particles on each energy level. From Eq. (B.22) in Appendix B, the thermodynamic probability for the distribution shown in Fig. 3.1b is

$$\Omega = \frac{N!}{N_0!N_1!N_2!\cdots} = \frac{N!}{\prod_{i=0}^{\infty} N_i!}$$

If degeneracy is included as shown in Fig. 3.4, then

$$\Omega_{\text{MB}} = N! \prod_{i=0}^{\infty} \frac{g_i^{N_i}}{N_i!} \quad (3.10)$$

- *The Bose-Einstein (BE) statistics:* Particles are indistinguishable and there is no limit for the number of particles in each quantum state; there are  $g_i$  quantum states on the  $i$ th energy level. From Eq. (B.23), the number of ways of placing  $N_i$  indistinguishable objects to  $g_i$  distinguishable boxes is  $\frac{(g_i + N_i - 1)!}{(g_i - 1)!N_i!}$ . Therefore, the thermodynamic probability for BE statistics is

$$\Omega_{\text{BE}} = \prod_{i=0}^{\infty} \frac{(g_i + N_i - 1)!}{(g_i - 1)!N_i!} \quad (3.11)$$

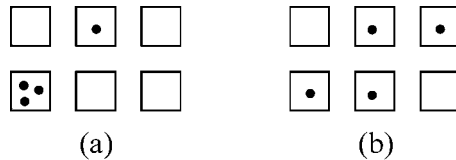
- *The Fermi-Dirac (FD) statistics:* Particles are indistinguishable and the energy levels are degenerate. There are  $g_i$  quantum states on the  $i$ th energy level, and each quantum state can be occupied by no more than one particle. Using Eq. (B.21), we obtain the thermodynamic probability for FD statistics as

$$\Omega_{\text{FD}} = \prod_{i=0}^{\infty} \frac{g_i!}{(g_i - N_i)!N_i!} \quad (3.12)$$

The three statistics are very important for understanding the molecular, electronic, crystalline, and radiative behaviors that are essential for energy transport processes in both small and large scales. MB statistics can be considered as the limiting case of BE or FD statistics. The thermodynamic relations and the velocity distribution of ideal molecular gases can be understood from MB statistics. BE statistics is important for the study of photons, phonons in solids, and atoms at low temperatures. It is the basis of Planck’s law of black-body radiation, the Debye theory for the specific heat of solids, and the Bose-Einstein condensation, which is important for superconductivity, superfluidity, and laser cooling of atoms. FD statistics can be used to model the electron gas and the electron contribution to the specific heat of solids. It is important for understanding the electronic and thermal properties of metals and semiconductors.

**Example 3-3.** Four indistinguishable particles are to be placed in two energy levels, each with a degeneracy of 3. Evaluate the thermodynamic probability of all arrangements, considering BE and FD statistics separately. What are the most probable arrangements?

**Solution.** There are two energy levels,  $g_0 = g_1 = 3$  and the total number of particles  $N = 4$ . The thermodynamic probability is  $\Omega = \Omega_0 \times \Omega_1$ , which depends on  $N_0$  and  $N_1$  ( $N_0 + N_1 = 4$ ). Figure 3.5 shows specific cases of the BE and FD distributions.



**FIGURE 3.5** Illustration of the arrangement for four particles on two energy levels, each with a degeneracy of 3. (a) Bose-Einstein statistics. (b) Fermi-Dirac statistics.

For BE statistics, we have

$$\begin{aligned} \Omega_{BE} &= \frac{(N_0 + g_0 - 1)!}{(g_0 - 1)!N_0!} \times \frac{(N_1 + g_1 - 1)!}{(g_1 - 1)!N_1!} \\ &= \frac{(N_0 + 2)(N_0 + 1)}{2} \times \frac{(6 - N_0)(5 - N_0)}{2} \end{aligned}$$

For FD statistics, we must have  $N_i \leq g_i$ ; therefore,  $1 \leq N_0 \leq 3$ , and

$$\Omega_{FD} = \frac{g_0!}{(g_0 - N_0)!N_0!} \times \frac{g_1!}{(g_1 - N_1)!N_1!} = \frac{6}{(3 - N_0)!N_0!} \times \frac{6}{(N_0 - 1)!(4 - N_0)!}$$

The results are summarized in the following table. Clearly, the most probable arrangement for both statistics in this case is  $N_0 = N_1 = 2$ .

$N_0$	0	1	2	3	4
$N_1$	4	3	2	1	0
$\Omega_{BE}$	15	30	36	30	15
$\Omega_{FD}$	–	3	9	3	–

For a given simple thermodynamics system of volume  $V$ , internal energy  $U$ , and total number of particles  $N$ , we wish to find the state (identified by the distribution  $N_0, N_1, N_2, \dots$ ) that maximizes  $\Omega$  or  $\ln \Omega$ , under the constraints given by Eq. (3.2) and Eq. (3.3), based on



the method of Lagrange multipliers (Appendix B). For MB statistics with degeneracy, from Eq. (3.10),

$$\ln \Omega = \ln N! + \sum_{i=0}^{\infty} N_i \ln g_i - \sum_{i=0}^{\infty} \ln N_i!$$

For a large number of particles, the Stirling formula gives  $\ln N! \approx N \ln N - N$  from Eq. (B.11). The above equation can be approximated as

$$\begin{aligned} \ln \Omega &\approx N \ln N - N + \sum_{i=0}^{\infty} N_i \ln g_i - \sum_{i=0}^{\infty} (N_i \ln N_i - N_i) \\ &= N \ln N - N + \sum_{i=0}^{\infty} N_i \left( \ln \frac{g_i}{N_i} + 1 \right) \end{aligned}$$

Notice that  $N$  and  $g_{i_s}$  are fixed and only  $N_{i_s}$  are variables, therefore,

$$d(\ln \Omega) = \sum_{i=0}^{\infty} \frac{\partial(\ln \Omega)}{\partial N_i} dN_i \approx \sum_{i=0}^{\infty} \left( \ln \frac{g_i}{N_i} + 1 - N_i \frac{1}{N_i} \right) dN_i = \sum_{i=0}^{\infty} \ln \frac{g_i}{N_i} dN_i = 0 \quad (3.13)$$

From the constraint equations, Eq. (3.2) and Eq. (3.3), we have

$$-\alpha \sum_{i=0}^{\infty} dN_i = 0 \quad (3.14a)$$

and

$$-\beta \sum_{i=0}^{\infty} \varepsilon_i dN_i = 0 \quad (3.14b)$$

where  $\alpha$  and  $\beta$  are Lagrangian multipliers and  $\varepsilon_{i_s}$  are treated as constants. Conventionally, negative signs are chosen because  $\alpha$  and  $\beta$  are generally nonnegative for molecular gases. By adding Eq. (3.14a) and Eq. (3.14b) to Eq. (3.13), we obtain

$$\sum_{i=0}^{\infty} \left( \ln \frac{g_i}{N_i} - \alpha - \beta \varepsilon_i \right) dN_i = 0$$

Because  $dN_i$  can be arbitrary, the above equation requires that  $\ln(g_i/N_i) - \alpha - \beta \varepsilon_i = 0$ . Hence,

$$N_i = \frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i}} \quad (3.15a)$$

or

$$\frac{N_i}{N} = \frac{g_i e^{-\alpha} e^{-\beta \varepsilon_i}}{\sum_{i=0}^{\infty} g_i e^{-\alpha} e^{-\beta \varepsilon_i}} \quad (3.15b)$$

This is the MB distribution. The physical meanings of  $\alpha$  and  $\beta$  will be discussed later. Using the same procedure described above, we can obtain the following for BE statistics,

$$N_i = \frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i} - 1} \quad (3.16)$$

which is the BE distribution. For FD statistics, we can obtain the FD distribution as follows

$$N_i = \frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i} + 1} \quad (3.17)$$

The results for all the three statistics are summarized in Table 3.1.

**Example 3-4.** Derive the BE distribution step by step. Under which condition can it be approximated by the MB distribution?

**Solution.** Using the thermodynamic probability of BE statistics in Eq. (3.11), we have

$$\begin{aligned} \ln \Omega &= \sum_{i=0}^{\infty} [\ln(g_i + N_i - 1)! - \ln(g_i - 1)! - \ln N_i!] \\ &\approx \sum_{i=0}^{\infty} [(g_i + N_i - 1) \ln(g_i + N_i - 1) - (g_i + N_i - 1) - \\ &\quad (g_i - 1) \ln(g_i - 1) + (g_i - 1) - N_i \ln N_i + N_i] \\ &= \sum_{i=0}^{\infty} [(g_i + N_i - 1) \ln(g_i + N_i - 1) - (g_i - 1) \ln(g_i - 1) - N_i \ln N_i] \end{aligned}$$

Hence, 
$$\begin{aligned} \frac{\partial \ln \Omega}{\partial N_i} &\approx \ln(g_i + N_i - 1) + (g_i + N_i - 1) \frac{1}{g_i + N_i - 1} - \ln N_i - N_i \frac{1}{N_i} \\ &= \ln \left( \frac{g_i + N_i - 1}{N_i} \right) \approx \ln \left( \frac{g_i}{N_i} + 1 \right), \quad \text{since } N_i \gg 1 \end{aligned}$$

To maximize  $\Omega$ , we set  $d(\ln \Omega) = 0$ , i.e.,

$$d(\ln \Omega) = \sum_{i=0}^{\infty} \frac{\partial(\ln \Omega)}{\partial N_i} dN_i \approx \sum_{i=0}^{\infty} \ln \left( \frac{g_i}{N_i} + 1 \right) dN_i = 0$$

By adding Lagrangian multipliers, Eq. (3.14a) and Eq. (3.14b), we have  $\sum_{i=0}^{\infty} [\ln(g_i/N_i + 1) - \alpha - \beta \varepsilon_i] dN_i = 0$ . Hence,  $N_i = g_i / (e^{\alpha} e^{\beta \varepsilon_i} - 1)$ , which is the BE distribution given in Eq. (3.16) and Table 3.1.

If  $\exp(\alpha + \beta \varepsilon_i) \gg 1$ , Eq. (3.16) and Eq. (3.17) reduce to the MB distribution, Eq. (3.15a). Under the limiting case of  $g_i \gg N_i \gg 1$ , we have

$$\frac{(g_i + N_i - 1)!}{(g_i - 1)! N_i!} = \frac{\overbrace{(g_i + N_i - 1) \cdots (g_i + 1) g_i}^{N_i \text{ terms}}}{N_i!} \xrightarrow{g_i \gg N_i \gg 1} \frac{g_i^{N_i}}{N_i!}$$

and 
$$\frac{g_i!}{(g_i - N_i)! N_i!} = \frac{\overbrace{g_i (g_i - 1) \cdots (g_i - N_i + 1)}^{N_i \text{ terms}}}{N_i!} \xrightarrow{g_i \gg N_i \gg 1} \frac{g_i^{N_i}}{N_i!}$$

That is to say that the thermodynamic probability for both the BE and FD statistics reduces to the MB statistics divided by  $N!$ , which is caused by the assumption of indistinguishable particles. Therefore,

$$\Omega_{\text{MB,corrected}} = \prod_{i=0}^{\infty} \frac{g_i^{N_i}}{N_i!} = \frac{\Omega_{\text{MB}}}{N!} \quad (3.18)$$

**TABLE 3.1** Summary of the Three Statistics

Statistics	Maxwell-Boltzmann (MB)	Bose-Einstein (BE)	Fermi-Dirac (FD)
Name of particles	Boltzons	Bosons	Fermions
Examples	Ideal gas molecules & in the limit of bosons and fermions	Photons & phonons	Electrons & protons
Distinguishability	Distinguishable	Indistinguishable	Indistinguishable
Degeneracy	Degenerate	Degenerate	Degenerate
Particles per quantum state	Unlimited	Unlimited	One
Thermodynamic probability $\Omega$	$N! \prod_{i=0}^{\infty} \frac{g_i^{N_i}}{N_i!}$	$\prod_{i=0}^{\infty} \frac{(g_i + N_i - 1)!}{(g_i - 1)! N_i!}$	$\prod_{i=0}^{\infty} \frac{g_i!}{(g_i - N_i)! N_i!}$
In the limit of $g_i \gg N_i$	$\Omega_{\text{MB}}$ (given above)	$\Omega_{\text{MB}}/N!$	$\Omega_{\text{MB}}/N!$
$\ln \Omega$	$N \ln N - N + \sum_{i=0}^{\infty} N_i [\ln(g_i/N_i) + 1]$	$\sum_{i=0}^{\infty} [(g_i + N_i - 1) \ln(g_i + N_i - 1) - N_i \ln N_i - (g_i - 1) \ln(g_i - 1)]$	$\sum_{i=0}^{\infty} [(g_i \ln g_i - N_i \ln N_i) - (g_i - N_i) \ln(g_i - N_i)]$
$d(\ln \Omega)$	$\sum_{i=0}^{\infty} \ln\left(\frac{g_i}{N_i}\right) dN_i$	$\sum_{i=0}^{\infty} \ln\left(\frac{g_i}{N_i} + 1\right) dN_i$	$\sum_{i=0}^{\infty} \ln\left(\frac{g_i}{N_i} - 1\right) dN_i$
$-\alpha \sum_{i=0}^{\infty} dN_i - \beta \sum_{i=0}^{\infty} \varepsilon_i dN_i$	$\ln\left(\frac{g_i}{N_i}\right) - \alpha - \beta \varepsilon_i = 0$	$\ln\left(\frac{g_i}{N_i} + 1\right) - \alpha - \beta \varepsilon_i = 0$	$\ln\left(\frac{g_i}{N_i} - 1\right) - \alpha - \beta \varepsilon_i = 0$
Distribution function $N_i$	$\frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i}}$	$\frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i} - 1}$	$\frac{g_i}{e^{\alpha} e^{\beta \varepsilon_i} + 1}$
Applications	Ideal gases; Maxwell's velocity distribution; limiting cases of BE and FD statistics	Planck's law; Bose-Einstein condensation; specific heat of solids	Electron gas; Fermi level; electron specific heat in metals

is called the “corrected” MB statistics. For ideal molecular gases at reasonably high temperatures,  $g_i \gg N_i$ . For this reason, the MB distribution may be considered as the limiting case of the BE or FD distribution (see Table 3.1).

## 3.2 THERMODYNAMIC RELATIONS

The thermodynamic properties and relations can be understood from the microscopic point of view. This includes the concept of heat and work, entropy, and the third law of thermodynamics. The partition function is key to the evaluation of thermodynamic properties.

### 3.2.1 Heat and Work

From Eq. (3.3), we have

$$dU = \sum_{i=0}^{\infty} \varepsilon_i dN_i + \sum_{i=0}^{\infty} N_i d\varepsilon_i \quad (3.19a)$$

The first term on the right is due to a redistribution of particles among the energy levels (which is related to a change in entropy), while the second is due to a shift in the energy levels associated with, e.g., a volume change. Consider a reversible quasi-equilibrium process for a closed system (such as a piston/cylinder arrangement). The work is associated to the volume change that does not change the entropy of the system, while heat transfer changes entropy of the system without affecting the energy levels. Therefore,

$$\delta Q = \sum_{i=0}^{\infty} \varepsilon_i dN_i \quad \text{and} \quad \delta W = - \sum_{i=0}^{\infty} N_i d\varepsilon_i \quad (3.19b)$$

In writing the above equation,  $\delta Q$  is positive for heat transferred to the system, and  $\delta W$  is positive for work done by the system. They are related to macroscopic quantities for simple systems by  $\delta Q = TdS$  and  $\delta W = PdV$ . Hence, we obtain the expression of the first law for a closed system,  $dU = \delta Q - \delta W$ . If the system is an open system, then  $\sum_{i=0}^{\infty} \varepsilon_i dN_i = dU + \delta W \neq \delta Q$ .

### 3.2.2 Entropy

The macroscopic property entropy is related to the thermodynamic probability by

$$S = k_B \ln \Omega \quad (3.20)$$

where  $k_B$  is the Boltzmann constant. Consider two separate systems A and B, and their combination as a system C. At a certain time, both A and B are individually in thermodynamic equilibrium. Denote the states as  $A_1$  and  $B_1$ , and the combined system as state  $C_1$ . The thermodynamic probability of system C at state  $C_1$  is related to those of  $A_1$  and  $B_1$  by

$$\Omega_1^C = \Omega_1^A \times \Omega_1^B$$

The entropy of  $C_1$  is then

$$S_1^C = k_B \ln \Omega_1^C = k_B \ln(\Omega_1^A \times \Omega_1^B) = k_B \ln \Omega_1^A + k_B \ln \Omega_1^B = S_1^A + S_1^B$$

Therefore, this definition of entropy meets the additive requirement.

The highest entropy principle states that the entropy of an isolated system will increase until it reaches a stable-equilibrium state (thermodynamic equilibrium), i.e.,  $\Delta S_{\text{isolated}} \geq 0$ . The microscopic understanding is that entropy is related to the probability of occurrence of a certain macrostate. For a system with specified  $U$ ,  $N$ , and  $V$ , the macrostate that corresponds to the thermodynamic equilibrium is the most probable state and, hence, its entropy is the largest. Any states, including those that deviate very slightly from the stable-equilibrium state, will have a much smaller thermodynamic probability. After the equilibrium state is reached, it is not possible for any macrostate, whose thermodynamic probability is much less than that of the equilibrium state, to occur within an observable amount of time.

### 3.2.3 The Lagrangian Multipliers

For all three types of statistics,  $d(\ln \Omega) = \alpha \sum_{i=0}^{\infty} dN_i + \beta \sum_{i=0}^{\infty} \varepsilon_i dN_i$ , where the first term is the change in the total number of particles and the second can be related to the net heat transfer for a closed system; therefore,  $d(\ln \Omega) = \alpha dN + \beta \delta Q$ . In a reversible process in which the total number of particles do not change (closed system),  $dN = 0$ ,  $d(\ln \Omega) = dS/k_B$ , and  $\delta Q = TdS$ . Hence, we have for all three statistics

$$\beta \equiv \frac{1}{k_B T} \quad (3.21)$$

To evaluate  $\alpha$ , we must allow the system to change its composition. In this case,

$$d(\ln \Omega) = \alpha \sum_{i=0}^{\infty} dN_i + \beta \sum_{i=0}^{\infty} \varepsilon_i dN_i = \alpha dN + \beta(dU + PdV)$$

or

$$TdS = k_B T \alpha dN + dU + PdV$$

Substituting the above equation into the definition of the Helmholtz function,  $dA = d(U - TS) = dU - TdS - SdT$ , we have

$$dA = -SdT - PdV - k_B T \alpha dN$$

Noting that the chemical potential  $\mu = (\partial A / \partial N)_{T,V} = -k_B T \alpha$ , we obtain

$$\alpha = -\frac{\mu}{k_B T} \quad (3.22)$$

where  $\mu$  is expressed in molecular quantity, and  $\alpha = -\mu/\bar{R}T$  if  $\mu$  is expressed in molar quantity.

### 3.2.4 Entropy at Absolute Zero Temperature

The third law of thermodynamics states that the entropy of any pure substance vanishes at the ground state (with absolute zero temperature); see Sec. 2.1.3. For BE statistics, we have

$$N = N_0 + N_1 + N_2 + \cdots = \frac{g_0}{e^{\alpha + \beta \varepsilon_0} - 1} + \frac{g_1}{e^{\alpha + \beta \varepsilon_1} - 1} + \frac{g_2}{e^{\alpha + \beta \varepsilon_2} - 1} + \cdots$$

At very low temperatures ( $T \rightarrow 0$ ),  $\beta = 1/k_B T \rightarrow \infty$ . Since  $\varepsilon_0 < \varepsilon_1 < \varepsilon_2 < \dots$ ,

$$\frac{N_i}{N_0} \approx \frac{g_i}{g_0} e^{-\beta(\varepsilon_i - \varepsilon_0)} \rightarrow 0 \text{ as } T \rightarrow 0 \text{ for } i \geq 1 \quad (3.23)$$

Hence,  $N \approx N_0$ ; that is, all particles will be at the lowest energy level (ground state). If  $g_0 = 1$ , as it is the case for a pure substance, then  $\Omega = 1$  and  $S = k_B \ln \Omega = 0$  as  $T \rightarrow 0$ ; this is consistent with the third law of thermodynamics. The occurrence for particles that obey BE statistics (bosons) to collapse to the ground state at sufficiently low temperatures is called the *Bose-Einstein condensation*. Such a state of matter is called the *Bose-Einstein condensate*, in which quantum effects dominate the macroscopic behavior.

Some important applications of the Bose-Einstein condensation are superfluidity and superconductivity. Liquid helium ( $^4\text{He}$ ) becomes a superfluid with no viscosity at temperatures below the  $\lambda$ -transition ( $T \approx 2.17$  K). The specific heat of helium at this temperature becomes infinitely large, suggesting that a phase transition occurs. Bose-Einstein condensate of atoms has been observed with laser cooling and trapping techniques.<sup>9</sup> Photons from the laser collide with the atoms. The absorption can be tuned using the Doppler shift so that only atoms traveling toward the laser can absorb the photons, resulting in reduced momentums in these atoms. Furthermore, the excited atoms will emit photons spontaneously in all directions. The net effect is a decrease in the velocity of the atoms, resulting in a kinetic temperature down to the nanokelvin range. In the last decade, the Nobel Prize in Physics was awarded for works related to the Bose-Einstein condensation four times: 1996, 1997, 2001, and 2003.

Although electrons are fermions (particles that obey FD statistics) that generally do not condense at zero temperature, they can form pairs at sufficiently low temperatures that behave like bosons. Below the critical temperature, pairs of electrons, called the Cooper pairs can travel freely without any resistance. This is the phenomenon called superconductivity, which was discovered at the beginning of the twentieth century. A large number of elements and compounds can be made superconducting at very low temperatures. Furthermore, some oxides become superconducting at temperatures above 90 K.<sup>10</sup> Superconductors have important applications in magnetic resonance imaging, high-speed and low-noise electronic devices, infrared sensors, and so forth. A similar phenomenon is the superfluidity in helium isotope  $^3\text{He}$ , which undergoes a phase transition at very low temperatures. The fermionic  $^3\text{He}$  atoms pair up to form bosonic entities that experience Bose-Einstein condensation at 3 mK.

For FD statistics, from Eq. (3.17), Eq. (3.21), and Eq. (3.22), we have

$$\frac{N_i}{g_i} = \frac{1}{e^{(\varepsilon_i - \mu)/k_B T} + 1} \quad (3.24)$$

As  $T \rightarrow 0$ , it is found that  $N_i/g_i = 1$  for all energy levels with  $\varepsilon_i < \mu$  and  $N_i/g_i = 0$  for energy levels with  $\varepsilon_i > \mu$ . That is, all quantum states are filled for  $i = 0, 1, 2, \dots, j$  (with  $\varepsilon_j < \mu$ ), and all quantum states are empty for  $i = j + 1, j + 2, \dots$  (with  $\varepsilon_{j+1} > \mu$ ), as schematically shown in Fig. 3.6. More discussions will be given in Chap. 5 on the behavior of free electrons. For now, it is sufficient to say that the thermodynamic probability  $\Omega = 1$  for FD statistics at absolute zero temperature. Therefore, the entropy  $S = 0$  at  $T \rightarrow 0$  K for both the BE and FD statistics. However, MB statistics does not satisfy the third law and is not applicable to very low temperatures.

### 3.2.5 Macroscopic Properties in Terms of the Partition Function

The *partition function* is an important quantity in statistical thermodynamics. Unlike the characteristic functions (such as the Helmholtz free energy and the Gibbs free energy defined in Chap. 2) used in macroscopic thermodynamics, the physical meaning of the partition function is not immediately clear. However, the introduction of the partition function

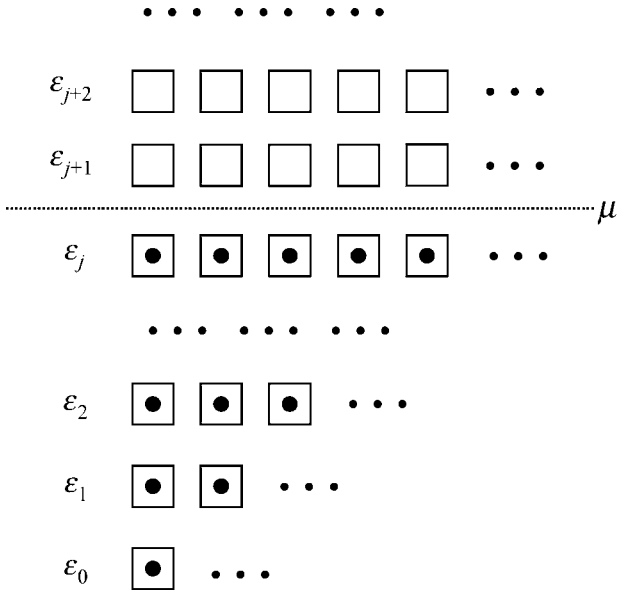


FIGURE 3.6 Schematic of the Fermi-Dirac distribution at 0 K.

allows the calculation of macroscopic thermodynamic properties from the microscopic representation. There are different types of partition functions. For MB statistics, the partition function is defined as

$$Z = N e^{\alpha} = \sum_{i=0}^{\infty} g_i e^{-\varepsilon_i/k_B T} \tag{3.25}$$

Therefore, 
$$N_i = \frac{N}{Z} g_i e^{-\varepsilon_i/k_B T} \tag{3.26}$$

Since 
$$\left[ \frac{\partial(\ln Z)}{\partial T} \right]_{V,N} = \frac{1}{Z} \left( \frac{\partial Z}{\partial T} \right)_{V,N} = \frac{\sum_{i=0}^{\infty} g_i e^{-\varepsilon_i/k_B T} \left( \frac{\varepsilon_i}{k_B T^2} \right)}{\sum_{i=0}^{\infty} g_i e^{-\varepsilon_i/k_B T}} = \frac{U e^{\alpha}}{k_B T^2} = \frac{U}{N e^{\alpha}} = \frac{U}{N k_B T^2}$$

we have 
$$U = N k_B T^2 \left[ \frac{\partial(\ln Z)}{\partial T} \right]_{V,N} \tag{3.27}$$

Using the corrected MB statistics given in Eq. (3.18), we can express the entropy as

$$\begin{aligned} S &= k_B \ln(\Omega_{MB}/N!) = k_B \sum_{i=0}^{\infty} N_i \left( 1 + \ln \frac{g_i}{N_i} \right) \\ &= k_B \sum_{i=0}^{\infty} N_i \left( 1 + \ln \frac{Z}{N} + \beta \varepsilon_i \right) = N k_B + N k_B \ln \frac{Z}{N} + k_B \beta U \end{aligned} \tag{3.28a}$$

Had we not divided  $\Omega_{\text{MB}}$  by  $N!$ , we would get  $S = Nk_B \ln Z + k_B \beta U$ , which is different from Eq. (3.28a) by a constant. After substituting  $\beta$  and  $U$  into Eq. (3.28a), we obtain

$$S = Nk_B \left\{ 1 + \ln \frac{Z}{N} + T \left[ \frac{\partial(\ln Z)}{\partial T} \right]_{V,N} \right\} \quad (3.28b)$$

The Helmholtz free energy is

$$A = U - TS = -Nk_B T \left( 1 + \ln \frac{Z}{N} \right) \quad (3.29)$$

The pressure is

$$P = - \left( \frac{\partial A}{\partial V} \right)_{T,N} = Nk_B T \left[ \frac{\partial(\ln Z)}{\partial V} \right]_{T,N} \quad (3.30)$$

The enthalpy  $H$  and the Gibbs free energy  $G$  can also be obtained. The partition function is now related to the macroscopic thermodynamic properties of interest for simple substances.

### 3.3 IDEAL MOLECULAR GASES

An important application of statistical mechanics is to model and predict the thermal properties of materials. In this section, the application of MB statistics to obtain the equation of state and the velocity distributions for ideal molecular gases is presented. The microscopic theories of the specific heat for ideal monatomic and polyatomic gases are given subsequently.

#### 3.3.1 Monatomic Ideal Gases

For a monatomic ideal gas at moderate temperatures, MB statistics can be applied, and the translational energies are

$$\varepsilon = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2) = \frac{1}{2}m\mathbf{v}^2 \quad (3.31)$$

Consider a volume element in the phase space,  $dx dy dz dp_x dp_y dp_z$ , where  $\mathbf{p} = m\mathbf{v}$  is the momentum of a molecule. The accuracy of specifying the momentum and the displacement is limited by  $\Delta x \Delta p_x \sim h$ , given by the uncertainty principle. The degeneracy, which is the number of quantum states (boxes of size  $h^3$ ) in a volume element of the phase space, is given by

$$dg = \frac{dx dy dz dp_x dp_y dp_z}{h^3} = \frac{m^3}{h^3} dx dy dz dv_x dv_y dv_z \quad (3.32)$$

Many useful results were obtained before quantum mechanics by assuming that  $h^3$  is some constant. A more rigorous proof of Eq. (3.32) will be given in Sec. 3.5. When the space between energy levels are sufficiently close, the partition function can be expressed in terms of an integral as  $Z_t = \int e^{-\varepsilon/k_B T} dg$  or

$$Z_t = \iiint dx dy dz \iiint \frac{m^3}{h^3} \exp \left[ -\frac{m}{2k_B T} (v_x^2 + v_y^2 + v_z^2) \right] dv_x dv_y dv_z \quad (3.33)$$



The space integration yields the volume  $V$ , and the velocity integration can be individually performed, i.e.,

$$\int_{-\infty}^{\infty} \exp\left(-\frac{mv_x^2}{2k_B T}\right) dv_x = \sqrt{\frac{2\pi k_B T}{m}} \quad (3.34)$$

Hence,

$$Z_t = V \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} \quad (3.35)$$

Therefore,

$$e^\alpha = \frac{V}{N} \left(\frac{2\pi m k_B T}{h^2}\right)^{3/2} \quad (3.36)$$

which is indeed much greater than unity at normal temperatures for most substances, suggesting that the MB statistics is applicable for ideal molecular gases. At extremely low temperatures, intermolecular forces cannot be neglected and the molecules are not independent anymore.

From Eq. (3.30), we have  $P = Nk_B T [\partial(\ln Z)/\partial V]_{T,N} = Nk_B T/V$ ; i.e.,

$$PV = Nk_B T \quad \text{or} \quad P = nk_B T \quad (3.37)$$

where  $n = N/V$  is the number density. The Boltzmann constant is the ideal (universal) gas constant on the molecular basis, i.e.,  $k_B = \bar{R}/N_A$ . The internal energy, the specific heats, and the absolute entropy can also be evaluated.

$$U = Nk_B T^2 \left[ \frac{\partial(\ln Z)}{\partial T} \right]_{V,N} = \frac{3}{2} Nk_B T \quad (3.38)$$

which is not a function of pressure. The molar specific internal energy is  $\bar{u} = \frac{3}{2} \bar{R}T$ , and the molar specific heats are

$$\bar{c}_v = \left( \frac{\partial \bar{u}}{\partial T} \right)_v = \frac{3}{2} \bar{R} \quad (3.39)$$

and

$$\bar{c}_p = \left( \frac{\partial \bar{h}}{\partial T} \right)_p = \frac{5}{2} \bar{R} \quad (3.40)$$

The above equations show that the specific heats of monatomic gases are independent of temperature, except at very high temperatures when electronic contributions become important. The molar specific heats do not depend on the type of molecules, but the same is not true for mass specific heats. Using Eq. (3.28b), the absolute entropy can be expressed as

$$S = Nk_B \left\{ \frac{5}{2} + \ln \left[ \frac{V}{N} \left( \frac{2\pi m k_B T}{h^2} \right)^{3/2} \right] \right\}$$

Therefore, the molar specific entropy is a function of  $T$  and  $P$ , i.e.,

$$\bar{s}(T,P) = \bar{R} \left\{ \frac{5}{2} + \ln \left[ \frac{k_B T}{P} \left( \frac{2\pi m k_B T}{h^2} \right)^{3/2} \right] \right\} \quad (3.41)$$

This is the *Sackur-Tetrode equation*.

### 3.3.2 Maxwell's Velocity Distribution

Rewrite  $N_i = g_i e^{-\alpha} e^{-\epsilon_i/k_B T}$  as  $dN = dg e^{-\alpha} e^{-\epsilon/k_B T}$ . In a volume  $V$  and from  $\mathbf{v}$  to  $\mathbf{v} + d\mathbf{v}$  (i.e.,  $v_x$  to  $v_x + dv_x$ ,  $v_y$  to  $v_y + dv_y$ , and  $v_z$  to  $v_z + dv_z$ ), the number of molecules  $dN$  per unit volume may be expressed as

$$\frac{dN}{V} = \frac{m^3}{h^3} dv_x dv_y dv_z \frac{N}{V} \left( \frac{h^2}{2\pi m k_B T} \right)^{3/2} \exp\left(-\frac{m}{2k_B T} \mathbf{v}^2\right) \quad (3.42)$$

or

$$f(\mathbf{v}) d\mathbf{v} = \frac{dN}{V} = n \left( \frac{m}{2\pi k_B T} \right)^{3/2} \exp\left(-\frac{m\mathbf{v}^2}{2k_B T}\right) d\mathbf{v} \quad (3.43)$$

where  $f(\mathbf{v})$  is the Maxwell velocity distribution in a unit volume. Notice that

$$F(\mathbf{v}) = \frac{f(\mathbf{v})}{n} = \left( \frac{m}{2\pi k_B T} \right)^{3/2} \exp\left(-\frac{m\mathbf{v}^2}{2k_B T}\right) \quad (3.44)$$

which is a Gaussian distribution. Notice that  $\mathbf{v}^2 = \mathbf{v} \cdot \mathbf{v} = v^2 = v_x^2 + v_y^2 + v_z^2$ . The distribution of velocity component is also Gaussian, such that

$$F(\mathbf{v}) = F(v_x)F(v_y)F(v_z) \quad (3.45)$$

Taking the  $x$  component as an example, we can write

$$F(v_x) = \left( \frac{m}{2\pi k_B T} \right)^{1/2} \exp\left(-\frac{mv_x^2}{2k_B T}\right) \quad (3.46)$$

The speed distribution may be obtained from the following by integrating the velocity distribution in a spherical shell (i.e., over the solid angle of  $4\pi$ ).

$$F(v)dv = \iiint_{4\pi} F(\mathbf{v})d\mathbf{v} = \iiint_{4\pi} \left( \frac{m}{2\pi k_B T} \right)^{3/2} \exp\left(-\frac{mv^2}{2k_B T}\right) v^2 d\Omega dv$$

Therefore,

$$F(v) = 4\pi \left( \frac{m}{2\pi k_B T} \right)^{3/2} v^2 \exp\left(-\frac{mv^2}{2k_B T}\right) \quad (3.47)$$

Figure 3.7 plots the speed distribution of He gas at 0, 300, and 800°C. When evaluating  $k_B T$ , we must convert  $T$  to absolute temperature. It can be seen that more molecules will be at higher speeds as the temperature increases. It should be noted that  $F(v=0) = 0$  but  $F(\mathbf{v})$  is maximum at  $v = 0$ . In the speed coordinate, an interval between  $v$  and  $v + dv$  corresponds to a spherical shell in the velocity space. Even though  $F(\mathbf{v})$  is maximum at  $v = 0$ ,

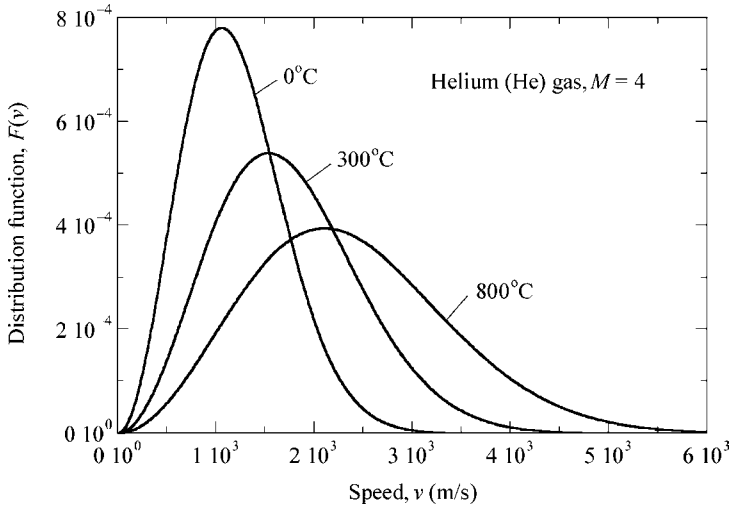


FIGURE 3.7 Speed distribution for helium gas at different temperatures.

the probability of finding a molecule per unit speed interval decreases to 0 as  $v \rightarrow 0$ , which is caused by the associated decrease in the volume of the spherical shell.

**Example 3-5.** Find the average speed and the root-mean-square speed for a He gas at  $200^\circ\text{C}$  at 100 kPa. What if the pressure is changed to 200 kPa? What are the most probable velocity and the most probable speed?

**Solution.** The average speed may be obtained from either the velocity distribution or the speed distribution. That is

$$\bar{v} = \iiint vF(\mathbf{v})d\mathbf{v} = \int_0^\infty vF(v)dv = \sqrt{\frac{8k_B T}{\pi m}} \quad (3.48)$$

The average of  $v^2$  is (see Appendix B.5)

$$\overline{v^2} = \iiint v^2F(\mathbf{v})d\mathbf{v} = \int_0^\infty v^2F(v)dv = \frac{3k_B T}{m} \quad (3.49a)$$

Therefore the root-mean-square speed is

$$v_{\text{rms}} = \sqrt{\overline{v^2}} = \sqrt{\frac{3k_B T}{m}} \quad (3.49b)$$

Plugging in the numerical values, we have  $\bar{v} = 1582$  m/s and  $v_{\text{rms}} = 1717$  m/s for He gas at  $200^\circ\text{C}$ . We also notice that the pressure has no effect on the speed distribution, unless it is so high that intermolecular forces cannot be neglected.

The most probable velocity  $\mathbf{v}_{\text{mp}} = 0$  because of the symmetry in the Gaussian distribution. We can obtain the most probable speed by setting  $F'(v) = 0$ , i.e.,

$$2v \exp\left(-\frac{mv^2}{2k_B T}\right) - v^2\left(\frac{mv}{k_B T}\right) \exp\left(-\frac{mv^2}{2k_B T}\right) = 0$$

The solution gives the most probable speed as  $v_{mp} = \sqrt{2k_B T/m}$ . For He gas at 200°C, it gives  $v_{mp} = 1402$  m/s. Note that  $v_{mp} : \bar{v} : v_{rms} = \sqrt{2} : \sqrt{8/\pi} : \sqrt{3} \approx 1.4 : 1.6 : 1.7$ .

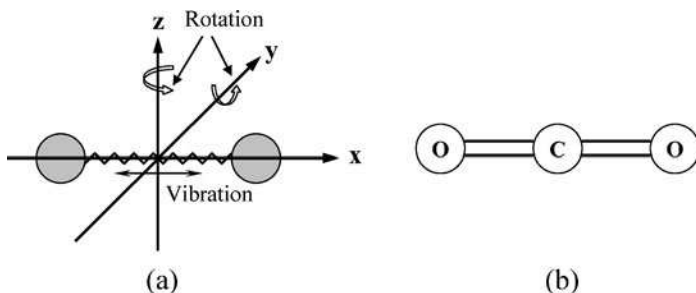
**Comment.** An important consequence for Eq. (3.49a) is that temperature is related to the mean kinetic energy of the molecule, i.e.,

$$\frac{1}{2}m\bar{v}_x^2 = \frac{1}{2}m\bar{v}_y^2 = \frac{1}{2}m\bar{v}_z^2 = \frac{1}{2}k_B T \quad (3.50)$$

The internal energy of a monatomic gas given in Eq. (3.38) is the sum of the kinetic energies of all molecules.

### 3.3.3 Diatomic and Polyatomic Ideal Gases

Additional degrees of freedom or energy storage modes must be considered for diatomic and polyatomic molecules, besides translation. The molecule may rotate about its center of gravity, and atoms may vibrate with respect to each other. For a molecule consisting of  $q$  atoms, each atom may move in all three directions, and there will be a total of  $3q$  modes. Consider the translation of the molecule as a whole; there are three translational degrees of freedom or modes, i.e.,  $\phi_t = 3$ . For diatomic molecules or polyatomic molecules whose atoms are arranged in a line (such as  $\text{CO}_2$ ), as shown in Fig. 3.8, there are two rotational

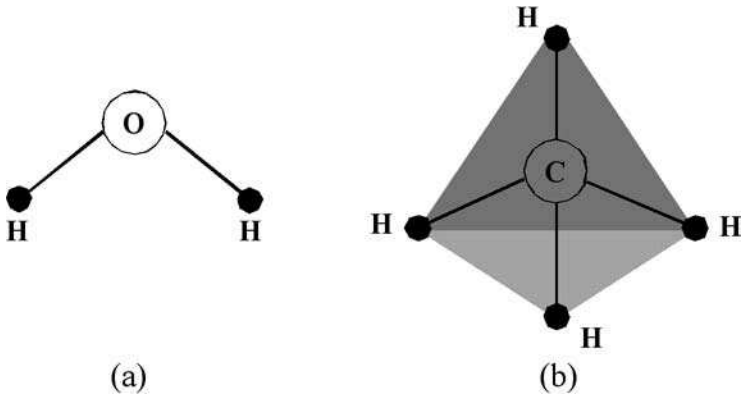


**FIGURE 3.8** (a) A diatomic molecule, showing two rotational and one vibrational degrees of freedom. (b)  $\text{CO}_2$  molecule, where the atoms are aligned.

degrees of freedom or modes, i.e.,  $\phi_r = 2$ . Therefore, there are  $\phi_v = 3q - 5$  vibrational modes, each consisting of two degrees of freedom corresponding to the kinetic energy and the potential energy. For polyatomic molecules whose atoms are not aligned (such as  $\text{H}_2\text{O}$  and  $\text{CH}_4$ , see Fig. 3.9), there are three rotational degrees of freedom, i.e.,  $\phi_r = 3$ . The vibrational modes are thus  $\phi_v = 3q - 6$ .

The total energy of a molecule may be expressed as the sum of translational, rotational, and vibrational energies, i.e.,  $\varepsilon = \varepsilon_t + \varepsilon_r + \varepsilon_v$ . For simplicity, we have neglected contributions from the electronic ground state and chemical dissociation, which can be included as additional terms in evaluating the internal energy and the entropy.<sup>1</sup> At high temperatures, the vibration mode can be coupled with the rotation mode. Here, however, it is assumed that these modes are independent. The partition function can be written as

$$Z = Z_t Z_r Z_v = \left( \sum g_t e^{-\varepsilon_t/k_B T} \right) \left( \sum g_r e^{-\varepsilon_r/k_B T} \right) \left( \sum g_v e^{-\varepsilon_v/k_B T} \right) \quad (3.51)$$



**FIGURE 3.9** (a) A  $\text{H}_2\text{O}$  molecule in which the atoms are not aligned. (b) The tetrahedral methane ( $\text{CH}_4$ ) molecule.

For polyatomic atoms, Eq. (3.31) through Eq. (3.36) hold for the translational modes.  $Z_t$  and  $Z_v$  are internal contributions that do not depend on volume; therefore, Eq. (3.37) also holds. Since the degrees of freedom are independent of each other, Maxwell's velocity and speed distributions discussed in Sec. 3.3.2 still hold for polyatomic gases. The problem now is to determine the rotational and vibrational energy levels and degeneracies. Generally speaking, there exists a certain characteristic temperature associated with each degree of freedom. The characteristic temperature for translation is very low for molecular gases. On the other hand, the characteristic temperature for rotation is slightly higher, and that for vibration is usually very high, as can be seen from Table 3.2 for selected diatomic molecules. If the temperature is much less than the characteristic temperature of a certain mode, then the contribution of that mode to the energy storage is negligible. For the temperature much higher than the characteristic temperature, however, there often exist some asymptotic approximations.

**Rotation.** A quantum mechanical analysis of a rigid rod, to be derived in Sec. 3.5.3, shows that the rotational energy levels are given by

$$\frac{\varepsilon_l}{k_B T} = l(l + 1) \frac{\Theta_r}{T} \quad (3.52)$$

**TABLE 3.2** Characteristic Temperatures of Rotation and Vibration for Some Diatomic Molecules

Substance	Symbol	$\Theta_r$ (K)	$\Theta_v$ (K)
Hydrogen	$\text{H}_2$	87.5	6320
Deuterium	$\text{D}_2$	43.8	4490
Hydrogen chloride	$\text{HCl}$	15.2	4330
Nitrogen	$\text{N}_2$	2.86	3390
Carbon monoxide	$\text{CO}$	2.78	3120
Nitric oxide	$\text{NO}$	2.45	2745
Oxygen	$\text{O}_2$	2.08	2278
Chloride	$\text{Cl}_2$	0.35	814
Sodium vapor	$\text{Na}_2$	0.08	140

Here,  $\Theta_r$  is the characteristic temperature for rotation and is given by  $\Theta_r = h^2/(8\pi^2 k_B I)$ , where  $I$  is the moment of inertia of the molecule about the center of mass. The larger the value of  $I$ , the smaller the characteristic temperature will be. This is clearly shown in Table 3.2. The degeneracy of rotational energy levels is

$$g_l = \frac{2l + 1}{\sigma} \quad (3.53)$$

where  $\sigma$  is a symmetry number that arises from molecular symmetry:  $\sigma = 1$  if the atoms are of different types (such as in a NO or CO molecule), and  $\sigma = 2$  if the atoms are the same (such as in a  $O_2$  or  $N_2$  molecule).

$$Z_r = \sum_{l=0}^{\infty} \frac{2l + 1}{\sigma} \exp\left[-l(l + 1) \frac{\Theta_r}{T}\right] \quad (3.54)$$

This series converges very fast for  $\Theta_r/T > 0.5$ , since

$$Z_r = \frac{1}{\sigma} \left[ 1 + 3 \exp\left(-\frac{2\Theta_r}{T}\right) + 5 \exp\left(-\frac{6\Theta_r}{T}\right) + 7 \exp\left(-\frac{12\Theta_r}{T}\right) + \dots \right]$$

For  $T/\Theta_r > 1$ , Eq. (3.54) may be expanded to give (see Problem 3.26)

$$Z_r = \frac{T}{\Theta_r \sigma} \left[ 1 + \frac{1}{3} \left(\frac{\Theta_r}{T}\right) + \frac{1}{15} \left(\frac{\Theta_r}{T}\right)^2 + \frac{4}{315} \left(\frac{\Theta_r}{T}\right)^3 + \dots \right] \quad (3.55)$$

At temperatures much higher than the characteristic temperature of rotation,  $T/\Theta_r \gg 1$ , the above equation reduces to

$$Z_r = \frac{T}{\sigma \Theta_r} \quad (3.56)$$

Under this limit, the contribution of the rotational energy to the internal energy becomes

$$U_r \approx N k_B T \quad (3.57)$$

The contribution to the molar specific heat by the two rotational degrees of freedom is

$$\bar{c}_{v,r} = \bar{R} \quad (3.58)$$

**Vibration.** The vibration in a molecule can be treated as a harmonic oscillator. For each vibration mode, the quantized energy levels are given in Sec. 3.5.5 as

$$\varepsilon_{v,i} = \left(i + \frac{1}{2}\right) h\nu, \quad i = 0, 1, 2, \dots \quad (3.59)$$

where  $\nu$  is the natural frequency of vibration, and the ground-state energy is  $\frac{1}{2}h\nu$ . The vibrational energy levels are not degenerated, i.e.,  $g_{v,i} = 1$ . Therefore, we can write

$$Z_v = \sum_{i=0}^{\infty} e^{-(i+1/2)h\nu/k_B T} = e^{-\Theta_v/2T} \sum_{i=0}^{\infty} e^{-i\Theta_v/T}$$

where  $\Theta_v = hv/k_B$  is a characteristic temperature for vibration and is listed in Table 3.2 for several diatomic molecules. The vibrational partition function becomes

$$Z_v = \frac{e^{-\Theta_v/2T}}{1 - e^{-\Theta_v/T}} = \frac{e^{\Theta_v/2T}}{e^{\Theta_v/T} - 1} \quad (3.60)$$

Its contribution to the internal energy and the specific heat can be written as

$$U_v = Nk_B\Theta_v\left(\frac{1}{2} + \frac{1}{e^{\Theta_v/T} - 1}\right) \quad (3.61)$$

and

$$\bar{c}_{v,v} = \bar{R} \frac{\Theta_v^2}{T^2} \frac{e^{\Theta_v/T}}{(e^{\Theta_v/T} - 1)^2} \quad (3.62)$$

At  $T \ll \Theta_v$ , the vibrational mode contributes to the internal energy but not to the specific heat. At  $T > 1.5\Theta_v$ ,  $U_v$  almost linearly depends on  $T$  and  $\bar{c}_{v,v} \approx \bar{R}$ . In classical statistical mechanics, it is believed that each degree of freedom contributes to the stored thermal energy with an amount of  $\frac{1}{2}k_B T$  and results in a specific heat of  $\frac{1}{2}k_B$  on the particle base. This is called the *equipartition principle*. The contribution of each vibrational mode is  $\bar{R}$  not  $\bar{R}/2$ , due to the fact that each vibrational mode includes a kinetic component and a potential component for energy storage and is generally considered as two degrees of freedom. It should be noted that the equipartition principle is only applicable at sufficiently high temperatures and for particles that obey MB statistics or, in some limiting cases, BE statistics. Because energy is additive, as is the specific heat, we can write

$$\bar{c}_v = \bar{c}_{v,t} + \bar{c}_{v,r} + \bar{c}_{v,v} \quad (3.63)$$

The result is schematically shown in Fig. 3.10. One can see that for a diatomic ideal gas,

$$\bar{c}_v = 2.5\bar{R} \quad \text{if} \quad \Theta_r \ll T \ll \Theta_v \quad (3.64)$$

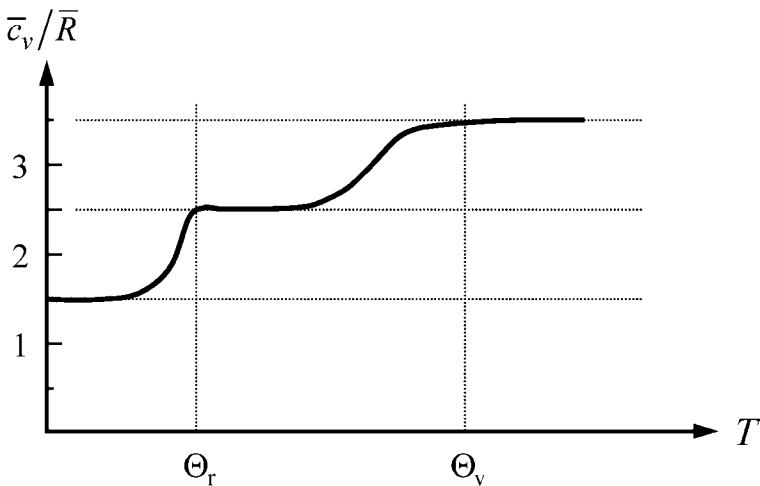


FIGURE 3.10 Typical specific heat curve of a diatomic ideal gas.

which happens to be near room temperature for many gases such as nitrogen and carbon monoxide; see Table 3.2. Figure 3.11 plots the specific heat for several real gases at sufficiently

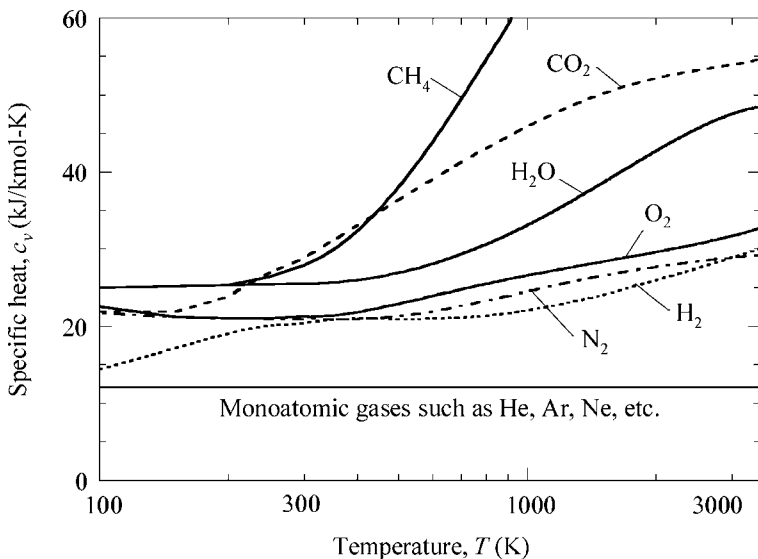


FIGURE 3.11 Specific heat at constant volume for several ideal gases.

low pressure so that the ideal gas model is applicable. It should be noted that, for hydrogen, nuclear spin is important and Eq. (3.54) needs to be modified to account for the spin degeneracy.<sup>1,2</sup> However, Eq. (3.57) and Eq. (3.58) predict the right trend and are applicable at temperatures much higher than  $\Theta_r$ . At extremely high temperatures (say 3000 K), electronic contributions and the coupling between rotation and vibration become important. Although Eq. (3.63) is the correct expression for the specific heat at moderate temperatures, two additional partition functions must be included to correctly evaluate the internal energy and the entropy (see Problem 3.22). We limit the derivations to the specific heat, which is closely related to heat transfer calculations.

The characteristic temperature for rotation is usually very small for polyatomic molecules because of their large moments of inertia. Therefore, the rotational degrees of freedom can be assumed as fully excited in almost any practical situation. Each rotational degree of freedom will contribute  $\bar{R}/2$  to the molar specific heat. For molecules whose atoms are aligned (such as  $\text{CO}_2$ ), the rotational contribution to the specific heat is  $\bar{R}$ , and

$$\bar{c}_v = \frac{5}{2}\bar{R} + \bar{R} \sum_{i=1}^{3q-5} \frac{\zeta_i^2 e^{\zeta_i}}{(e^{\zeta_i} - 1)^2}, \quad \zeta_i = \Theta_{v,i}/T \quad (3.65)$$

If  $T \gg \Theta_{v,i}$ , then  $\bar{c}_v \rightarrow \bar{R}(3q - 2.5)$ . For molecules such as  $\text{H}_2\text{O}$  and  $\text{CH}_4$  whose atoms are not aligned, we have,

$$\bar{c}_v = 3\bar{R} + \bar{R} \sum_{i=1}^{3q-6} \frac{\zeta_i^2 e^{\zeta_i}}{(e^{\zeta_i} - 1)^2} \quad (3.66)$$

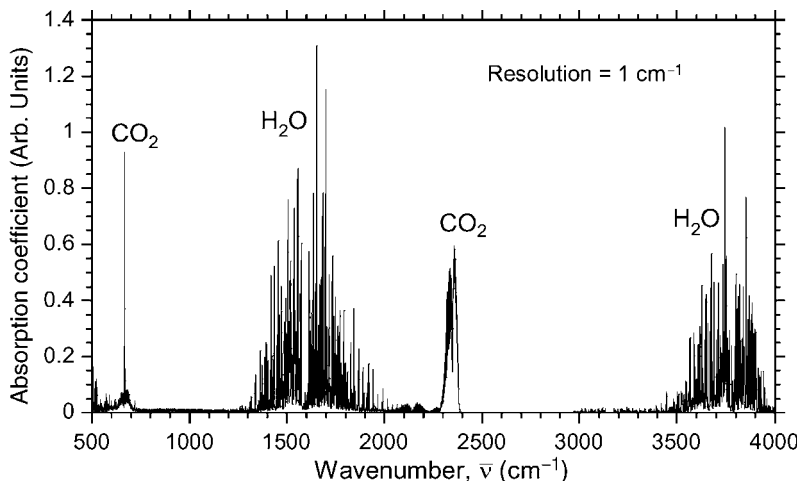


**TABLE 3.3** Vibrational Modes of Several Gases, Where the Integer in the Parentheses Indicates the Number of Degenerate Modes

Type	$\text{cm}^{-1}$	$\text{cm}^{-1}$	$\text{cm}^{-1}$	$\text{cm}^{-1}$	Total $f_v$
CO <sub>2</sub>	667 (2)	1343	2349	–	4
H <sub>2</sub> O	1595	3657	3756	–	3
CH <sub>4</sub>	1306 (3)	1534 (2)	2916	3019 (3)	9

In this case,  $\bar{c}_v \rightarrow \bar{R}(3q - 3)$  at  $T \gg \Theta_{v,i}$ . Again, electronic contribution may be significant at very high temperatures. Table 3.3 lists the vibrational frequencies for several commonly encountered gases. The unit of frequency is given in inverse centimeter ( $\text{cm}^{-1}$ ), which is often used in spectroscopic analyses. Note that  $\Theta_v = hv/k_B = hc_0\bar{\nu}/k_B$ , where  $\bar{\nu}$  is the wavenumber in  $\text{cm}^{-1}$  if we take  $c_0 = 3 \times 10^{10}$  cm/s. That is  $\Theta_v$  (K) =  $1.44 \bar{\nu}$  ( $\text{cm}^{-1}$ ). One can use this table to estimate the specific heat of these gases based on Eq. (3.65) or Eq. (3.66).

In reality, vibration-rotation interactions result in multiple absorption lines around each vibration mode, which can be observed through infrared absorption spectroscopy. Figure 3.12

**FIGURE 3.12** Infrared absorption spectrum of ambient air obtained with a Fourier-transform infrared spectrometer.

shows the molecular absorption spectra of CO<sub>2</sub> and H<sub>2</sub>O measured with a Fourier-transform infrared spectrometer. The absorption spectra were obtained by comparing the spectrum when the measurement chamber is open with that when the chamber is purged with a nitrogen gas, which does not absorb in the mid-infrared region. The concentrations of H<sub>2</sub>O and CO<sub>2</sub> in the experiments were not controlled since the purpose is to demonstrate the infrared absorption frequencies only. While the resolution of  $1 \text{ cm}^{-1}$  is not high enough to resolve very fine features, the absorption bands near  $670 \text{ cm}^{-1}$  due to degenerate bending modes and near  $2350 \text{ cm}^{-1}$  due to asymmetric stretching mode in CO<sub>2</sub> can be clearly seen. Note that the symmetric vibration mode of CO<sub>2</sub> at  $1343 \text{ cm}^{-1}$  is infrared inactive, i.e., it does not show up in the absorption spectrum but can be observed with Raman spectroscopy. Furthermore, the vibration-rotation interactions cause multiple lines in the water vapor absorption bands from  $1300$  to  $2000 \text{ cm}^{-1}$  and from  $3500$  to  $4000 \text{ cm}^{-1}$ .

**Example 3-6.** How many rotational degrees of freedom are there in a silane ( $\text{SiH}_4$ ) molecule? If a low-pressure silane gas is raised to a temperature high enough to completely excite its rotational and vibrational modes, find its specific heats.

**Solution.** For  $\text{SiH}_4$ , there will be three translational degrees of freedom, i.e.,  $\phi_t = 3$ , three rotational degrees of freedom, i.e.,  $\phi_r = 3$ , and  $\phi_v = 3q - 6 = 9$  vibrational degrees of freedom. If all the modes are excited, the specific heat for constant volume will be  $c_v = 1.5R + 1.5R + 9R = 12R$ . Given that  $M = 32$ , we find  $c_v = 3.12 \text{ kJ}/(\text{kg} \cdot \text{K})$ ,  $c_p = 3.38 \text{ kJ}/(\text{kg} \cdot \text{K})$ , and  $\gamma = 13R/12R = 1.083$ . The actual specific heats would be much smaller at moderate temperatures.

### 3.4 STATISTICAL ENSEMBLES AND FLUCTUATIONS

---

We have finished the discussion about statistical thermodynamics of independent particles without mentioning ensembles. In a system of independent particles, there is no energy associated with particle-particle interactions or the configuration of the particles. For dependent particles or dense fluids, the previous analysis can be extended by using *statistical ensembles*, which was pioneered by J. Willard Gibbs (1839–1903) in the late nineteenth century in his 1902 book, *Elementary Principles of Statistical Mechanics*. Statistical ensembles are a large set of macroscopically similar systems. When the properties are averaged over a properly chosen ensemble, the macroscopic properties can be considered as the same as the time-averaged quantity of the same system. There are three basic types of ensembles: microcanonical ensemble, canonical ensemble, and grand canonical ensemble.<sup>1,5</sup>

A *microcanonical ensemble* is composed of a large set of identical systems. Each system in the ensemble is isolated from others by rigid, adiabatic, and impermeable walls. The energy, volume, and number of particles in each system are constant. The results obtained using the microcanonical ensemble for independent particles are essentially the same as what we have obtained in previous sections. It is natural to ask the question as to what extent the statistical mechanics theory presented in previous sections will be valid for nanosystems. If the equilibrium properties are defined based on a large set of microcanonical ensembles and considered as the time-averaging properties of the system, there will be sufficiently large number of particles in the whole ensemble to guarantee the basic types of statistics, and the thermodynamics relations derived in Secs. 3.1 and 3.2 are still applicable. On the other hand, the difference between the energy levels due to quantization may be large enough to invalidate the substitution of summation with integration. We will discuss the energy level quantization further in Sec. 3.5. In deriving the properties of ideal gases in Sec. 3.3, the consideration of the translational, rotational, and vibrational degrees of freedom is on the basis of individual molecules. Therefore, the conclusions should be applicable to systems under thermodynamic equilibrium.

In a *canonical ensemble*, each system is separated from others by rigid and impermeable walls, which are diathermal. All systems have the same volume and number of particles. However, the systems can exchange energy. At equilibrium, the temperature  $T$  will be the same for all systems. An important result of applying the canonical ensemble is that the energy fluctuation (i.e., the standard deviation of energy of the system) is proportional to  $1/\sqrt{N}$ , where  $N$  is the total number of independent particles.

In a *grand canonical ensemble*, each system is separated from others by rigid, diathermal, and permeable walls. While the volume is fixed and is the same for each system, the number of particles as well as the energy of each system can vary. The temperature and the chemical potential must be the same for all systems at equilibrium. This allows the study of density fluctuations for each system. The result for monatomic molecules yields that the density fluctuation is also proportional to  $1/\sqrt{N}$ .

The canonical and grand canonical ensembles are essential for the study of complex thermodynamic systems, such as mixtures, chemical equilibria, dense gases, and liquids, which will not be further discussed in this text. Interested readers can find more details from Tien and Lienhard<sup>1</sup> and Carey.<sup>5</sup> A simple theory based on independent particles of phonons and electrons will be discussed in Chap. 5. While the partition function can also be used to study the thermodynamic relations of solids, the approach used in solid state physics will be adopted in a detailed study of the properties of solids presented in Chap. 7.

### 3.5 BASIC QUANTUM MECHANICS

So far we have largely avoided the derivations and equations involving quantum mechanics, by using the conclusions from quantum theory on a need basis without proof. In this section, we shall present the basics of quantum mechanics to enhance the understanding of the materials already presented and to provide some background for future chapters.

In classical mechanics, the state of a system is completely described by giving the position and the momentum of each particle in the system at any given time. The equation of motion is given in Eq. (3.1), which is also the basis for molecular dynamics. The position and the momentum of each particle are precisely determined using the initial values and the forces exerted on it afterward. According to the wave-particle duality, particles also have wave characteristics. The results are described in quantum mechanics by the Schrödinger wave equation. The solution of the Schrödinger equation is given in the form of a *wave-function*, which describes the probabilities of the possible outcome rather than the exact position and momentum of the particle. Another important aspect in quantum mechanics is the use of operators in mathematical manipulations.

#### 3.5.1 The Schrödinger Equation

Consider the following equation that describes a wave in the  $x$  direction (see Appendices B.6 and B.7):

$$\Psi(x,t) = \tilde{A}e^{i(2\pi x/\lambda - 2\pi\nu t)} \quad (3.67)$$

where  $\tilde{A} = A' + iA''$  is a complex constant,  $\lambda$  is the wavelength, and  $\nu$  is the frequency. One can take the real part of  $\Psi$ , i.e.,

$$\text{Re}(\Psi) = A' \cos(2\pi x/\lambda - 2\pi\nu t) - A'' \sin(2\pi x/\lambda - 2\pi\nu t)$$

which is a cosine function of  $x$  for any given  $t$ . The complex notation is convenient for obtaining derivatives. If Eq. (3.67) is used to describe a moving particle, with a mass  $m$  and a momentum  $p$ , it can be shown that

$$-i\hbar \frac{\partial}{\partial x} \Psi = \frac{h}{\lambda} \Psi = p\Psi \quad (3.68a)$$

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi = \frac{p^2}{2m} \Psi = E_K \Psi \quad (3.68b)$$

and 
$$i\hbar \frac{\partial}{\partial t} \Psi = h\nu \Psi = \varepsilon \Psi \quad (3.68c)$$

where  $\hbar = h/2\pi$ , which is the Planck constant divided by  $2\pi$ ,  $E_K$  is the kinetic energy of the particle, and  $\varepsilon$  is the total energy of the particle. In writing Eq. (3.68), we have applied

the concept of wave-particle duality to relate  $p = h/\lambda$  and  $\varepsilon = h\nu$ . If the particle possesses only the kinetic and potential energies, we have

$$\varepsilon = E_K + E_P = \frac{p^2}{2m} + \Phi(\mathbf{r}) \quad (3.69a)$$

where  $\Phi(\mathbf{r}) = \Phi(x,y,z)$  is the potential function that depends on the position of the particle. Define the Hamiltonian operator in the 3-D case as

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + \Phi(\mathbf{r}) \quad (3.69b)$$

It can be seen that  $\hat{H}\Psi = \varepsilon\Psi$ . Hence,

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + \Phi(\mathbf{r})\Psi = i\hbar\frac{\partial\Psi}{\partial t} \quad (3.70)$$

which is the time-dependent Schrödinger equation.<sup>8</sup> From  $\varepsilon\Psi = i\hbar\frac{\partial\Psi}{\partial t}$ , one can obtain

$$\Psi(\mathbf{r},t) = \Psi_0(\mathbf{r})e^{-ie_0t/\hbar} \quad (3.71a)$$

The general time dependence for different energy eigenvalues can be written as a summation:

$$\Psi(\mathbf{r},t) = A_1\Psi_{01}(\mathbf{r})e^{-ie_1t/\hbar} + A_2\Psi_{02}(\mathbf{r})e^{-ie_2t/\hbar} + \dots \quad (3.71b)$$

Therefore, the key to solve the Schrödinger equation becomes how to obtain the initial wavefunctions. For this reason, Eq. (3.70) can be rewritten as follows:

$$-\frac{\hbar^2}{2m}\nabla^2\Psi + \Phi(\mathbf{r})\Psi = \varepsilon\Psi \quad (3.72)$$

which is called the time-independent Schrödinger equation. The solution gives the wavefunction  $\Psi(\mathbf{r})$ , which is often expressed in terms of a set of eigenfunctions,  $\Psi_1, \Psi_2, \Psi_3, \dots$ , each with an eigenvalue energy,  $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$ , respectively. The solution, or the wavefunction, must satisfy

$$\int_V \Psi\Psi^* dV = 1 \quad (3.73)$$

where the superscript \* denotes the complex conjugate since the wavefunction is in general complex, and the integration is over the whole volume. The physical significance is that the probability of finding the particle in the volume must be 1. The wavefunction is also called a state function because it describes the quantum state of the particle, and  $\Psi\Psi^*$  is called the probability density function. The average or expectation value of any physical quantity  $\eta$  is calculated by

$$\langle \eta \rangle = \int_V \Psi^* \hat{\eta} \Psi dV \quad (3.74)$$

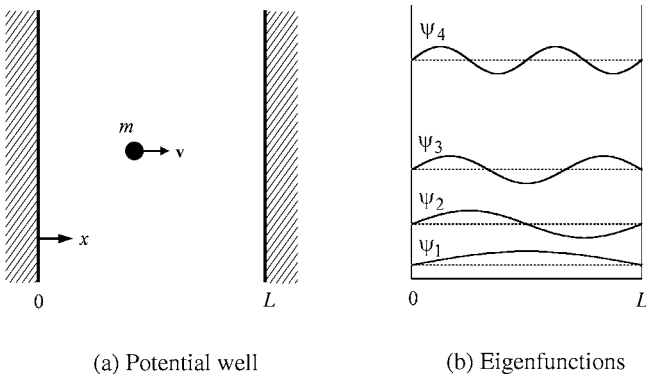
where  $\hat{\eta}$  signifies an operator of  $\eta$ . For example, the average energy of the particle is

$$\langle \varepsilon \rangle = \int_V \Psi^* \hat{H} \Psi dV \quad (3.75)$$

Several examples are discussed in the following sections to show how to obtain the wavefunctions and the physical significance of the solutions.

### 3.5.2 A Particle in a Potential Well or a Box

The 1-D potential well is illustrated in Fig. 3.13a, where a particle is confined within a physical space between  $0 < x < L$  and the particle can move parallel to the  $x$  axis only. This is



**FIGURE 3.13** Illustration of (a) a 1-D potential well and (b) the eigenfunctions.

equivalent of saying that the potential energy is zero inside and infinite outside the potential well, i.e.,

$$\Phi(x) = \begin{cases} 0, & \text{for } 0 < x < L \\ \infty, & \text{at } x = 0 \text{ and } L \end{cases} \quad (3.76)$$

The Schrödinger equation becomes

$$-\frac{\hbar^2}{2m} \nabla^2 \Psi = \varepsilon \Psi \quad (3.77)$$

whose solutions are  $\Psi(x) = A \cos(kx) + B \sin(kx)$ , where  $k = \sqrt{2m\varepsilon/\hbar^2}$ . Because the particle is confined inside the well, the wavefunction must be zero outside the potential well. Another requirement for the wavefunction is that it must be continuous. Thus, we must have  $\Psi(0) = \Psi(L) = 0$ . This requires that  $A = 0$  and, by taking only the positive  $k$  values, we have

$$kL = n\pi, \quad n = 1, 2, 3 \dots \quad (3.78)$$

The eigenfunctions are therefore  $\Psi_n(x) = B_n \sin(n\pi x/L)$ , which can be normalized by letting

$$\int_0^L \Psi_n(x) \Psi_n^*(x) dx = 1 \text{ to get}$$

$$\Psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right) \quad (3.79)$$

Therefore, the solution requires the particle to possess discretized energy values, i.e., its energy cannot be increased continuously but with finite differences between neighboring states. It can easily be seen that

$$\varepsilon_n = \frac{\hbar^2 n^2}{8mL^2} \quad (3.80)$$

The quantized energy eigenvalues are called energy levels for each quantum state, and the index  $n$  is called a quantum number. The eigenfunctions are standing waves as shown in Fig. 3.13b for the first four quantum states. For molecules, the difference between energy levels is very small and the energy distribution can often be approximated as a continuous distribution. For electrons at very small distances,  $L \rightarrow 10$  nm for example, quantization may be important. The effects of quantum confinement take place when the quantum well thickness becomes comparable to the de Broglie wavelength of the particle, such as electrons or holes in a semiconductor. Quantum wells can be formed by a sandwiched structure of heterogeneous layers, such as AlGaAs/GaAs/AlGaAs. The bandgap of the outer layers is larger than that of the inner layers to form an effective potential well. These structures are used for optoelectronic applications such as lasers and radiation detectors. The thickness of the active region can be a few nanometers. In some cases, multiple quantum wells are formed with periodic layered structures, called *superlattices*, which have unique optical, electrical, and thermal properties.

**Example 3-7.** Derive the uncertainty principle. Suppose the wavefunction is given by Eq. (3.79) for a particle with energy  $\varepsilon_n$  given in Eq. (3.80).

**Solution.** To find the average position of the particle, we use

$$\langle x \rangle = \int_0^L \Psi^* x \Psi dx = \frac{2}{L} \int_0^L x \sin^2 \left( \frac{n\pi x}{L} \right) dx = \frac{L}{2}$$

The variance of  $x$ ,  $\sigma_x^2 = \langle x - \langle x \rangle \rangle^2 = \langle x^2 \rangle - 2\langle x \rangle^2 + \langle x \rangle^2 = \langle x^2 \rangle - \langle x \rangle^2$ . With

$$\langle x^2 \rangle = \frac{2}{L} \int_0^L x^2 \sin^2 \left( \frac{n\pi x}{L} \right) dx = \frac{L^2}{3} - \frac{L^2}{2n^2\pi^2}$$

we obtain the standard deviation of  $x$  as

$$\sigma_x = L \left( \frac{1}{12} - \frac{1}{2n^2\pi^2} \right)^{1/2}$$

For the momentum, we use the operator  $p \rightarrow -i\hbar(\partial/\partial x)$ . Hence,

$$\langle p \rangle = \int_0^L \Psi^* \left( -i\hbar \frac{d\Psi}{dx} \right) dx = -i\hbar \frac{2n\pi}{L^2} \int_0^L \sin \left( \frac{n\pi x}{L} \right) \cos \left( \frac{n\pi x}{L} \right) dx = 0$$

and

$$\langle p^2 \rangle = \int_0^L \Psi^* (-\hbar^2) \frac{d^2\Psi}{dx^2} dx = \left( \frac{n\pi\hbar}{L} \right)^2$$

We have  $\sigma_p = n\pi\hbar/L$  and obtain the following expression:

$$\sigma_x \sigma_p = \frac{\hbar}{2} \left( \frac{\pi^2 n^2}{3} - 2 \right)^{1/2} \quad (3.81)$$

Taking the smallest quantum number,  $n = 1$ , we get  $\sigma_x \sigma_p \approx 0.5678\hbar > \hbar/2$ , which is a proof of the uncertainty principle given in Eq. (3.9).

Next, consider a free particle in a 3-D box,  $0 < x < a$ ,  $0 < y < b$ ,  $0 < z < c$ . It can be shown that the (normalized) eigenfunctions are

$$\Psi_{x,y,z} = \sqrt{\frac{8}{abc}} \sin\left(\frac{n_x \pi x}{a}\right) \sin\left(\frac{n_y \pi y}{b}\right) \sin\left(\frac{n_z \pi z}{c}\right) \quad (3.82)$$

with the energy eigenvalues:

$$\varepsilon_{x,y,z} = \frac{\hbar^2}{8m} \left( \frac{n_x^2}{a^2} + \frac{n_y^2}{b^2} + \frac{n_z^2}{c^2} \right) \quad (3.83)$$

where  $n_x, n_y, n_z = 1, 2, 3, \dots$ . When  $a = b = c = V^{1/3}$ , Eq. (3.83) can be simplified as

$$\varepsilon_{x,y,z} = \frac{\hbar^2}{8mV^{2/3}} (n_x^2 + n_y^2 + n_z^2) \quad (3.84)$$

Let  $\eta = (n_x^2 + n_y^2 + n_z^2)^{1/2}$ , then we can evaluate the number of quantum states between  $\eta$  and  $\eta + d\eta$ , which is nothing but the degeneracy. For sufficiently large  $V$ , the quantum states are so close to each other that the volume within the spherical shell between  $\eta$  and  $\eta + d\eta$  is equal to the number of quantum states. Only one-octant of the sphere is considered in Eq. (3.84) because  $n_x > 0$ ,  $n_y > 0$ ,  $n_z > 0$ . The total volume is therefore one-eighth of the spherical shell; hence,

$$dg = \frac{1}{8} 4\pi \eta^2 d\eta = \frac{2\pi V (2m)^{3/2}}{\hbar^3} \varepsilon^{1/2} d\varepsilon \quad (3.85)$$

With  $\varepsilon = \frac{1}{2}mv^2$  and  $d\varepsilon = mv dv$ , we obtain

$$dg = \frac{m^3 V}{\hbar^3} 4\pi v^2 dv \quad (3.86)$$

This equation is essentially the same as Eq. (3.32), with  $dx dy dz = V$  and  $dv_x dv_y dv_z = 4\pi v^2 dv$ . Equation (3.86) provides a rigid proof of Eq. (3.32), which is the translational degeneracy. It should be noted that the classical statistical mechanics results in the same expression for  $U$  and  $p$ , as well as the Maxwell velocity distribution for ideal gases. However, the constant  $h$  must be included to correctly express  $S$  as in Eq. (3.41). Equation (3.86) will also be used in Chap. 5 to study the free electron gas in metals. When using the momentum  $p = mv$  as the variable, we have

$$dg = \frac{V}{\hbar^3} 4\pi p^2 dp \quad (3.87)$$

Because Eq. (3.87) does not involve mass, it is also applicable to phonons and photons as will be discussed in Chaps. 5 and 8.

### 3.5.3 A Rigid Rotor

The rigid rotor model can be used to study the rotational movement of diatomic molecules as well as the movement of an electron in a hydrogen atom. Consider two particles separated by a fixed distance  $r_0 = r_1 + r_2$  as shown in Fig. 3.14. The masses of the particles are

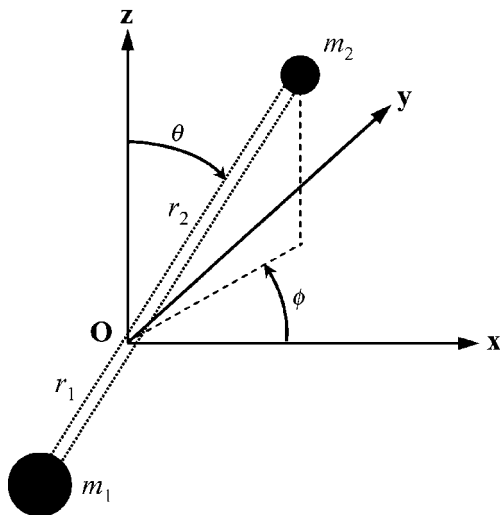


FIGURE 3.14 Schematic of a rotor consisting of two particles.

$m_1$  and  $m_2$ , respectively. Since the center of mass is at the origin, we have  $m_1 r_1 = m_2 r_2$ . The moment of inertia is

$$I = m_1 r_1^2 + m_2 r_2^2 = m_r r_0^2 \quad (3.88)$$

where  $m_r = m_1 m_2 / (m_1 + m_2)$  is the reduced mass. We can study the rotational movement of the particles by considering a particle with a mass of  $m_r$  that rotates around at a fixed distance  $r = r_0$  from the origin in the  $\theta$  and  $\phi$  directions. In the spherical coordinates,

$$\nabla^2 = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial}{\partial r} \right) + \frac{1}{r^2 \sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial}{\partial \theta} \right) + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \quad (3.89)$$

Because  $r \equiv r_0$ , the derivative with respect to  $r$  vanishes. The potential energy is zero for free rotation. By setting the mass to be  $m_r$  and  $\Phi = 0$  in Eq. (3.72) and noticing that  $m_r r_0^2 = I$ , we obtain

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \Psi}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \Psi}{\partial \phi^2} = -\frac{2I\epsilon}{\hbar^2} \Psi \quad (3.90)$$

This partial differential equation can be solved by separation of variables. We get two ordinary differential equations by letting  $\Psi(\theta, \phi) = P(\theta)\psi(\phi)$ , i.e.,

$$\frac{d^2 \psi}{d\phi^2} = -m^2 \psi \quad (3.91)$$

and

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{dP}{d\theta} \right) + \left( \frac{2I\epsilon}{\hbar^2} - \frac{m^2}{\sin^2 \theta} \right) P = 0 \quad (3.92)$$

Here,  $m$  is a new eigenvalue, and the periodic boundary conditions shall be applied to  $P$  and  $\psi$ , respectively. The solution of Eq. (3.91) is readily obtained as

$$\psi(\phi) = A e^{im\phi} \quad (3.93)$$



with  $m = 0, \pm 1, \pm 2, \dots$ , to satisfy the periodic boundary conditions:  $\psi(\phi) = \psi(2\pi + \phi)$ . A transformation,  $\cos \theta = \xi$ , can be used so that Eq. (3.92) becomes

$$(1 - \xi^2) \frac{d^2 P}{d\xi^2} - 2\xi \frac{dP}{d\xi} + \left( \frac{2I\varepsilon}{\hbar^2} - \frac{m^2}{1 - \xi^2} \right) P = 0 \quad (3.94)$$

Because  $\theta$  is defined from 0 and  $\pi$ , we have  $-1 \leq x \leq 1$ . In order for Eq. (3.94) to have solutions that are bounded at  $x = \pm 1$ ,  $2I\varepsilon/\hbar^2 = l(l + 1)$ , where  $l$  is an integer that is greater than or at least equal to the absolute value of  $m$ . Therefore, the energy eigenvalues are

$$\varepsilon_l = \frac{\hbar^2}{2I} l(l + 1), \quad l = |m|, |m| + 1, |m| + 2, \text{ etc.} \quad (3.95)$$

Equation (3.94) is called the associated Legendre differential equation. The solutions are the associated Legendre polynomials given as

$$P_l^m(\xi) = \frac{(1 - \xi^2)^{m/2}}{l! 2^l} \frac{d^{m+1}}{d\xi^{m+1}} (\xi^2 - 1)^l \quad (3.96)$$

Finally, after normalization, the standing wavefunctions can be expressed as

$$\Psi_l^m(\theta, \phi) = \frac{1}{\sqrt{2\pi}} \left[ \frac{(2l + 1)(l - m)!}{2(m + 1)!} \right]^{1/2} P_l^m(\cos \theta) e^{im\phi} \quad (3.97)$$

**Discussion.** It can be seen that Eq. (3.95) is identical to Eq. (3.52). The energy level is determined by the principal quantum number  $l$ . On the other hand, for each  $l$ , there are  $2l + 1$  quantum states corresponding to each individual  $m$ , because  $m$  can take 0,  $\pm 1$ ,  $\pm 2$  up to  $\pm l$ . This means that the degeneracy  $g_l = 2l + 1$ . When the two atoms are identical, such as in a nitrogen molecule, the atoms are indistinguishable when they switch positions. The degeneracy is reduced by a symmetry number, as given in the expression of Eq. (3.53). It should be noted that the nuclear spin degeneracy is important for hydrogen (see Problem 3.27).<sup>1</sup>

### 3.5.4 Atomic Emission and the Bohr Radius

A hydrogen atom is composed of a proton and an electron. Since the mass of the proton is much greater than that of the electron, it can be modeled as the electron moving around the nucleus. The mass of the electron is  $m_e = 9.11 \times 10^{-31}$  kg, and the position of the electron can be described in the spherical coordinates as  $\mathbf{r} = (r, \theta, \phi)$ . The force exerted on the electron is Coulomb's force, which gives a potential field

$$\Phi(r) = -\frac{C_1}{r} \quad (3.98)$$

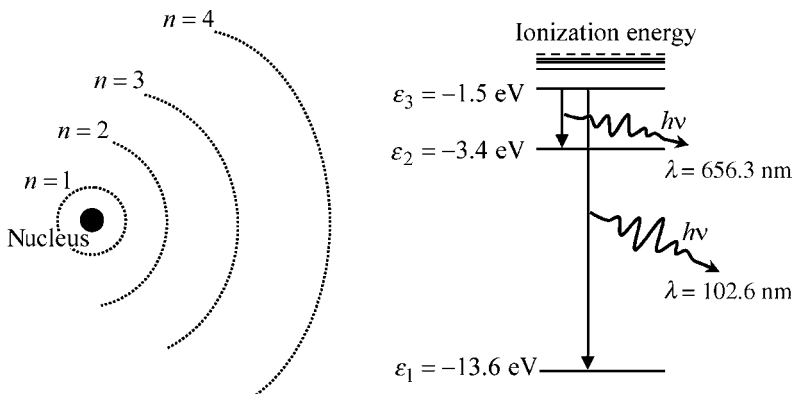
where  $C_1 = e^2/(4\pi\varepsilon_0) = 2.307 \times 10^{-28} \text{ N} \cdot \text{m}^2$ , with the electron charge  $e = 1.602 \times 10^{-19} \text{ C}$  and the dielectric constant  $\varepsilon_0 = 8.854 \times 10^{-12} \text{ F/m}$ . Let  $\Psi(r, \theta, \phi) = R(r)P(\theta)\psi(\phi)$ . In doing the separation of variables, we notice that the potential  $\Phi$  is independent of  $\theta$  and  $\phi$ , and the total energy is equal to the sum of the rotational energy and the energy associated with  $r$ . The eigenvalues for the rotational energy are given in Eq. (3.95). Using Eq. (3.72) and Eq. (3.89), we can write the equation for  $R(r)$  as follows:

$$\frac{\hbar}{2m_e} \frac{d}{dr} \left( r^2 \frac{dR}{dr} \right) + \left( \frac{C_1}{r} + \varepsilon - \frac{l(l + 1)\hbar^2}{2I} \right) R = 0 \quad (3.99)$$

which is the associated Laguerre equation, and its solutions are the associated Laguerre polynomials. The solutions give the energy eigenvalues as<sup>5,8</sup>

$$\varepsilon_n = -\frac{m_e C_1^2}{2\hbar^2 n^2} \quad (3.100)$$

where the negative values are used for convenience to show that the energy increases with the principal quantum number  $n$ . For  $n = 1$ ,  $-m_e C_1^2/2\hbar^2 = -13.6$  eV, as shown in Fig. 3.15.



**FIGURE 3.15** Electron orbits (left) and energy levels (right) in a hydrogen atom. The *ionization energy* is the energy required for an electron to escape the orbit.

Note that  $1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$ . When the electron is in a higher energy state, it has a tendency of relaxing to a lower energy state by spontaneously emitting a photon, with precisely the same energy as given by the energy difference between the two energy levels:

$$h\nu = \varepsilon_i - \varepsilon_j = \frac{m_e C_1^2}{2\hbar^2} \left( \frac{1}{n_j^2} - \frac{1}{n_i^2} \right) \quad (3.101)$$

The emission or absorption of photons by electrons is called *electronic transitions*. When  $i = 3$  and  $j = 1$ , we have  $h\nu = 12.1$  eV, corresponding to the wavelength of 102.6 nm (ultraviolet), which is the second line in the Lyman series. When  $i = 3$  and  $j = 2$ , we have  $h\nu = 1.89$  eV, corresponding to the wavelength of 656.4 nm (red), which is the first line in the Balmer series. A more detailed description of the atomic emission lines can be found from Sonntag and van Wylen.<sup>2</sup>

The next question is: What is the radius of a particular electron orbit? This is an important question because it gives us a sense of how small an atom is. When a particle is in an orbit, the classical force balance gives that

$$\frac{C_1}{r^2} = m_e \left( \frac{v^2}{r} \right) \quad (3.102)$$

which is to say that  $E_K = m_e v^2/2 = C_1/2r$ , and the sum of the kinetic and potential energies is

$$\varepsilon = E_K + E_P = \frac{C_1}{2r} - \frac{C_1}{r} = -\frac{C_1}{2r} \quad (3.103)$$

Equations (3.100) and (3.103) can be combined to give discrete values of the radius of each orbit in the following:

$$r_n = \frac{\hbar^2}{m_e C_1} n^2 = a_0 n^2 \quad (3.104)$$

When the electron is in the innermost orbit, the radius is given by  $a_0 = \epsilon_0 \hbar^2 / (\pi m_e e^2) = 0.0529$  nm, which is called the *Bohr radius*. Therefore, the hydrogen atom in its ground state can be considered as having a diameter of approximately  $1 \text{ \AA}$  (Angstrom), or  $0.1$  nm. Niels Bohr (1885–1962) was a Danish physicist who received the Nobel Prize in Physics in 1922 for his contributions to the understanding of the structure of atoms and quantum physics. One should accept the quantum interpretation of the electron radius as a characteristic length, not the exact distance that the electron would rotate around the nucleus in the same manner a planet rotates around a star.

### 3.5.5 A Harmonic Oscillator

The last example of quantum mechanics is the linear spring as shown in Fig. 3.16. Consider a 1-D oscillator with a mass  $m$  and the spring force  $F(x) = -Kx$ . The origin can be selected such that  $F(0) = 0$ . It can be shown that the potential is

$$\Phi(x) = - \int_0^x F(x) dx = \frac{1}{2} Kx^2 \quad (3.105)$$

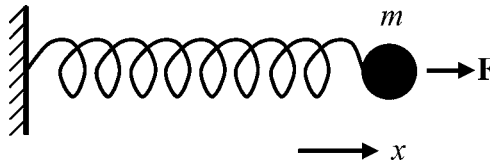


FIGURE 3.16 A linear spring.

From classical mechanics, we can solve Newton's equation  $m\ddot{x} + Kx = 0$  to obtain the solution

$$x = A \sin(\omega t + \phi_0) \quad (3.106)$$

where constant  $A$  is the amplitude, constant  $\phi_0$  is the initial phase, and parameter  $\omega = \sqrt{K/m}$  is the angular resonance frequency.

It can be shown that the total energy  $\epsilon = E_K + E_P = KA^2/2$  is a constant and the maximum displacement is  $A$ . The velocity is the largest at  $x = 0$  and zero at  $x = \pm A$ .

The Schrödinger wave equation can be written as

$$\frac{\hbar^2}{2m} \frac{d^2\Psi}{dx^2} + \left( \epsilon - \frac{Kx^2}{2} \right) \Psi = 0 \quad (3.107)$$

with the boundary condition being  $\Psi(x) = 0$  at  $x \rightarrow \pm \infty$ . The constants can be grouped by using  $\alpha = 2m\epsilon/\hbar^2$  and  $\beta = \sqrt{Km}/\hbar$ . Then Eq. (3.107) can be transformed by using  $\xi = \sqrt{\beta}x$  and  $\Psi(x) = Q(\xi) \exp(-\xi^2/2)$  to

$$\frac{d^2Q}{d\xi^2} - 2\xi \frac{dQ}{d\xi} + \left( \frac{\alpha}{\beta} - 1 \right) Q = 0 \quad (3.108)$$

This is the Hermite equation, and the solutions are Hermite polynomials given by

$$H_n(\xi) = (-1)^n e^{\xi^2} \frac{d^n}{d\xi^n} \left( e^{-\xi^2} \right) \quad (3.109)$$

when  $\alpha$  and  $\beta$  must satisfy the eigenvalue equation:

$$\frac{\alpha}{\beta} - 1 = 2n, \quad n = 0, 1, 2, \dots \quad (3.110)$$

The normalized wavefunctions can be written as

$$\Psi_n(x) = \left( \frac{\sqrt{\beta/\pi}}{n!2^n} \right)^{1/2} H_n(\beta^{1/2}x) \exp\left(-\frac{\beta x^2}{2}\right) \quad (3.111)$$

The energy eigenvalues can be obtained from Eq. (3.110) as

$$\varepsilon_n = \left(n + \frac{1}{2}\right) \hbar \sqrt{K/m} = \left(n + \frac{1}{2}\right) \hbar \omega \quad (3.112)$$

The above equation was used to study the vibrational contributions in diatomic molecules; see Eq. (3.59). The  $1/2$  term was not included in Planck's original derivation of the black-body radiation function. The significance lies in that if the ground-state energy is zero, both its kinetic energy and potential energy must be zero, suggesting that both the position and the momentum must be zero. This would violate the uncertainty principle. As mentioned earlier, in classical mechanics, the particle is limited to the region  $-A < x < A$ , where  $A$  is the amplitude given in Eq. (3.106). This is not the case in the quantum theory, as shown in Fig. 3.17, for the first few energy levels and the associated wavefunctions. Notice that probability density function  $\Psi^2$  is nonzero even though the absolute value of  $x$  exceeds  $\sqrt{2\varepsilon/K}$ .

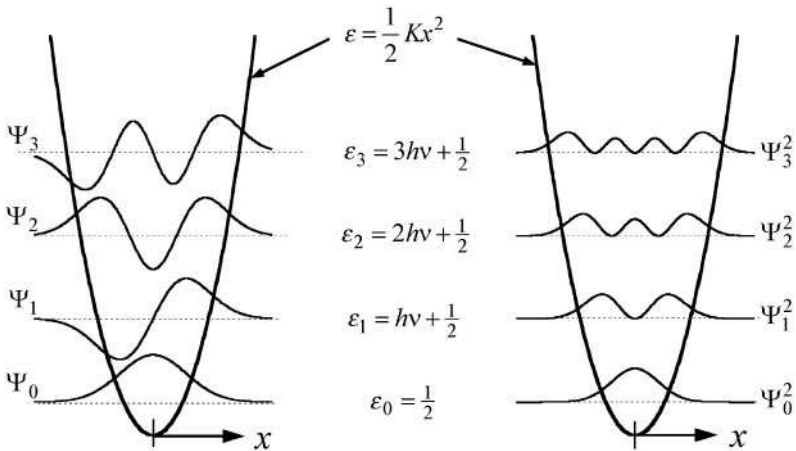


FIGURE 3.17 Wavefunctions and probability density functions for vibration energy levels.

The application of quantum theory allows us to predict the specific heat of ideal gases. In deriving the equations shown in Sec. 3.3.3, we have largely neglected nonlinear and anharmonic vibration, electronic contribution, and dissociation. These factors may become important at very high temperatures. The degeneracy due to the coupling of rotation and vibration can cause multiple absorption/emission lines in the infrared in polyatomic molecular gases, as shown in Fig. 3.12.

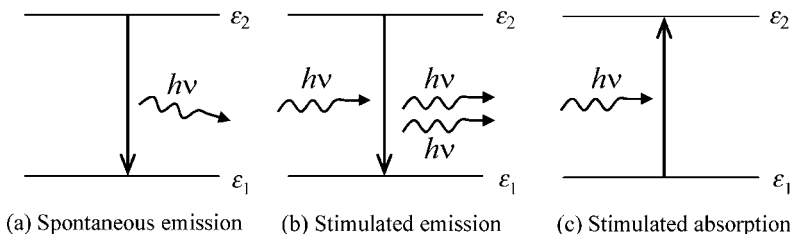
### 3.6 EMISSION AND ABSORPTION OF PHOTONS BY MOLECULES OR ATOMS

---

We have learned that the emission of photons is associated with transitions from a higher energy level to a lower energy level that reduces the total energy of the molecular system. The reverse process is the absorption of photons that increases the energy of the system through transitions from lower energy levels to higher energy levels. As discussed earlier, an electronic transition requires a large amount of energy, and the emitted or absorbed photons are at frequencies from deep ultraviolet ( $\lambda \approx 100$  nm) to slightly beyond the red end of the visible region ( $\lambda \approx 1 \mu\text{m}$ ). On the other hand, vibration or rotation-vibration modes lie in the mid-infrared ( $2.5 \mu\text{m} < \lambda < 25 \mu\text{m}$ ), while their overtones or higher-order harmonics lie in the near-infrared region ( $0.8 \mu\text{m} < \lambda < 2.5 \mu\text{m}$ ). Rotational modes alone may be active in the far-infrared and microwave regions ( $\lambda > 25 \mu\text{m}$ ). Transitions between different energy levels of the molecules or atoms are called *bound-bound transitions*, because these energy states are called *bound states*. Bound-bound transitions happen at discrete frequencies due to quantization of energy levels. Dissociation or ionization can also occur at high temperatures. The difference between adjacent energy levels is very small because the electrons can move freely (i.e., not bound to the atom or the molecule). Therefore, *free-free* or *bound-free transitions* happen in a broadband of frequencies. In gases, these broader transitions occur only at extremely high temperatures.

If a molecule at elevated energy states were placed in a surrounding at zero absolute temperature (i.e., empty space), it would lower its energy states by emitting photons in all directions until reaching its ground state. However, the emission processes should occur spontaneously regardless of the surroundings. Suppose the molecule is placed inside an isothermal enclosure, after a long time, the energy absorbed must be equal to that emitted to establish a thermal equilibrium with its surroundings. The thermal fluctuation of oscillators is responsible for the equilibrium distribution, i.e., Planck's law developed in 1900. Einstein examined how matter and radiation can achieve thermal equilibrium in a fundamental way and published a remarkable paper, "On the quantum theory of radiation" in 1917.<sup>11</sup> The interaction of radiation with matter is essentially through emission or absorption at the atomistic dimension, although solids or liquids can reflect radiation and small particles can scatter radiation. Einstein noticed that spontaneous emission and pure absorption (i.e., transition from a lower level to a higher level by absorbing the energy from the incoming radiation) alone would not allow an equilibrium state of an atom to be established with the radiation field. He then hypothesized the concept of *stimulated* or *induced emission*, which became the underlying principle of lasers. In a stimulated emission process, an incoming photon interacts with the atom. The interaction results in a transition from a higher energy state to a lower energy state by the emission of another photon of the same energy toward the same direction as the incoming photon. Saying in other words, the stimulated photon is a *clone* of the stimulating photon with the same energy and momentum. Whether an incoming photon will be absorbed, will stimulate another, or will pass by without any effect on the atom is characterized by the probabilities of these events. Understanding the emission and absorption processes is important not only for coherent emission but also for thermal radiation.<sup>12</sup> While more detailed treatments will be given in later chapters, it is important to gain a basic understanding of the quantum theory of radiative transitions and microscopic description of the radiative properties.

Consider a canonical ensemble of single molecules or atoms, with two nondegenerate energy levels,  $\varepsilon_1$  and  $\varepsilon_2$  ( $\varepsilon_2 > \varepsilon_1$ ), in thermal equilibrium with an enclosure or cavity at temperature  $T$ . Suppose the total number of particles is  $N$ , and let  $N_1$  and  $N_2$  be the number of particles at the energy level corresponding to  $\varepsilon_1$  and  $\varepsilon_2$ , respectively. These particles do not interact with each other at all. The concept of canonical ensemble can be understood as if each cavity has only one atom, but there are  $N$  single-atom cavities with one atom in each cavity. As shown in Fig. 3.18, there are three possible interaction mechanisms, i.e., spontaneous



**FIGURE 3.18** Illustration of the emission and absorption processes. (a) Spontaneous emission. (b) Stimulated emission. (c) Stimulated absorption.

emission, stimulated emission, and *stimulated or induced absorption*. Here, stimulated absorption refers to the process that the energy of the photon is absorbed, and consequently, the transition occurs from the lower energy level to the higher energy level. In a stimulated absorption process, the number of photons before the process is 1 and after the process is  $1 - 1 = 0$ . In a stimulated emission process, the number of photons beforehand is 1 and afterward is  $1 + 1 = 2$ . Therefore, stimulated emission is regarded also as *negative absorption*. Each of the photons involved in this process will have an energy equal to  $h\nu = \varepsilon_2 - \varepsilon_1$  and a momentum  $h\nu/c$ .

Transition from the higher energy level to the lower energy level cannot take place if the population of atoms on the higher energy level,  $N_2 = 0$ , and vice versa. Einstein further assumed that the probability of transition is proportional to the population at the initial energy level, and spontaneous transition should be independent of the radiation field. Hence, the rate of transition from  $\varepsilon_2$  to  $\varepsilon_1$  due to spontaneous emission can be written as

$$\left(\frac{dN_1}{dt}\right)_A = -\left(\frac{dN_2}{dt}\right)_A = AN_2 \quad (3.113)$$

where  $A$  is *Einstein's coefficient* of spontaneous emission. On the other hand, the transition rate due to stimulated emission should also be proportional to the energy density of the radiation field  $u(\nu, T)$ . Thus,

$$\left(\frac{dN_1}{dt}\right)_B = BN_2u(\nu, T) \quad (3.114)$$

Stimulated absorption will cause a transition rate that is proportional to  $N_1$  and  $u(\nu, T)$ :

$$\left(\frac{dN_1}{dt}\right)_C = -CN_1u(\nu, T) \quad (3.115)$$

In Eq. (3.114) and Eq. (3.115), constants  $B$  and  $C$  are Einstein's coefficients of stimulated emission and absorption, respectively. The combination of these processes must maintain a zero net change of the populations at equilibrium. Thus,

$$AN_2 + BN_2u(\nu, T) - CN_1u(\nu, T) = 0 \quad (3.116)$$

Atoms or molecules in a thermal equilibrium are described by the Maxwell-Boltzmann statistics of molecular gases given by Eq. (3.26):  $N_1/N_2 = e^{(\varepsilon_2 - \varepsilon_1)/k_b T} = e^{h\nu/k_b T}$ . Therefore, Eq. (3.116) can be rewritten as

$$u(\nu, T) = \frac{A/B}{(C/B)e^{h\nu/k_b T} - 1} \quad (3.117)$$

Comparing this equation with Planck's distribution, Eq. (8.41) in Chap. 8, we see that  $B = C$  and  $A/B = 8\pi h\nu^3/c^3$ . The two-level system can easily be generalized to arbitrary energy levels to describe the fundamental emission and absorption processes. The emission and absorption processes not only exchange energy between the field and the atom but also transfer momentum. How will an atom move inside a cavity? The phenomenon of a molecule or atom in a radiation field is like the Brownian motion, in which the radiation quanta exert forces on the molecule or the atom as a result of momentum transfer during each emission or absorption process. Consequently, the molecule or the atom will move randomly following Maxwell's velocity distribution at the same temperature as the radiation field. The equilibrium radiation field, which obeys the quantum statistics (i.e., BE statistics) that was not realized until 1924, and the motion of a molecular gas, which obeys classical statistics, can be coupled to each other to become mutual equilibrium. Einstein also asserted that each spontaneously emitted photon must be directional, while the probability of spontaneous emission should be the same in all directions. In fact, Einstein's 1917 paper complemented Planck's 1900 paper on radiation energy quanta and his own 1905 paper on photoelectric emission and, thus, provided a complete description of the quantum nature of photons, although the name "photon" was not coined until 1928.

At moderate temperatures, the population at higher energy states is too small for stimulated emission to be of significance for optical and thermal radiation. Thus, the absorption comes solely from induced absorption. When stimulated emission is important, the contributions of stimulated emission and stimulated absorption cannot be separated by experiments. The effect is combined to give an effective absorption coefficient by taking stimulated emission as negative absorption, whereas the emission of radiation includes solely the spontaneous emission.<sup>12</sup> The effective absorption coefficient is proportional to the population difference,  $N_1 - N_2$ . On the other hand, if a *population inversion* can be created and maintained such that  $N_2 > N_1$ , the material is called a *gain medium* or *active medium*. In an active medium, stimulated emission dominates stimulated absorption so that more and more photons will be cloned and the radiation field be amplified coherently. The principle of stimulated emission was applied in 1950s and early 1960s for the development of maser, which stands for *microwave amplification by stimulated emission of radiation*, and laser, which stands for *light amplification by stimulated emission of radiation*.<sup>13</sup> Lasers have become indispensable to modern technologies and daily life.

### 3.7 ENERGY, MASS, AND MOMENTUM IN TERMS OF RELATIVITY

---

Special theory of relativity or *special relativity* predicts that energy and mass can be converted to each other. If we retain the definition of mass as in the classical theory, only energy conservation is the fundamental law of physics. The mass does not have to be conserved. On the other hand, for processes that do not involve changes below the atomic level or inside the nuclei, the mass can indeed be considered as conserved. According to the special relativity, the rest energy of a free particle is related to its mass and the speed of light by

$$E_0 = mc^2 \quad (3.118)$$

The rest energy is simply the energy when the particle is not moving relative to the reference frame. Suppose the free particle is moving at a velocity  $v$  in a given reference frame, then its momentum is given by<sup>14</sup>

$$p = \frac{mv}{\sqrt{1 - v^2/c^2}} \quad (3.119)$$

When  $v \ll c$ , Eq. (3.119) reduces to the classical limit, i.e.,  $p = mv$ . It can be seen that for a particle with nonzero mass, its momentum would increase as  $v \rightarrow c$  without any bound. There is no way we could accelerate a particle to the speed of light. If there is anything that travels with the speed of light, it has to be massless, i.e.,  $m = 0$ . An example of massless particles is the light quanta or photons. The kinetic energy can be evaluated by integrating the work needed to accelerate a particle, i.e.,

$$E_K = \int_0^x F dx = \int_0^x \frac{dp}{dt} dx = \int_0^x \frac{dp}{dv} \frac{dv}{dt} dx = \int_0^v \frac{dp}{dv} v dv$$

Using Eq. (3.119), we find that

$$E_K = \frac{mc^2}{\sqrt{1 - v^2/c^2}} - mc^2 \quad (3.120)$$

When  $v \ll c$ , we have  $1/\sqrt{1 - v^2/c^2} \approx 1 + v^2/2c^2$  so that  $E_K = mv^2/2 = p^2/2m$  in the low-speed limit. In the relativistic limit, however,  $E_K$  will be on the order of  $mc^2$ . Because energy is additive, the total energy of a moving free particle is

$$E = E_K + E_0 = \frac{mc^2}{\sqrt{1 - v^2/c^2}} \quad (3.121)$$

Obviously, the energy of a particle would become infinite if its speed approaches the speed of light, unless its mass goes to zero. It can be shown that  $E^2 - E_0^2 = m^2c^4/(1 - v^2/c^2) - m^2c^4 = p^2c^2$ , where  $p$  is given in Eq. (3.119). This gives another expression of energy in terms of the rest energy, the momentum, and the speed of light as follows:

$$E^2 = m^2c^4 + p^2c^2 \quad (3.122)$$

It should be noted that, in general,  $pc$  is not equal to the kinetic energy. For  $v \ll c$ , the total energy is approximately the same as the rest energy. Comparing Eq. (3.119) and Eq. (3.121), we notice that  $E = pc(c/v)$ . Therefore, when  $v \rightarrow c$ , we see that  $E \rightarrow pc$  (which is unbounded unless  $m = 0$ ). For a photon that travels at the speed of light, in order for the above equations to be meaningful, we must set its mass to zero. From Eq. (3.122), we have for photons that

$$p = \frac{E}{c} = \frac{h\nu}{c} \quad (3.123)$$

which is the same as Eq. (3.7) in Sec. 3.1.3. By noting that  $\lambda\nu = c$ , we obtain

$$\lambda = \frac{h}{p} \quad (3.124)$$

The kinetic energy of a photon is  $pc$  or  $h\nu$  since its rest energy is zero. One should not attempt to calculate the kinetic energy of a photon by  $\frac{1}{2}mv^2$ , because photons are not only massless but also *relativistic particles*, for which the energy and momentum must be evaluated according to the above mentioned equations. While photons do not have mass, it has been observed that photons can be used to create particles with nonzero mass or vice versa, as in *creation* or *annihilation reactions*. High energy physics has proven that mass is not always conserved. Furthermore, energy and mass can be interconverted. A small amount of mass can be converted into a large amount of energy, as in a nuclear reaction.



### 3.8 SUMMARY

---

This chapter started with very basic independent particle systems to derive the three major statistics, i.e., the Maxwell-Boltzmann, Bose-Einstein, and Fermi-Dirac statistics. The classical and quantum statistics were then applied to thermodynamic systems, providing microscopic interpretations of the first, second, and third laws of thermodynamics, as well as Bose-Einstein condensate. The velocity distribution and specific heat of ideal gases were explained based on the semi-classical statistics, followed by a brief description of quantum mechanics to understand the quantization of translational, rotational, and vibrational modes. The fundamental emission and absorption processes of molecules or atoms were discussed along with the concept of stimulated emission. Finally, matter-energy conversion was described within the framework of the relativistic theory. While most of the explanations in this chapter are semi-classical and somewhat oversimplified, it should provide a solid background to those who did not have a formal education in statistical mechanics and quantum physics. These materials will be frequently referenced in the rest of the book.

### REFERENCES

---

1. C. L. Tien and J.H. Lienhard, *Statistical Thermodynamics*, Hemisphere, New York, 1985.
2. R. E. Sonntag and G.J. van Wylen, *Fundamentals of Statistical Thermodynamics*, Wiley, New York, 1966.
3. J. E. Lay, *Statistical Mechanics and Thermodynamics of Matter*, Harper Collins Publishers, New York, 1990.
4. C. E. Hecht, *Statistical Thermodynamics and Kinetic Theory*, W. H. Freeman and Company, New York, 1990.
5. V. P. Carey, *Statistical Thermodynamics and Microscale Thermophysics*, Cambridge University Press, Cambridge, UK, 1999.
6. F. C. Chou, J. R. Lukes, X. G. Liang, K. Takahashi, and C. L. Tien, "Molecular dynamics in microscale thermophysical engineering," *Annu. Rev. Heat Transfer*, **10**, 144–176, 1999.
7. S. Maruyama, "Molecular Dynamics Method for Microscale Heat Transfer," in *Advances in Numerical Heat Transfer*, W. J. Minkowycz and E. M. Sparrow (eds.), Vol. 2, pp. 189–226, Taylor & Francis, New York, 2000.
8. D. J. Griffiths, *Introduction to Quantum Mechanics*, 2nd ed., Prentice Hall, New York, 2005.
9. H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping*, Springer, New York, 1999.
10. G. Burns, *High-Temperature Superconductivity: An Introduction*, Academic Press, Boston, MA, 1992.
11. A. Einstein, "Zur quantentheorie der strahlung," *Phys. Z.*, **18**, 121–128, 1917; English translation in *Sources of Quantum Mechanics*, B. L. Van der Waerden (ed.), North-Holland Publishing Company, Amsterdam, the Netherlands, 1967.
12. H. P. Baltes, "On the validity of Kirchhoff's law of heat radiation for a body in a nonequilibrium environment," *Progress in Optics*, **13**, 1–25, 1976.
13. J. P. Gordon, H. J. Zeiger, and C. H. Townes, "The maser—New type of microwave amplifier, frequency standard, and spectrometer," *Phys. Rev.*, **99**, 1264–1274, 1955; A. L. Schawlow and C. H. Townes, "Infrared and optical masers," *Phys. Rev.*, **112**, 1940–1949, 1958.
14. R. Wolfson and J. M. Pasachoff, *Physics with Modern Physics for Scientists and Engineers*, 3rd ed., Addison-Wesley, Reading, MA, 1999.

### PROBLEMS

---

- 3.1. For a rectangular prism (i.e., a cuboid) whose three sides are  $x$ ,  $y$ , and  $z$  if  $x + y + z = 9$ , find the values of  $x$ ,  $y$ , and  $z$  so that the volume of the prism is maximum.

- 3.2.** Make a simple computer program to evaluate the relative error of Stirling's formula:  $\ln x! \approx x \ln x - x$  for  $x = 10, 100,$  and  $1000$ .
- 3.3.** For each of the following cases, determine the number of ways to place 25 books on 5 shelves (distinguishable by their levels). The order of books within an individual shelf is not considered.
- The books are distinguishable, and there is no limit on how many books can be put on each shelf.
  - Same as (a), except that all the books are the same (indistinguishable).
  - The books are distinguishable, and there are 5 books on each shelf.
  - The books are distinguishable, and there are 3 books on the 1st shelf, 4 on the 2nd, 5 on the 3rd, 6 on the 4th, and 7 on the 5th.
- 3.4.** For each of the following cases, determine the number of ways to put 4 books on 10 shelves (distinguishable by their levels). Disregard their order on each shelf.
- The books are distinguishable, and there is no limit on how many books you can place on each shelf.
  - Same as (a), but there is a maximum of 1 book on any shelf.
  - Same as (a), except that the books are identical (indistinguishable).
  - Same as (b), except that the books are identical.
- 3.5.** A box contains 5 red balls and 3 black balls. Two balls are picked up randomly. Determine the following:
- What's the probability that the second ball is red?
  - What's the probability that both are red?
  - If the first one is black, what is the probability that the second is red?
- 3.6.** Suppose you toss two dice, what's the probability of getting a total number (a) equal to 5 and (b) greater than 5?
- 3.7.** Draw 5 cards from a deck of 52 cards.
- What is the probability of getting a royal flush?
  - What is the probability of getting a full house? [A royal flush is a hand with A, K, Q, J, and 10 of the same suit. A full house is a hand with three of one kind and two of another (a pair).]
- 3.8.** For a Gaussian distribution function,  $f(x) = a \exp[-(x - \mu)^2]$ , where  $a$  and  $\mu$  are positive constants.
- Find the normalized distribution function  $F(x)$ .
  - Show that the mean value  $\bar{x} = \mu$ .
  - Determine the variance  $u_{\text{var}}$  and the standard deviation  $\sigma$ .
- 3.9.** The speed distribution function for  $N$  particles in a fixed volume is given by  $f(V) = AV(B - V)/B^3$ , where  $V (> 0)$  is the particle speed, and  $A$  and  $B$  are positive constants. Determine:
- The probability density function  $F(V)$ .
  - The number of particles  $N$  in the volume.
  - The minimum speed  $V_{\text{min}}$  and maximum speed  $V_{\text{max}}$ .
  - The most probable speed where the probability density function is the largest.
  - The average speed  $\bar{V}$  and the root-mean-square average speed  $V_{\text{rms}} = \sqrt{\bar{V}^2}$ .
- 3.10.** Six bosons are to be placed in two energy levels, each with a degeneracy of two. Evaluate the thermodynamic probability of all arrangements. What is the most probable arrangement?
- 3.11.** Four fermions are to be placed in two energy levels, each with a degeneracy of four. Evaluate the thermodynamic probability of each arrangement. What is the most probable arrangement?
- 3.12.** Derive the Fermi-Dirac distribution step by step. Clearly state all assumptions. Under which condition, can it be approximated by the Maxwell-Boltzmann distribution?
- 3.13.** What is the Boltzmann constant and how is it related to the universal gas constant? Show that the ideal gas equation can be written as  $P = nk_B T$ . What is the number density of air at standard conditions (1 atm and 25°C)?
- 3.14.** How many molecules are there per unit volume (number density) for a nitrogen gas at 200 K and 20 kPa? How would you estimate the molecular spacing (average distance between two adjacent molecules)?
- 3.15.** Use Eq. (3.28a) and  $1/T = (\partial S / \partial U)_{V,N}$  to show that  $\beta = 1/k_B T$ .
- 3.16.** Show that  $\beta = 1/k_B T$  and  $\alpha = -\mu/k_B T$  for all the three statistics. [Hint: Follow the discussion in Sec. 3.2 with a few more steps.]
- 3.17.** Consider 10 indistinguishable particles in a fixed volume that obey the Bose-Einstein statistics. There are three energy levels with  $\epsilon_0 = 0.5$  eu,  $\epsilon_1 = 1.5$  eu, and  $\epsilon_2 = 2.5$  eu, where "eu" refers to a certain energy unit. The degeneracies are  $g_0 = 1, g_1 = 3,$  and  $g_2 = 5,$  respectively.

- (a) If the degeneracy were not considered, in how many possible ways could you arrange the particles on the three energy levels?
- (b) You may notice that different arrangements may result in the same energy. For example, both the arrangement with  $N_1 = 9, N_2 = 0, N_3 = 1$  and the arrangement with  $N_1 = 8, N_2 = 2, N_3 = 0$  yield an internal energy  $U = 7$  eu. How many arrangements are there with  $U = 9$  eu? Calculate the thermodynamic probability for all macrostates with  $U = 9$  eu.
- (c) The ground state refers to the state corresponding to the lowest possible energy of the system. Determine the ground-state energy and entropy. What is the temperature of this system at the ground state?
- (d) How many microstates are there for the macrostate with  $U = 25$  eu?

**3.18.** Consider a system of a single type of constituents, with  $N$  particles (distinguishable from the statistical point of view) and only two energy levels  $\epsilon_0 = 0$  and  $\epsilon_1 = \epsilon$  (nondegenerate).

- (a) What is the total number of microstates in terms of  $N$ . How many microstates are there for the macrostate that has energy  $U = (N - 1)\epsilon$ ? Show that the energy of the most probable macrostate is  $N\epsilon/2$ .
- (b) What are the entropies of the states with  $U = 0$  and  $U = (N - 1)\epsilon$ . Sketch  $S$  as a function of  $U$ . Comment on the negative temperature,  $1/T = (\partial S/\partial U)_{V,N} < 0$ . Is it possible to have a system with a negative absolute temperature?

**3.19.** A system consists of six indistinguishable particles that obey Bose-Einstein statistics with two energy levels. The associated energies are  $\epsilon_0 = 0$  and  $\epsilon_1 = \epsilon$ , and the associated degeneracies are  $g_0 = 1$  and  $g_1 = 3$ . Answer the following questions:

- (a) How many possible macrostates are there? How many microstates corresponding to the macrostate with three particles on each energy level?
- (b) What is the most probable macrostate, and what are its corresponding energy  $U$  and thermodynamic probability  $\Omega$ ?
- (c) Show that at 0 K, both the energy and the entropy of this system are zero. Also, show that for this system the entropy increases as the energy increases.

**3.20.** From the Sackur-Tetrode equation, show that  $s_2 - s_1 = c_p \ln(T_2/T_1) - R \ln(P_2/P_1)$ .

**3.21.** Write  $U, p, A$ , and  $S$  in terms of the partition function  $Z$ . Express  $H$  and  $G$  in terms of the partition function  $Z$ . For an ideal monatomic gas, express  $H$  and  $G$  in terms of  $T$  and  $P$ .

**3.22.** For an ideal diatomic gas, the partition function can be written as  $Z = Z_t Z_r Z_v Z_e Z_D$ , where  $Z_e = g_{e0}$  is the degeneracy of the ground electronic level, and  $Z_D = \exp(-D_0/k_B T)$  is the chemical partition function that is associated with the reaction of formation. Here,  $g_{e0}$  and  $D_0$  can be regarded as constants for a given material. Contributions to the partition function beside the translation are due to internal energy storage and thus are called the *internal contribution*, i.e.,  $Z_{\text{int}} = Z_r Z_v Z_e Z_D$ . Find the expressions of  $U, P, A, S, H$ , and  $G$  in terms of  $N, T$ , and  $P$  (or  $V$ ) with appropriate constants, assuming that the temperature  $T \gg \Theta_r$  and is comparable with  $\Theta_v$ .

**3.23.** For an ideal molecular gas, derive the distribution function in terms of the kinetic energy  $\epsilon = mv^2/2$ , i.e.,  $f(\epsilon)$ .

**3.24.** Prove Eq. (3.48), Eq. (3.49a) and Eq. (3.50).

**3.25.** Evaluate and plot the Maxwell speed distribution for Ar gas at 100, 300, and 900 K. Tabulate the average speed, the most probable speed, and the rms speed at these temperatures.

**3.26.** A special form of the Euler-Maclaurin summation formula is

$$\sum_{j=a}^{\infty} f(j) = \int_a^{\infty} f(x) dx + \frac{1}{2} f(a) - \frac{1}{12} f'(a) + \frac{1}{720} f^{(3)}(a) - \frac{1}{30,240} f^{(5)}(a) + \dots$$

Consider the rotational partition function,

$$Z_r = \sum_{j=0}^{\infty} (2j + 1) \exp\left[-j(j + 1) \frac{\Theta_r}{T}\right]$$

and show that

$$Z_r \approx \frac{T}{\Theta_r} \left[ 1 + \frac{1}{3} \frac{\Theta_r}{T} + \frac{1}{15} \left(\frac{\Theta_r}{T}\right)^2 + \dots \right]$$

which is Eq. (3.55) for  $\sigma = 1$ .

**3.27.** Because of the nuclear spin degeneracy, hydrogen  $H_2$  gas is consistent of two different types: *ortho-hydrogen* and *para-hydrogen*. The rotational partition functions can be written, respectively, as

$$Z_{r,\text{ortho}} = 3 \sum_{l=0,2,4,\dots} (2l+1) \exp\left[-l(l+1)\frac{\Theta_r}{T}\right]$$

and

$$Z_{r,\text{para}} = \sum_{l=1,3,5,\dots} (2l+1) \exp\left[-l(l+1)\frac{\Theta_r}{T}\right]$$

so that  $Z_{r,H_2} = 3 \sum_{l=0,2,4,\dots} (2l+1) \exp\left[-l(l+1)\frac{\Theta_r}{T}\right] + \sum_{l=1,3,5,\dots} (2l+1) \exp\left[-l(l+1)\frac{\Theta_r}{T}\right]$ .

Evaluate the temperature-dependent specific heat of each of the two types of hydrogen, which can be separated and stay separated for a long time before the equilibrium distribution is restored. Calculate the specific heat of hydrogen in the equilibrium distribution as a function of temperature. The ratio  $Z_{r,\text{ortho}}/Z_{r,\text{para}}$  is the same as the equilibrium ratio of the two types and varies from 0 at very low temperatures to 3 near room temperature.

**3.28.** Calculate the specific heat and the specific heat ratio  $\gamma = c_p/c_v$  for nitrogen  $N_2$  at 30, 70, 300, and 1500 K. Assume the pressure is sufficiently low for it to be an ideal gas.

**3.29.** Calculate the specific heat and the specific heat ratio  $\gamma = c_p/c_v$  for oxygen  $O_2$  at 50, 100, 300, and 2000 K. Assume the pressure is sufficiently low for it to be an ideal gas.

**3.30.** Estimate the mole and mass specific heats of CO gas at 100, 300, and 3000 K. Show in a specific heat versus temperature graph the contributions from different modes.

**3.31.** (a) How many rotational degrees of freedom are there in a  $CO_2$  molecule and in a  $H_2O$  molecule? (b) If the temperature of a low-pressure  $CO_2$  gas is raised high enough to completely excite its rotational and vibrational modes, what will be its specific heats  $c_v$  and  $c_p$ ? Express your answer in both  $\text{kJ}/(\text{kg} \cdot \text{K})$  and  $\text{kJ}/(\text{kmol} \cdot \text{K})$ .

**3.32.** Compute and plot the temperature-dependent specific heat for the following ideal gases and compare your results with tabulated data or graphs: (a)  $CO_2$ , (b)  $H_2O$ , and (c)  $CH_4$ .

**3.33.** Do a literature search to discuss the following topics: (a) the significance of partition functions, (b) the different types of statistical ensembles, and (c) statistical fluctuations.

**3.34.** We have discussed the translational degeneracy  $dg$  in a 3-D space with a volume  $V$ , as given in Eq. (3.85). Consider the situation when the particle is confined in a 2-D square potential well. Find the proper wavefunctions and the energy eigenvalues. Assuming the area  $A$  is very large, find the translational degeneracy  $dg$  in terms of  $A$ ,  $m$ ,  $\epsilon$ , and  $d\epsilon$ .

**3.35.** Estimate the speed an electron needs in order to escape from the ground state of a hydrogen atom. What is the de Broglie wavelength of the electron at the initial speed? If a photon is used to knock out the electron in the ground state, what would be the wavelength of the photon? Why is it inappropriate to consider the electron movement in an atom as an analogy to the movement of Mars in the solar system?

**3.36.** For the harmonic oscillator problem discussed in Sec. 3.5.5. Show that Eq. (3.111) is a solution for Eq. (3.107) for  $n = 0, 1$ , and 2. Plot  $\Psi_0^2$ ,  $\Psi_1^2$ , and  $\Psi_2^2$  to discuss the differences between classical mechanics and quantum mechanics.

*This page intentionally left blank*

---

# CHAPTER 4

---

## KINETIC THEORY AND MICRO/NANOFLUIDICS

---

Statistical mechanics involves determination of the most probable state and equilibrium distributions, as well as evaluation of the thermodynamic properties in the equilibrium states. Kinetic theory deals with the local average of particle properties and can be applied to nonequilibrium conditions to derive transport equations.<sup>1-7</sup> Kinetic theory, statistical mechanics, and molecular dynamics are based on the same hypotheses; they are closely related and overlap each other in some aspects. Knowledge of kinetic theory is important to understanding gas dynamics, as well as electronic and thermal transport phenomena in solid materials.

In this chapter, we first introduce the simple kinetic theory of ideal gases based on the mean-free-path approximation. While it can help us obtain the microscopic formulation of several familiar transport equations and properties, the simple kinetic theory is limited to local equilibrium and, hence, is good only for time durations much longer than the mechanistic timescale, called the relaxation time. The advanced kinetic theory is based on the Boltzmann transport equation (BTE), which will also be presented in this chapter. The BTE is an integro-differential equation of the distribution function in terms of space, velocity, and time. It takes into account changes in the distribution function caused by external forces and collisions between particles. Many macroscopic phenomenological equations, such as Fourier's law of heat conduction, the Navier-Stokes equation for viscous flow, and the equation of radiative transfer for photons and phonons, can be derived from the BTE, under the assumption of local equilibrium. Finally, in the last section of this chapter, we present the application of kinetic theory to the flow of dilute gases in micro/nanostructures and the associated heat transfer. The application of kinetic theory to heat conduction in metals and dielectrics will be discussed in forthcoming chapters.

---

### 4.1 KINETIC DESCRIPTION OF DILUTE GASES

---

In this section, we will introduce the simple kinetic theory of ideal molecular gases. The purpose is to provide a step-by-step learning experience leading to more advanced topics. There are several hypotheses and assumptions in kinetic theory of molecules.

- *Molecular hypothesis*: Matter is composed of small discrete particles (molecules or atoms); any macroscopic volume contains a large number of particles. At 25°C and 1 atm, 1- $\mu\text{m}^3$  space of an ideal gas contains 27 million molecules.

- *Statistic hypothesis*: Time average is often used since any macroscopic observation takes much longer than the characteristic timescale of molecular motion (such as the average time lapse between two subsequent collisions of a given molecule).
- *Kinetic hypothesis*: Particles obey the laws of classical mechanics.
- *Molecular chaos*: The velocity and position of a particle are uncorrelated. The velocities of any two particles are not correlated.
- *Ideal gas assumptions*: Molecules are rigid spheres resembling billiard balls. Each molecule has a diameter  $d$  and a mass  $m$ . All collisions are elastic and conserve both energy and momentum. Molecules are widely separated in space (i.e., a dilute gas). Intermolecular forces are negligible except during molecular collisions. The duration of collision is negligible compared with the time between collisions. No collision can occur with more than two particles.

The general molecular *distribution function* is  $f(\mathbf{r}, \mathbf{v}, t)$ , which is a function of space, velocity, and time. The distribution function gives the particle (number) density in the phase space at any time. Therefore, the number of particles in a volume element of the phase space is

$$dN = f(\mathbf{r}, \mathbf{v}, t) dx dy dz dv_x dv_y dv_z = f(\mathbf{r}, \mathbf{v}, t) dV d\varpi \quad (4.1)$$

where we have used  $\varpi$  for the velocity space ( $d\varpi = dv_x dv_y dv_z$ ). Integrating Eq. (4.1) over the velocity space gives the number of particles per unit volume, or the number density, as

$$n(\mathbf{r}, t) = \frac{dN}{dV} = \int_{\varpi} f(\mathbf{r}, \mathbf{v}, t) d\varpi \quad (4.2)$$

Note that the density is  $\rho(\mathbf{r}, t) = m \cdot n(\mathbf{r}, t)$ , where  $m$  is the mass of a particle. The total number of particles inside the volume  $V$  as a function of time is then

$$N(t) = \iiint_{V, \varpi} f(\mathbf{r}, \mathbf{v}, t) dV d\varpi \quad (4.3)$$

In a thermodynamic equilibrium state,

$$f(\mathbf{r}, \mathbf{v}, t) = f(\mathbf{v}) \quad (4.4)$$

which is independent of space and time. Any intensive property will be the same everywhere.

#### 4.1.1 Local Average and Flux

Let  $\psi = \psi(\mathbf{r}, \mathbf{v}, t)$  be any additive property of a single molecule, such as kinetic energy and momentum. Note that  $\psi$  may be a scalar or a vector. The *local average* or simply the *average* of the property  $\psi$  is defined as

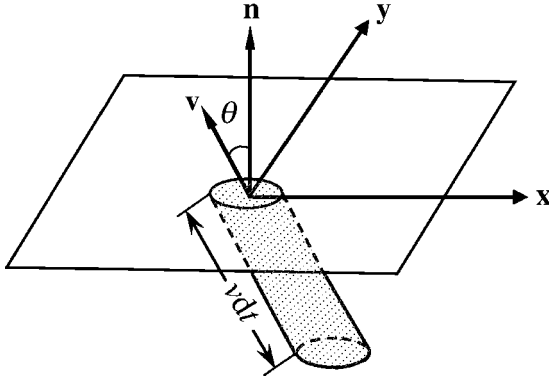
$$\bar{\psi} = \int_{\varpi} f \psi d\varpi / \int_{\varpi} f d\varpi = \frac{1}{n} \int_{\varpi} f \psi d\varpi \quad (4.5)$$

which is a function of  $\mathbf{r}$  and  $t$ . The *ensemble average* is the average over the phase space, i.e.,

$$\langle \psi \rangle = \frac{1}{N} \iiint_{V, \varpi} f \psi dV d\varpi \quad (4.6)$$

For a uniform gas, the local average and the ensemble average are the same.

The transfer of  $\psi$  across an area element  $dA$  per unit time per unit area is called the *flux* of  $\psi$ . As shown in Fig. 4.1, particles having velocities between  $\mathbf{v}$  and  $\mathbf{v} + d\mathbf{v}$  that will pass



**FIGURE 4.1** Illustration of the flux of particles and quantities through a surface.

through the area  $dA$  in the time interval  $dt$  must be contained in the inclined cylinder, whose volume is  $dV = v dt \cos \theta dA = \mathbf{v} \cdot \mathbf{n} dA dt$ . It is assumed that  $dt$  is sufficiently small such that particle-particle collisions can be neglected. The number of particles with velocities between  $\mathbf{v}$  and  $\mathbf{v} + d\mathbf{v}$  within the inclined cylinder can be calculated by

$$f(\mathbf{r}, \mathbf{v}, t) dV d\varpi = f(\mathbf{r}, \mathbf{v}, t) \mathbf{v} \cdot \mathbf{n} dA dt d\varpi \quad (4.7)$$

The flux of the property  $\psi$  is then

$$\text{flux of } \psi \text{ within } d\varpi = \frac{\psi f(\mathbf{r}, \mathbf{v}, t) \mathbf{v} \cdot \mathbf{n} d\varpi dA dt}{dA dt}$$

Integrating over all velocities yields the total flux of  $\psi$ :

$$J_\psi = \int_{\varpi} \psi f \mathbf{v} \cdot \mathbf{n} d\varpi \quad (4.8)$$

Equation (4.8) gives the net flux since it is evaluated for all  $\varpi$ , or over a solid angle of  $4\pi$  in the spherical coordinates. Very often the integration is performed over the hemisphere with  $\mathbf{v} \cdot \mathbf{n} = v \cos \theta > 0$  for positive flux or  $\mathbf{v} \cdot \mathbf{n} = v \cos \theta < 0$  for negative flux. When  $\psi = 1$ , Eq. (4.8) gives the particle flux:

$$J_N = \int_{\varpi} f \mathbf{v} \cdot \mathbf{n} d\varpi \quad (4.9)$$

In an equilibrium state, this integration can be evaluated using the spherical coordinates. Noting that  $\mathbf{v} \cdot \mathbf{n} = v \cos \theta$  and  $f = f(\mathbf{v})$ , which is independent of the direction (isotropic), we can obtain the particle flux in the positive  $z$  direction by integrating over the hemisphere in the velocity space, i.e.,

$$J_N = \int_{v=0}^{\infty} \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} f(\mathbf{v}) v^3 \cos \theta \sin \theta d\theta d\phi dv = \pi \int_0^{\infty} f(\mathbf{v}) v^3 dv \quad (4.10)$$



In writing Eq. (4.10), we have kept the vector variable in  $f(\mathbf{v})$  to signify that it is a velocity distribution. One should bear in mind that the last expression is based on the fact that  $f(\mathbf{v})$  is not a function of  $\theta$  and  $\phi$ . For an ideal molecular gas,  $f(\mathbf{v})$  is given by the Maxwell velocity distribution, i.e., Eq. (3.43) in Chap. 3. If the integration in Eq. (4.10) is performed over the whole sphere with  $\theta$  from 0 to  $\pi$ , we would obtain the net flux of particles, which is zero in the equilibrium case. The average speed can be evaluated using Eq. (4.5); hence,

$$\bar{v} = \frac{1}{n} \int_{\mathcal{O}} f(\mathbf{v})v d\mathcal{O} = \frac{1}{n} \iiint_{v,\phi,\theta} f(\mathbf{v})v^3 \sin\theta d\theta d\phi dv = \frac{4\pi}{n} \int_0^{\infty} f(\mathbf{v})v^3 dv \quad (4.11)$$

Here, we have assumed an isotropic distribution function to obtain the last expression. The above equation is evaluated over the solid angle of  $4\pi$  to obtain the average of all velocities. Comparing Eq. (4.10) and Eq. (4.11), we can see that

$$J_N = \frac{n\bar{v}}{4} \quad (4.12a)$$

For an ideal gas, since  $f(\mathbf{v})$  is given by the Maxwell velocity distribution, Eq. (3.44), we obtain

$$J_N = \frac{n\bar{v}}{4} = n\sqrt{\frac{k_B T}{2\pi m}} \quad (4.12b)$$

Because each particle has the same mass, the mass flux is given by

$$J_m = m \int_{\mathcal{O}} f \mathbf{v} \cdot \mathbf{n} d\mathcal{O} = \frac{\rho\bar{v}}{4} \quad (4.13)$$

Substituting  $\psi = mv^2/2$  into Eq. (4.8), one obtains the kinetic energy flux  $J_{KE}$ . In an equilibrium state with an isotropic distribution, the kinetic energy flux in the positive  $z$  direction is  $J_{KE} = (\pi m/2) \int_0^{\infty} f(\mathbf{v})v^5 dv$ , whereas the net kinetic energy flux is zero. Note that Eq. (4.8) is a general equation that is also applicable to nonequilibrium and anisotropic distributions.

When  $\psi = m\mathbf{v}$ , the momentum flux is a vector, which is often handled by considering individual components. Note that the rate of transfer of momentum across a unit area is equal to the force that the area must exert upon the gas to sustain the equilibrium. Furthermore, the surface may be projected to three orientations, yielding a nine-component tensor in the momentum flux:

$$P_{ij} = \int_{\mathcal{O}} (mv_j) f v_i d\mathcal{O}, \quad i, j = 1, 2, 3 \quad (4.14a)$$

Here,  $(v_1, v_2, v_3)$  and  $(v_x, v_y, v_z)$  are used interchangeably. Let  $P = \rho\bar{v}_i^2$ , which is always positive, and  $\tau_{ij} = \rho\overline{v_j v_i}$  for  $i \neq j$  and 0 for  $i = j$ . We can rewrite the above equation as

$$P_{ij} = n\overline{mv_j v_i} = \rho\overline{v_j v_i} = P\delta_{ij} + \tau_{ij} \quad (4.14b)$$

where  $\delta_{ij}$  is the Kronecker delta, which is equal to 1 when  $i = j$  and 0 when  $i \neq j$ . It can be seen that  $P$  is the normal stress or static pressure and  $\tau_{ij}$  ( $i \neq j$ ) is the shear stress, which is zero in a uniform, stationary gas (without bulk motion). Notice that the velocity distribution in the

vicinity of the wall is the same as that away from the wall because of the reflection by the wall. The pressure is now related to the momentum flux, i.e.,  $3P = \rho(\overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}) = \rho\overline{v^2}$ , or

$$\frac{P}{\rho} = \frac{1}{3} \overline{v^2} \tag{4.15}$$

which is Boyle’s law. Compared with the ideal gas equation, the right-hand side must be related to temperature. In kinetic theory, temperature is associated to the mean translational kinetic energy of the molecule, i.e.,

$$\frac{3}{2} k_B T = \frac{1}{2} m\overline{v^2} = \frac{1}{2} m\overline{v_x^2} + \frac{1}{2} m\overline{v_y^2} + \frac{1}{2} m\overline{v_z^2} \tag{4.16}$$

We have derived this equation from statistical mechanics in Chap. 3. The temperature defined based on the kinetic energy of the particles is sometimes referred to as the *kinetic temperature*. Combining Eq. (4.15) and Eq. (4.16), we get the ideal gas equation,  $P = nk_B T$ , as expected. From the above discussion, one can see clearly how the macroscopic properties such as pressure and temperature are related to the particle distribution function. For ideal gases at equilibrium, we have derived the Maxwell velocity and speed distributions in Chap. 3.

**Example 4-1.** Show that  $P = \rho\overline{v_n^2}$ , where  $v_n$  is the velocity component normal to the wall, and  $P = \rho\overline{v^2}/3$  for equilibrium distribution.

**Solution.** Consider the horizontal plane shown in Fig. 4.1 as the wall, below which is a gas in equilibrium. Multiplying Eq. (4.7) by  $m\mathbf{v}$  gives the momentum of the particles with velocities between  $\mathbf{v}$  and  $\mathbf{v} + d\mathbf{v}$ , impinging on the wall:  $m\mathbf{v}f(\mathbf{v})\mathbf{v} \cdot \mathbf{n} dAdtd\varpi$ , which of course is equal to the impulse on the wall:  $d\mathbf{F}dt$ . The normal component  $v_n = \mathbf{v} \cdot \mathbf{n} = v\cos\theta$  contributes to an impulse on the wall:  $mv_n^2 f(\mathbf{v})dAdtd\varpi$ , that is always positive regardless of the sign of  $v_n$ . However, the contributions of all parallel components cancel out due to isotropy. The pressure can be evaluated by integrating over all velocities, i.e.,  $P = \int_{\varpi} mv_n^2 f(\mathbf{v}) d\varpi = mn\overline{v_n^2} = \rho\overline{v_n^2}$ . We have used the definition of local average given by Eq. (4.5). If the distribution is isotropic, then  $P = m \int_0^\infty \int_0^{2\pi} \int_0^\pi f(\mathbf{v})v^4 \cos^2\theta \sin\theta d\theta d\phi dv = (4\pi m/3) \int_0^\infty f(\mathbf{v})v^4 dv$  since  $v_n = v\cos\theta$ . Compared with  $\overline{v^2} = (1/n) \int_0^\infty \int_0^{2\pi} \int_0^\pi f(\mathbf{v})v^4 \sin\theta d\theta d\phi dv = (4\pi/n) \int_0^\infty f(\mathbf{v})v^4 dv$ , we obtain  $P = mn\overline{v^2}/3 = \rho\overline{v^2}/3$ . The distribution function is uniform inside the container; hence, the wall may be a physical wall or merely an imaginary one since pressure exists everywhere in the fluid.

### 4.1.2 The Mean Free Path

The *mean free path*, defined as the average distance the particle travels between two subsequent collisions, is a very important concept. It is often used to determine whether a given phenomenon belongs to the macroscale (continuum) regime or otherwise falls in the microscale regime when the governing equations derived under the assumption of local equilibrium break down. One of the applications is in microfluidics, to be discussed later in this chapter, and another is in the electrical and heat conduction in solids, which will be studied in Chap. 6.

Consider the case in Fig. 4.2: a particle of diameter  $d$  moving at an average velocity  $\bar{v}$  (assuming all other particles are at rest). During a time interval  $dt$ , the volume swept by the particle within  $d$  from the centerline is  $dV = \pi d^2 \bar{v} dt$ . The  $n dV$  particles, whose centers are inside this volume element, will collide with the moving particle. Therefore, the frequency of collisions, i.e., number of collisions per unit time is  $\pi n d^2 \bar{v}$ . The time between two subsequent collisions,  $\tau$ , is the inverse of the frequency of collision. The mean free path  $\Lambda$  is

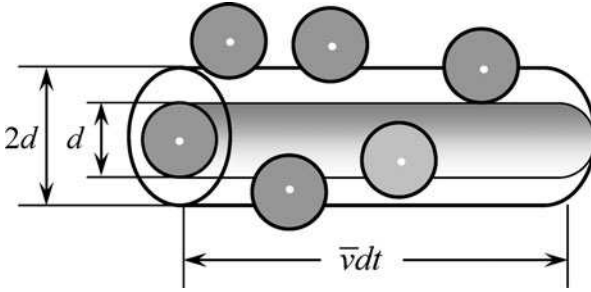


FIGURE 4.2 Schematic used for a simple derivation of the mean free path.

the average distance that a particle travels between two subsequent collisions and is equal to the ratio of the average velocity to the frequency of collision. Therefore,

$$\Lambda = \bar{v}\tau \approx (\pi nd^2)^{-1} \quad (4.17)$$

and depends only on the particle size and the number density. The average time between two subsequent collisions  $\tau$  is termed the *relaxation time*, and the average frequency of collision  $\tau^{-1}$  is the *scattering rate* or *collision rate*. The scattering rate is the average number of collisions an individual particle experiences per unit time.

For electrons whose diameters are negligible compared with that of the other particles that scatter them, the mean free path is

$$\Lambda_{\text{electron (or } \Lambda_{\text{photon)}}} = \frac{1}{nA_c} \quad (4.18)$$

where  $A_c$  is the scattering cross-sectional area and  $n$  is the number density of the scatter, such as phonons or defects. Equation (4.18) also applies for the case of photons that can be scattered by particles, such as molecules in the atmosphere. The photon mean free path is also called the *radiation penetration depth*, as will be discussed in Chap. 8.

When the relative movement of particles is considered based on the Maxwell velocity distribution, Eq. (4.17) is modified slightly for an ideal gas as follows:

$$\Lambda \approx \frac{1}{\sqrt{2}\pi nd^2} = \frac{k_B T}{\sqrt{2}\pi d^2 P} \quad (4.19)$$

The scattering rate, or the collision frequency, is

$$\tau^{-1} = \bar{v}/\Lambda \quad (4.20)$$

Notice that the relaxation time  $\tau$  is an important characteristic time. It tells how quickly the system will restore to equilibrium (at least locally), if disturbed. Table 4.1 lists the diameters for typical molecules.

**Example 4-2.** Calculate the mean free path for air at 25°C and 1 atm. How does it compare with the average spacing between molecules? Find the relaxation time and the number of collisions a molecule experiences per second. What is the speed of sound in the air? Explain why we can smell odor far away from its source quickly.

**Solution.**  $n = P/(k_B T) = 1.0133 \times 10^5 / 1.381 \times 10^{-23} / 298.15 = 2.46 \times 10^{25} \text{ m}^{-3}$ . The average spacing between molecules can be calculated from  $L_0 = n^{-1/3} = 3.4 \text{ nm}$ . The mean free path

**TABLE 4.1** Molecular Diameter for Selected Molecules<sup>1</sup>

Gas type	Molecular weight, $M$ (kg/kmol)	Diameter, $d$ ( $10^{-10}$ m, or Å)
H <sub>2</sub>	2	2.74
He	4	2.19
O <sub>2</sub>	32	3.64
N <sub>2</sub>	28	3.78
Air	29	3.72
CH <sub>4</sub>	16	4.14
NH <sub>3</sub>	17	4.43
H <sub>2</sub> O	18	4.58
CO <sub>2</sub>	44	4.64

calculated from Eq. (4.19) is  $\Lambda = (\sqrt{2}\pi nd^2)^{-1} = 66$  nm ( $d = 0.37$  nm from Table 4.1), which is about 20 times longer than the molecular spacing. The speed of sound can be calculated from  $v_a = \sqrt{\gamma RT} = 345$  m/s using  $\gamma = c_p/c_v = 1.4$  and  $R = k_B/m$ . The average speed is  $\bar{v} = \sqrt{8k_B T/\pi m} = 466$  m/s. Therefore, the relaxation time is  $\tau = \Lambda/\bar{v} = 0.14$  ns. On the average, each molecule experiences  $\tau^{-1}$  collisions, i.e., more than 7 billion collisions per second. Although the mean free path is very small, molecules may travel for a long (absolute) distance because of the high average speed. It does not take many molecules for the nose to detect an odor. The odor source usually contains numerous individual molecules.

Let  $p(\xi)$  be the probability that a molecule travels at least  $\xi$  between collisions. The probability for the particle to collide within an element distance  $d\xi$  is  $d\xi/\Lambda$ . Thus, the probability for a free path greater than  $\xi + d\xi$  is less than  $p(\xi)$  by the probability of collision between  $\xi$  and  $\xi + d\xi$ , i.e.,

$$p(\xi + d\xi) = p(\xi) \left(1 - \frac{d\xi}{\Lambda}\right) \quad (4.21)$$

Therefore,  $dp(\xi)/p(\xi) = -d\xi/\Lambda$ . Since  $p(0) = 1$ , integrating from 0 to  $\xi$  yields

$$p(\xi) = e^{-\xi/\Lambda} \quad (4.22)$$

The probability density function (PDF) for the free path is given by

$$F(\xi) = -\frac{dp(\xi)}{d\xi} = \frac{1}{\Lambda} e^{-\xi/\Lambda} \quad (4.23)$$

One can verify that  $\int_0^\infty F(\xi)d\xi = 1$  and  $\bar{\xi} = \int_0^\infty F(\xi)\xi d\xi = \Lambda$ . Therefore, Eq. (4.23) is indeed the free-path PDF. The probability for molecules to have a free path less than  $\xi$  is given as

$$p(<\xi) = 1 - p(\xi) = \int_0^\xi F(\xi)d\xi = 1 - e^{-\xi/\Lambda} \quad (4.24)$$

Figure 4.3 shows the free-path distribution functions. It is an exponential function. In dealing with radiation or photons, the mean free path is called the radiation penetration depth. Radiation will decay exponentially with distance in an absorbing medium. The fraction of photons that will transmit through a distance equal to the penetration depth is  $1/e \approx 37\%$ .

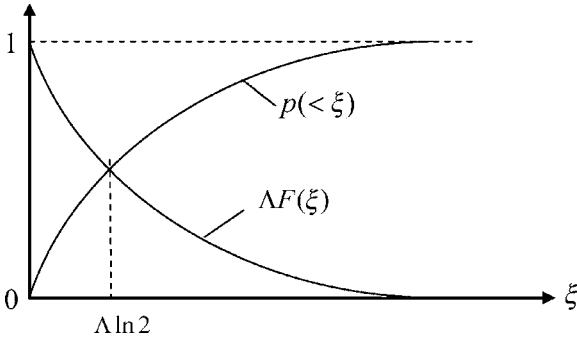


FIGURE 4.3 Free-path distribution functions.

**4.2 TRANSPORT EQUATIONS AND PROPERTIES OF IDEAL GASES**

Consider a molecular gas at steady state but not at equilibrium, with a 1-D gradient of some macroscopic properties. Under the assumption of local equilibrium,

$$f(\mathbf{r}, \mathbf{v}, t) = f(\xi, \mathbf{v}) \tag{4.25}$$

where  $\xi$  is the coordinate along which the gradient occurs. The *average collision distance*  $\Lambda_a$  is defined as the separation of the planes at which particles, on the average, across a plane located at  $\xi_0$  will experience the next collision, as shown in Fig. 4.4a. It may be assumed that particles that will cross the plane before the next collision are located in a hemisphere of radius equal to the mean free path  $\Lambda$ . The problem is how to obtain the average projected length ( $\Lambda \cos \theta$ ) in the  $\xi$ -coordinate, as shown in Fig. 4.4b. A simple calculation yields

$$\Lambda_a = \frac{\int_0^{2\pi} \int_0^{\pi/2} \Lambda \cos \theta dA \cos \theta \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi/2} dA \cos \theta \sin \theta d\theta d\phi} = \frac{2\pi \frac{\Lambda}{3} dA}{\pi dA} = \frac{2}{3} \Lambda \tag{4.26}$$

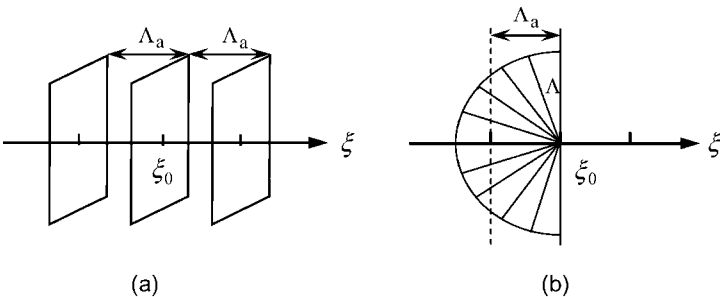


FIGURE 4.4 Illustration of the concepts of (a) average planes of collision and (b) average collision distance  $\Lambda_a$ , with respect to the mean free path.

Note that the projected area  $dA \cos \theta$  is used to account for the particle flux. One can consider the free-path distribution and integrate over all free paths. The resulting  $\Lambda_d/\Lambda$  is the same.<sup>12</sup>

#### 4.2.1 Shear Force and Viscosity

Consider a gas flowing in the  $x$  direction with a velocity gradient in the  $y$  direction, as shown in Fig. 4.5. Here,  $v_B$  is the *average* or *bulk velocity*, which has a nonzero component

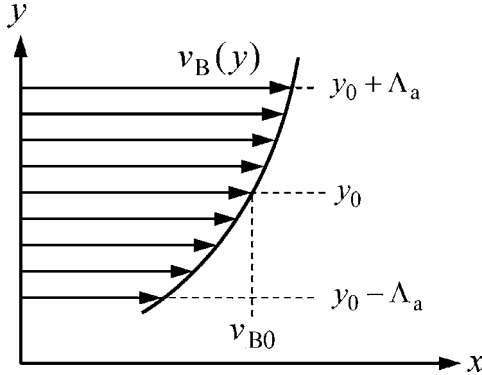


FIGURE 4.5 Schematic of a fluid moving with a bulk velocity  $v_B(y)$  that varies in the  $y$  direction.

only in the  $x$  direction. The velocity due to random motion is sometimes called *thermal velocity*, which follows certain equilibrium distribution with an average equal to zero. The fact that the equilibrium distribution is followed everywhere is based on the assumption of local equilibrium. Molecular random motion will cause an exchange of momentum between the upper layer and the lower layer. The net effect is a tendency to accelerate the flow in the lower layer and decelerate the flow in the upper layer. In other words, the flow below the  $y = y_0$  plane will exert a shear force to the flow above the  $y = y_0$  plane and vice versa. The average momentum of the particles is a function of  $y$  only, i.e.,  $p_x(y) = mv_B(y)$ . The momentum flux across the  $y = y_0$  plane can be evaluated using the concept of mean planes above and below  $y = y_0$ . It may be assumed that all the molecules going upward across the  $y = y_0$  plane are from the  $y = (y_0 - \Lambda_a)$  plane. Therefore, the momentum flux in the positive  $y$  direction is

$$J_p^+ = \frac{n\bar{v}}{4} m \left( v_{B0} - \Lambda_a \left. \frac{dv_B}{dy} \right|_{y_0} \right) \quad (4.27a)$$

where  $n\bar{v}/4$  is the molecular flux, with  $\bar{v}$  as the average speed without considering bulk motion, and  $v_{B0} = v_B(y_0)$ . Similarly, the momentum flux downward is

$$J_p^- = \frac{n\bar{v}}{4} m \left( v_{B0} + \Lambda_a \left. \frac{dv_B}{dy} \right|_{y_0} \right) \quad (4.27b)$$

The net momentum flux, which is equal to the shear force  $P_{yx} = \tau_{yx} = J_p^+ - J_p^-$ , is therefore

$$\tau_{yx} = -\frac{1}{3}\rho\bar{v}\Lambda \left. \frac{dv_B}{dy} \right|_{y_0} \quad (4.28)$$

Comparing Eq. (4.28) with the Newton's law of shear stress in Eq. (2.36), i.e.,  $\tau_{yx} = -\mu \left. \frac{dv_B}{dy} \right|_{y_0}$ , dynamic viscosity is obtained from the simple kinetic theory as

$$\mu = \frac{1}{3}\rho\bar{v}\Lambda \quad (4.29)$$

The above equation provides an order-of-magnitude estimate. While the density is proportional to pressure, the mean free path is inversely proportional to the number density, or density. The average velocity is a function of temperature only. Therefore, the viscosity depends only on temperature and the type of molecules, but not on pressure. The result from more detailed calculations and experiments suggests that Eq. (4.29) be multiplied by 3/2, i.e.,

$$\mu = \frac{1}{2}\rho\bar{v}\Lambda = \frac{m\bar{v}}{2\sqrt{2}\pi d^2} = \frac{1}{\pi d^2} \sqrt{\frac{mk_B T}{\pi}} \quad (4.30)$$

Equation (4.30) is recommended for use in the exercises to estimate the viscosity. It should be noted that the above discussion is based on the simple ideal gas model that each molecule is a rigid (or hard) sphere and all collisions are elastic. Additional modifications have been made to correctly account for the temperature dependence. These models will not be discussed here, and interested readers can find them in the literature.<sup>1-7</sup>

## 4.2.2 Heat Diffusion

Heat conduction is due to the temperature gradient inside the medium. In an ideal molecular gas, the random motion of molecules transports thermal energy from place to place. Sometimes we call the particles that are responsible for thermal energy transport *heat carriers*. Similar to the argument for momentum transfer, it is straightforward to illustrate heat diffusion in a 1-D temperature-gradient system at steady state and under local equilibrium, using Fig. 4.6. The net energy flux across the  $x = x_0$  plane is given by

$$J_E = J_E^+ - J_E^- = \frac{n\bar{v}}{4}[\bar{\varepsilon}(x_0 - \Lambda_a) - \bar{\varepsilon}(x_0 + \Lambda_a)] = -\frac{1}{3}n\bar{v}\Lambda \left. \frac{d\bar{\varepsilon}}{dx} \right|_{x_0} \quad (4.31)$$

where  $\bar{\varepsilon}$  is the average thermal energy per molecule and, hence, is a function of temperature. Based on the definition of specific heat

$$n \left. \frac{d\bar{\varepsilon}}{dx} \right|_{x_0} = n \left. \frac{d\bar{\varepsilon}}{dT} \frac{dT}{dx} \right|_{x_0} = nmc_v \left. \frac{dT}{dx} \right|_{x_0}$$

The heat flux is related to the temperature gradient as

$$q_x'' = J_E = -\frac{1}{3}\rho c_v \bar{v} \Lambda \left. \frac{dT}{dx} \right|_{x_0} \quad (4.32)$$

which is Fourier's law with the thermal conductivity:

$$\kappa = \frac{1}{3}\rho c_v \Lambda \bar{v} \quad (4.33)$$

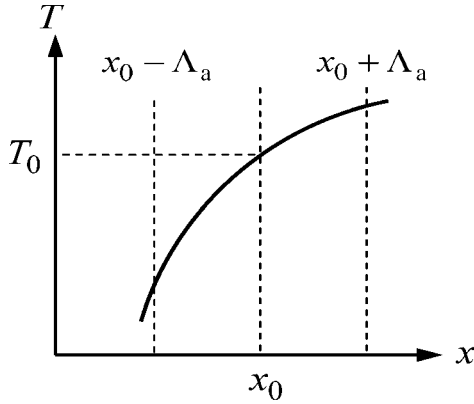


FIGURE 4.6 One-dimensional heat diffusion.

Because  $c_v$  and  $\bar{v}$  are functions of temperature only and  $\Lambda$  is inversely proportional to  $\rho$ , the thermal conductivity of a given ideal gas is a function of temperature and independent of pressure.

Comparing Eq. (4.30) with Eq. (4.33), we have  $\kappa = 0.667\mu c_v$ . The calculated results are consistently lower than the tabulated values for real gases. The reason is the assumption that the average collision distance is the same for both momentum transport and energy transfer. Generally speaking, molecules with a larger speed travel farther than those with a smaller speed. Once the molecules pass the mean plane, they will persist a little while before collision. The persistence effect is larger for energy transfer because the translational kinetic energy of a molecule is proportional to the square of the speed, while that of momentum is proportional to the velocity components. In gases, the average collision distance is greater for energy transfer and depends on the type of gas. Extensive studies of the similarity between  $\mu$  and  $\kappa$  have resulted in a more accurate expression for calculating the thermal conductivity of ideal gases than the one given in Eq. (4.33). Eucken's formula relates the Prandtl number  $Pr \equiv \nu/\alpha = c_p\mu/\kappa$  to the specific heat ratio as follows:<sup>6</sup>

$$Pr = \frac{4\gamma}{9\gamma - 5} \quad (4.34)$$

Based on Eucken's formula, the following equation is recommended to replace Eq. (4.33) in predicting the thermal conductivity of ideal gases:

$$\kappa = \frac{9\gamma - 5}{4} c_v \mu \quad (4.35)$$

where  $\mu$  can be calculated from Eq. (4.30). For a monatomic gas,  $\gamma = 5/3$  and

$$\kappa = 2.5\mu c_v = 1.25\rho c_v \Lambda \bar{v} \quad (4.36a)$$

For a diatomic gas at intermediate temperatures when the translational and rotational modes are fully excited but no vibrational modes have been excited, we have  $\gamma = 1.4$  and

$$\kappa = 1.9\mu c_v = 0.95\rho c_v \Lambda \bar{v} \quad (4.36b)$$



The results calculated from Eq. (4.35) agree reasonably well with the tabulated thermal conductivity values of typical gases. Additional corrections are required when the temperature deviates significantly from the room temperature. More complicated formulations are needed to better account for the temperature dependence.<sup>4</sup>

**Example 4-3.** Calculate the viscosity and the thermal conductivity of air at 300 K and 100 kPa. How will your answers change if the temperature is increased to 306 K and the pressure is decreased to 50 kPa?

**Solution.** From Eq. (4.30), we have

$$\begin{aligned}\mu &= \frac{1}{\pi d^2} \sqrt{\frac{mk_B T}{\pi}} = \frac{1}{\pi(3.72 \times 10^{-10})^2} \left( \frac{29}{6.022 \times 10^{26}} \cdot \frac{1.381 \times 10^{-23} \times 300}{\pi} \right)^{1/2} \\ &= 1.83 \times 10^{-5} \text{ N} \cdot \text{s/m}^2\end{aligned}$$

It is within 1% of the measured value. From Eq. (4.36b) and  $c_v = R/(\gamma - 1) = 716.6 \text{ J/(kg} \cdot \text{K)}$ ,  $\kappa = 1.9\mu c_v = 0.025 \text{ W/(m} \cdot \text{K)}$ . This is within 5% of the measured value. Notice that  $\mu$  and  $\kappa$  depend on temperature only. If the change in specific heat is neglected, then  $\kappa \propto \mu \propto \sqrt{T}$ . When the temperature is increased by 2% to 306 K, both  $\kappa$  and  $\mu$  will increase by 1% regardless of the pressure.

### 4.2.3 Mass Diffusion

Consider a small duct linking two gas tanks containing different types of ideal gases at the same temperature and pressure, as shown in Fig. 4.7a. The total number density of the mixture will be the same as that in either tank, i.e.,  $n = n_A(x) + n_B(x)$ , as illustrated in Fig. 4.7b. Therefore,

$$\frac{dn_A}{dx} = -\frac{dn_B}{dx} \quad (4.37)$$

Fick's law states that

$$J_{N_A} = -D_{AB} \frac{dn_A}{dx} \quad (4.38a)$$

or

$$J_{m_A} = -D_{AB} \frac{d\rho_A}{dx} \quad (4.38b)$$

where  $D_{AB}$  in  $\text{m}^2/\text{s}$  is called the *binary diffusion coefficient* or *diffusion coefficient* between A and B. Notice that the molecular transfer rate  $\dot{N}_A = J_{N_A} A$  and the mass transfer rate  $\dot{m}_A = J_{m_A} A$ . Similarly, we can write Fick's law for type B molecules as

$$J_{N_B} = -D_{BA} \frac{dn_B}{dx} \quad (4.39a)$$

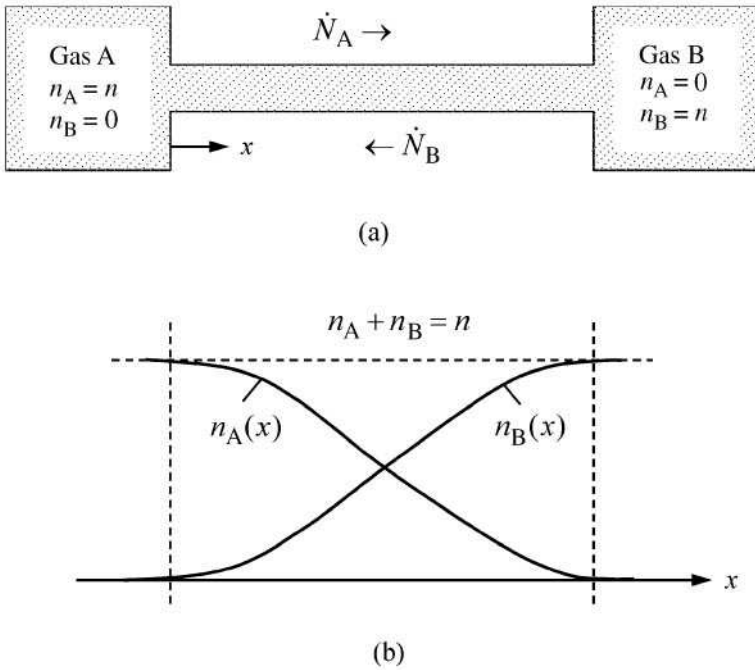
or

$$J_{m_B} = -D_{BA} \frac{d\rho_B}{dx} \quad (4.39b)$$

Because the flux of type B molecules must balance that of type A molecules to maintain a uniform pressure, we have

$$J_{N_A} = -J_{N_B} \quad (4.40)$$

Equations (4.37) through (4.40) imply that  $D_{BA} = D_{AB}$ .



**FIGURE 4.7** Schematic of binary diffusion between ideal gases. (a) Two reservoirs of different types of gas molecules connected through a duct. (b) Concentration distributions in terms of the number densities.

Using the microscopic descriptions of mass diffusion, one can write the positive and negative flux at a certain location  $x_0$  using the average distance concept discussed earlier. Hence,

$$J_{N_A} = \frac{\bar{v}}{4} \left[ n_A(x_0) - \Lambda_a \frac{dn_A}{dx} \Big|_{x_0} \right] - \frac{\bar{v}}{4} \left[ n_A(x_0) + \Lambda_a \frac{dn_A}{dx} \Big|_{x_0} \right]$$

The result is

$$J_{N_A} = -\frac{1}{3} \Lambda \bar{v} \frac{dn_A}{dx} \Big|_{x_0} \quad (4.41)$$

Comparing Eq. (4.41) with Eq. (4.38a), we have

$$D_{AB} = \frac{1}{3} \Lambda \bar{v} \quad (4.42)$$

In the case of similar molecules (such as isotopes),  $D_{AA'} = (2/3) \sqrt{k_B T / \pi m} / (n \pi d^2)$ , which is often called the self-diffusion coefficient. The calculation for the mean free path and the average velocity for a mixture of dissimilar molecules is certainly more involved. However, a simple expression can be obtained using the central distance  $\bar{d} = (d_A + d_B)/2$  and the reduced mass  $m_r = m_A m_B / (m_A + m_B)$ ; that is

$$D_{AB} = \frac{3}{8} \frac{1}{n \bar{d}^2} \sqrt{\frac{k_B T}{2 \pi m_r}} \quad (4.43)$$

Equation (4.43) is recommended for calculation of the binary diffusion coefficient. Recall that the Schmidt number is the ratio of the momentum diffusivity to the mass diffusivity, i.e.,

$$Sc \equiv \frac{\nu}{D_{AB}} \quad (4.44)$$

The Lewis number is defined as the ratio of the mass diffusivity to the thermal diffusivity as follows:

$$Le \equiv \frac{D_{AB}}{\alpha} = \frac{Pr}{Sc} \quad (4.45)$$

Heat and mass transfer analogy provides a convenient way to calculate convective mass transfer in a boundary layer. The mass transfer rate is related to the convective mass transfer coefficient  $h_m$  by

$$\dot{m}_B = h_m A_s (\rho_{B,s} - \rho_{B,\infty}) \quad (4.46)$$

where  $A_s$  is the surface area,  $\rho_B$  is the density of species B, and subscripts s and  $\infty$  signify that the quantity is at the surface and in the free stream, respectively. Heat and mass transfer analogy gives

$$h_m = \frac{h}{\rho c_p} Le^{-2/3} \quad (4.47)$$

Equations (4.46) and (4.47) are very useful for calculating the heat transfer during evaporation demonstrated in the following example:

**Example 4-4.** Dry air at 30°C flows at a speed of 2 m/s over a flat plate, with an area of  $3 \times 3 \text{ m}^2$ , which is maintained at 24°C. A thin layer of water is formed on the top surface where convection occurs. Determine the heat transfer rate from the plate to the air. For water at 24°C, the saturation pressure  $P_{\text{sat}} = 3 \text{ kPa}$  and the heat of evaporation  $h_{\text{fg}} = 2445 \text{ kJ/kg}$ .

**Solution.** Neglect the temperature gradient inside the water layer and radiative heat transfer. We first evaluate air properties at 300 K and 100 kPa, as in Examples 4-2 and 4-3. The results are  $\rho = P/RT = 1.163 \text{ kg/m}^3$ ,  $\mu = 1.83 \times 10^{-5} \text{ N} \cdot \text{s/m}^2$ ,  $c_p = 1003 \text{ J/(kg} \cdot \text{K)}$ ,  $Pr = 0.737$ , and  $\kappa = 0.025 \text{ W/(m} \cdot \text{K)}$ . Hence,  $Re_L = 3.8 \times 10^5$ . From Eq. (2.40),  $\bar{h} = 0.664 (\kappa/L) Re_L^{1/2} Pr^{1/3} = 3.08 \text{ W/(m}^2 \cdot \text{K)}$  and  $\dot{Q}_{\text{conv}} = \bar{h} A_s (T_s - T_\infty) = -166.2 \text{ W}$ . The negative sign indicates that the convection heat transfer is from the air to the surface.

To calculate the mass transfer rate, we assume that  $\rho_{B,\infty} = 0$  (dry air) and  $\rho_{B,s} = P_{\text{sat}} M_{H_2O} / (\bar{R} T_s) = 3000 \times 18 / (8314 \times 297) = 0.022 \text{ kg/m}^3$  (saturated water vapor). Using Eq. (4.43) with  $T = 300 \text{ K}$ , we can estimate the binary diffusion coefficient between air and water to be  $D_{AB} = 1.7 \times 10^{-5} \text{ m}^2/\text{s}$ , which is about two-thirds of the measured value:  $D_{AB} = 2.56 \times 10^{-5} \text{ m}^2/\text{s}$ . Considering the simplifications made in deriving the diffusion coefficient, the agreement is reasonable. Using the measured  $D_{AB}$ , we find  $Le = D_{AB}/\alpha = 1.2$  and  $h_m = 0.00234 \text{ m/s}$  from Eq. (4.47). The mass transfer rate  $\dot{m}_B = h_m A_s \rho_{B,s} = 0.46 \text{ g/s}$ , and the heat transfer rate by evaporation is  $\dot{Q}_{\text{evap}} = \dot{m}_B h_{\text{fg}} = 1124.7 \text{ W}$ . The total heat transfer rate is the sum of evaporation and convection, i.e.,  $\dot{Q}_{\text{evap}} + \dot{Q}_{\text{conv}} = 1124.7 - 166.2 = 958.5 \text{ W}$ . This example suggests that evaporative cooling is an important mechanism of heat transfer at wetted surfaces.

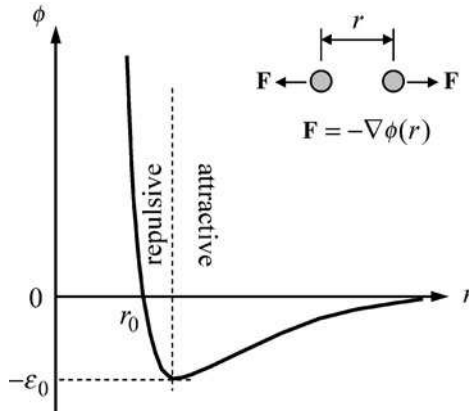
#### 4.2.4 Intermolecular Forces

Although the mean-free-path method is simple and can predict the temperature and pressure dependence of the transport coefficient correctly, the rigid-elastic-sphere model does not represent the actual collision process. Collision between molecules does not necessarily occur by contact, as in the case with billiard balls. It is a force field described by the intermolecular potential that governs the collision process between molecules. For a pair of molecules, there exists an attractive force (i.e., the van der Waals force) between them as a result of the fluctuating dipoles in the two molecules. This force generally varies with  $1/r^7$  for sufficiently large  $r$ , where  $r$  is the center-to-center distance between the two molecules. On the other hand, when the distance between the molecules becomes very small, a strong repulsive internuclear force arises because of the overlap of electronic orbits in the atoms. The combination of these potentials is modeled by some semi-empirical function, such as the Lennard-Jones <6,12> potential, expressed as follows:

$$\phi_{ij}(r_{ij}) = -4\epsilon_0 \left[ \left( \frac{r_0}{r_{ij}} \right)^6 - \left( \frac{r_0}{r_{ij}} \right)^{12} \right] \quad (4.48)$$

where  $\phi_{ij}$  is the intermolecular potential,  $r_{ij}$  is the distance between the  $i$ th and  $j$ th particles,  $\epsilon_0$  is a constant, and  $r_0$  is a characteristic length. Notice that the potential has a minimum  $\phi_{ij} = -\epsilon_0$  at  $r_{ij} \approx 1.12r_0$ , where the attractive and repulsive forces balance each other. For typical gas molecules,  $r_0$  ranges from 0.25 to 0.4 nm. The potential function is illustrated in Fig. 4.8. The force between the molecules is the negative gradient of the potential, i.e.,

$$\mathbf{F}_{ij} = -\nabla\phi_{ij} = \frac{24\epsilon_0}{r_0} \left[ 2\left(\frac{r_0}{r_{ij}}\right)^{13} - \left(\frac{r_0}{r_{ij}}\right)^7 \right] \frac{\mathbf{r}_{ij}}{r_{ij}} \quad (4.49)$$



**FIGURE 4.8** Illustration of the intermolecular potential  $\phi(r)$  as a function of the distance  $r$  between two molecules. The subscripts  $i$  and  $j$  used in Eq. (4.48) and Eq. (4.49) are dropped for simplicity.

The combination of Eq. (4.49) with Eq. (3.1) allows computer simulation of the trajectory of each molecule when the initial position and velocity are prescribed. Although molecular dynamics is a powerful tool for dense phases and for the study of phase change problems, it is not very effective in dealing with dilute gases. The direct simulation

Monte Carlo (DSMC) method is an alternative to the deterministic method and has been used extensively in gas dynamics. Additional discussions about these numerical techniques will be given in Sec. 4.4 on microfluidics. In the next section, a more sophisticated kinetic theory based on the Boltzmann transport equation will be presented.

### 4.3 THE BOLTZMANN TRANSPORT EQUATION

In addition to the rigid-sphere assumption, the simple kinetic theory is based on local equilibrium and cannot be used to study nonequilibrium processes that happen at a timescale much less than the relaxation time or at a length scale less than the mean free path. The Boltzmann transport equation (BTE) is the basis of classical transport theories of molecular and atomic systems. It is not limited to local equilibrium and can be applied to small length scales and small timescales. The equation formulated by Ludwig Boltzmann in his original investigation of the dynamics of gases over 130 years ago has been extended to the study of electron and phonon transport in solids, as well as radiative transfer in gases. Macroscopic conservation and rate equations can be derived from the BTE under appropriate assumptions. A brief introduction of the BTE is given in this section. More detailed coverage of the history, formulation, and solution techniques of the BTE can be found from Chapman and Cowling,<sup>5</sup> Tien and Lienhard,<sup>6</sup> and Cercignani.<sup>7</sup>

Suppose at time  $t$ , a particle at the spatial location  $\mathbf{r}$  moves with a velocity  $\mathbf{v}$ . At  $t + dt$ , without collision, the particle will move to  $\mathbf{r} + d\mathbf{r} = \mathbf{r} + \mathbf{v}dt$  and its velocity becomes  $\mathbf{v} + d\mathbf{v} = \mathbf{v} + \mathbf{a}dt$ . Here,  $\mathbf{a} = \mathbf{F}/m$  is the acceleration in a body force field. Therefore, in the absence of collision, the probability of finding a particle in the phase space does not change with time. Therefore,

$$\frac{f(\mathbf{r} + \mathbf{v}dt, \mathbf{v} + \mathbf{a}dt, t + dt) - f(\mathbf{r}, \mathbf{v}, t)}{dt} = \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = 0 \quad (4.50a)$$

where

$$\frac{\partial f}{\partial \mathbf{r}} = \nabla f = \frac{\partial f}{\partial x} \hat{\mathbf{x}} + \frac{\partial f}{\partial y} \hat{\mathbf{y}} + \frac{\partial f}{\partial z} \hat{\mathbf{z}}$$

is the gradient, and

$$\frac{\partial f}{\partial \mathbf{v}} = \nabla_{\mathbf{v}} f = \frac{\partial f}{\partial v_x} \hat{\mathbf{x}} + \frac{\partial f}{\partial v_y} \hat{\mathbf{y}} + \frac{\partial f}{\partial v_z} \hat{\mathbf{z}}$$

can be considered as the gradient defined in the velocity space. Equation (4.50a) is the Liouville equation in classical mechanics. In the absence of both body force and collision, the substantial derivative of the distribution function is

$$\frac{Df}{Dt} \equiv \frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} = 0 \quad (4.50b)$$

Generally speaking, particles in random motion collide with each other at very high frequencies unless the density is extremely low. A major advance in the kinetic theory of gases is the introduction of the collision term proposed by Boltzmann in the 1870s. The BTE can be written as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} + \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} = \left[ \frac{\partial f}{\partial t} \right]_{\text{coll}} \quad (4.51)$$

where the collision term can be separated into a source term and a sink term such that

$$\left[ \frac{\partial f}{\partial t} \right]_{\text{coll}} = \Gamma_+ - \Gamma_- = \sum_{\mathbf{v}'} [W(\mathbf{v}, \mathbf{v}')f(r, \mathbf{v}', t) - W(\mathbf{v}', \mathbf{v})f(r, \mathbf{v}, t)] \quad (4.52)$$

Here,  $W(\mathbf{v}, \mathbf{v}')$  is called the scattering probability, which can be understood as the fraction of particles with a velocity  $\mathbf{v}'$  that will change their velocity to  $\mathbf{v}$  per unit time due to collision. The function  $W$  depends on the nature of the scatters and is usually a complicated nonlinear function of the velocities.

The BTE is a nonlinear integro-differential equation that cannot be solved exactly. Approximations are usually used to facilitate the solution for given applications. The *relaxation time approximation* provides an easier way to solve the BTE under conditions not too far away from the equilibrium. It gives a linear collision term:

$$\left[ \frac{\partial f}{\partial t} \right]_{\text{coll}} = \frac{f_0 - f}{\tau(\mathbf{v})} \quad (4.53)$$

where  $f_0$  is the equilibrium distribution and the relaxation time  $\tau$  is often treated as independent of the velocity. The solution of Eq. (4.53) gives  $f(t) - f_0 = [f(t_1) - f_0]\exp[-(t - t_1)/\tau]$ , where  $t_1$  is the initial time when the system deviates somewhat from the equilibrium. This suggests that an equilibrium will be reached at a timescale  $\Delta t = t - t_1$  on the order of  $\tau$ . Furthermore, it is collision that restores a system from a nonequilibrium state to an equilibrium state. David Enskog proposed a successive approximation method to include higher-order scattering term by introducing a small perturbation to the equilibrium distribution. This is the well-known Chapman-Enskog method.<sup>5-7</sup>

### 4.3.1 Hydrodynamic Equations

The continuity, momentum, and energy equations can be derived from the BTE. Multiplying the BTE by a molecular quantity  $\psi$  and integrating it over all velocities, we have

$$\int_{\mathcal{W}} \psi \frac{\partial f}{\partial t} d\mathcal{W} + \int_{\mathcal{W}} \psi \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} d\mathcal{W} + \int_{\mathcal{W}} \psi \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} d\mathcal{W} = \int_{\mathcal{W}} \psi (\Gamma_+ - \Gamma_-) d\mathcal{W} \quad (4.54)$$

Using the definition of local average  $\bar{\psi} = \frac{1}{n} \int_{\mathcal{W}} f \psi d\mathcal{W}$  from Eq. (4.5), the first term in the above equation becomes

$$\int_{\mathcal{W}} \psi \frac{\partial f}{\partial t} d\mathcal{W} = \frac{\partial}{\partial t} \int_{\mathcal{W}} \psi f d\mathcal{W} - \int_{\mathcal{W}} f \frac{\partial \psi}{\partial t} d\mathcal{W} = \frac{\partial(n\bar{\psi})}{\partial t} - n \frac{\partial \bar{\psi}}{\partial t} \quad (4.55a)$$

Note that  $\nabla \cdot (\psi \mathbf{v}) = \mathbf{v} \cdot \nabla \psi + \psi \nabla \cdot \mathbf{v} = \mathbf{v} \cdot \nabla \psi$  since the velocity components are independent variables in the phase space. For the second term, we have

$$\int_{\mathcal{W}} \psi (\mathbf{v} \cdot \nabla f) d\mathcal{W} = \nabla \cdot \int_{\mathcal{W}} \psi \mathbf{v} f d\mathcal{W} - \int_{\mathcal{W}} f \nabla \cdot (\psi \mathbf{v}) d\mathcal{W} = \nabla \cdot (n\bar{\psi \mathbf{v}}) - n \bar{\mathbf{v}} \cdot \nabla \bar{\psi} \quad (4.55b)$$

The third term is

$$\int_{\mathcal{W}} \psi \mathbf{a} \cdot \frac{\partial f}{\partial \mathbf{v}} d\mathcal{W} = \mathbf{a} \cdot \left[ (\psi f) \Big|_{v_x, v_y, v_z = -\infty}^{v_x, v_y, v_z = \infty} - \int_{\mathcal{W}} f \frac{\partial \psi}{\partial \mathbf{v}} d\mathcal{W} \right] = -n \mathbf{a} \cdot \frac{\partial \bar{\psi}}{\partial \mathbf{v}} \quad (4.55c)$$

Substituting Eq. (4.55) into Eq. (4.54), we obtain

$$\frac{\partial}{\partial t}(n\bar{\psi}) + \nabla \cdot (n\bar{\psi}\mathbf{v}) - n\left(\frac{\partial\bar{\psi}}{\partial t} + \mathbf{v} \cdot \nabla\bar{\psi} + \mathbf{a} \cdot \frac{\partial\bar{\psi}}{\partial\mathbf{v}}\right) = \Phi_+ - \Phi_- \quad (4.56)$$

where the right-hand side contains a source term and a sink term. When  $\psi$  is proportional to the velocity to the  $j$ th power ( $j = 0, 1, 2$ ), or the  $j$ th moment, the source and sink terms in Eq. (4.56) cancel out when reaction is not considered, and the gas particles can be treated as rigid spheres.

We can substitute  $\psi = m$ , the zeroth moment, into Eq. (4.56) to get the mass balance as

$$\frac{\partial\rho}{\partial t} + \nabla \cdot (\rho\mathbf{v}_B) = 0 \quad \text{or} \quad \frac{D\rho}{Dt} + \rho\nabla \cdot \mathbf{v}_B = 0 \quad (4.57a)$$

where  $\mathbf{v}_B = \bar{\mathbf{v}}$  is the *bulk velocity*. This is exactly the same as Eq. (2.41). One can extend the above derivation to a system of multiple gas species involving chemical reaction. For the  $i$ th species, it can be shown that

$$\frac{D\rho_i}{Dt} + \rho_i\nabla \cdot \mathbf{v}_{i,B} = \Phi_{i,\text{net}} \quad (4.57b)$$

where  $\Phi_{i,\text{net}}$  represents the net rate of creation due to reaction.

To derive the momentum equation, substitute the first moment  $\psi = m\mathbf{v}$  into Eq. (4.56). The first term becomes  $\partial(\rho\mathbf{v}_B)/\partial t$ . The second term is more complicated. We can separate the velocity as  $\mathbf{v} = \mathbf{v}_B + \mathbf{v}_R$ , where  $\mathbf{v}_R$  is due to the random motion and is called *thermal velocity*, whose average is zero. Therefore,  $(\mathbf{v}_B + \mathbf{v}_R)(\mathbf{v}_B + \mathbf{v}_R) = \mathbf{v}_B\mathbf{v}_B + \mathbf{v}_R\mathbf{v}_R$ , where  $\mathbf{v}_R\mathbf{v}_R$  is a dyadic whose array is a second-order tensor. In fact,  $\rho\mathbf{v}_R\mathbf{v}_R$  is nothing but the stress tensor given in Eq. (4.14a). Because  $\psi = m\mathbf{v}$  and the velocity is an independent variable, both  $\partial\psi/\partial t$  and  $\nabla\psi$  vanish. The last term is simply  $\rho\mathbf{a}$ . The combination of all the terms gives

$$\rho \frac{\partial\mathbf{v}_B}{\partial t} + \mathbf{v}_B \frac{\partial\rho}{\partial t} + \mathbf{v}_B\nabla \cdot (\rho\mathbf{v}_B) + \rho(\mathbf{v}_B \cdot \nabla)\mathbf{v}_B + \nabla \cdot \{P_{ij}\} - \rho\mathbf{a} = 0$$

Applying the mass balance equation, we can simplify the momentum equation as

$$\frac{D\mathbf{v}_B}{Dt} = -\frac{1}{\rho}\nabla \cdot \{P_{ij}\} + \mathbf{a} \quad (4.58)$$

The stress tensor can be obtained from Eq. (4.14b). When Stokes' hypothesis is used to simplify the constitutive relations between the stresses and the velocity gradients of a viscous fluid, we have

$$P_{ij} = \begin{cases} P - 2\mu\frac{\partial v_i}{\partial x_i} + \frac{2}{3}\mu\nabla \cdot \mathbf{v}_B, & i = j \\ -\mu\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right), & i \neq j \end{cases} \quad (4.59)$$

where  $x_i = x, y, z$  (for  $i = 1, 2, 3$ ), and  $v_i$  ( $i = 1, 2, 3$ ) is the velocity component of the bulk velocity  $\mathbf{v}_B$ . Substituting Eq. (4.59) into Eq. (4.58), one obtains exactly the same result as Eq. (2.42). The derivation is left as an exercise (see Problem 4.12).

Next, we derive the energy equation for viscous flow of a monatomic gas, using the second moment.  $\psi = \varepsilon = \frac{1}{2}m\mathbf{v}_R^2$  because only the random motion contributes to the internal energy. The first term in Eq. (4.46) becomes  $\partial(\rho u)/\partial t$ , where  $u$  is the mass specific internal energy. The second term  $\nabla \cdot (n\bar{\psi}\mathbf{v}) = \frac{1}{2}\nabla \cdot (\rho\mathbf{v}_B\mathbf{v}_R^2) + \frac{1}{2}\nabla \cdot (\rho\mathbf{v}_R\mathbf{v}_R^2) = \nabla \cdot (\rho u\mathbf{v}_B) + \nabla \cdot \mathbf{J}_E$ ,

where  $\mathbf{J}_E = n \int_{\mathcal{D}} f \mathbf{v}_R \varepsilon d\mathcal{D}$  is the energy flux vector to be discussed further in Sec. 4.3.2. Notice that  $\overline{\partial(\frac{1}{2} m \mathbf{v}_R^2) / \partial t} = 0$  and  $\overline{n \mathbf{v} \cdot \nabla(\frac{1}{2} m \mathbf{v}_R^2)} = \rho \mathbf{v} \cdot [\mathbf{v}_R \cdot \nabla(\mathbf{v} - \mathbf{v}_B)] = -\rho \mathbf{v} \cdot (\mathbf{v}_R \cdot \nabla \mathbf{v}_B) = \{P_{ij}\} : \nabla \mathbf{v}_B$ , which can be considered the product of the momentum flux and the bulk velocity gradient. This tensor product can be calculated according to  $\{P_{ij}\} : \nabla \mathbf{v}_B = \sum_i \sum_j P_{ij} (\partial v_i / \partial x_j)$ . For the force term, we have  $n \mathbf{a} \overline{\partial(\frac{1}{2} m \mathbf{v}_R^2) / \partial \mathbf{v}} = \rho \mathbf{a} \cdot \overline{\mathbf{v}_R} = 0$ . The energy conservation equation can be expressed as

$$\frac{\partial}{\partial t} (\rho u) + \nabla \cdot (\rho u \mathbf{v}_B) + \nabla \cdot \mathbf{J}_E + \{P_{ij}\} : \nabla \mathbf{v}_B = 0$$

After it is simplified using the continuity equation, we have

$$\rho \frac{Du}{Dt} = -\nabla \cdot \mathbf{J}_E - \{P_{ij}\} : \nabla \mathbf{v}_B \quad (4.60)$$

The left-hand side consists the transient term and the advection term. Among the two terms on the right-hand side, the first one corresponds to the energy transfer by heat diffusion, and the second one includes the pressure effect as well as the viscous dissipation. It can be shown that Eq. (4.60) is the same as Eq. (2.43) (see Problem 4.13). In a stationary medium with  $\mathbf{v}_B = 0$ , Eq. (4.60) reduces to the heat diffusion equation,  $\kappa \nabla^2 T = \rho c_p (\partial T / \partial t)$  (see Problem 4.14). In the earlier derivations, the velocity  $\mathbf{v}$  is taken as an independent variable in the distribution function. Another way of deriving the macroscopic conservation equations is to take the random velocity  $\mathbf{v}_R$  as the independent variable and modify the distribution function to a new one,  $f(\mathbf{r}, \mathbf{v}_R, t)$ .<sup>5</sup>

In deriving the macroscopic conservation equations, it is assumed that  $f(\mathbf{r}, \mathbf{v}, t)$  obeys certain equilibrium distribution at any given location. This is the local-equilibrium assumption, which is only valid when the mean free path is much smaller than the characteristic length. For systems with dimensions comparable to or smaller than the mean free path, the local-equilibrium assumption breaks down, as will be discussed in Sec. 4.4 and forthcoming chapters.

### 4.3.2 Fourier's Law and Thermal Conductivity

The transport equations and coefficients can be obtained based on the BTE. Here, as an example, the 1-D Fourier's law will be derived. When the characteristic time  $t_c$  is much greater than the relaxation time and the length scale is much greater than the mean free path, we write the BTE under the relaxation time approximation Eq. (4.53) as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} = \frac{f_0 - f}{\tau(\mathbf{v})}$$

Assume that the temperature gradient is in the  $x$  direction and the medium is stationary. If the medium moves with a bulk velocity, we can set the coordinate to move at the bulk velocity so that the local average velocity is zero. The distribution function will vary with  $x$  only, and at steady state, we have  $v_x (df/dx) = (f_0 - f)/\tau$ . We further assume that  $f$  is not very far away from equilibrium so that  $df/dx \approx df_0/dx$ , which is the condition of *local equilibrium*. Therefore,

$$f \approx f_0 - \tau v_x \frac{df_0}{dT} \frac{dT}{dx} \quad (4.61)$$



The heat flux in the  $x$  direction is

$$J_{E,x} = q_x'' = \int_{\varpi} f \varepsilon v_x d\varpi = \int_{\varpi} \left( f_0 - \tau v_x \frac{df_0}{dT} \frac{dT}{dx} \right) \varepsilon v_x d\varpi \quad (4.62a)$$

Some explanation is needed about the equilibrium distribution. Let us use Maxwell's velocity distribution Eq. (3.43) as an example. The distribution function can be viewed as a function of  $v_x$ ,  $v_y$ , and  $v_z$  for a given  $T$  and other parameters when integrating over the velocity space. On the other hand, it can be viewed as a function of  $T$  only by fixing  $v_x$ ,  $v_y$ ,  $v_z$ , and all other parameters. This allows us to obtain  $df_0/dT$ , which in turn, can be viewed as a function of  $v_x$ ,  $v_y$ , and  $v_z$  in order to carry out the integration. Note that  $\int_{\varpi} f_0 \varepsilon v_x d\varpi = 0$  because  $f_0$  is the equilibrium distribution. It should also be noted that the integration over  $v_x^2$  is the same as the integration over  $v_y^2$  or  $v_z^2$ . Hence, the integration over  $v_x^2$  equals 1/3 of the integration over  $v^2$ . After some manipulations, we can write

$$q_x'' = -\kappa \frac{dT}{dx} \quad (4.62b)$$

which is Fourier's law with the thermal conductivity expressed as

$$\kappa = \frac{1}{3} \int_{\varpi} \frac{df_0}{dT} \tau v^2 \varepsilon d\varpi \quad (4.63a)$$

The above integral is often converted to integration over the energy, which gives

$$\kappa = \frac{1}{3} \int_0^{\infty} \frac{df_0}{dT} \tau v^2 \varepsilon D(\varepsilon) d\varepsilon \quad (4.63b)$$

where  $D(\varepsilon)$  is the density of states, which can be considered as the volume in the velocity space per unit energy interval. If we take both the relaxation time  $\tau$  and the velocity  $v$  as their average values that can be moved out of the integral, we have  $\kappa \approx \frac{1}{3} \tau \bar{v}^2 \rho C_v = \frac{1}{3} \Lambda \bar{v} \rho C_v$ , which is identical to Eq. (4.33). If we assume only  $\tau$  is independent of frequency, we can use Maxwell's velocity distribution Eq. (3.43) to evaluate  $\kappa = (\tau/3) \int_{\varpi} (df_0/dT) v^2 \varepsilon d\varpi$  for a monatomic gas ( $\varepsilon = \frac{1}{2} m v^2$ ) (see Problem 4.15). The result  $\kappa = 1.31 \Lambda \bar{v} \rho C_v$  is in good agreement with Eq. (4.36a), considering the assumption of a constant relaxation time.

Under local-equilibrium assumption and by applying the relaxation time approximation, we can write the 3-D Fourier's law as

$$\mathbf{q}'' = \mathbf{J}_E = \int_{\varpi} f \mathbf{v} \varepsilon d\varpi = -\kappa \nabla T \quad (4.64)$$

where  $\kappa$  is already given in Eq. (4.63a) and  $\mathbf{v}$  is the thermal velocity. Equation (4.64) proves that the first term on the right-hand side of Eq. (4.60) is indeed associated with heat diffusion.

#### 4.4 MICRO/NANOFLUIDICS AND HEAT TRANSFER

---

A large number of microdevices involving fluid flow in microstructures have been designed and built since the late 1980s. Examples are microsensors, actuators, valves, heat pipes, and microducts used in heat engines and heat exchangers.<sup>8-10</sup> Micro/nanofluidics research is an active area with applications in biomedical diagnosis (lab-on-a-chip) and drug delivery, MEMS/NEMS sensors and actuators, micropumps for ink-jet printing, and microchannel heat sinks for electronic cooling. Many researchers are also studying fluid flow inside nanostructures, such as nanotubes, and developing unique devices, such as nanojets.

Under the continuum assumption, matter is continuous and indefinitely divisible. Properties are defined as the average over elements much larger than the microscopic structure of the fluid but much smaller than the macroscopic device scale. For flow inside micro/nanostructures, the mean free path of the fluid molecules may be comparable to or smaller than the characteristic dimensions. The continuum assumption is often not valid since the interaction between the molecules and the solid surfaces becomes important. In his seminal paper in 1946, Tsien drew the attention of aerodynamicists to the study of non-continuous fluid mechanics, for applications in high-altitude flights and vacuum systems, which subsequently formed the field of *rarefied gas dynamics*.<sup>11</sup> In the same paper, he delineated the realms from conventional gas dynamics (i.e., continuum regime): *slip flow*, "blank" (which was later called *transition flow*), and *free molecule flow* based on the ratio of the mean free path to the characteristic length, i.e., the Knudsen number as will be discussed in the next section.

Some of the earlier studies are still valid and can help understand fluid flow in microstructures.<sup>12,13</sup> On the other hand, there are several aspects that are unique to microfluidics, making it distinctly different from the rarefied fluid dynamics. In microstructures, surface-to-volume ratio is much greater than that in macrostructures, and hence, surface forces become dominant over body forces. One of the direct impacts is a significant pressure drop and a greater mass flow rate than that predicted with the continuum theory.<sup>9,10</sup> Because of the large pressure drop, the velocity is usually not very high. The Reynolds number is significantly smaller due to the small dimensions and relatively low velocity. The axial heat conduction, which is negligible for macroflow, may become important for micro/nanoflow. Due to the large pressure drop, compressibility is another issue that needs to be considered even though the speed is much less than the speed of sound. A change in the density further complicates the pressure distribution, making it nonlinear along the streamline. Liquid is also used in many applications such as microchannel cooling. Furthermore, the phase change by evaporation and condensation is another important aspect in a number of microdevices, such as micro-heat pipes.

Although measurements in micro/nanoflow are challenging, a large number of miniaturized flow and temperature sensors have been developed and integrated into the microdevices to perform measurements with a high spatial resolution. Submicron polysilicon hot-wire anemometers, hot-film shear-stress sensors, piezoresistive and diaphragm-type pressure sensors, and submicron thermocouples are some examples.<sup>9</sup> For flow visualization, both x-ray and caged-dye techniques have been used to image the flow field. Micro-particle image velocimetry (PIV) is a powerful technique for flow visualization and sometimes for thermal measurements. In micro-PIV, small particles imbedded into the fluid scatter pulsed laser light. A microscopic system allows the illumination and collection of the scattered light into a CCD camera. The flow is illuminated at two times, and the velocity vectors are determined based on the displacement of particles. The temperature field can be determined based on the Brownian motion, i.e., random fluctuation of the particles.<sup>14</sup>

The next section focuses on gas flow, which can be categorized into different regimes based on the range of the Knudsen number. Examples of slip flow and free molecule conduction are provided to illustrate the effect of rarefaction. More detailed research on microfluidics and microflow devices can be found from the monographs.<sup>14,15</sup> Reviews of recent studies on the heat transfer in microstructures involving liquids, evaporation, and condensation can be found from Peterson et al.<sup>16</sup>, Garimella and Sobhan<sup>17</sup>, and Poulikakos et al.<sup>18</sup>

#### 4.4.1 The Knudsen Number and Flow Regimes

The continuum model is no longer valid when one of the geometric dimensions, called the characteristic dimension  $L_c$ , is comparable to the mechanistic length, such as the mean free path  $\Lambda$ . This can happen when the gas is at very low pressure (rarefied) or when the characteristic dimension is extremely small: from a few micrometers down to several nanometers in micro- and nanochannels. As a result, boundary scattering becomes significant and the gas molecules have a large chance to collide with the wall as compared to the collision between molecules.

The ratio of the mean free path to the characteristic length defines an important dimensionless parameter, called the Knudsen number:

$$Kn \equiv \frac{\Lambda}{L} \quad (4.65)$$

Recall the definition of the Reynolds number and the Mach number, which are  $Re_L = \rho v_\infty L / \mu$  and  $Ma = v_\infty / v_a$ , where  $v_\infty$  is free stream velocity and  $v_a = \sqrt{\gamma RT}$  is the speed of sound in the gas. When internal flow is considered,  $v_\infty$  should be replaced by the bulk velocity  $v_m$ . From Eq. (4.30),  $\mu = \rho \Lambda \sqrt{2RT/\pi}$ . Therefore,

$$Kn = \sqrt{\frac{\pi \gamma}{2}} \frac{Ma}{Re_L} \quad (4.66)$$

An example is in the boundary layer for flow over a flat plate at length  $x$ . The characteristic length here is the boundary layer thickness  $\delta$  rather than  $x$ . For a laminar flow,  $L = \delta \propto x / \sqrt{Re_x}$ , or  $Re_L = (\delta/x) Re_x \propto \sqrt{Re_x}$ ; hence in the boundary layer,

$$Kn \propto \frac{Ma}{\sqrt{Re_x}} \quad (4.67)$$

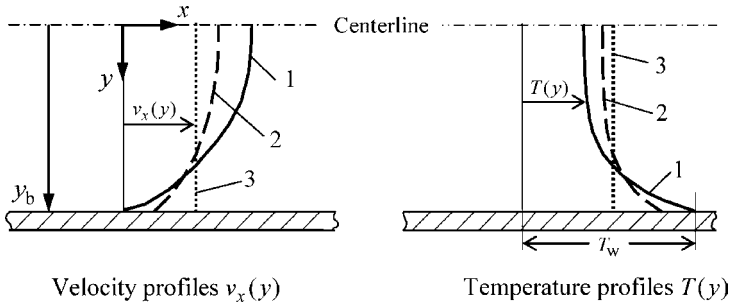
The physics of fluid flow depends much on the magnitude of  $Kn$ . The local  $Kn$  determines the degree of rarefaction and the degree of deviation from the continuum assumption. The regimes are divided based on  $Kn$  in Table 4.2. The regime boundaries are instructive rather than exact because they depend on more parameters about the fluid conditions. A small  $Kn$

**TABLE 4.2** Flow Regimes Based on the Knudsen Number<sup>10</sup>

Regime	Method of calculation	$Kn$ range
Continuum	Navier-Stokes and energy equations with no-slip/ no-jump boundary conditions	$Kn \leq 0.001$
Slip flow	Navier-Stokes and energy equations with slip/ jump boundary conditions, DSMC	$0.001 < Kn \leq 0.1$
Transition	BTE, DSMC	$0.1 < Kn \leq 10$
Free molecule	BTE, DSMC	$Kn > 10$

generally corresponds to a continuum flow ( $Kn < 0.001$ ). In this regime, the Navier-Stokes equations are applicable, the velocity of the fluid at the boundary is the same as that of the wall, and the temperature of the fluid adjacent to the wall is the same as the surface temperature. Care must be taken in regard to the compressibility. Conventionally, the flow can be assumed incompressible if  $Ma < 0.3$ . However, in some microdevices where pressure changes drastically, density change can be significant and thus compressibility must be taken into consideration.

When  $Kn$  is increased from about 0.001 to 0.1, noncontinuum (slip) boundary conditions must be applied. Slip flow refers to the situation when the velocity of the fluid at the wall is not the same as the wall velocity, as shown in Fig. 4.9. In the heat transfer problem,



**FIGURE 4.9** Illustration of the velocity and temperature profiles for internal flow, in the three regimes: 1: continuum, 2: velocity slip and temperature jump, and 3: free molecule.

the temperature of the fluid adjacent to the wall is different from that of the wall, as shown on the right of Fig. 4.9. This is called *temperature jump*. In the slip/jump regime, the Navier-Stokes equations can still be used for the flow with modified boundary conditions, as will be discussed in the next section.

If  $Kn > 10$ , the flow is called free molecule flow that is dominated by ballistic scattering between the molecules and the surfaces. The continuum assumption breaks down completely. No local velocity or temperature of the gas can be defined for the fluid. The “slip” velocity is the same as the velocity of the mainstream, i.e., the fluid velocity will be the same regardless of the distance from the wall, as shown clearly in Fig. 4.9. The same is true for the fluid temperature: no gradient exists near the wall even through there is heat transfer between the wall and the gas. Molecular-based models, such as the BTE or the DSMC, are the best to solve problems in this regime, as well as in the transition regime between slip flow and the free molecule flow.<sup>19</sup>

In the continuum regime, numerical solution techniques include finite element method, finite difference method, boundary element method, and so forth. In recently years, flexible mesh schemes, such as the unstructured grids or mesh-free technique, have become popular. Commercial computational fluid dynamics (CFD) software is often available and can be applied to complex geometries. For numerical solutions of the Boltzmann equations and modeling fluid flow at the molecular level, both deterministic and stochastic methods have been developed. The challenge lies in how to handle the collision terms. Relaxation time approximation and higher-order approximations with nonlinear terms have been applied. Lattice Boltzmann (LB) method based on mesoscopic kinetic equations has emerged as a promising numerical technique for simulating single-phase and multiphase flows involving complex interfacial dynamics and geometries.<sup>20</sup> In the LB method, each grid is a volume element that consists of a collection of particles described by the

Boltzmann distribution function. The fluid particles collide with each other as they move under the applied force at each discrete time step. By developing a simplified version of the kinetic equation, the LB method avoids solving the full BTE and thus reduces computational time and memory. Direct simulation of the molecular movements can be carried out in two ways, as discussed in the following.

Molecular Dynamics (MD) considers the position and the velocity of each particle at any time by using a deterministic approach. The molecules are assumed to obey Newton's laws of motion in Eq. (3.1), and their interactions are governed by the intermolecular potentials. An example is the Lennard-Jones  $\langle 6,12 \rangle$  potential given in Eq. (4.48) that is commonly used directly or with some modifications. In the MD simulation, the first step is called *initialization*, which randomly assigns  $N$  molecules in a region of space and sets their velocities according to some equilibrium distribution. After the initial statistical assignment, all the remaining steps are deterministic. The time evolution of the position and the velocity of each particle is found by integrating Newton's equations of motion, numerically, using a small time step. Periodic boundary conditions are often used to simulate the inlet and the outlet of the flow. Statistical averaging, called ensemble averaging, is used to calculate the internal energy, effective temperature, pressure, and other properties at a given time. The internal energy is the sum of the total kinetic and potential energies. The temperature is based on the average kinetic energy (for monatomic gases). The pressure is calculated using the *virial theorem*.<sup>18</sup> Usually the simulation time step is on the order of femtoseconds, and it requires thousands of time steps to simulate a process for a few picoseconds in real time. The required computational time is proportional to the square of the number of particles  $N$  in the simulation. Therefore, the MD method provides complete information of the trajectories of all particles at a great computational expense. This method is best suited for dense gases, liquids, and solids for which a large number of particles are confined within a small volume. The MD method is particularly useful at the nanoscale as the number of particles becomes reasonably small and the total time steps are manageable. It can also be used to simulate boiling and vaporization, as well as the ablation process. Note that the MD method is often the only method available for the study of some nanoscale phenomena because no experiments could be conducted at that time.

Considering the inefficiency in modeling dilute gases using the MD method, Bird in the 1960s established a statistical technique to model rarefied gas flow and transport processes.<sup>19</sup> This method is called the direct simulation Monte Carlo (DSMC) method that has matured as a powerful simulation tool, especially for transition flow and free molecule flow. Some have combined it with continuum models to form a hybrid method for multiscale simulation.<sup>15</sup> The principle of the DSMC method is the same as that of the MD method; however, intermolecular interactions are dealt with entirely on a probabilistic basis rather than the deterministic basis. In the DSMC method, the space is divided into cells, each with a large number of molecules that mimic but do not follow exactly the motion of real molecules. The motion of particles and collisions between them are simulated via a probabilistic process using a time step smaller than the relaxation time. The interaction between the molecules and the boundary is also simulated according to certain statistical models. Since only a small portion of particles are actually simulated at each time step to represent the actual molecules, the computational time is proportional to  $N$  rather than  $N^2$  as in the MD method. This greatly reduces the required computational resources, although the DSMC method is not so efficient for low  $Kn$  flow, where continuum theory or direct solution of the BTE is more effective.

#### 4.4.2 Velocity Slip and Temperature Jump

The interaction between the gas molecules and the wall plays a critical role when the gas becomes rarefied. However, a fundamental understanding of such interaction is often not available. When a molecule impinges on the wall, it will be reflected (or reemitted) after collision with the molecules near the surface of the wall (if adsorption is neglected). If the

reflection is specular, the tangential momentum (or velocity) will remain the same whereas the normal momentum will be reversed. If all the molecules are specularly reflected, there will not be any shear force or friction between the gas and the wall. However, this is not the case for most engineering applications. Another extreme is the diffuse reflection case, in which the molecule will acquire mutual equilibrium with the wall and be reemitted randomly into the hemisphere. For a stream of molecules, the effect is such that the reflected molecules will follow the Maxwell velocity distribution at the wall temperature. The *momentum accommodation coefficients* can be defined as

$$\alpha_v = \frac{p_i - p_r}{p_i - p_w} \Big|_{\parallel}, \text{ for tangential components} \quad (4.68)$$

and

$$\alpha_{v'} = \frac{p_i - p_r}{p_i - p_w} \Big|_{\perp}, \text{ for normal components} \quad (4.69)$$

where  $p = mv$  is the momentum, the subscripts  $i$  and  $r$  represent the incident and the reflected, and the subscript  $w$  refers to the Maxwell velocity distribution corresponding to the surface temperature  $T_w$ .<sup>2,3</sup> Clearly,  $\alpha_v = \alpha_{v'} = 0$  for specular reflection, and  $\alpha_v = \alpha_{v'} = 1$  for diffuse reflection. Similarly, the *thermal accommodation coefficient* can be defined based on the ratio of energy differences as

$$\alpha_T = \frac{\varepsilon_i - \varepsilon_r}{\varepsilon_i - \varepsilon_w} \quad (4.70a)$$

where  $\varepsilon$  is the average energy of a molecule and  $\varepsilon_w$  is the energy when the molecules are in thermal equilibrium with the wall. For diffuse reflection, the molecule is completely accommodated by the wall, and  $\varepsilon_r = \varepsilon_w$ , i.e.,  $\alpha_T = 1$ . On the other hand, if the reflection is specular, the molecule is not accommodated at all and the reflected energy will be the same as the incident energy, i.e.,  $\alpha_T = 0$ . For monatomic molecules, thermal accommodation coefficient involves translational kinetic energy only, and the kinetic energy is proportional to the absolute temperature. Hence, we can write the thermal accommodation coefficient in terms of temperatures as

$$\alpha_T = \frac{T_i - T_r}{T_i - T_w} \quad (4.70b)$$

For polyatomic molecules, it is reasonable to think that the accommodation coefficients for translational, rotational, and vibrational degrees of freedom may be different. However, due to the lack of information on the nature of interaction between the gas molecules and the wall, usually no distinction is made between the accommodation coefficients for different degrees of freedom. In addition, Eq. (4.70b) is often extended to polyatomic molecules with the assumption that the temperature difference is sufficiently small for the specific heat to be independent of temperature. The thermal accommodation coefficient depends on the nature of the molecules, the molecular structure of the solid wall, the surface roughness and cleanliness, the temperature, and the degree of rarefaction. Saxena and Joshi provide a comprehensive review and data compilation of earlier works.<sup>21</sup> The values of  $\alpha_T$  for air-aluminum and air-steel systems range from 0.87 to 0.97. However,  $\alpha_T$  can be less than 0.02 between pure He gas and clean metallic surfaces. Earlier measurements showed that for most engineering surfaces,  $\alpha_v$  ranges from 0.87 to 1 for air. Arkilic et al. measured tangential momentum accommodation coefficients for  $N_2$ , Ar, and  $CO_2$  in silicon microchannels and found that  $\alpha_v$  is between 0.75 and 0.85.<sup>22</sup> This is possibly due to the relatively smooth crystalline silicon surfaces. Generally speaking,  $\alpha_{v'}$  is not very important and can be assumed the same as  $\alpha_v$ .

Slip flow is an important regime for microchannel flows and MEMS devices. The velocity slip and temperature-jump boundary conditions are presented in this section,

together with some analytical solutions for simple cases. If the wall is not moving, the slip boundary condition based on the geometry shown in Fig. 4.9 reads

$$v_x(y_b) = -\frac{2 - \alpha_v}{\alpha_v} \Lambda \left( \frac{\partial v_x}{\partial y} \right)_{y_b} + 3\sqrt{\frac{R}{8\pi T}} \Lambda \left( \frac{\partial T}{\partial x} \right)_{y_b} \quad (4.71)$$

All the derivatives and the fluid properties are evaluated at  $y = y_b$ . The first term on the right is proportional to the velocity gradient perpendicular to the flow direction, and the second term is known as *thermal creep* due to the temperature gradient along the flow direction. It should be noted that the net mass transfer (creep) is from cold region to hot region. It can be shown that the first term goes with  $Kn$  and the second term goes with the square of  $Kn$ . Higher-order terms can be included by expressing them as  $Kn$  raised to higher powers and higher-order derivatives.<sup>14</sup> The temperature-jump boundary condition reads

$$T(y_b) - T_w = -\frac{2 - \alpha_T}{\alpha_T} \frac{2\gamma}{\gamma + 1} \frac{\Lambda}{Pr} \left( \frac{\partial T}{\partial y} \right)_{y_b} + \frac{v_x^2(y_b)}{4R} \quad (4.72)$$

Equation (4.72) suggests that the temperature of the fluid at the wall will not be the same as the wall temperature, as shown in Fig. 4.9. The second term on the right is due to viscous dissipation caused by the slip velocity and is usually negligibly small.

Let us consider the fluid flow through a channel between two fixed parallel plates, i.e., the Poiseuille flow, as shown in Fig. 4.10. It is assumed that  $W \geq 2H$  and the edge effect

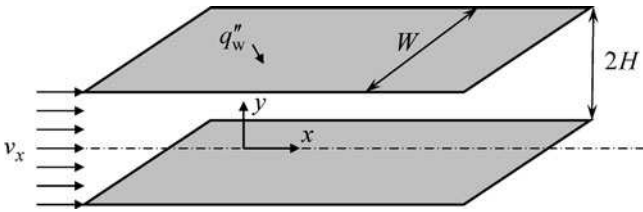


FIGURE 4.10 Micro/nanoscale Poiseuille flow with heat transfer.

can be neglected. When  $Kn = \Lambda/2H$  is less than 0.1 or so, slip flow with temperature-jump boundaries can be applied together with the Navier-Stokes equation and the energy equation to obtain the velocity and temperature distributions. For simplicity, assume the fluid is incompressible and fully developed with constant properties. The momentum equation can be written as

$$\frac{d^2 v_x}{d\eta^2} = \frac{H^2}{\mu} \frac{dP}{dx} \quad (4.73)$$

where  $\eta = y/H$ . The symmetry requires  $(dv_x/d\eta)_{\eta=0} = 0$ . The slip condition given by Eq. (4.71) can be simplified by neglecting thermal creep and higher-order terms, i.e.,

$$v_x(\eta = 1) = -2\beta_v \left( \frac{dv_x}{d\eta} \right)_{\eta=1} \quad (4.74)$$

where

$$\beta_v = \frac{2 - \alpha_v}{\alpha_v} Kn \quad (4.75)$$

The solution gives the fully developed velocity distribution as

$$\frac{v_x(\eta)}{v_m} = \frac{3}{2} \frac{1 + 4\beta_v - \eta^2}{1 + 6\beta_v} \quad (4.76)$$

where  $v_m$  is the bulk velocity, which can be expressed as

$$v_m = \int_0^1 v_x(\eta) d\eta = (1 + 6\beta_v) \frac{H^2}{3\mu} \left( -\frac{dP}{dx} \right) \quad (4.77)$$

Define the *velocity slip ratio*  $\zeta = v_x(\eta = 1)/v_m = 6\beta_v/(1 + 6\beta_v)$ , which is the ratio of the velocity of the fluid at the wall to the bulk velocity. The velocity distribution can be rewritten as

$$\frac{v_x(\eta)}{v_m} = \frac{3 - \zeta}{2} - \frac{3(1 - \zeta)}{2} \eta^2 \quad (4.78)$$

The energy equation is simplified based on Eq. (2.43) without dissipation as follows:

$$\rho c_p v_x \frac{\partial T}{\partial x} = \kappa \left( \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) \quad (4.79a)$$

Let's consider the case with a uniform wall heat flux  $q_w''$  at both plates. For thermally full development,  $\partial T/\partial x$  must not depend on  $x$  and  $y$ ; hence, the term  $\partial^2 T/\partial x^2$  can be dropped out. Applying the energy balance for an elementary control volume inside the fluids, we can rewrite Eq. (4.79a) after some tedious derivations as follows:

$$\frac{\partial^2 \Theta}{\partial \eta^2} = \frac{v_x}{v_m} \quad (4.79b)$$

where  $\Theta(\eta) = (\kappa/H)(T - T_w)/q_w''$  is a dimensionless temperature. Integrating Eq. (4.79b) yields

$$\Theta(\eta) = \frac{3 - \zeta}{4} \eta^2 - \frac{1 - \zeta}{8} \eta^4 + C_1 \eta + C_2 \quad (4.80)$$

The symmetry at  $\eta = 0$  requires that  $\Theta'(\eta = 0) = 0$ . When the second term in Eq. (4.72) due to viscous dissipation is neglected, the nondimensional boundary condition becomes

$$\Theta(\eta = 1) = -2\beta_T \left( \frac{\partial \Theta}{\partial \eta} \right)_{\eta=1} \quad (4.81)$$

where

$$\beta_T = \frac{2 - \alpha_T}{\alpha_T} \frac{2\gamma}{\gamma + 1} \frac{Kn}{Pr} \quad (4.82)$$

Applying the boundary conditions, we obtain  $C_1 = 0$  and  $C_2 = (\zeta - 5)/8 - 2\beta_T$ . From Eq. (4.81) and Fig. 4.10, the heat flux from the surface to the fluid can be expressed as

$$q_w'' = \kappa \left( \frac{\partial T}{\partial y} \right)_{y=H} = \kappa \frac{T_w - T(y = H)}{2\beta_T H} \quad (4.83)$$

Here,  $2\beta_T H$  is called the *temperature-jump distance*, which can be thought as an effective length for heat conduction between the wall and the fluid. With the assumption of constant properties, the dimensionless bulk temperature can be calculated by

$$\Theta_m = \int_0^1 \frac{v_x(\eta)}{v_m} \Theta(\eta) d\eta \quad (4.84)$$



The Nusselt number is defined based on the hydraulic diameter  $D_h = 4H$  for parallel plates, i.e.,

$$Nu = \frac{hD_h}{\kappa} = \frac{q_w''}{T_w - T_m} \frac{4H}{\kappa} = -\frac{4}{\Theta_m}$$

Substituting the integration of Eq. (4.84), one obtains after some manipulations<sup>23</sup>

$$Nu = \frac{140}{17 - 6\zeta + (2/3)\zeta^2 + 70\beta_T} \quad (4.85)$$

The above equation approaches to  $Nu = 140/17$  when both  $\zeta$  and  $\beta_T$  become negligibly small, i.e., in the continuum limit. Furthermore, the Nusselt number decreases monotonically as  $\beta_T$  increases. Note that  $\beta_T$  will be increased if the mean free path increases or if  $\alpha_T$  decreases. On the other hand, the Nusselt number increases slightly as  $\zeta$  increases, e.g., with a smaller  $\alpha_v$ . In any case, both  $\zeta$  and  $\beta_T$  should be much smaller than unity for the slip-jump conditions to hold. If one of the plates is insulated, while the other plate is maintained at a uniform heat flux  $q_w''$ , the velocity distribution is the same, and the Nusselt number can be calculated from Inman as<sup>23</sup>

$$Nu = \frac{140}{26 - 3\zeta + (1/3)\zeta^2 + 70\beta_T} \quad (4.86)$$

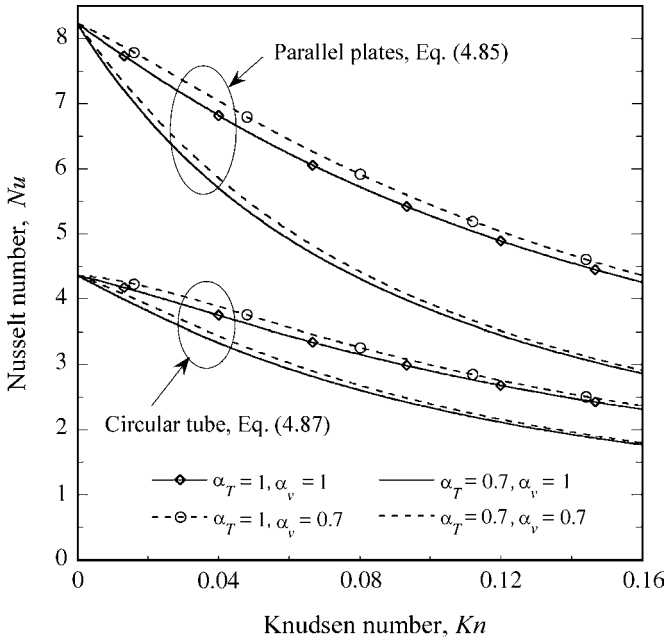
with one insulated wall. Because there is no heat transfer, temperature jump does not occur at the insulated surface. For a circular tube of inner diameter  $D$ , Sparrow and Lin derived the Nusselt number for constant heat flux, which can be expressed as

$$Nu_D = \frac{q_w''}{T_w - T_m} \frac{D}{\kappa} = \frac{48}{11 - 6\zeta + \zeta^2 + 48\beta_T} \quad (4.87)$$

where  $\zeta = 8\beta_v/(1 + 8\beta_v)$ .<sup>24</sup> The expressions for  $\beta_v$  and  $\beta_T$  are the same as in the case with parallel plates, i.e., Eq. (4.75) and Eq. (4.82), except that  $Kn = \Lambda/D$  for a circular tube.

Figure 4.11 illustrates the variation of the Nusselt number as the Knudsen number changes, for air at near room temperature with a uniform heat flux, assuming different accommodation coefficients. Note that  $Kn = \Lambda/2H$  for Poiseuille flow, and  $Kn = \Lambda/D$  for a circular tube. The change in the Knudsen number can be considered as the combined effect of the pressure and the channel dimension. It should be noted that the slip-jump conditions impose an upper limit on the velocity or the temperature gradient near the boundary. In the continuum limit, the shear stress and the Nusselt number are infinite at the entrance and decrease with  $x$  until the flow is fully developed. Assume that the velocity and the temperature are uniform at the entrance. From Eq. (4.74) and Eq. (4.81), we obtain correspondingly  $\tau_s = -\mu(\partial v_x/\partial y)_{y=H} \leq \mu v_m(2H\beta_v) = \tau_{s,\max}$  and  $Nu \leq 2/\beta_T = Nu_{\max}$ , which are the values at the entrance. For a circular tube, it can be shown that  $Nu_{\max} = \beta_T^{-1}$  at the entrance (see Problem 4.22).

Yu and Ameen presented analytical solutions for a rectangular channel with constant wall heat flux on all surfaces using an integral transform method.<sup>25</sup> Hadjiconstantinou and Simek provided an extensive review of the literature dealing with slip channel flow with constant wall temperature.<sup>26</sup> Most of the works did not consider the effect of axial conduction. This assumption is good only for large values of the *Peclet number*, defined as the product of the Reynolds number and the Prandtl number ( $Pe = RePr$ ). As the channel dimensions become very small,  $Re$  will decrease but  $Kn$  will increase. It is possible for  $Kn$  to be large enough for slip and jump to occur at a relatively small  $Re$ . Axial conduction enhances the heat transfer between the fluid and the wall, and thus increases  $Nu$  especially when  $Kn$  is small. In the no-slip case when  $Kn = 0$ , it is well-known that  $Nu = 7.54$  for parallel plates and  $Nu = 3.66$  for circular tubes without axial conduction, i.e.,  $Pe \rightarrow \infty$ . In the



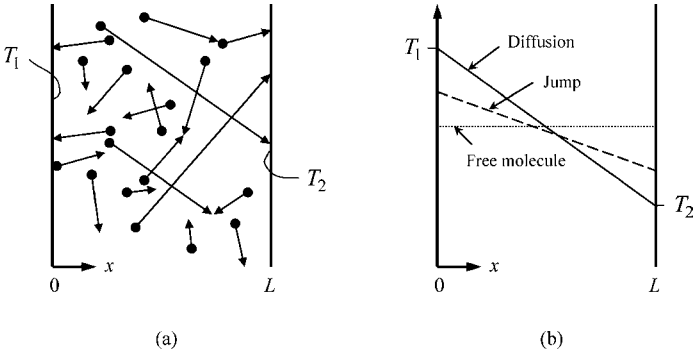
**FIGURE 4.11** Calculated Nusselt number as a function of the Knudsen number for air ( $\gamma = 1.4$  and  $Pr = 0.7$ ) with different accommodation coefficients.

extreme when  $Pe \rightarrow 0$ ,  $Nu$  becomes 8.12 (7.7% increase) and 4.18 (14.2% increase), respectively. These values are much closer to the case of constant heat flux, i.e.,  $Nu = 8.23$  for parallel plates and 4.36 for a circular tube as shown in Fig. 4.11. When both  $\alpha_v$  and  $\alpha_T$  are unity, it can be shown that the Nusselt number is reduced to about 50% of the value when  $Kn$  is varied from 0 to about 0.16, similar to the constant heat flux case. The Nusselt number goes down significantly with decreasing  $\alpha_T$  and goes up somewhat with decreasing  $\alpha_v$ . The lack of sufficient knowledge of the actual behavior of fluids near the wall makes it difficult to precisely determine the accommodation coefficients. Many of the surfaces used in earlier systems are quite different from those used in MEMS and NEMS, where highly pure crystalline dielectric surfaces are commonly used.

#### 4.4.3 Gas Conduction—from the Continuum to the Free Molecule Regime

Free molecule flow is important for flight at high altitudes and often associated with chemical reactions and shock waves. The heat transfer aspects of high-speed flow can be found from Rohsenow and Choi,<sup>12</sup> or Eckert and Drake.<sup>13</sup> In this section, we use a simple case to illustrate the heat transfer regimes for gas conduction. Consider the conduction by gas between two large plates at temperatures  $T_1$  and  $T_2$ , respectively. The plates are separated at a distance  $L$ , and the space in between is filled with an ideal gas, as shown in Fig. 4.12. Neglect radiative and convective heat transfer (bulk motion). The heat conduction of a gas is dominated by diffusion, when  $Kn = \Lambda/L \ll 1$ , where  $\Lambda$  is obtained at some effective mean temperature between  $T_1$  and  $T_2$ . In this case, the heat flux can be calculated by applying Fourier's law,

$$q_{DF}'' = \kappa \frac{T_1 - T_2}{L} \quad (4.88)$$



**FIGURE 4.12** Heat conduction between two large parallel surfaces filled with an ideal gas. (a) Schematic of the gas molecules. (b) Illustration of the temperature distributions.

where  $\kappa$  can be evaluated using Eq. (4.35) at an effective mean temperature defined as

$$T_{m,DF} = \left( \frac{2}{3} \frac{T_1^{3/2} - T_2^{3/2}}{T_1 - T_2} \right)^2 \quad (4.89)$$

The above equation takes into consideration the fact that  $\kappa \propto \sqrt{T}$ , with the assumption that the specific heat is a constant at temperatures between  $T_1$  and  $T_2$ . As long as the density is sufficiently low for the ideal gas model to be valid, the thermal conductivity does not depend on the pressure. The temperature distribution can be obtained by integrating  $q'' = -\kappa(T) (dT/dx)$ , i.e.,

$$T(x) = \left[ T_1^{3/2} - (T_1^{3/2} - T_2^{3/2}) \frac{x}{L} \right]^{2/3} \quad (4.90)$$

which deviates somewhat from a linear relationship. When  $Kn = \Lambda/L \gg 1$ , however, the chance for molecules to collide with the wall is much larger than for them to collide with each other. The actual distance a molecule can travel will be less than the mean free path due to collision with the boundary. In the extreme case, one can completely neglect the collisions between molecules and analyze the heat transfer by the molecules, bouncing back and forth between the two plates. The molecules can be sorted into a forward flux and a backward flux, each at a certain equilibrium temperature, determined by the thermal accommodation coefficients. Assume that the thermal accommodation coefficients  $\alpha_T$  are the same at both walls. The flux temperatures are

$$T_{1'} = \frac{T_1 + (1 - \alpha_T)T_2}{2 - \alpha_T} \quad \text{and} \quad T_{2'} = \frac{T_2 + (1 - \alpha_T)T_1}{2 - \alpha_T} \quad (4.91)$$

The effective mean temperature of the gas in the free molecule regime is defined as

$$T_{m,FM} = \frac{4T_{1'}T_{2'}}{(\sqrt{T_{1'}} + \sqrt{T_{2'}})^2} \quad (4.92)$$

The net heat flux between the two plates can be expressed as<sup>2</sup>

$$q''_{FM} = \frac{T_1 - T_2}{\frac{(2 - \alpha_T)\sqrt{8\pi RT_{m,FM}}}{\alpha_T(\gamma + 1)c_v P}} \quad (4.93)$$

In the free molecule regime, the heat flux is proportional to the pressure  $P$  but independent of  $L$  for the given boundary temperatures. This is because the heat transfer rate is proportional to the number density of particles. For intermediate values of  $Kn$ , the two equations derived under the extreme cases can be combined by adding the thermal resistances such that

$$q'' = \frac{\kappa(T_1 - T_2)}{L \left( 1 + Kn \frac{2 - \alpha_T}{\alpha_T} \frac{9\gamma - 5}{\gamma + 1} \sqrt{\frac{T_{m,FM}}{T_{m,DF}}} \right)} \tag{4.94}$$

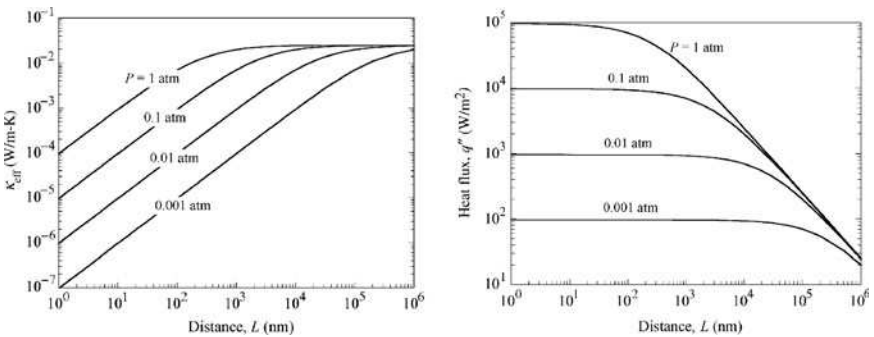
In writing the above equation, we have applied Eq. (4.35) with  $\mu = 2\Lambda P/\sqrt{2\pi RT_{m,DF}}$  from Eq. (4.30). The mean free path  $\Lambda$  used in  $\kappa$  and  $Kn$  should be evaluated at  $T_{m,DF}$ . When the temperature difference between the surfaces is smaller than the absolute temperature of the cooler surface,  $T_{m,DF} \approx T_{m,FM} \approx (T_1 + T_2)/2$ . The physical interpretation of Eq. (4.94) is a temperature jump near the surfaces, due to ballistic interaction of the particles with each surface, and a diffusive middle layer, due to particle-particle collisions. For this reason, Eq. (4.94) is called the temperature-jump approximation, which approaches the diffusion limit when  $Kn \ll 1$  and the free molecule limit when  $Kn \gg 1$ . In the transition region, when  $\Lambda$  is on the same order as  $L$ , Eq. (4.94) may be considered a reduction in the mean free path due to boundary scattering that yields a decrease in the thermal conductivity  $\kappa$  from the bulk or diffusion value. This approach will be further explored in the study of the size effect on the thermal conductivity of thin solid films in the next chapter.

**Example 4-5.** Calculate the heat flux per 1 K temperature difference near room temperature between two large parallel plates filled with air, assuming  $\alpha_T = 0.9$ . Plot the results as a function of distance  $L$  and pressure  $P$ . How will you determine the effective thermal conductivity?

**Solution.** Let  $T_1 = 300.5$  K and  $T_2 = 299.5$  K. The effective temperatures calculated from Eq. (4.89) and Eq. (4.92) are very close to the arithmetic mean temperature of 300 K. Note that  $\gamma = 1.4$  for air at room temperature. Because  $\kappa$  is independent of pressure, we have  $\kappa = 0.025$  W/m<sup>2</sup> · K from Example 4-3. The mean free path obtained from Eq. (4.19) is  $\Lambda/\Lambda_0 = P_0/P$ , where  $\Lambda_0 = 66.5$  nm

at 300 K and the atmospheric pressure  $P_0 = 1$  atm. The effective thermal conductivity can be defined as  $\kappa_{\text{eff}} = q''L/(T_1 - T_2)$ ; hence

$$\kappa_{\text{eff}} = \kappa \left( 1 + Kn \frac{2 - \alpha_T}{\alpha_T} \frac{9\gamma - 5}{\gamma + 1} \right)^{-1} = \kappa \left( 1 + 3.87 \frac{\Lambda_0 P_0}{LP} \right)^{-1}$$



**FIGURE 4.13** Distance dependence of the effective thermal conductivity (left) and the heat flux (right).

It can be seen that  $\kappa_{\text{eff}}$  depends on the product  $LP$  (which is proportional to  $Kn^{-1}$ ). However, the same cannot be said for the heat flux. The calculated results are shown in Fig. 4.13 for the effective thermal conductivity and heat flux as a function of the separation distance. In the diffusion limit,  $\kappa_{\text{eff}}$  is independent of the distance and the pressure, whereas  $q''$  increases as  $L$  is reduced (proportional to  $1/L$ ). At 1 atm, microscale heat transfer becomes important when  $L < 1.5 \mu\text{m}$  (or  $Kn > 0.03$ ), as  $\kappa_{\text{eff}}$  starts to drop, and the dependence of  $q''$  on  $1/L$  becomes nonlinear. In the free molecule limit,  $\kappa_{\text{eff}}$  decreases linearly with both  $L$  and  $P$  (i.e., the  $Kn$ ), whereas  $q''$  is independent of  $L$  but depends linearly on  $P$ . Note that there exists an upper limit of  $q''$  for any given pressure. These trends are clearly demonstrated by Fig. 4.13.

The heat transfer calculation mentioned above is important to cryogenic and low-pressure applications. In recent years, atomic force microscopy has become a versatile tool for probing and manipulating, lithography, and thermal manufacturing and measurements at the nanoscales. The heat transfer between the tip and the surface at several nanometers may be governed by free molecule flow even at ambient conditions (see Problem 4.27). Radiation heat transfer may increase tremendously when the spacing is less than the characteristic wavelength, which is about  $10 \mu\text{m}$  at 300 K. Hence, radiative heat transfer may be a dominating effect. More details on the nanoscale radiative heat transfer will be given in Chap. 10.

## 4.5 SUMMARY

---

The simple kinetic theory was introduced based on the ideal gas model, providing a microscopic description of the transport coefficients, such as viscosity, thermal conductivity, and mass diffusion coefficient. This allows one to gain an intuitive understanding of the macroscopic phenomenological or semi-empirical equations, which are important for heat conduction and convection. The complete Boltzmann transport equation (BTE) was then presented from the microscopic point of view. It was shown that the classical transport equations, such as the Fourier law and Navier-Stokes equations can be derived from the BTE under appropriate assumptions. Similar derivations also apply to electron and phonon systems, which will be studied in Chap. 5. The effect of the Knudsen number on the microchannel flow with ideal gases was discussed. The equations for slip flow were solved for simple geometries to provide modified convection heat transfer correlations. Finally, ballistic heat conduction in free molecule flow was described, and a simplified equation was presented that links the continuum region to free molecule flow in the case of conduction between solid walls filled with an ideal gas. The pressure and distance effects on the thermal conductivity and heat flux were clearly demonstrated. The principles discussed in this chapter not only have applications to microfluidics and convection heat transfer applications but also are important to the subsequent chapters on heat transfer in solid micro/nanostructures.

## REFERENCES

---

1. J. H. Jeans, *The Dynamical Theory of Gases*, 4th ed., Cambridge University Press, Cambridge, UK, 1925.
2. E. H. Kennard, *Kinetic Theory of Gases*, McGraw-Hill, New York, 1938.
3. S. A. Schaaf and P. L. Chambre, *Flow of Rarefied Gases*, Princeton University Press, Princeton, NJ, 1961.
4. J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, 2nd ed., Wiley, New York, 1964.
5. S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-uniform Gases*, 3rd ed., Cambridge University Press, London, UK, 1971.

6. C. L. Tien and J. H. Lienhard, *Statistical Thermodynamics*, Hemisphere, New York, 1985.
7. C. Cercignani, *The Boltzmann Equations and Its Applications*, Springer, New York, 1988.
8. M. J. Madou, *Fundamentals of Microfabrication: The Science of Miniaturization*, 2nd ed., CRC Press, Boca Raton, FL, 2002.
9. C.-M. Ho and Y.-C. Tai, "Micro-electro-mechanical-systems (MEMS) and fluid flows," *Annu. Rev. Fluid Mech.*, **30**, 579–612, 1998.
10. M. Gad-el-Hak, "The fluid mechanics of microdevices—the freeman scholar lecture," *J. Fluids Eng.*, **121**, 5–33, 1999.
11. H.-S. Tsien, "Superaerodynamics, mechanics of rarefied gases," *J. Aeronautical Sci.*, **13**, 653–664, 1946.
12. E. H. Rohsenow and H. Y. Choi, *Heat, Mass, and Momentum Transfer*, Prentice-Hall, Englewood Cliffs, NJ, 1961.
13. E. R. G. Eckert and R. M. Drake, *Analysis of the Heat and Mass Transfer*, McGraw-Hill, New York, 1972.
14. N.-T. Nguyen and S. T. Wereley, *Fundamentals and Applications of Microfluidics*, Artech House, Boston, MA, 2002.
15. G. E. Karniadakis and A. Beskok, *Micro Flows: Fundamentals and Simulation*, Springer-Verlag, New York, 2002.
16. G. P. Peterson, L. W. Swanson, and F. M. Gerner, "Micro heat pipes," in *Microscale Energy Transport*, C. L. Tien, A. Majumdar, and F. M. Gerner (eds.), Taylor & Francis, Washington DC, pp. 295–338, 1998.
17. S. V. Garimella and C. B. Sobhan, "Transport in microchannels—a critical review," *Annu. Rev. Heat Transfer*, **13**, 1–50, 2003.
18. D. Poulikakos, S. Arcidiacono, and S. Maruyama, "Molecular dynamics simulation in nanoscale heat transfer: a review," *Microscale Thermophys. Eng.*, **7**, 181–206, 2003.
19. G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*, Oxford University Press, Oxford, UK, 1994.
20. S. Chen and G. D. Doolen, "Lattice Boltzmann method for fluid flows," *Annu. Rev. Fluid Mech.*, **30**, 329–364, 1998.
21. S. C. Saxena and R.K. Joshi, *Thermal Accommodation and Adsorption Coefficients of Gases*, Hemisphere, New York, 1989.
22. E. B. Arkilic, K. S. Breuer, and M. A. Schmidt, "Mass flow and tangential momentum accommodation in silicon micromachined channels," *J. Fluid Mech.*, **437**, 29–43, 2001.
23. R. Inman, *Laminar Slip Flow Heat Transfer in a Parallel Plate Channel or a Round Tube with Uniform Wall Heating*, NASA TN D-2393, 1964.
24. E. M. Sparrow and S. H. Lin, "Laminar heat transfer in tubes under slip-flow conditions," *J. Heat Transfer*, **84**, 363–369, 1962.
25. S. Yu and T. A. Ameel, "Slip flow convection in isoflux rectangular microchannels," *J. Heat Transfer*, **124**, 346–355, 2002.
26. N. G. Hadjiconstantinou and O. Simek, "Constant-wall-temperature Nusselt number in micro and nano-channels," *J. Heat Transfer*, **124**, 356–364, 2002.

## PROBLEMS

---

- 4.1. (a) Determine the mean free path  $\Lambda$ , average molecular spacing  $L_0$ , and the frequency of collision  $\tau^{-1}$  for air at sea level (15°C and 1 atm).  
 (b) Determine the root-mean-square free path.  
 (c) What is the probability of finding a free path greater than  $4\Lambda$ ?  
 (d) Calculate  $\Lambda$ ,  $L_0$ , and  $\tau^{-1}$  for air at 200 miles above sea level with  $M = 17.3$ ,  $P/P_0 = 5.9 \times 10^{-11}$ , and  $\rho/\rho_0 = 10^{-11}$ , where the subscript 0 signifies properties at sea level.  
 (e) What is the kinetic temperature at this altitude? Explain the reason why  $M$  changes with the altitude.
- 4.2. Use the mean-free-path distribution to answer the following questions:  
 (a) What is the root-mean-square free path in terms of the mean free path  $\Lambda$ ?

- (b) What is the most probable free path?  
 (c) What is the probability of finding a free path greater than  $\Lambda$ ?
- 4.3.** Air is pumped to a pressure  $P = 10$  Pa at  $25^\circ\text{C}$ . Calculate the following quantities: the average distance between adjacent molecules  $L_0$ , the molecular mean free path  $\Lambda$ , the number of collisions that a molecule experiences every second, the molecule flux  $J_N$  on any surface, the most probable speed of the molecules, the most probable velocity of the molecules, and the average kinetic energy of each molecule.
- 4.4.** Hydrogen gas is cooled to  $100$  K, while the pressure is reduced to  $0.1$  Pa. Determine the mean free path  $\Lambda$  and the average frequency of collision. What are the rms speed and the average kinetic energy of a molecule? What is the momentum flux of the gas on the container? What are the most probable free path, the most probable speed, and the most probable velocity?
- 4.5.** What is the dependence of  $\mu$  and  $\kappa$  on pressure and temperature? How does  $D_{AB}$  depend on pressure? For water vapor and air,  $D_{AB} = 2.56 \times 10^{-5}$  m<sup>2</sup>/s at  $298$  K and  $100$  kPa. Plot  $D_{AB}$  as a function of temperature at  $P = 10, 20, 50,$  and  $100$  kPa.
- 4.6.** Calculate  $\mu, c_v, \kappa,$  and  $Pr$  for oxygen and nitrogen at  $100, 300,$  and  $1000$  K, and  $1$  atm. Compare your calculated results with the values tabulated in most heat transfer textbooks to estimate the relative differences.
- 4.7.** A chamber containing  $\text{O}_2$  at  $100$  K and  $10^{-3}$  atm is placed in the outer space. The oxygen leaks to the outer space through a small hole,  $1 \mu\text{m}$  diameter, in the chamber wall.  
 (a) Estimate the number of molecules that escape from the container per unit time.  
 (b) What is the mass flux? What is the mass flow rate?  
 (c) Evaluate the flux of kinetic energy,  $J_{KE}$  using Eq. (4.8). How is your answer compared with  $J_N \times m\bar{v}^2/2$ ? Why are the results different?  
 (d) If the diameter of the hole is increased to  $1$  cm, is the basis of your calculation still valid?
- 4.8.** A tube connects a  $\text{CH}_4$  line to the air. Assuming both ends of the tube are at  $1$  atm and  $25^\circ\text{C}$ , calculate the binary diffusion coefficient between  $\text{CH}_4$  and air. Find the mass flow rate of  $\text{CH}_4$  to the air and that of air to the  $\text{CH}_4$  line, given the tube has an inner diameter of  $5$  mm and a length of  $7$  m. Sketch the concentration distributions in the tube line.
- 4.9.** A tube connects an  $\text{O}_2$  container to a  $\text{N}_2$  container. Assume that the temperature is  $200^\circ\text{C}$  and the pressure is  $2$  atm inside the containers and the tube. Calculate the mass exchange rates of  $\text{O}_2$  and  $\text{N}_2$  from one container to the other, assuming that the tube has an inner diameter of  $5$  mm and a length of  $3$  m.
- 4.10.** Dry air at  $34^\circ\text{C}$  flows over a flat plate of length  $L = 0.1$  m with a velocity of  $15$  m/s. The width of the plate is  $1$  m. The surface of the plate is covered with a thin soaked fabric, and electric power is applied to the plate to maintain its surface temperature at  $20^\circ\text{C}$ .  
 (a) Assuming that the bottom of the plate is insulated, determine the required electric power.  
 (b) After a long period of operation, the fabric is completely dry. Neglect the changes in the convection coefficient and the electric power. What will be the steady-state surface temperature?  
 (c) Is it a good assumption to neglect the radiative heat transfer?
- 4.11.** Use  $r_0 = 2.869 \times 10^{-10}$  m and  $\varepsilon_0/k_B = 10.22$  K for He to calculate and plot the Lennard-Jones potential. Set one molecule at a fixed (pinned) position on the  $x$ -axis, say at  $x = 5$  nm. The other molecule starts at the origin with an initial velocity  $\mathbf{v}_0 = v_0(\hat{x} \cos \beta + \hat{y} \sin \beta)$ , where  $\beta$  is a small angle between  $\mathbf{v}_0$  and the  $x$ -axis. Develop a computer program to calculate the trajectory of the moving particle in the  $x$ - $y$  plane, for various  $v_0$  and  $\beta$ , based on (a) the rigid-elastic-sphere assumption and (b) the intermolecular force field. Comment on the differences between the results obtained from the two models.
- 4.12.** Using Eq. (4.59), show that Eq. (4.58) is identical to Eq. (2.42). Hint:
- $$\nabla \cdot \{P_{ij}\} = \left( \frac{\partial P_{xx}}{\partial x} + \frac{\partial P_{yx}}{\partial y} + \frac{\partial P_{zx}}{\partial z} \right) \hat{x} + \left( \frac{\partial P_{xy}}{\partial x} + \frac{\partial P_{yy}}{\partial y} + \frac{\partial P_{zy}}{\partial z} \right) \hat{y} + \left( \frac{\partial P_{xz}}{\partial x} + \frac{\partial P_{yz}}{\partial y} + \frac{\partial P_{zz}}{\partial z} \right) \hat{z}.$$
- 4.13.** Derive the viscous dissipation term in Eq. (2.43) based on Eq. (4.60).
- 4.14.** From Eq. (4.60), derive the heat diffusion equation:  $\kappa \nabla^2 T = \rho c_p (\partial T / \partial t)$ .

**4.15.** Assuming  $\tau$  is independent of the frequency, use the Maxwell velocity distribution, Eq. (3.43),

to evaluate  $\kappa = (\pi/3) \int_{\mathfrak{w}} v^2 \varepsilon (df_0/dT) d\mathfrak{w}$  for a monatomic gas, where  $\varepsilon = \frac{1}{2} m v^2$ .

**4.16.** Consider an isothermal gas flow in the  $x$  direction with a bulk velocity distribution  $v_B(y) = \bar{v}_x(y)$  as shown in Fig. 4.5. The velocity distribution is not very far from the equilibrium so that  $f = f_0 - \tau v_y (df_0/dv_B) (dv_B/dy)$ . Find an expression of the dynamic viscosity  $\mu$ . Hint:  $\tau_{yx} = \int_{\mathfrak{w}} (m v_x) v_y f d\mathfrak{w}$  according to Eq. (4.14a); the answer is  $(m k_B T / \pi)^{1/2} / (4d^2)$ .

**4.17.** What is the continuum assumption, and when does the continuum assumption break down? Define the Knudsen number, and what is its physical significance? What are the unique issues related to microfluidics? What are the applications of microfluidics?

**4.18.** What happens at the boundary layer for a fluid moving over a large plate during slip flow? Describe both the velocity distribution and the temperature distribution near the wall. Write the slip-flow boundary conditions, and discuss the significance of each term.

**4.19.** Integrate Eq. (4.84) to find the dimensionless bulk temperature  $\Theta_m$ ; and then use the definition of Nusselt number to prove Eq. (4.85).

**4.20.** Find the temperature distribution for slip flow between two parallel plates when the bottom plate is insulated and the top plate is heated at a uniform heat flux. Continue on to verify Eq. (4.86).

**4.21.** Find the velocity and temperature distributions for slip flow through a circular tube with a uniform wall heat flux. Continue on to verify Eq. (4.87).

**4.22.** For slip flow with temperature jump in a circular tube, show that there exists a maximum Nusselt number at the entrance, given by  $Nu_{\max} = 1/\beta_T$ .

**4.23.** For Poiseuille flow with velocity slip, calculate the friction coefficient  $C_f = \tau_s / (\rho v_m^2 / 2)$  at the entrance and for fully developed gas flow.

**4.24.** For fully developed gas flow in a circular tube, develop an expression for the ratio of the required pump powers with slip and without slip.

**4.25.** A heat sink contains 100 microchannels, each 1 mm long with a  $1 \mu\text{m} \times 30 \mu\text{m}$  cross section. Cold air at  $22^\circ\text{C}$  flows in at 2 atm with a velocity of 4 m/s. The sides of the channel are well insulated, and a constant wall flux  $q_w'' = 40 \text{ W/m}^2$  is removed by the flow. Neglecting the entry region, what will be the exit temperature of the air? What will be the wall temperature at the exit? (Assume that  $\alpha_v = \alpha_T = 0.8$ .)

**4.26.** For the same fluid, entrance conditions, and wall heat flux as in Problem 4.24, estimate the convection coefficient for fully developed flow in a circular tube as a function of the tube diameter. Take  $D = 300 \text{ nm}$ ,  $3 \mu\text{m}$ , and  $300 \mu\text{m}$ .

**4.27.** Model the cantilever tip of an atomic force microscope (AFM) as a flat disk, with a diameter of 100 nm, that is above a flat surface at 300 K. If the tip is heated to 400 K, calculate the heat flux from the tip to the surface when the distance varies from 10 to 1 nm, assuming that the tip and the sample surface are surrounded by dry air at ambient pressure. How will your calculation change if the pressure is reduced to 1 torr? [1 torr = 1 mmHg = 133.3 Pa.]

**4.28.** Team Project 1: Derive the Nusselt number for constant wall temperature for a laminar slip flow either in a circular tube or between two parallel plates.

**4.29.** Team Project 2: Develop a computer program to evaluate the Nusselt number in the entry region for uniform wall heat flux in a circular flow.

**4.30.** Team Project 3: Perform a simulation using the DSMC method for gas conduction between two plates, with different Knudsen numbers. Compare your results with Eq. (4.94).



*This page intentionally left blank*

---

# CHAPTER 5

---

## THERMAL PROPERTIES OF SOLIDS AND THE SIZE EFFECT

---

One of the thrust areas of research in micro/nanoscale heat transfer is related to transport processes in solid state devices. In the early 1990s, much research had been done to identify the regimes when the microscale effect must be considered in dealing with problems occurring at small length scales and/or timescales.<sup>1,2</sup> Cahill et al. provided a more recent survey on the thermal phenomena and measurement techniques associated with solid state devices.<sup>3</sup> The critical dimensions of integrated circuits have continued to shrink during the past few decades, with printing features currently already below 100 nm; some are approaching the 10-nm limit of most available fabrication technologies. Overheating caused by thermal energy generation is a major source of device failure, and it often occurs in very small regions, known as hot spots. A remarkable number of micro/nanostructured materials and systems have temperature-dependent figures of merit. Therefore, understanding the thermophysical properties, thermal transport physics, and thermal metrology from the micrometer down to the nanometer length scales is critically important for future development of microelectronic devices and nanobiotechnology.

This chapter focuses on simple phonon theory and electronic theory of the specific heat, thermal conductivity, and thermoelectricity of metals and insulators. The Boltzmann transport equation (BTE) has been used to facilitate the understanding of microscopic behavior, together with the quantum statistics of phonons and electrons. The quantum size effect on phonon specific heat is extensively covered. Examples are given to analyze direct thermoelectric conversion for temperature measurement, power generation, and refrigeration. Furthermore, a detailed treatment of classical size effect on the thermal conductivity is presented. Finally, the concepts of quantum electrical conductance and thermal conductance are introduced.

---

### 5.1 SPECIFIC HEAT OF SOLIDS

---

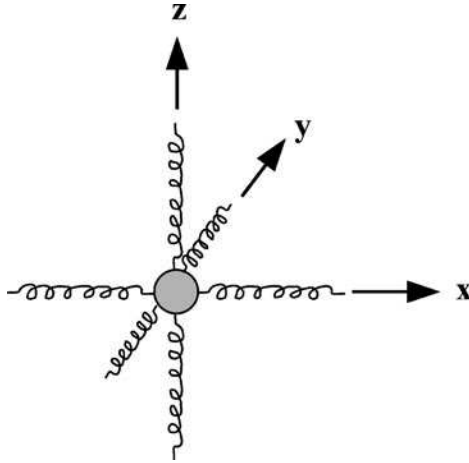
In this section, simple models of the specific heat of bulk solids are described considering the contribution by lattice vibrations and for metals the additional contribution by free electrons. The purpose is to understand the macroscopic behavior from the microscopic point of view and to prepare students for further study on the quantum size effect to be discussed in subsequent sections.

#### 5.1.1 Lattice Vibration in Solids: The Phonon Gas

The atoms in solids are close to each other, and interatomic forces keep them in position. Atoms cannot move around except for vibrations near their equilibrium positions. In

crystalline solids, atoms are organized into periodical arrays, and each identical structural unit is called a lattice. Lattice vibrations contribute to thermal energy storage and heat conduction. In metals, electrons are responsible for electrical transport and heat conduction but are less important for storing thermal energy except at very low temperatures.

The simple oscillator model treats each atom as a harmonic oscillator, which vibrates along all three axes (see Fig. 5.1). If the vibrational degrees of freedom were completely



**FIGURE 5.1** The harmonic oscillator model of an atom in a solid.

excited, we would expect the high-temperature limit of the specific heat of elementary (monatomic) solids to be

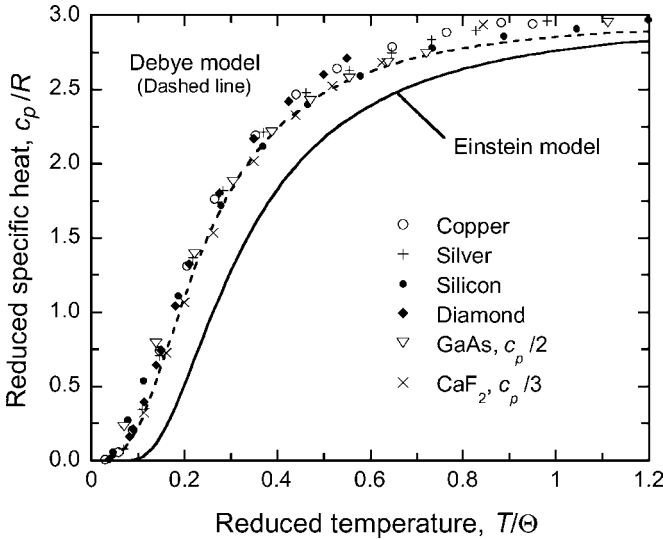
$$\bar{c}_v = 3\bar{R} \quad (5.1)$$

which is called the Dulong-Petit law, named after Pierre-Louis Dulong and Alexis-Thérèse Petit in 1819. The Dulong-Petit law can be understood in terms of the equipartition principle in classical statistics. However, it cannot predict low-temperature behavior, and even above the room temperature, this model significantly overpredicts the specific heats for diamond, graphite, and boron.

Einstein in 1907 proposed a simple harmonic oscillator model and its quantized energy levels  $(i + \frac{1}{2})h\nu$ ,  $i = 1, 2, \dots$ , to obtain the specific heat as a function of temperature. Here, the frequency  $\nu$  is a characteristic vibration frequency of the solid material. The procedure is similar to the analysis of vibration energies for diatomic gas molecules, e.g., Eq. (3.59) to Eq. (3.62). The resulting specific heat for a monatomic solid is

$$\bar{c}_v(T) = 3\bar{R} \frac{\Theta_E^2}{T^2} \frac{e^{\Theta_E/T}}{(e^{\Theta_E/T} - 1)^2} \quad (5.2)$$

where the factor 3 accounts for oscillation in all three directions and  $\Theta_E = h\nu/k_B$  is called the Einstein temperature.<sup>4,5</sup> It can be shown that  $\bar{c}_v \rightarrow 0$  as  $T \rightarrow 0$  and  $\bar{c}_v \rightarrow 3\bar{R}$  at  $T \gg \Theta_E$ . In the intermediate temperature range, however, the Einstein specific heat is significantly lower than the experimental data. This can be seen from Fig. 5.2, where the experimental results of the constant-pressure specific heat are taken from Ashcroft and Mermin.<sup>5</sup> It should be noted that  $c_p = c_v$  for a solid under the incompressible assumption. The reduced



**FIGURE 5.2** Comparison of model predictions with experimental data of the specific heat for several crystalline solids.

temperature is the ratio of the temperature to the characteristic temperature. The experimental data were plotted using the Debye temperature given in Table 5.1. The reason that the specific heat of diamond is far from  $3R$  near room temperature is because of its very high characteristic temperature (or frequency of vibration).

In the Einstein model, each atom is treated as an independent oscillator and all atoms are assumed to vibrate at the same frequency. In 1912, Max Born and Theodore von Kármán first realized that the bonding in a solid prevents independent vibrations. Therefore, a collection of vibrations must be considered under the force-spring interactions of the nearby atoms. To avoid the complicated calculations, Peter Debye in 1912 simplified the model by assuming that the velocity of sound is the same in all crystalline directions and for all frequencies. In addition, there is a high-frequency cutoff and no vibration can occur beyond this frequency. As will be seen from subsequent sections, the Debye model is a great success and has been prevailing even though more advanced and realistic theories have been developed.

### 5.1.2 The Debye Specific Heat Model

The Debye model for the specific heat of solids includes a large number of closely spaced frequencies of vibration up to a certain upper bound  $\nu_m$ , which is determined by the total number of vibration modes  $3N$ , where  $N$  is the number of atoms. The high-frequency limit is indeed plausible because the shortest wavelength of the lattice wave should be on the order of the interatomic distances, or the lattice constants. Rather than treating each atom as an individual oscillator, the Debye model assumes that vibrations are inside the whole crystal just like standing waves. For elastic vibrations, there are longitudinal waves (e.g., sound waves) and transverse waves (with two polarizations) in a crystal. In analogy to electromagnetic waves and photons, the quanta of lattice waves are called *phonons*. The energy of a phonon is  $\varepsilon = h\nu$ , where  $\nu$  is the vibration frequency. The momentum of a phonon is

**TABLE 5.1** The Debye Temperature, Melting Temperature, and Other Properties for Selected Solids. The Data are Mainly Taken from Kittel<sup>4</sup> and Ashcroft and Mermin.<sup>5</sup> The Reported Densities are for 22°C Except for Ar

Element/ compound	Symbol/ formula	$M$ (kg/kmol)	$\Theta_D$ (K)	$T_{\text{melt}}$ (K)	$n_a$ ( $10^{28} \text{ m}^{-3}$ )	$\rho$ ( $10^3 \text{ kg/m}^3$ )
Argon	Ar	40	92	84	2.66 (4 K)	1.77 (4 K)
Mercury	Hg	200.6	72	234	4.26	14.26
Sodium	Na	23	158	371	2.65	1.013
Lithium	Li	6.9	344	454	4.7	0.542
Lead	Pb	207	105	601	3.3	11.34
Zinc	Zn	65.4	327	692	6.55	7.13
Magnesium	Mg	24.3	400	922	4.30	1.74
Aluminum	Al	27	428	934	6.03	2.7
Calcium	Ca	40	230	1113	2.30	1.53
Silver	Ag	108	225	1235	5.85	10.5
Copper	Cu	63.5	340	1358	8.45	8.93
Gold	Au	197	165	1338	5.90	19.3
Iron	Fe	56	470	1811	8.50	7.87
Silicon	Si	28	645	1687	5.0	2.33
Diamond	C	12	2000	3620	17.6	3.52
Potassium bromide	KBr	119	177	1007		2.75
Sodium chloride	NaCl	58.5	281	1074		2.17
Gallium arsenide	GaAs	144.6	360	1511		5.32
Calcium fluoride	CaF <sub>2</sub>	78	474	1696		3.18

$p = h\nu/v_p = h/\lambda$ , where  $v_p = \lambda\nu$  is the propagation speed for the given phonon mode, or frequency, and  $\lambda$  is the phonon wavelength. It should be noticed that the propagation speeds of longitudinal and transverse waves are different. So far, we have related lattice vibrations to lattice waves and to the translational movement of the phonon gas, which follows the Bose-Einstein statistics. However, the total number of phonons is not conserved since it depends on temperature. Thus, we do not need to apply the constraint given in Eq. (3.2) and can simply set  $\alpha = 0$  in Eq. (3.16). The result is

$$\frac{N_i}{g_i} = \frac{1}{e^{\epsilon_i/k_b T} - 1} \quad (5.3)$$

Suppose the energy levels are closely spaced; we can write Eq. (5.3) in terms of a continuous function called the Bose-Einstein distribution function:

$$f_{\text{BE}}(\nu) = \frac{dN}{dg} = \frac{1}{e^{h\nu/k_b T} - 1} \quad (5.4)$$

The *degeneracy* for phonons is the number of quantum states per unit volume in the phase space. For a given volume  $V$  and within a spherical shell in the momentum space (from  $p$  to  $p + dp$ ), we have from Eq. (3.87) that  $dg = 4\pi V p^2 dp / h^3 = 4\pi V \nu^2 d\nu / v_p^3$ . Hence,

$$\frac{dg}{V} = \frac{g(\nu)d\nu}{V} = D(\nu)d\nu = \frac{4\pi\nu^2}{v_p^3} d\nu \quad (5.5)$$

Here, we have introduced the *density of states* of phonons,  $D(\nu)$ , which is the number of quantum states per unit volume per unit frequency or energy ( $h\nu$ ) interval. The number density in terms of the density of states can be expressed as

$$n = \int_0^{\infty} f_{\text{BE}}(\nu) D(\nu) d\nu \quad (5.6)$$

Because there exist one longitudinal and two transverse waves, the phonon density of states in a large spherical shell of the momentum space can be written as

$$D(\nu) = 4\pi\nu^2 \left( \frac{1}{v_l^3} + \frac{2}{v_t^3} \right) = \frac{12\pi\nu^2}{v_a^3} \quad (5.7)$$

where  $v_l$  is the speed of the longitudinal wave,  $v_t$  is the speed of the transverse wave, and  $v_a$  is a weighted average defined in the above equation. The total number of quantum states must be equal to  $3N$ . Using integration in place of summation, we have

$$\frac{3N}{V} = \int_0^{\infty} D(\nu) d\nu = \int_0^{\nu_m} \frac{12\pi\nu^2}{v_a^3} d\nu \quad (5.8)$$

where  $\nu_m$  is an upper limit of the frequency that can be obtained from Eq. (5.8) as

$$\nu_m = \left( \frac{3n_a}{4\pi} \right)^{1/3} v_a \quad (5.9)$$

Here,  $n_a = N/V$  is the number density of atoms.

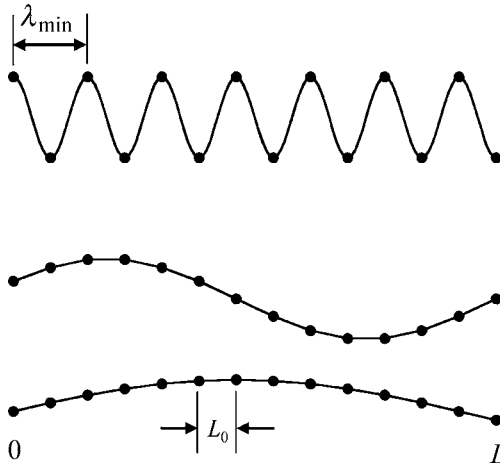
The Debye temperature is defined as

$$\Theta_D = \frac{h\nu_m}{k_B} = \frac{h}{k_B} \left( \frac{3n_a}{4\pi} \right)^{1/3} v_a \quad (5.10)$$

The Debye temperature and the number density for various solids are listed in Table 5.1 together with some other properties. The listed values of the Debye temperature were based on the experimentally measured specific heat at very low temperatures, rather than that calculated from the speed of sound. The result of the Debye specific heat theory agrees fairly well with the experimental data for several crystalline solids in a large temperature range, as can be seen from Fig. 5.2. The high-temperature limit of the specific heat is  $6\bar{R}$  for GaAs and  $9\bar{R}$  for  $\text{CaF}_2$ , because the number of atoms in a unit cell of the lattice is 2 and 3, respectively.

**Example 5-1.** The average speed of the longitudinal waves is  $v_l = 8970$  m/s and that of the transverse waves is  $v_t = 5400$  m/s in silicon. Find the average propagation speed, the maximum frequency, the Debye temperature, and the minimum wavelength  $\lambda_{\text{min}}$ . How does  $\lambda_{\text{min}}$  compare with the average distance between atoms?

**Solution.** Since  $v_a^3 = 3/(v_l^{-3} + 2v_t^{-3})$ , we have  $v_a = 5972$  m/s. Given  $n_a = 5.0 \times 10^{28} \text{ m}^{-3}$ , we obtain  $\nu_m = 1.36 \times 10^{13} \text{ Hz} = 13.6 \text{ THz}$  from Eq. (5.9) and  $\Theta_D = 655 \text{ K}$  from Eq. (5.10), which is a little bit higher than the experimental value of 645 K listed in Table 5.1. The experimental value was obtained by fitting the low-temperature specific heat with the Debye model. The minimum wavelength is estimated by  $\lambda_{\text{min}} = v_a/\nu_m = 0.44 \text{ nm} = 4.4 \text{ \AA}$ . The average spacing between atoms can be estimated by  $L_0 = n_a^{-1/3} = 0.27 \text{ nm}$  or  $2.7 \text{ \AA}$ , suggesting that  $\lambda_{\text{min}} \approx 2L_0$ . The maximum wavelength of the lattice wave will be twice the extension of the solid. For a cubic solid with each side  $L$ , we have  $\lambda_{\text{max}} \approx 2L$ . The lattice waves are illustrated in Fig. 5.3 in a 1-D case.



**FIGURE 5.3** Illustration of the minimum wavelength  $\lambda_{\min} = 2L_0$  and the maximum wavelength  $\lambda_{\max} = 2L$  associated with lattice vibrations in a solid with a dimension  $L$  and with a periodic array of atoms (dots).

The distribution function for phonons can now be written as

$$f(\nu) = \frac{1}{V} \frac{dN}{d\nu} = D(\nu) f_{\text{BE}}(\nu) = \frac{12\pi\nu^2}{v_a^3 (e^{h\nu/k_B T} - 1)} = \frac{9n_a \nu^2}{v_m^3 (e^{h\nu/k_B T} - 1)}, \nu \leq \nu_m \quad (5.11)$$

The vibration contribution to the internal energy can be written as

$$U - U_0 = \int_0^\infty f(\nu) h\nu d\nu \quad (5.12a)$$

where  $U_0$  is the internal energy at 0 K when no vibration modes are excited. The result after some manipulation becomes

$$U - U_0 = 9Nk_B T \left( \frac{T}{\Theta_D} \right)^3 \int_0^{x_D} \frac{x^3}{e^x - 1} dx \quad (5.12b)$$

where  $x_D = \Theta_D/T$ . The molar specific heat is then

$$\bar{c}_v(T) = \left( \frac{\partial \bar{u}}{\partial T} \right)_V = 9\bar{R} \left( \frac{T}{\Theta_D} \right)^3 \int_0^{x_D} \frac{x^4 e^x}{(e^x - 1)^2} dx \quad (5.13)$$

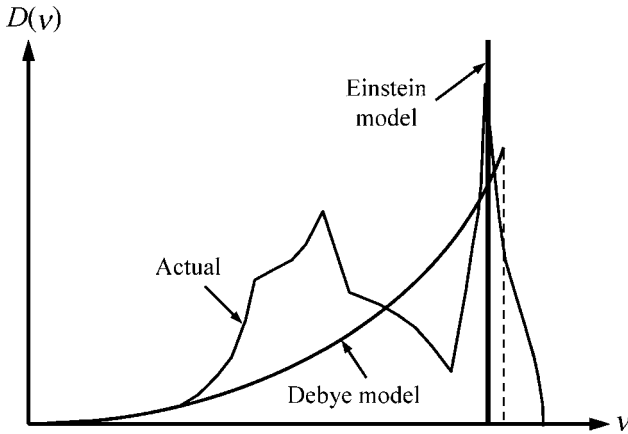
The specific heat predicted by the Debye theory agrees very well with experimental data of many solids. Notice that  $\int_0^{x_D} [x^4 e^x / (e^x - 1)^2] dx = 4 \int_0^{x_D} [x^3 / (e^x - 1)] dx - x_D^4 / (e^{x_D} - 1)$ . When  $T \gg \Theta_D$ ,  $x_D \rightarrow 0$  and  $e^x - 1 \approx x$ . Thus,  $\int_0^{x_D} [x^3 / (e^x - 1)] dx \rightarrow x_D^3 / 3$ , and the Debye specific heat approaches  $3\bar{R}$  in the high-temperature limit. The relative difference is about

5% at  $T = \Theta_D$ . Using Eq. (B.9), it can be shown that at  $T \ll \Theta_D$ , Eq. (5.13) can be approximated by

$$\bar{c}_v \approx \frac{12\pi^4}{5} \bar{R} \left( \frac{T}{\Theta_D} \right)^3 \propto T^3 \quad (5.14)$$

which is known as the  $T^3$  law, and it agrees with experiments within a few percents for  $T/\Theta_D < 0.1$ .

In essence, the Einstein specific heat theory assumed that all oscillations are at the same frequency, and it implied that the density of states has a sharp peak at that frequency and is zero at all other frequencies. On the other hand, the Debye theory is based on a parabolic density of states function,  $D(\nu) \propto \nu^2$ . More detailed studies have revealed that the actual phonon density of states is a complicated function of the frequency,<sup>5,7</sup> as illustrated in Fig. 5.4



**FIGURE 5.4** Illustration of the phonon density of states in the Einstein model and the Debye model as compared with the actual behavior of metals.

for aluminum and copper according to neutron scattering measurements. There are different phonon branches in a real crystal that affect the density of states in different frequency regions. A detailed discussion will be deferred to Chap. 6 when we take a deeper look into the crystalline structures and phonon dispersion relations. In general, the Debye theory predicts correctly the low-temperature behavior when only the low-frequency phonon modes are excited; this is probably the most significant contribution of the Debye model. At higher temperatures, the Debye model can be considered as a first-order approximation, as shown in Fig. 5.2.

### 5.1.3 Free Electron Gas in Metals

The translational motion of free electrons within the solid is largely responsible to the electrical and thermal conductivities of metals. Sometimes, the free electrons are called electron gas that draws an analogy between electrons and monatomic molecules. However, there are distinct differences between electrons in a solid and molecules in an ideal gas. The number of free electrons is on the order of the number of atoms. For Au, Cu, and Ag, we shall



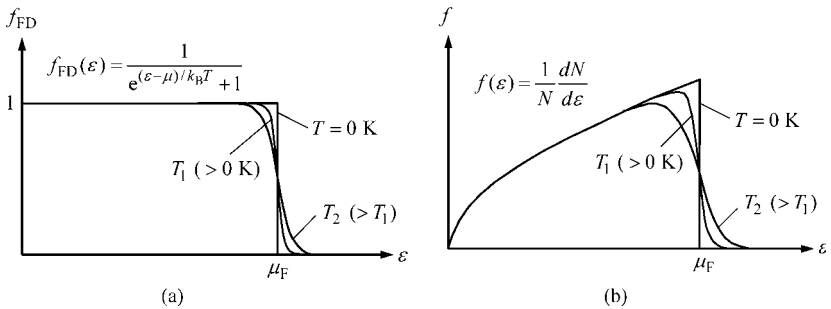
**TABLE 5.2** Electronic Properties of Selected Metals; Data Mainly from Kittel<sup>4</sup>

	Li	Na	K	Cu	Ag	Au	Mg	Ca	Zn	Al	Pb
$\mu_F$ (eV)	4.72	3.23	2.12	7.0	5.51	5.5	7.13	4.68	9.39	11.6	9.37
$n_e$ ( $10^{28} \text{ m}^{-3}$ )	4.7	2.65	1.4	8.45	5.85	5.90	8.60	4.60	13.1	18.1	13.2
Electrons/atom	(1)	(1)	(1)	(1)	(1)	(1)	(2)	(2)	(2)	(3)	(4)
$r_c$ ( $\mu\Omega \cdot \text{cm}$ ) at 22°C	9.32	4.75	7.19	1.70	1.61	2.20	4.30	3.60	5.92	2.74	21.0

assume there is 1 free electron per atom but there are 3 electrons per atom for Al and 4 electrons per atom for Pb (see Table 5.2). Electrons obey the Fermi-Dirac distribution given in Eq. (3.24). A continuous function called the *Fermi function* can be defined as

$$f_{\text{FD}}(\varepsilon) = \frac{dN}{dg} = \frac{1}{e^{(\varepsilon-\mu)/k_B T} + 1} \quad (5.15)$$

The Fermi function is plotted in Fig. 5.5a, where  $\mu_F = \mu$  at  $T = 0 \text{ K}$  is called the *Fermi energy*. It will be shown later that  $\mu$  changes little when the temperature is not very high.

**FIGURE 5.5** (a) The Fermi function and (b) the distribution function of free electrons in a metal.

At the absolute temperature of 0 K,  $f_{\text{FD}} = 1$  when  $\varepsilon < \mu_F$ , and  $f_{\text{FD}} = 0$  when  $\varepsilon > \mu_F$ . As the temperature increases, the function falls less sharply.

The degeneracy for electrons is further increased by 2, due to the existence of positive and negative spins. In a volume  $V$  of a spherical shell in the momentum space, we have  $dg = 8\pi V(m_e/h)^3 v^2 dv$  from Eq. (3.86) by considering the spin degeneracy. Hence, the distribution function in terms of the electron speed is

$$f(v) = \frac{1}{V} \frac{dN}{dv} = 8\pi \left( \frac{m_e}{h} \right)^3 \frac{v^2}{e^{(\varepsilon-\mu)/k_B T} + 1} \quad (5.16)$$

Using  $f(v)dv = f(\varepsilon)d\varepsilon$  and  $\varepsilon = m_e v^2/2$ , we obtain the distribution function in terms of the kinetic energy of the electrons as

$$f(\varepsilon) = \frac{1}{V} \frac{dN}{d\varepsilon} = 4\pi \left( \frac{2m_e}{h^2} \right)^{3/2} \frac{\sqrt{\varepsilon}}{e^{(\varepsilon-\mu)/k_B T} + 1} \quad (5.17)$$

This equation is plotted in Fig. 5.5*b*. Note that  $f(\varepsilon) = f_{\text{FD}}(\varepsilon)D(\varepsilon)$ , where  $D(\varepsilon)$  is the density of states for free electrons and is expressed as

$$D(\varepsilon) = 4\pi \left( \frac{2m_e}{h^2} \right)^{3/2} \sqrt{\varepsilon} \quad (5.18)$$

Now, we are ready to evaluate the Fermi energy  $\mu_F$ . At  $T \rightarrow 0$ , the number density of electrons becomes

$$n_e = \frac{N_e}{V} = \lim_{T \rightarrow 0} \int_0^{\mu} 4\pi \left( \frac{2m_e}{h^2} \right)^{3/2} \frac{\sqrt{\varepsilon}}{e^{(\varepsilon - \mu)/k_B T} + 1} d\varepsilon \quad (5.19)$$

which gives

$$\mu_F = \frac{h^2}{8m_e} \left( \frac{3n_e}{\pi} \right)^{2/3} \quad (5.20)$$

Typical values of  $\mu_F$  range from 2 to 12 eV. Table 5.2 lists the Fermi energy, the electron number density, the number of electrons per atom, and the electrical resistivity of various metals. The temperature dependence of  $\mu$  for electrons is given by the Sommerfeld expansion:<sup>5</sup>

$$\mu(T) = \mu_F \left[ 1 - \frac{1}{3} \left( \frac{\pi k_B T}{2\mu_F} \right)^2 + \dots \right] \quad (5.21a)$$

It can be seen that  $\mu(T) \approx \mu_F$  at moderate temperatures. Arnold Sommerfeld (1868–1951) was a German physicist and one of the founders of quantum mechanics. As a professor at the University of Munich, he advised a large number of doctorate students who became famous in their own right, including Peter Debye, Wolfgang Pauli, Werner Heisenberg, among others. Sommerfeld applied the FD statistics to study free electrons in metals and resolved the difficulty in the classical theory for electron specific heat. As discussed in Chap. 3, electrons tend to fill all the quantum states up to a certain energy level. In many texts,  $\mu(T)$  is called the Fermi level or the Fermi energy, which is temperature dependent. As the temperature increases, only those electrons near the Fermi level will be redistributed. Because of the importance of the Sommerfeld expansion for the integration involving the FD function, some useful equations are summarized in Appendix B.8. By noticing that the difference between  $\mu(T)$  and  $\mu_F$  is small, we can use Eq. (B.74) and Eq. (B.78) to derive the electron number density as follows:

$$n_e = \int_0^{\infty} D(\varepsilon) f_{\text{FD}}(\varepsilon, T) d\varepsilon \approx \int_0^{\mu_F} D(\varepsilon) d\varepsilon + (\mu - \mu_F) D(\mu_F) + \frac{\pi^2 (k_B T)^2}{6} D'(\mu_F)$$

where the first term is the same as the right-hand side of Eq. (5.19). Since the number density is independent of temperature, we must have

$$(\mu - \mu_F) D(\mu_F) + \frac{\pi^2 (k_B T)^2}{6} D'(\mu_F) = 0 \quad (5.21b)$$

which proves Eq. (5.21a) since  $D(\varepsilon)/D'(\varepsilon) = 2\varepsilon$ .

**Example 5-2.** Calculate  $\mu$  at 300 K and 3000 K for copper using  $\mu_F = 7$  eV. Find the maximum speed (Fermi velocity) and the average speed of electrons for copper at 0 K. How will the Fermi velocity change if the temperature is changed to  $T = 300$  K?

**Solution:** Note that  $k_B = 1.381 \times 10^{-23}/1.602 \times 10^{-19} = 8.62 \times 10^{-5} \text{ eV/K}$ . From Eq. (5.21a), we have

$$\frac{\mu(T) - \mu_F}{\mu_F} \approx -\frac{1}{3} \left( \frac{\pi k_B T}{2\mu_0} \right)^2 = -1.24 \times 10^{-10} T^2$$

which is about 0.0011% at 300 K and 1.2% at 10,000 K. The change in  $\mu$  is indeed very small. At  $T = 0$ ,  $\mu_F = \frac{1}{2} m_e v_{\text{max}}^2 = \frac{1}{2} m_e v_F^2$ . Hence,

$$v_{\text{max}} = v_F = \sqrt{2\mu_F/m_e} \quad (5.22a)$$

$$\bar{\varepsilon} = \frac{1}{2} m_e \bar{v}^2 = \frac{U}{N} = \int_0^{\mu_F} f(\varepsilon) \varepsilon d\varepsilon / \int_0^{\mu_F} f(\varepsilon) d\varepsilon = \frac{3}{5} \mu_F \quad (5.22b)$$

$$v_{\text{rms}} = \sqrt{\frac{2\bar{\varepsilon}}{m_e}} = \sqrt{\frac{6\mu_F}{5m_e}} \quad (5.22c)$$

Electrons are constantly moving even at absolute zero temperature. For copper, we get  $v_F = 1.57 \times 10^6 \text{ m/s}$  and  $v_{\text{rms}} = 1.22 \times 10^6 \text{ m/s}$ , which is about three quarters of  $v_F$ . The classical model based on the equipartition principle or the Maxwell-Boltzmann distribution would give  $\frac{3}{2} k_B T = \frac{1}{2} m_e \bar{v}^2$  or  $v_{\text{rms}} = \sqrt{3k_B T/m_e} = 0$  at absolute zero temperature. Because  $\mu$  changes little from 0 to 300 K, the Fermi velocity at 300 K is essentially the same as that obtained at 0 K.

**Discussion:** If we use the rms velocity to calculate the de Broglie wavelength as in Example 3-2, we obtain  $\lambda_{\text{DB}} = 0.6 \text{ nm}$ . If an electron is accelerated in vacuum to 50 keV, the velocity will be greater than one-third of that of light, and the de Broglie wavelength will be extremely small ( $\lambda_{\text{DB}} \approx 0.0066 \text{ nm}$ ). The resolutions in conventional optical microscopy and photolithography are usually limited by  $\lambda/2$  (the diffraction limit), which is on the order of 200 nm for visible light. Electron microscopy can have a much higher resolution (down to 0.1 nm), and e-beam nanolithography allows the manufacturing of features just a few nanometers.

In order to find out the specific heat of electrons, we first calculate the internal energy:

$$U = V \int_0^{\infty} \varepsilon f_{\text{FD}}(\varepsilon) D(\varepsilon) d\varepsilon \quad (5.23)$$

Because the distribution function does not vary significantly except near  $\varepsilon = \mu$ , the Sommerfeld expansion can be used to express the integration [see Eq. (B.78) in Appendix B]. Hence,

$$\frac{U}{V} \approx \int_0^{\mu_F} \varepsilon D(\varepsilon) d\varepsilon + \mu_F (\mu - \mu_F) D(\mu_F) + \frac{\pi^2 (k_B T)^2}{6} \mu_F D'(\mu_F) + \frac{\pi^2 (k_B T)^2}{6} D(\mu_F)$$

One can see from Eq. (5.21b) that the two middle terms on the right side cancel out. It should also be noted that  $D(\mu_F) = 3n_e/2\mu_F$ . Therefore,

$$U \approx \frac{3}{5} N \mu_F \left[ 1 + \frac{5\pi^2}{12} \left( \frac{k_B T}{\mu_F} \right)^2 + \dots \right] \quad (5.24)$$

The specific heat of free electrons can then be obtained as

$$\bar{c}_{v,e} = \left( \frac{\partial \bar{u}}{\partial T} \right)_V = \frac{\pi^2 k_B T}{2 \mu_F} R \quad (5.25)$$

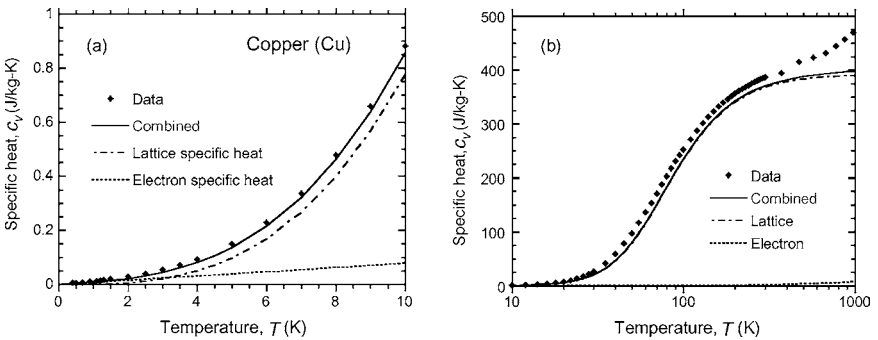
which is much smaller than  $\frac{3}{2}\bar{R}$  as we would obtain if electrons were behaving as an ideal monatomic molecular gas. Another way of obtaining Eq. (5.25) is to use integration, which is left as an exercise (see Problem 5.6). Electronic contribution to the specific heat of solids is negligible except at very low temperatures (a few kelvins or less). The specific heat of metals at very low temperatures can thus be expressed as

$$c_v(T) = \gamma_s T + BT^3 \tag{5.26}$$

where the linear term is the electronic contribution and the cubic term is the lattice contribution for which  $B$  can be obtained from Eq. (5.14). The coefficient  $\gamma_s$  is known as the *Sommerfeld constant*, which can be obtained from Eq. (5.25). The experimental values of  $\gamma_s$  generally agree with those predicted by the free-electron model given in Eq. (5.25) for most alkali metals (e.g., Na, K) and noble metals (e.g., Cu, Ag, Au). For transition metals with magnetic properties, such as Fe and Mn, the measured  $\gamma_s$  value can be an order of magnitude greater than the predicted. On the other hand, for semimetals like Bi, the measured  $\gamma_s$  value can be an order of magnitude smaller than the predicted. Further discussions can be found from the text of Ashcroft and Mermin.<sup>5</sup>

**Example 5-3.** Calculate and plot the specific heat of copper, and compare with the data in Touloukian and Buyco.<sup>6</sup> Discuss the contribution of electrons and lattice vibrations.

**Solution:** From Table 5.1, the Debye temperature for Cu is  $\Theta_D = 340$  K. At  $T < 30$  K, we can apply the  $T^3$  law given in Eq. (5.14) to find the coefficient  $B$  in Eq. (5.26) to be  $5.95 \times 10^{-6} \bar{R} [\text{K}^{-3}]$ . Using  $\mu_F = 7$  eV from Table 5.2, the Sommerfeld coefficient can be calculated from Eq. (5.25) as  $\gamma_s = 6.08 \times 10^{-5} \bar{R} [\text{K}^{-1}]$ . Therefore the two contributions will be equal at  $T = 3.2$  K. The results are plotted in Fig. 5.6a at temperatures below 10 K. At higher temperatures, as shown in Fig. 5.6b,



**FIGURE 5.6** Electron and lattice contributions to the specific heat of Cu (a) at low temperatures and (b) from 10 to 1000 K.

the electronic contribution is much smaller compared with the lattice specific heat: about 0.3% at 100 K, 0.6% at 300 K, and 2% at 1000 K. The data show much higher specific heat values than those predicted by the Debye model. The addition of the electronic contributions cannot fully account for the difference. Noting that  $R = \bar{R}/M = 130.9 \text{ J}/(\text{kg} \cdot \text{K})$  at 1000 K, the specific heat calculated from the Debye model of  $c_v = 390.6 \text{ J}/(\text{kg} \cdot \text{K})$  is 99.5% of  $3R$  given by the Dulong-Petit law. There are several reasons that may be responsible to the deviation between the Debye model and measurements at high temperatures. The first is the anharmonic vibration that was not considered in the simple models with harmonic vibrations. The contribution of anharmonic vibrations becomes more important at higher temperatures since the amplitude of vibration increases with temperature. Secondly, thermal expansion cannot be ignored at high temperatures. The variation of the distance between atoms may change the potential function and thus increase the specific heat. Additionally, when thermal expansion is not negligibly small, the specific heat at constant pressure may be greater than that at constant volume. Interested readers are referred to the literature for further discussions.<sup>8,9</sup>

## 5.2 QUANTUM SIZE EFFECT ON THE SPECIFIC HEAT

The above discussion assumes that the physical dimensions are much larger than the lattice constant. In nanoscale devices and structures, such as 2-D thin films and superlattices, 1-D nanowires and nanotubes, or 0-D quantum dots and nanocrystals, substitution of summation by integration is no longer appropriate. Note that a 2-D thin film is confined in one dimension, a 1-D wire is confined in two dimensions, and a 0-D quantum dot is confined in all three dimensions. In nanostructures, it is necessary to consider quantization of the energy levels. The specific heat becomes a function of the actual dimensions. Experimental demonstrations of quantum size effect on specific heat have been made on Pb particles,<sup>10</sup> carbon nanotubes,<sup>11</sup> and titanium dioxide nanotubes,<sup>12</sup> to name a few. To analyze the quantum size effect on the lattice specific heat, we begin with a wavelike treatment of the vibrational modes in this section.

### 5.2.1 Periodic Boundary Conditions

Consider a 1-D chain of  $N + 1$  atoms as sketched in Fig. 5.3, where the end nodes are fixed in position. The solution should be a standing wave with the following eigenfunctions:

$$\sin\left(\frac{\pi x}{L}\right), \sin\left(\frac{2\pi x}{L}\right), \sin\left(\frac{3\pi x}{L}\right), \dots, \sin\left(\frac{\pi x}{L_0}\right)$$

where  $L/L_0 = N$ , which is the total number of vibration modes within a length of  $L$ . Another approach is based on the Born-von Kármán periodic boundary conditions.<sup>5</sup> Instead of treating the solid as a bounded specimen whose atoms are fixed at each boundary, the Born-von Kármán lattice model takes the medium as an infinite extension with periodic boundary conditions. For a solid whose dimensions are  $L_x$ ,  $L_y$ , and  $L_z$ , in the Cartesian coordinates, the standing wave solutions are

$$\exp(ik_x x), \exp(ik_y y), \exp(ik_z z) \quad (5.27)$$

where  $\mathbf{k} = (k_x, k_y, k_z)$  is called the *lattice wavevector* with  $k^2 = k_x^2 + k_y^2 + k_z^2$ . The allowed discretized values are

$$k_x = 0, \pm \frac{2\pi}{L_x}, \pm \frac{4\pi}{L_x}, \pm \frac{6\pi}{L_x}, \dots, \pm \frac{(N_x - 1)\pi}{L_x}, + \frac{N_x\pi}{L_x} \quad (5.28a)$$

$$k_y = 0, \pm \frac{2\pi}{L_y}, \pm \frac{4\pi}{L_y}, \pm \frac{6\pi}{L_y}, \dots, \pm \frac{(N_y - 1)\pi}{L_y}, + \frac{N_y\pi}{L_y} \quad (5.28b)$$

$$k_z = 0, \pm \frac{2\pi}{L_z}, \pm \frac{4\pi}{L_z}, \pm \frac{6\pi}{L_z}, \dots, \pm \frac{(N_z - 1)\pi}{L_z}, + \frac{N_z\pi}{L_z} \quad (5.28c)$$

where the last term has “+” term only and should be included only if the number of atoms along each direction  $N_x$ ,  $N_y$ , or  $N_z$  is an even number. The central distance between adjacent atoms is  $L_x/N_x$ ,  $L_y/N_y$ , or  $L_z/N_z$  in the given direction. The individual components of the lattice wavevector may be negative or zero in this case. In the 1-D case, it can be seen that the total number of modes is the same as the total number of atoms along the 1-D chain. However, the infinite medium representation with periodic boundary conditions is advantageous not only in mathematical derivations but also for the physical interpretation of lattice dynamics.

## 5.2.2 General Expressions of Lattice Specific Heat

The general expression of the lattice vibrational energy in a solid is given as

$$u(T) = u_0 + \sum_P \sum_K \hbar\omega \left( \frac{1}{e^{\hbar\omega/k_B T} - 1} + \frac{1}{2} \right) \quad (5.29)$$

where  $u_0$  accounts for the static energy at absolute zero temperature, the first term in the parenthesis is the Bose-Einstein distribution  $f_{\text{BE}}(\omega, T)$  given in Eq. (5.4), and the second term in the parenthesis corresponds to the *zero-point energy* that is associated with the  $\frac{1}{2}\hbar\nu$ , due to *quantum fluctuation* or *vacuum fluctuation*, in the vibrational energy levels. We use  $h\nu$  and  $\hbar\omega$  interchangeably whichever is more convenient. The summation is over all phonon branches in terms of the wavevector index  $K$  and the polarization index  $P$ . A phonon branch (sometimes also called a phonon mode) describes the behavior of a type of phonons with a continuous frequency rather than a discrete frequency. The concept of phonon branches will be presented in detail in the subsequent chapter. The lattice specific heat can be expressed as<sup>4</sup>

$$c_v(T) = \sum_P \sum_K \hbar\omega \frac{\partial}{\partial T} f_{\text{BE}}(\omega, T) \quad (5.30)$$

Upon introducing the density of states, we can replace the summation over  $k$ -space with an integration as follows:

$$c_v(T) = \sum_P \int_0^\infty \hbar\omega \frac{\partial f_{\text{BE}}}{\partial T} D(\omega) d\omega = k_B \sum_P \int_0^\infty \left( \frac{\hbar\omega}{k_B T} \right)^2 \frac{e^{\hbar\omega/k_B T}}{(e^{\hbar\omega/k_B T} - 1)^2} D(\omega) d\omega \quad (5.31)$$

Since the density of states is expressed as the number of modes per unit volume, Eq. (5.31) gives the specific heat per unit volume. Neutron scattering and Raman scattering are common ways of determining the density of states from the relationship between  $\omega$  and the lattice wavevector  $\mathbf{k}$  along selected crystal directions. The function  $\omega = \omega(\mathbf{k})$  is called a *dispersion relation*. If discretized values are expressed using the Delta functions in the expression of  $D(\omega)$ , Eq. (5.31) is equivalent to Eq. (5.30), and both the equations can be considered as the general expressions of the specific heat due to lattice vibrations. For a nanostructure with very few atoms in a particular direction, Eq. (5.30) may be more convenient to use. On the other hand, in directions with a large number of atoms, Eq. (5.31) should be the preferable choice.

## 5.2.3 Dimensionality

The method of periodic boundary conditions allows one to determine the density of states for simple dispersion relations easily. Figure 5.7 shows the  $k$ -space, or the *reciprocal lattice space*, in the 2-D case. Each individual block of area  $4\pi^2/(L_x L_y)$  represents a mode, and the number of modes up to a certain value of  $k$  is equal to the total number of blocks inside the circle. One can also use this graph to visualize the 3-D case. Each box of volume  $8\pi^3/(L_x L_y L_z)$  represents a mode, and the number of modes for a given upper limit  $k$  is equal to the total number of boxes within a sphere of radius  $k$ , i.e.,

$$N = \frac{4\pi}{3} k^3 / \left( \frac{8\pi^3}{L_x L_y L_z} \right) = \frac{V k^3}{6\pi^2} \quad (5.32)$$

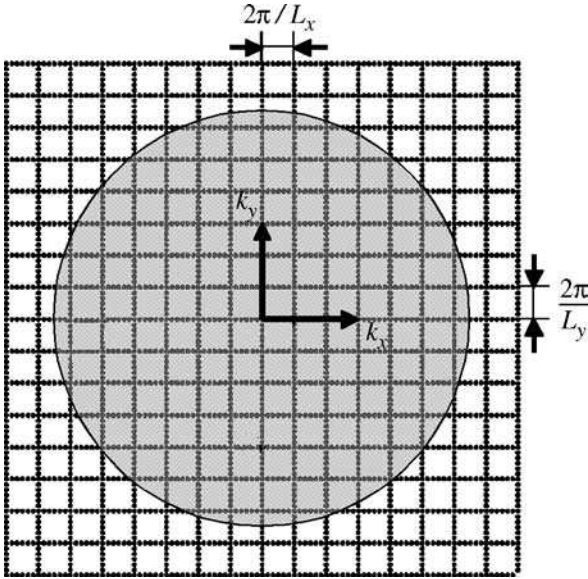


FIGURE 5.7 Schematic of the reciprocal lattice space, or  $k$ -space.

When the dimensions are large enough, the density of states can be expressed as

$$D(\omega) = \frac{1}{V} \frac{dN}{d\omega} = \frac{k^2}{2\pi^2} \frac{dk}{d\omega} \quad (5.33)$$

Assume the dispersion relation is linear, i.e.,

$$\omega = v_a k \quad (5.34)$$

where  $v_a$  is the average speed of the longitudinal and transverse waves as in Eq. (5.7). We can rewrite Eq. (5.33) as

$$D(\omega) = \frac{\omega^2}{2\pi^2 v_a^3} \quad (5.35)$$

This expression is equivalent to Eq. (5.7) for a single polarization. Equations (5.32) and (5.34) can be combined to obtain the high-frequency limit by setting  $N$  equal to the number of atoms. The result is the same as Eq. (5.9). When Eq. (5.35) is substituted into Eq. (5.31), the Debye expression of the specific heat given in Eq. (5.13) is readily obtained.

If the number of atoms is very small in a particular direction, there will be only a few values for the particular wavevector component. The dimensionality will be reduced, and the wavevector component can be assumed as zero in that direction. For a 2-D solid (such as a thin film or a quantum well), the density of states is defined as the number of quantum states per unit area. By assuming a linear dispersion relation, we obtain

$$N = \frac{\pi k^2}{4\pi^2/L_x L_y} = \frac{A k^2}{4\pi} \quad (5.36)$$

and

$$D(\omega) = \frac{1}{A} \frac{dN}{d\omega} = \frac{k}{2\pi} \frac{dk}{d\omega} = \frac{\omega}{2\pi v_a^2} \quad (5.37)$$

For a 1-D solid (such as a nanowire or a nanotube), by noting that  $N = 2k/(2\pi/L_x) = Lk/\pi$ , we find the density of states to be

$$D(\omega) = \frac{1}{\pi v_a} \quad (5.38)$$

which is independent of the frequency. It can be shown that, in the low-temperature limit, the specific heat for a 2-D solid is proportional to  $T^2$  and that for a 1-D solid is proportional to  $T$ .<sup>13</sup> Experimental evidence of the dimensionality change has been known for a long time in graphite, which has a layered lattice structure with a strong bonding between atoms within each layer and a weak interactive force between layers. The specific heat of graphite is approximately proportional to  $T^2$  at low temperatures.<sup>14</sup> On the other hand, the linear temperature dependence of specific heat has been observed in carbon nanotubes.<sup>11</sup>

It can be seen from Eq. (5.31) that when  $\hbar\omega \gg k_B T$ , the integrand approaches zero. Therefore, the contribution to the specific heat is negligibly small when the phonon energy is much higher than  $k_B T$ . The speed of lattice waves ranges from 1000 to 10,000 m/s, the phonon wavelength corresponding to  $k_B T$  is called *thermal phonon wavelength*, which can be calculated from  $\lambda_{th} = v_a \hbar / k_B T$ . At room temperature,  $\lambda_{th}$  is approximately 0.3 nm for  $v_a = 2000$  m/s and 1 nm for  $v_a = 6000$  m/s. At 10 K,  $\lambda_{th} \approx 10$  nm for  $v_a = 2000$  m/s, and  $\lambda_{th} \approx 30$  nm for  $v_a = 6000$  m/s. It is expected that the quantum size effect will become more significant at low temperatures, because the thermal phonon wavelength may be greater than the smallest physical length, such as the thickness of the film and the diameter of the wire.

#### 5.2.4 Thin Films Including Quantum Wells

Thin films, including quantum wells, are important components for microelectronic and photonic devices. We will use the following example to elucidate the effect of film thickness and temperature on the specific heat of thin films.

**Example 5-4.** Evaluate the low-temperature behavior of the specific heat of a thin film made of a monatomic solid. Assume that the film thickness is  $L$ , which has  $q$  monatomic layers, i.e.,  $L = qL_0$ . The average acoustic speed  $v_a$  may be assumed to be independent of temperature. Values of silicon given in Example 5-2 may be used in the numerical evaluation.

**Solution.** The molar specific heat can be expressed as

$$\bar{c}_v(T) = \frac{3V\bar{R}}{Nk_B} \sum_{k_x, k_y, k_z} \hbar\omega \frac{\partial f_{BE}}{\partial T} \quad (5.39)$$

where the number 3 accounts for the three polarizations. Assume the dimension perpendicular to the film is the  $z$  direction. The allowable modes in the  $z$  direction are given by  $k_z = 0, \pm 2\pi/L, \pm 4\pi/L, \dots$ . In order for the total number of modes in the  $z$  direction to be equal to  $q$  for all  $q$  values, we shall use the following limits:

$$k_z = \begin{cases} 0, \pm \frac{2\pi}{L}, \pm \frac{4\pi}{L}, \dots, \pm \frac{(q-1)\pi}{L}, & \text{for } q = 1, 3, 5 \dots \\ 0, \pm \frac{2\pi}{L}, \dots, \pm \frac{(q-2)\pi}{L}, + \frac{q\pi}{L}, & \text{for } q = 2, 4, 6 \dots \end{cases} \quad (5.40)$$

Assume that the lattice is infinitely extended in the directions parallel to the film. We can substitute the summation with an integration in the parallel directions using cylindrical coordinates. Therefore,

$$\bar{c}_v(T) = \frac{3V\bar{R}}{(2\pi)^3 Nk_B} \sum_{k_z} \left( \int_0^{\sqrt{k_0^2 - k_z^2}} \hbar\omega \frac{\partial f_{BE}}{\partial T} 2\pi \eta d\eta \right) \Delta k_z \quad (5.41)$$



where  $\eta^2 = k_x^2 + k_z^2$  and  $\Delta k_z = 2\pi/L$ . The cutoff value  $k_D$  is determined by setting the total number of modes equal to the number of atoms per unit area. Equation (5.36) can be used to evaluate the number of modes for each  $k_z$  and then summed up over all  $k_z$  values. Hence,

$$\frac{N}{A} = \sum_{k_z} \frac{\eta^2}{4\pi} = \sum_{k_z} \frac{k_D^2 - k_z^2}{4\pi} \quad (5.42a)$$

Noting that  $N/A = q/L_0^2$  and there are a total of  $q$  terms in the summation, we can solve Eq. (5.42a) for  $k_D$  as follows:

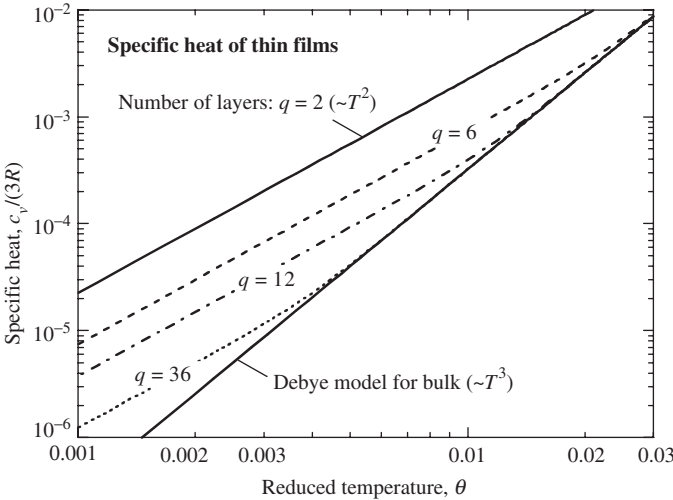
$$k_D = \left( \frac{4\pi}{L_0^2} + \sum_{k_z} \frac{k_z^2}{q} \right)^{1/2} \quad (5.42b)$$

In the limit of a single atomic layer,  $k_D = 2\sqrt{\pi}/L_0 = 3.54/L_0$ ; when  $q \rightarrow \infty$ ,  $k_D \rightarrow 3.982/L_0$ , a value close to  $k_D = \sqrt{6\pi^2}/L_0 = 3.90/L_0$  in the 3-D case. Note that the value of  $k_D$  normalizes the specific heat so that it approaches to the high-temperature limit of  $3R$ . At low temperatures, when the quantum size effect is significant, a slightly different  $k_D$  does not alter the results much.

Using the linear dispersion relation,  $\omega = v_s k = v_s \sqrt{\eta^2 + k_z^2}$ , and the transformation relation,  $\zeta^2 = \eta^2 + k_z^2$ , we can rewrite Eq. (3.41) as follows:

$$\bar{c}_v(T) = \frac{3\bar{R}A}{2\pi N} \left( \frac{k_B T}{\hbar v_s} \right)^2 \sum_{k_z} \int_{x_z}^{x_D} \frac{x^3 e^x}{(e^x - 1)^2} dx \quad (5.43)$$

where  $x = \hbar v_s k/k_B T$ , and  $x_D$  and  $x_z$  correspond to  $k_D$  and  $k_z$ , respectively. The  $T^2$  dependence at low temperatures is evident when  $q = 1$  or  $k_z = 0$  only. The modes associated with  $k_z = 0$  are parallel to the interface and are called *planar modes*. We have carried out a numerical evaluation of Eq. (5.42a) and Eq. (5.43) for different values of  $q$  to see when the departure from bulk behavior will occur. The results are plotted in Fig. 5.8 as a function of the reduced temperature, defined as



**FIGURE 5.8** Quantum size effect on the specific heat of thin films, where the reduced temperature is defined as  $\theta = Tk_B L_0 / \hbar v_s$ .

$\theta = Tk_B L_0 / \hbar v_s = L_0 / \lambda_{th}$ . Note that  $\hbar v_s / k_B L_0$  is on the same order of the Debye temperature for bulk materials,  $\Theta_D$ . When  $q$  and  $T$  are sufficiently large, the result from Eq. (5.43) is the same as that predicted by the Debye model for bulk materials. It can be clearly seen that the departure occurs at low temperatures and especially for small  $q$ . Except for very small values of  $q$ , the departure occurs when  $q\theta = L/\lambda_{th} \ll 1$ .

The procedure used for this example is similar to that used by Prasher and Phelan,<sup>15</sup> except that we have considered the planar modes ( $k_z = 0$ ) in evaluating Eq. (5.43). The result is an increase in the specific heat in the microscopic regime. By excluding these modes, previous studies predicted a reduction in the specific heat for small  $q$  at low temperatures.<sup>15</sup> This example clarifies that planar modes are critically important when the thickness is small, especially at low temperatures. As mentioned earlier, due to the layerlike structures, the specific heat of graphite exhibits 2-D behavior at low temperatures. More detailed theory on the specific heat of thin films and graphite can be found from Nicklow et al.<sup>16</sup>, and Tomic et al.<sup>17</sup>

### 5.2.5 Nanocrystals and Carbon Nanotubes

To illustrate the quantum size effect on a nanocrystal, consider a cubic solid whose side is  $L = qL_0$ . The total number of vibrational modes are  $q^3$ . When  $q$  is small, we cannot use the integration and must use the summation in Eq. (5.30) to calculate the specific heat. The summation is over a spherical  $k$ -space for all allowed components as expressed in Eq. (5.28). We can estimate an upper bound  $k_D = \sqrt[3]{6\pi^2/L_0}$  based on the spherical volume measured by the volume element  $(2\pi/qL_0)^3$  to ensure that the total number of modes is equal to  $q^3$ , as we did earlier with the help of Fig. 5.7. For  $k_z = 0$ , the volume in the  $k$ -space for all allowable  $k_x$  and  $k_y$  can be estimated by the surface area multiplied by the height  $(2\pi/qL_0)$ , or  $\pi k_D^2(2\pi/qL_0)$ . The fraction of the planar modes, including all three planes, is approximately equal to  $3\pi k_D^2/(2\pi/qL_0)^2/q^3 \approx 3.63/q$ , which is inversely proportional to  $q$ . As  $q$  decreases or the particle becomes smaller, there will be more planar modes. The conventional argument is that when the size gets smaller, the surface to volume ratio becomes larger. Thus, surface modes will become significant for small-sized particles. However, we prefer to use planar modes following our previous discussion about size effect on the specific heat of thin films because the phonons with the planar modes do not have to propagate along the surfaces. When there are more planar modes, a quadratic function of the specific heat is anticipated at low temperatures. At even lower temperatures, the axial modes (i.e., when only one component of the lattice wavevector is nonzero) may also be important. The axial modes correspond to a linear temperature dependence at low temperatures. Because of the quantization, the specific heat of quantum dots or nanocrystals is a discontinuous function of temperature, at low temperatures. However, by combining the different contributions, an approximate function that describes the temperature dependence of specific heat can be expressed as follows:

$$c_v(T) \approx a_3 T^3 + \frac{a_2 T^2}{L} + \frac{a_1 T}{L^2} \quad (5.44a)$$

where  $a_1$ ,  $a_2$ , and  $a_3$  are positive constants. Clearly, the specific heat becomes size dependent and will be enhanced at low temperatures. Transitions to 2-D and 1-D dimensionalities should occur subsequently as the temperature is lowered. The results are the same for a spherical grain when the length  $L$  in Eq. (5.44a) is replaced by the diameter  $d$  of the sphere.<sup>18</sup> A recent study of the surface and size effects on the specific heat of nanoparticles can be found from Wang et al. (*Int. J. Thermophys.*, **27**, 139, 2006).

When the temperature is very low, however, only the lowest frequency modes can be excited and a *second quantum size effect* will occur. This means that among the  $q^3$  modes, we are left with a few axial modes only, which are  $\mathbf{k} = (\pm 2\pi/L, 0, 0)$ ,  $(0, \pm 2\pi/L, 0)$ , and  $(0, 0, \pm 2\pi/L)$ . These modes have the longest phonon wavelength. From Eq. (5.30), the specific heat can be expressed as

$$c_v(T \rightarrow 0) = \frac{a}{T^2} \exp\left(-\frac{b}{T}\right) \quad (5.44b)$$

where  $a$  and  $b$  are positive constants. Because Eq. (5.44b) converges to zero faster than  $T^3$ , the second quantum size effect will reduce the specific heat at extremely low temperatures.<sup>18</sup> Experiments were made in the early 1970s on lead particles as small as 2.2 nm diameter.<sup>10</sup>

At temperatures below 15 K, the specific heat of these particles is much greater than that for the bulk material. However, as the temperature is reduced to about 2 K, the difference diminishes. The combination of Eq. (5.44a) and Eq. (5.44b) provides a physically plausible explanation of the experimental observations.<sup>18</sup> The lowest temperature limit in Eq. (5.44b) does not apply to the thin-film case discussed earlier because there can exist a large number of modes with very small wavevector components in the directions parallel to the plane.

Unlike diamond, which contains 3-D tetrahedral structures, graphite crystallizes in the hexagonal system with sheetlike structures. While diamond and graphite are each a polymorph of the element carbon, they exhibit dramatically different properties due to their different crystalline structures. Diamond is hard, transparent, and an electrical insulator. On the contrary, graphite is quite soft, opaque, and a good electrical conductor. Graphene is a single atomic layer of carbon atoms packed into a benzene-ring structure. Carbon nanotubes may be considered as rolled from a graphene sheet into a hollow cylinder, with one or both of its ends capped with half a fullerene molecule. The discovery of the  $C_{60}$  and other fullerenes by Robert Curl, Harold Kroto, and Richard Smalley was recognized through the 1996 Nobel Prize in Chemistry conferred on them. The diameter of single-walled carbon nanotubes (SWNTs) can be as small as 0.4 nm with a typical diameter 1 to 2 nm and as long as 100  $\mu\text{m}$  or so. Multi-walled carbon nanotubes (MWNTs) and nanotube ropes can have a diameter from 10 to 200 nm.

As mentioned earlier, graphite has a 2-D structure and exhibits  $T^2$  dependence at low temperatures. For an isolated graphene sheet, the in-plane or parallel transverse acoustic phonon mode or branch has a velocity of  $v_{\text{TA-p}} = 15,000$  m/s and the longitudinal acoustic phonon mode has a velocity of  $v_{\text{LA}} = 24,000$  m/s. On the other hand, the out-of-plane or perpendicular transverse phonon branch is described by a quadratic dispersion relation,  $\omega \propto k^2$ , which is the dominant mode for the specific heat at low temperatures. According to the dimensionality and the dispersion relation, the specific heat of a graphene sheet depends almost linearly on  $T$  at lower temperatures (see Problem 5.11) and on  $T^2$  as the temperature is raised above 100 K or so.

The four acoustic phonon modes or branches are expected to be the dominant contributions to the specific heat of isolated SWNTs at low temperatures. These include two (degenerate) transverse modes, one longitudinal mode, and a twisting mode or torsional mode associated with the rigid rotation around the nanotube axis. The dispersion relation is linear for all four modes at low frequencies.<sup>19</sup> Therefore, because of the 1-D structure, the specific heat is expected to be linearly dependent on temperature. As the temperature is raised, however, higher frequency modes are excited and the 2-D characteristics of carbon nanotubes come into play. Watt de Heer has written an elegant article on this topic.<sup>20</sup> There are significant differences between SWNTs, MWNTs, and nanotube ropes or bundles; the actual temperature dependence can be more complicated and dependent on the diameter.

In nanostructures, the electron density of states is also subject to quantization. The theory for the electronic contribution to the specific heat is more complicated. The electron-electron and electron-phonon interactions as well as the distribution of energy levels and the Fermi energy need to be considered in a detailed model.<sup>21,22</sup> The electronic specific heat of small particles is still a linear function of temperature. Generally speaking, the electronic contribution to the specific heat is negligibly small unless the temperature is below about 1 K. Therefore, we will not discuss the electronic size effect on the specific heat any further.

### 5.3 ELECTRICAL AND THERMAL CONDUCTIVITIES OF SOLIDS

---

In this section, we use kinetic theory to study the electron and phonon transport properties of metals and insulators in the bulk form. The coupling between electrical current and heat

flux due to electric field and temperature gradient will be studied in the following section, followed by a discussion of the size effect on the electrical and thermal conductivities.

### 5.3.1 Electrical Conductivity

We start with the simple kinetic theory approach based on the Drude free-electron model, also known as the Drude-Lorentz theory. As shown in Fig. 5.9, the electrical resistance of

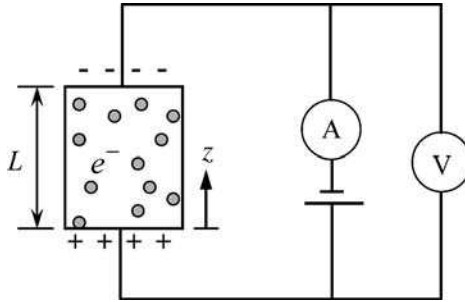


FIGURE 5.9 Illustration of electrical conduction.

a resistor is  $R = r_e L/A_c = L/\sigma A_c$ , where  $r_e$  is the resistivity; its inverse  $\sigma$  is the conductivity,  $L$  is the length, and  $A_c$  is the cross-sectional area. Ohm's law relates the voltage drop  $\Delta V$  and the current  $I$  by  $\Delta V = IR$ , which can be rearranged as

$$\frac{I}{A_c} = \sigma \frac{\Delta V}{L} \quad (5.45)$$

Notice that  $J = I/A_c$  is the *current density* (charge per unit cross-sectional area per unit time), and  $E = \Delta V/L$  is the electric field (note that the electric field is in the direction of decreasing voltage). Rewriting it in the vector form, we have

$$\mathbf{J} = \sigma \mathbf{E} \quad (5.46)$$

The above equation may be considered as the microscopic Ohm's law. An electron of charge  $-e$  is accelerated in an electric field according to Newton's law as

$$\mathbf{F} = -e\mathbf{E} = m_e \frac{d\mathbf{v}}{dt} \quad (5.47)$$

Due to collisions, electrons cannot move completely freely. The velocity change of an electron during a relaxation time  $\tau$  (the average traveling time between collisions) due to an external field is called the *drift velocity*  $\mathbf{u}_d$ . The probability that a traveling particle will collide with another particle or a defect during an infinitesimal time  $dt$  is given by  $dt/\tau$ . The acceleration term in Eq. (5.47) can then be approximated by  $\mathbf{u}_d/\tau$ . Another way of viewing is that there exists a damping force that is proportional to the drift velocity, i.e.,  $m_e \gamma \mathbf{u}_d$ , where  $\gamma$  is the damping coefficient that happens to be the electron scattering rate  $1/\tau$ . At steady state, the damping force must balance the external electrical force, i.e.,  $-e\mathbf{E} = m_e \mathbf{u}_d/\tau$ . The current density is related to the drift velocity by  $\mathbf{J} = -en_e \mathbf{u}_d$ , hence,

$$\mathbf{J} = \frac{n_e e^2 \tau}{m_e} \mathbf{E} \quad (5.48)$$

Comparing the above equation with Eq. (5.46), we obtain the Drude-Lorentz expression:

$$\sigma = \frac{n_e e^2}{m_e} \tau \quad (5.49)$$

The preceding equation is often used to obtain the relaxation time  $\tau$  from the measured electrical conductivity  $\sigma$ . At moderate temperatures, it can be assumed that the characterization velocity of electrons is the Fermi velocity  $v_F$ , and the mean free path of electrons can be written as

$$\Lambda_e = v_F \tau \quad (5.50)$$

The electron scattering mechanisms are illustrated in Fig. 5.10. Electron-electron scattering is inelastic and usually negligible compared with electron-phonon scattering, which

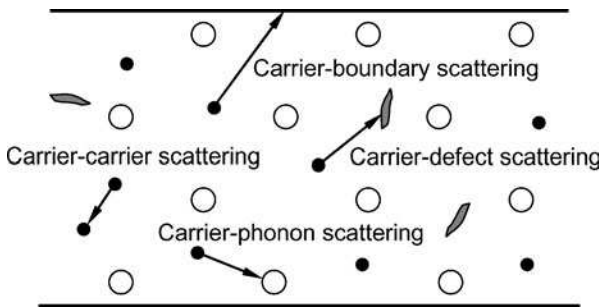


FIGURE 5.10 Schematic of various carrier scattering mechanisms.

is also inelastic. Because lattice vibrations are enhanced as temperature increases, electron-phonon scattering is expected to be dominant at high temperatures. Defect or impurity scattering, on the other hand, is important at low temperatures. For bulk materials that are large enough, boundary scattering is negligible. According to Matthiessen's rule, the scattering rate of independent scattering events can be added to yield the total scattering rate. For a bulk material, we have

$$\frac{1}{\tau} = \frac{1}{\tau_{e-e}} + \frac{1}{\tau_{e-ph}} + \frac{1}{\tau_{e-d}} \approx \frac{1}{\tau_{e-ph}} + \frac{1}{\tau_{e-d}} \quad (5.51a)$$

where the subscripts e-e, e-ph, and e-d are for electron-electron, electron-phonon, and electron-defect scattering. Using Eq. (5.50), we can write the above equation in terms of the mean free path as follows:

$$\frac{1}{\Lambda_e} = \frac{1}{\Lambda_{e-ph}} + \frac{1}{\Lambda_{e-d}} \quad (5.51b)$$

In semiconductors, both electrons and holes can carry currents. The scattering mechanisms can be considered separately. Boundary scattering becomes important when the characteristic dimension  $L$  is comparable to the mean free path of the bulk material  $\Lambda_e$ . Here,  $L$  can be the thickness of a thin film or the diameter of a thin wire. An effective mean free path can be defined for the evaluation of the scattering rate and the conductivity, i.e.,

$$\frac{1}{\Lambda_{e,\text{eff}}} = \frac{1}{\Lambda_e} + \frac{1}{\Lambda_{e-b}} \quad (5.52)$$

where the subscript e-b is for electron-boundary scattering. It can be seen that when boundary scattering is important, the effective mean free path will be suppressed, or the scattering rate will increase, when Eq. (5.52) is substituted into Eq. (5.50). The electrical conductivity will be reduced, and the reduction is size dependent. This is similar to the molecular heat transfer discussed in Chap. 4 when the  $Kn$  number, i.e.,  $\Lambda/L$ , is comparable or larger than 1. Further discussion of the size effect on the conductivities of solids will be given in Sec. 5.5.

The Bloch formula for electrical resistivity due to electron-phonon scattering gives

$$r_{e-ph} = 4r_0 \left( \frac{T}{\Theta} \right)^5 \int_0^{\Theta/T} \frac{x^5 e^x}{(e^x - 1)^2} dx \tag{5.53}$$

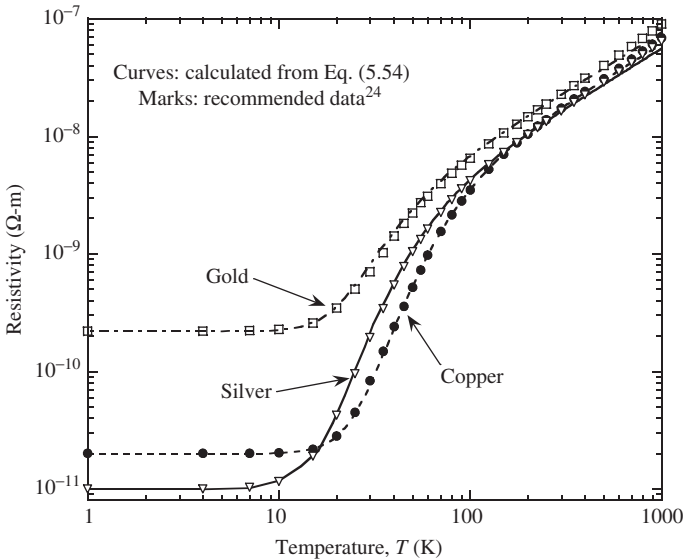
where  $r_0$  is a constant, and  $\Theta$  is a characteristic temperature that is very close to the Debye temperature.<sup>23</sup> The derivation of the above equation requires a careful treatment of the electron-phonon interaction within the framework of the electron band theory considering both the  $N$  process and the  $U$  process, which will be discussed in Chap. 6. The Bloch formula predicts that the electrical resistivity approaches zero as the temperature approaches absolute zero for a pure metal. When  $T \ll \Theta$ , the low-temperature approximation of the lattice resistivity can be written as

$$r_{e-ph} \approx 500r_0 T^5 / \Theta^5 \tag{5.53a}$$

Because of impurities, electron-defect scattering gives a residual resistivity  $r_{e-d}$  that is important at low temperatures and its value is independent of temperature. Adding the scattering rates using Matthiessen's rule,<sup>23</sup> the electrical resistivity is obtained as

$$r_e = r_{e-ph} + r_{e-d} \tag{5.54}$$

Figure 5.11 compares the model with the electrical resistivity data recommended for high-purity bulk metals after annealing.<sup>24</sup> Take the electrical resistivity of gold as an example,



**FIGURE 5.11** Comparison of the measured electrical resistivity data<sup>24</sup> of 99.999% pure copper, gold, and silver with the model considering electron-phonon scattering and electron-defect scattering.

it can be seen that phonon scattering dominates the electrical resistivity at high temperatures and results in  $r_{e\text{-ph}} \approx r_0 T/\Theta$ , which is proportional to  $T$ . It should be noted that  $\Theta_D$  listed in Table 5.1 can be used to approximate  $\Theta$  in most cases. The constant  $r_0$  can be determined using the resistivity values at 22°C, or 295 K, given in Table 5.2. At very low temperatures,  $r_e \approx r_{e\text{-d}}$ , which is independent of temperature but depends strongly on the impurity concentration.

**Example 5-5.** Consider a large copper specimen of high purity with a very small defect scattering rate of  $1/\tau_{e\text{-d}} = 5 \times 10^8$  rad/s at the liquid-helium temperature of 4.2 K. Find the electrical resistivity, the electron scattering time, and the mean free path of this specimen at 1, 295, and 590 K.

**Solution.** We first use Eq. (5.49) to evaluate the residual resistivity at 1 K by assuming that the scattering rate is the same at 4.2 and 1 K. This yields an electrical resistivity  $r_e \approx r_{e\text{-d}} = m_e(n_e e^2 \tau_{e\text{-d}}) = 2.1 \times 10^{-5} \mu\Omega \cdot \text{cm}$ , or an electrical conductivity of  $4.76 \times 10^{12} (\Omega \cdot \text{m})^{-1}$ . The electrical resistivity at 295 K is given in Table 5.2 to be  $r_{e\text{-ph}} \approx r_e = 1.7 \mu\Omega \cdot \text{cm}$ . Because the Debye temperature for Cu is 340 K, we can approximate the resistivity at 590 K to be twice that of the resistivity at 295 K, i.e.,  $3.4 \mu\Omega \cdot \text{cm}$ . The scattering time is approximately  $2 \times 10^{-9}$  s at 1 K,  $2.47 \times 10^{-14}$  s at 295 K, and  $1.24 \times 10^{-14}$  s at 590 K since the number density is assumed to be temperature independent. Using Eq. (5.50) and the Fermi velocity of  $v_F = 1.57 \times 10^6$  m/s from Example 5-2, we have the mean free path  $\Lambda_e = 3.14$  mm at 1 K, 38.8 nm at 295 K, and 19.4 nm at 590 K. The conductivity of a copper film with a thickness less than 100 nm may be affected by boundary scattering. At low temperatures, however, boundary scattering may be dominant for low-dimensional structures even at the micrometer length scale. For metals, electrons are also responsible for thermal transport. Knowledge of the electrical transport is critical to the understanding of thermal properties. The effect of boundary scattering on transport properties is called the classical size effect.<sup>1,2</sup> Quantum size effect can modify the density of states of electrons and hence the electrical and thermal properties, as will be discussed in Sec. 5.5.3.

### 5.3.2 Thermal Conductivity of Metals

Free electrons are the thermal energy carriers in metals. As discussed in Chap. 4, kinetic theory predicts that the thermal conductivity is

$$\kappa = \frac{1}{3} \rho c_{v,e} v_F \Lambda_e \quad (5.55a)$$

where  $\rho = n_e m_e$  is the mass of electrons per unit volume and  $c_{v,e}$  is the mass specific heat of the electrons. Note that  $\rho c_{v,e}$  is the volumetric specific heat of electrons and can be expressed as  $\rho c_{v,e} = n_e \pi^2 k_B^2 T / 2\mu_F$  using the electron specific heat formula given in Eq. (5.24). Substituting the expression for  $\rho c_{v,e}$  and  $v_F \Lambda_e = v_F^2 \tau \approx 2\mu_F \tau / m_e$  into Eq. (5.55a), we obtain the thermal conductivity of a given metal as follows:

$$\kappa = \frac{n_e \pi^2 k_B^2 T}{3m_e} \tau \quad (5.55b)$$

which is proportional to  $\tau T$ . The Wiedemann-Franz law can be obtained by comparing this equation with the expression for the electrical conductivity given in Eq. (5.49), viz.,

$$Lz \equiv \frac{\kappa}{\sigma T} = \frac{1}{3} \left( \frac{\pi k_B}{e} \right)^2 = 2.44 \times 10^{-8} \text{ W} \cdot \Omega / \text{K}^2 \quad (5.56)$$

where  $Lz$  is called *the Lorentz number*. The measured  $Lz$  value for most conductors is between 2.2 and  $2.7 \times 10^{-8} \text{ W} \cdot \Omega / \text{K}^2$  at room temperature. The derivations given above were based on the simple kinetic theory, which is consistent with the solution of the BTE

under the assumptions of local equilibrium and the relaxation time approximation. The actual scattering process may result in some differences in the effectiveness of transferring momentum and energy during electron-phonon scattering. More detailed theories and experiments have shown that the thermal conductivity of metals is independent of temperature at moderate and high temperatures.<sup>23</sup> The Wiedemann-Franz law is therefore valid near and above room temperature for most metals. As the temperature is lowered, electron-phonon scattering yields a thermal resistance (or  $1/\kappa$ ) that is proportional to  $T^2$ , not  $T^4$  as one would obtain by combining Eq. (5.53a) and Eq. (5.56). Recall that in the intermediate region, approximately between 10 and 100 K, the Wiedemann-Franz law is not valid. At very low temperatures, defect scattering dominates and, because defect scattering is elastic, the Wiedemann-Franz law is valid again so that  $\kappa \propto T$ . Therefore, the thermal conductivity at cryogenic temperatures can be expressed as

$$\frac{1}{\kappa(T)} = \frac{A}{T} + BT^2 \tag{5.57}$$

where  $A$  and  $B$  are positive constants. The first term on the right dominates at very low temperatures, when the thermal conductivity is proportional to  $T$ . As the temperature increases, the thermal conductivity reaches a peak and then falls down proportional to  $1/T^2$ . As the temperature approaches the room temperature, the thermal conductivity changes little with temperature until the melting point is reached. Figure 5.12 plots the thermal conductivity

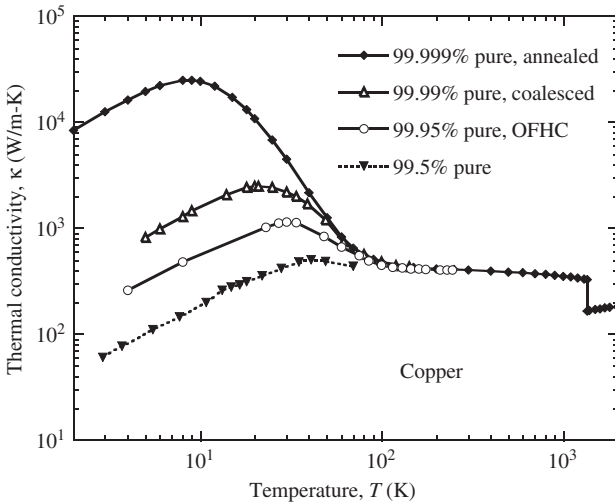


FIGURE 5.12 Thermal conductivity of copper with different impurities.<sup>25</sup>

of copper with different impurity concentrations.<sup>25</sup> The highest purity annealed copper has a residual resistivity of  $5.79 \times 10^{-12} \Omega \cdot \text{m}$ . Oxygen free high conductivity (OFHC) copper is commonly used in absolute cryogenic radiometers to build the cavity receiver. Even 0.5% impurity concentration will make the conductivity dramatically decrease at lower temperatures. On the other hand, the thermal conductivity is less sensitive to the impurity concentration above 100 K and changes little until the melting temperature of 1358 K. Beyond the melting point, the thermal conductivity values are for liquid copper.



### 5.3.3 Derivation of Conductivities from the BTE

So far, we have used simple kinetic theory to discuss the electrical and thermal conductivities of metals. It is hoped that these discussions have provided some insights into basic phenomena. To understand the detailed mechanisms, we now present the approaches based on the BTE under two assumptions: local equilibrium and relaxation time approximation. Recall from Chap. 4 that the distribution function can be expressed in terms of  $f(\mathbf{r}, \mathbf{v}, t)$  or  $f(\mathbf{r}, \mathbf{p}, t)$ , where  $\mathbf{p} = m_e \mathbf{v}$  for electrons. In describing the phonon specific heat, we have extensively used the phonon wavevector  $\mathbf{k}$  as well as the  $k$ -space. The advanced electronic theory or band theory, which is to be discussed in Chap. 6, is also based on the  $k$ -space. Using the magnitude relations:  $p = h/\lambda$  and  $k = 2\pi/\lambda$ , we have  $\mathbf{k} = \mathbf{p}/\hbar$ . Therefore, the distribution function can be written in terms of  $\mathbf{k}$  or  $f(\mathbf{r}, \mathbf{k}, t)$ . The energy of an electron is related to its wavevector by  $\varepsilon = p^2/2m_e = \hbar^2 k^2/2m_e$ . Under the local-equilibrium condition, the distribution function can be written in terms of temperature  $T(\mathbf{r}, t)$  and energy  $\varepsilon$  such that

$$f(\mathbf{r}, \mathbf{k}, t) d\mathbf{k} = f_1(\varepsilon, T) \frac{d\mathbf{k}}{d\varepsilon} d\varepsilon = f_1(\varepsilon, T) D(\varepsilon) d\varepsilon \quad (5.58)$$

where  $D(\varepsilon) = d\mathbf{k}/d\varepsilon$  is the density of states, and  $f_1(\varepsilon, T)$  is such that  $n(\mathbf{r}, t) = \int_0^\infty f_1(\varepsilon, T) D(\varepsilon) d\varepsilon$  and  $\bar{\varepsilon}(\mathbf{r}, t) = \int_0^\infty \varepsilon f_1(\varepsilon, T) D(\varepsilon) d\varepsilon$ . For the equilibrium distribution of free electrons,  $f_1(\varepsilon, T)$  is nothing but the Fermi-Dirac function given in Eq. (5.15). When the distribution function is isotropic in the  $k$ -space, the density of states is given in Eq. (5.18) since  $d\mathbf{k} = dk_x dk_y dk_z = 4\pi k^2 dk$  and  $d\varepsilon = \hbar^2 k dk / m_e$ . As discussed earlier, free electrons will occupy all the quantum states below the Fermi level. The Fermi level corresponds to a maximum  $k$  in all directions in the  $k$ -space, which is a spherical surface. All the electron quantum states are included in this sphere called the Fermi sphere. The argument is similar to the Debye model of phonons, where there is an upper bound of the wavevector and the distribution is assumed to be isotropic. We will see in Chap. 6 that the Fermi surface even for monatomic solids with the simplest crystalline structures is not exactly spherical. This is because the electrons in solids are not really independent particles. For simplicity, a spherical Fermi surface is assumed in this section.

Suppose there is a constant electric field  $E$  along with a temperature gradient in the  $z$  direction. The function  $f_1(\varepsilon, T)$  is a nonequilibrium distribution that depends on  $z$ . At steady state under the relaxation time approximation, we can rewrite Eq. (4.51) as follows:

$$f_1(\varepsilon, T) = f_0(\varepsilon, T) + \tau(\varepsilon) \left( \frac{eE}{m_e} \frac{\partial f_1}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial v_z} - v_z \frac{\partial f_1}{\partial T} \frac{dT}{dz} \right) \quad (5.59)$$

where  $f_0(\varepsilon, T)$  corresponds to the equilibrium distribution, which for electrons is the Fermi-Dirac function  $f_{\text{FD}}$ . The relaxation time is not taken as a constant; rather, it is assumed to be dependent on the wavevector or the energy. Note that  $\partial \varepsilon / \partial v_z = (\partial \varepsilon / \partial v)(\partial v / \partial v_z) = m_e v (v_z / v) = m_e v_z$ . As discussed in Chap. 4, under local equilibrium, we also assume that

$$\frac{\partial f_1}{\partial \varepsilon} \approx \frac{\partial f_0}{\partial \varepsilon} \quad \text{and} \quad \frac{\partial f_1}{\partial T} \approx \frac{\partial f_0}{\partial T} \quad (5.60)$$

We will consider the effect of applied field and temperature gradient separately. When there is no temperature gradient, the current density can be written as

$$J_c = -eJ_N = -e \int_0^\infty v_z \left( f_{\text{FD}} + \tau v_z eE \frac{\partial f_{\text{FD}}}{\partial \varepsilon} \right) D(\varepsilon) d\varepsilon \quad (5.61a)$$

The first term  $-\int_0^\infty e v_z f_{\text{FD}}(\varepsilon, T) D(\varepsilon) d\varepsilon$  is zero; and therefore,

$$J_e = -e^2 E \int_0^\infty \tau(\varepsilon) v_z^2 \frac{\partial f_{\text{FD}}}{\partial \varepsilon} D(\varepsilon) d\varepsilon \quad (5.61b)$$

Because the integration is over the equilibrium distribution, it is one-third of the integration if  $v_z^2$  is replaced by  $v^2 = 2\varepsilon/m_e$ . The electrical conductivity can be expressed as

$$\sigma = -\frac{2e^2}{3m_e} \int_0^\infty \frac{\partial f_{\text{FD}}}{\partial \varepsilon} \tau(\varepsilon) \varepsilon D(\varepsilon) d\varepsilon \quad (5.62)$$

Note that  $\partial f_{\text{FD}}/\partial \varepsilon \approx -\delta(\varepsilon - \mu)$ , where  $\delta(\varepsilon - \mu)$  is the Dirac delta function with a sharp peak at  $\varepsilon = \mu$  and essentially zero when  $\varepsilon \neq \mu$ . Furthermore,  $\int_{-\infty}^\infty f(x)\delta(x - a)dx = f(a)$ . Consequently, the only active electrons are those around the Fermi level. This small fraction of electrons, however, is responsible to the conduction of electricity and heat in metals. We have by assuming  $\mu(T) \approx \mu_F$  that

$$\sigma = -\frac{2e^2}{3m_e} \tau_F \mu_F D(\mu_F) \quad (5.63)$$

which is the same as Eq. (5.49) since  $D(\mu_F) = 3n_e/(2\mu_F)$  according to Eq. (5.18) and Eq. (5.20). The relaxation time is not the average of all electrons but the average of only those electrons near the Fermi surface.

To evaluate the thermal conductivity, we set the applied field to be zero. Note that for an open system of fixed volume,  $dU = \delta Q - \mu dN$ , i.e., the heat flux is equal to the energy flux *minus* the product of the chemical potential and the particle flux. Hence,

$$q_z'' = J_E - \mu J_N = \int_0^\infty v_z(\varepsilon - \mu) \left( f_{\text{FD}}(\varepsilon, T) - \tau(\varepsilon) v_z \frac{\partial f_{\text{FD}}}{\partial T} \frac{dT}{dz} \right) D(\varepsilon) d\varepsilon \quad (5.64)$$

It should be noted that the integration of the equilibrium distribution function in Eq. (5.64) is zero, as noticed earlier. Furthermore, the integration for  $v_z^2$  can be converted to the integration for  $v^2 = 2\varepsilon/(3m_e)$ . After some manipulations, it can be shown that the thermal conductivity is

$$\kappa = \frac{2}{3m_e} \int_0^\infty \tau(\varepsilon) (\varepsilon - \mu) \varepsilon \frac{\partial f_{\text{FD}}}{\partial T} D(\varepsilon) d\varepsilon \quad (5.65a)$$

Using Eq. (B.82) from Appendix B.8, i.e.,  $\frac{\partial f_{\text{FD}}}{\partial T} = -\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \left( \frac{\varepsilon - \mu}{T} \right)$ , we obtain after applying Eq. (B.80) that

$$\kappa = -\frac{2}{3m_e T} \int_0^\infty \tau(\varepsilon) (\varepsilon - \mu)^2 \varepsilon \frac{\partial f_{\text{FD}}}{\partial \varepsilon} D(\varepsilon) d\varepsilon = \frac{2}{3m_e T} \tau(\mu_F) \mu_F D(\mu_F) \frac{\pi^2 (k_B T)^2}{3} \quad (5.65b)$$

This is essentially the same expression as in Eq. (5.55b) for the electron thermal conductivity obtained from the simple kinetic theory. The previous discussion based on the Fermi-Dirac distribution not only confirms the simple kinetic theory but also explains why  $v_F$  should be used in Eq. (5.50) and Eq. (5.55a) rather than the rms velocity of electrons. A familiarity with the BTE will help the study of the classical size effect due to boundary scattering and thermoelectricity phenomena to be discussed in subsequent sections.

The derivation above has confirmed the electrical conductivity and thermal conductivity expressions, and explained that the scattering rate corresponds to electrons with energy equal to the Fermi energy. Therefore, the Wiedemann-Franz law is also confirmed since the scattering rates for the electron (momentum) transport and that for energy transport cancel each other. Electron-phonon scattering must satisfy the energy and momentum conservations. When the amount of energy change of electrons before and after collision is comparable with  $k_B T$ , the scattering is inelastic and the two scattering processes can differ significantly. This happens at lower temperatures since  $k_B T$  is small. At very low temperatures, since electron-defect scattering is elastic, the transport of electron momentum is as effective as the transport of energy. As discussed earlier, the result in the low-temperature region for electron-phonon scattering is such that the electrical resistivity follows  $T^5$ , while  $1/\kappa$  follows  $T^2$ . In order for Eq. (5.55a) and Eq. (5.55b) to be valid, it is often thought as if the relaxation time for thermal conductivity is somewhat different than that for electrical conductivity. In essence, it is not the scattering time that is different; it is the relaxation time approximation that is not valid. By using two relaxation times, one can simplify the scattering process. The relaxation time for momentum transfer retains its meaning of the relaxation time, as in Eq. (5.49) for the electrical conductivity. On the other hand, the relaxation time in Eq. (5.55b) is sometimes called the *energy relaxation time*, which is taken as a weighted average to approximate the difference in the scattering effectiveness for energy exchange.

### 5.3.4 Thermal Conductivity of Insulators

Conduction in insulators is dominated by lattice waves or phonons. This class of materials includes diamond, quartz, glass, as well as semiconductor materials like silicon and GaAs. Kinetic theory predicts the thermal conductivity of dielectric materials or electrical insulators as follows:

$$\kappa = \frac{1}{3} \rho c_v v_a \Lambda_{\text{ph}} \quad (5.66)$$

where  $\rho c_v$  is the lattice volumetric specific heat,  $v_a$  is the average speed of corresponding acoustic waves or phonons, and  $\Lambda_{\text{ph}}$  is the phonon mean free path and is related to the scattering rate by  $\Lambda_{\text{ph}} = v_a \tau$ . When  $v_a$  is used, it is often assumed that the dispersion relation is linear, i.e.,  $v_g = v_p$ . For crystalline solids, the acoustic speed is on the order of 5000 m/s and depends little on temperature; however, it may depend on the polarization. The density decreases slightly as temperature increases due to thermal expansion but the change is negligibly small. The specific heat  $c_v$  is a function of temperature as predicted by the Debye theory, and it is nearly constant at temperatures close to or higher than the Debye temperature. The mean free path can be evaluated based on phonon-phonon scattering and phonon-defect scattering.

Before discussing further the temperature dependence of the scattering rate, we would like to derive Eq. (5.66) from the relaxation time approximation based on the Debye theory. The assumption is that the phonon velocity can be taken as a constant that is averaged over all three modes according to Eq. (5.7), which describes the density of states. For phonons, the distribution function can be conveniently converted into the frequency  $\nu$  domain. Suppose there is a temperature gradient in the  $z$  direction; using the procedure similar to that used in the previous section, the thermal conductivity can be expressed as

$$\kappa = \iiint_{\nu, \phi, \theta} v_z \hbar \nu \tau v_z \frac{\partial f_{\text{BE}}}{\partial T} \frac{D(\nu)}{4\pi} \sin \theta d\theta d\phi d\nu \quad (5.67)$$

where  $D(\nu)/4\pi$  is the density of states per unit solid angle. Noting that  $v_z = v_a \cos \theta$  and the distribution function is independent of the direction, we can integrate Eq. (5.67) over all angles first to get  $\int_0^{2\pi} \int_0^\pi \cos^2 \theta \sin \theta d\theta d\phi = 4\pi/3$ . With the upper limit of the frequency  $\nu_m$ , determined by Eq. (5.9), we can rewrite Eq. (5.67) as in the following:

$$\kappa = \frac{1}{3} \int_0^{\nu_m} \tau v_a^2 h \nu \frac{\partial f_{BE}}{\partial T} D(\nu) d\nu \tag{5.68}$$

The integration over the spherical coordinates offers a different way for deriving the 1/3 term in the kinetic expression of thermal conductivity obtained earlier for a molecular gas and an electron gas. In addition to the assumption that the acoustic velocity is independent of the frequency, we further assume that the scattering rate is independent of the frequency. Hence, both  $\tau$  and  $v_a$  can be taken out of the integrand. The remaining part is the specific heat per unit volume, defined in Eq. (5.31). It is clear that Eq. (5.66) can be obtained based on the assumption that phonon speed, relaxation time, and mean free path are independent of frequency.

Using Matthiessen’s rule, the phonon mean free path can be expressed as

$$\frac{1}{\Lambda_{ph}} = \frac{1}{\Lambda_{ph-ph}} + \frac{1}{\Lambda_{ph-d}} \tag{5.69}$$

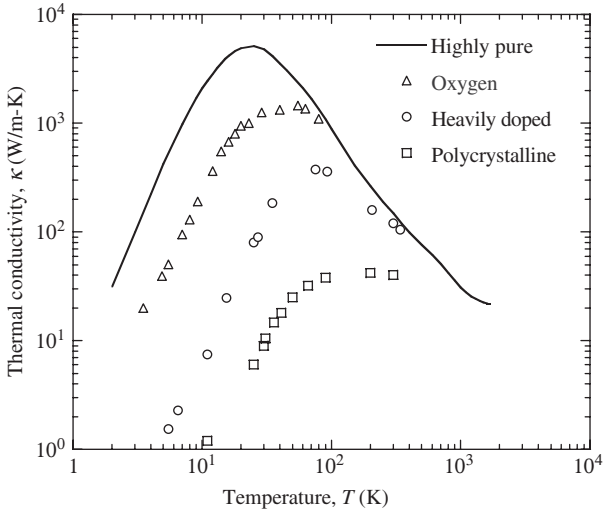
where ph-ph and ph-d stand for phonon-phonon scattering and phonon-defect scattering, respectively. The inverse of the mean free path can be added because the number of collisions can be added. The scattering rate due to phonon-phonon scattering is inversely proportional to temperature at relatively high temperatures, i.e.,  $\Lambda_{ph-ph}$  decreases as temperature increases. This causes a reduction in thermal conductivity as temperature goes up. To the first-order approximation, we can say that the thermal conductivity is inversely proportional to temperature in the high-temperature limit. At low temperatures, scattering on defects dominates and the scattering rate is more or less constant. The thermal conductivity depends on the specific heat and should also vary with  $T^3$ . In addition, the size of the sample affects the mean free path and hence the thermal conductivity. Also, as the temperature is reduced, phonons with lower frequencies play an important role in the thermal transport and storage. Thus, boundary scattering is expected to be more important at low temperatures. The effective mean free path can be defined similar to that for electron scattering as

$$\frac{1}{\Lambda_{ph,eff}} = \frac{1}{\Lambda_{ph}} + \frac{1}{\Lambda_{ph-b}} \tag{5.70}$$

Figure 5.13 shows the thermal conductivity of silicon with different impurity concentrations. For highly pure single-crystal silicon, the thermal conductivity is comparable with a good electrical conductor such as aluminum. As the impurity becomes more important, the scattering rate is increased and the mean free path is reduced, resulting in a reduction in the thermal conductivity. The contribution of free electrons is less important compared with lattice conduction.

**Example 5-6.** Estimate the mean free path and the phonon scattering rate of pure silicon at 5, 10, 20, 100, 300, and 1000 K. Also, calculate the corresponding thermal diffusivity  $\alpha = \kappa/\rho c_p$ .

**Solution.** The purpose of this example is to give some quantitative information of the mean free path and its temperature dependence. The calculation is straightforward using Eq. (5.68) by assuming that the density and the phonon velocity are independent of temperature. From Example 5-1, we have  $v_a \approx 6000$  m/s, and the density is found from Table 5.1 to be 2330 kg/m<sup>3</sup>. The specific heat can be calculated from the Debye model, and the thermal conductivity can be found from Fig. 5.12. The computed results are tabulated in the following table. The mean free path and the thermal diffusivity increase dramatically as the temperature is lowered. Because the crystal is highly pure, there is



**FIGURE 5.13** Data of thermal conductivity of silicon taken from Touloukian et al.<sup>25</sup> The fitted curve is for highly pure silicon with a dopant concentration less than  $10^{16} \text{ cm}^{-3}$ ; triangles are for  $p$ -type single-crystal silicon with an oxygen concentration of  $2 \times 10^{17} \text{ cm}^{-3}$ ; circles are for a heavily doped  $n$ -type silicon with a phosphorus concentration of  $2 \times 10^{19} \text{ cm}^{-3}$ ; and squares are for a  $p$ -type polycrystalline silicon with a boron concentration of  $3 \times 10^{20} \text{ cm}^{-3}$ .

very little scattering at low temperatures. The decrease in the conductivity is caused by the reduction in the specific heat. At high temperatures, the specific heat does not change significantly, and the decrease in the thermal conductivity is due to the increasing phonon-phonon scattering rate. It should be mentioned that at very high temperatures, thermally activated free electrons or holes will also increase the impurity scattering.

Temperature (K)	5	10	20	100	300	1000
Thermal conductivity $\kappa$ [W/(m · K)]	424	2110	4940	884	148	31.2
Specific heat $c_p$ [J/(kg · K)]	0.034	0.28	3.43	260	712	921
Mean free path $\Lambda$ (m)	$2.7 \times 10^{-3}$	$1.6 \times 10^{-3}$	$3.1 \times 10^{-4}$	$7.3 \times 10^{-7}$	$4.5 \times 10^{-8}$	$7.3 \times 10^{-9}$
Scattering rate $1/\tau$ (rad/s)	$2.2 \times 10^6$	$3.7 \times 10^6$	$1.9 \times 10^7$	$8.2 \times 10^9$	$1.3 \times 10^{11}$	$8.3 \times 10^{11}$
Thermal diffusivity $\alpha$ (m <sup>2</sup> /s)	5.4	3.3	0.62	$1.5 \times 10^{-3}$	$8.9 \times 10^{-5}$	$1.5 \times 10^{-6}$

When the phonon mean free path is comparable with the smallest dimension so that  $Kn \equiv \Lambda/L > 1$ , boundary scattering or classical size effect should be considered, as will be discussed in Sec. 5.5. When  $Kn \gg 1$ , ballistic or phonon-boundary scattering becomes dominant compared with phonon-phonon and phonon-defect scattering. As in the case of free molecule flow, Fourier's law is applicable only in the diffusion limit. When ballistic scattering is significant, the temperature at the boundary is discontinuous. Since phonons obey the same statistics as photons, the transfer process is more radiative than conductive. Even at the steady state, the 1-D temperature distribution without heat generation is non-linear. We will study the equation of phonon radiative transfer (EPRT) in Chap. 7 along with other equations that should be used for small timescales or length scales, where Fourier's law of heat conduction breaks down. This is especially important at low temperatures, for small structures, and/or in rapid processes such as during a short laser pulse.

So far, we have studied the basics of phonon contributions to the thermal conductivity under the relaxation time approximation, i.e., by assuming that  $\tau$  is independent of the vibration frequency. Furthermore, we have taken the average acoustic velocity and assumed that it is also independent of the vibration frequency. A further assumption is made that the phonon dispersion relations are isotropic and linear up to a maximum frequency. Real crystals behave very differently from the simple pictures just presented. To understand this, we must study the phonon dispersion relations for all phonon branches, with different polarizations and along different crystal directions. While the study of crystalline structures and phonon dispersion relations will be deferred to Chap. 6, we can write the general expression for thermal conductivity under the local-equilibrium condition in two forms. The summation form reads as

$$\kappa(\hat{\mathbf{n}}) = \sum_P \sum_K \hbar \omega(\mathbf{k}) \frac{\partial f_{BE}}{\partial T} \tau(\mathbf{k}) v_{g,n}^2(\mathbf{k}) \tag{5.71}$$

where the summation is over the wavevector index  $K$  and the polarization index  $P$ ,  $v_{g,n}(\mathbf{k})$  is the phonon group velocity for the given polarization in the direction  $\hat{\mathbf{n}}$  along which the thermal conductivity is to be evaluated. The integration form reads as

$$\kappa(\hat{\mathbf{n}}) = k_B \sum_P \int_0^\infty \tau(\omega) v_{g,n}^2(\omega) \left( \frac{\hbar \omega}{k_B T} \right)^2 \frac{e^{\hbar \omega/k_B T}}{(e^{\hbar \omega/k_B T} - 1)^2} D(\omega) d\omega \tag{5.72}$$

where  $D(\omega)$  is the density of states for an individual polarization. For isotropic distribution in the  $k$ -space,

$$D(\omega) = \frac{1}{(2\pi)^3} \frac{d\mathbf{k}}{d\omega} = \frac{1}{2\pi^2} \frac{k^2}{d\omega/dk} = \frac{\omega^2}{2\pi^2 v_p^2 v_g}$$

where  $v_p(\mathbf{k}) = \omega/k$  and  $v_g(\mathbf{k}) = d\omega/dk$  are the phase and group velocities for the corresponding polarization. It should be noted that when  $d\mathbf{k}$  is used in the numerator, it is the elemental volume in the  $k$ -space, i.e.,  $d\mathbf{k} = dk_x dk_y dk_z = k^2 \sin \theta dk d\theta d\phi$ . In some equations, we place the derivative of a vector in the denominator to obtain the gradient, as in Eq. (4.50) and Eq. (B.58). If the density of states is properly handled so that it contains information about a particular microstructure, Eq. (5.72) would be identical to Eq. (5.71). Otherwise, Eq. (5.72) is the approximation of Eq. (5.71) for large systems. For a large system with isotropic dispersion, we have

$$\kappa = \frac{k_B}{6\pi^2} \left( \frac{k_B T}{\hbar} \right)^3 \sum_P \int_0^{x_m} \tau(x) \frac{v_g(x)}{v_p^2(x)} \frac{x^4 e^x}{(e^x - 1)^2} dx \tag{5.73}$$

where the upper limit corresponds to the maximum frequency of each phonon polarization or branch. Equation (5.73) helps us understand low-temperature behavior of thermal conductivity of insulators.

For the same frequency, while the energy of a phonon is the same as that of a photon  $h\nu$ , the acoustic wave has a much shorter wavelength than the electromagnetic wave because of the small speed  $v_a$  compared with the speed of light. Thus, the momentum of a phonon will be much greater than that of a photon of the same frequency. As an example, our ears sense sound waves in the frequency range from 20 to 20,000 Hz. Assume  $v_a = 1000$  m/s; then, the wavelength range is 50 m to 5 cm. However, these are not the most important frequencies for thermal energy transfer in solids. The smallest vibration wavelength is roughly  $\lambda_{\min} = 2L_0 \approx 0.5$  nm. With a typical velocity of  $v_a = 5000$  m/s in crystalline solids, the highest frequency  $\nu_m$  is on the order of 10 THz. Note that 1 THz (terahertz) =  $10^{12}$  Hz. Compared with electromagnetic wave spectrum, this frequency falls in the mid-infrared spectral region. Therefore, electromagnetic radiation can interact with such phonons, and the resulting absorption is called lattice absorption or phonon absorption. High-frequency phonons are called *optical phonons*. On the other hand, acoustic phonons refer to the frequency range from 0 to 10 THz. By setting  $k_B T = h\nu$ , we find that the frequency corresponding to the thermal energy of translational motion of a particle is on the order of  $\nu = k_B T/h = 6$  THz at 300 K (where  $k_B T = 26$  meV). The thermal phonon wavelength  $\lambda_{th}$  is therefore on the order of 1 nm with  $v_a \approx 5000$  m/s. On the other hand, low-frequency phonons are responsible for energy storage and transfer in crystalline solids at cryogenic temperatures. The shift in the dominant frequency for phonon transport resembles Wien's displacement law for blackbody radiation because phonons and photons are governed by the same statistics. The phonon wave effect and quantum size effect are expected to become important when the characteristic dimension is on the order of the thermal wavelength, as illustrated earlier in the study of specific heat of solids.

## 5.4 THERMOELECTRICITY

Solid state energy conversion devices are very important, and it is hoped that nanotechnology may offer solutions for improving the efficiency of these devices, such as thermoelectric refrigerators and power generators. An understanding of thermoelectricity is useful for further development of these solid state energy conversion devices. To illustrate the *thermoelectric effect*, assume there are an electric field  $\mathbf{E}$  and a temperature gradient  $\nabla T$  along the  $z$  direction of a conductor. We can substitute Eq. (5.59) and Eq. (5.60) into Eq. (5.61) for the electrical current density and into Eq. (5.64) for the heat flux. By dropping the integration for the equilibrium distribution and using Appendix B.8, we can write the 3-D vector forms of the current density and the heat flux as (see Problem 5.21)

$$\mathbf{J}_e = L_{11} \left( \mathbf{E} + \frac{\nabla\mu}{e} \right) - L_{12} \nabla T \quad (5.74)$$

and

$$\mathbf{q}'' = L_{21} \left( \mathbf{E} + \frac{\nabla\mu}{e} \right) - L_{22} \nabla T \quad (5.75)$$

$$\text{where} \quad L_{11} = -e^2 \Psi_0, \quad L_{12} = \frac{e}{T} \Psi_1, \quad L_{21} = T L_{12} = e \Psi_1, \quad \text{and} \quad L_{22} = -\frac{1}{T} \Psi_2 \quad (5.76)$$

Here, the function  $\Psi_n$  is defined as

$$\Psi_n = \frac{1}{3} \int_0^\infty (\epsilon - \mu)^n \tau v^2 \frac{\partial f_{FD}}{\partial \epsilon} D(\epsilon) d\epsilon \quad (5.77)$$

In writing this equation, we have used Eq. (B.81) and converted  $(d\mu/dT) \nabla T = \nabla\mu$  in order to consider the spatial dependence of  $\mu$ . Let

$$\mathbf{E} + \frac{\nabla\mu}{e} = -\nabla\Phi \quad (5.78)$$

where  $\Phi$  is called the *electrochemical potential* because it is the combination of the electrostatic potential and the chemical potential. For metals at low or intermediate temperatures, the variation in  $\mu$  is relatively small, and the terms involving  $\nabla\mu$  in Eq. (5.74) and Eq. (5.75) can be dropped out. For semiconductors, changing the dopant or impurity concentration as well as the temperature may cause a large gradient of  $\mu$ , and thus  $\nabla\mu$  cannot be neglected. When there is no temperature gradient, we can easily find the electrical conductivity of metals to be

$$\sigma = L_{11} \quad (5.79)$$

The thermal conductivity is defined by  $\mathbf{q}'' = -\kappa\nabla T$  when no electric current flows. By setting  $J_e = 0$  and combining Eq. (5.74) and Eq. (5.75), we find that the thermal conductivity is related to the coefficients by

$$\kappa = L_{22} - L_{12}L_{21}/L_{11} \quad (5.80)$$

For metals, the second term on the right-hand side is much smaller than the first one so that we can approximate  $\kappa \approx L_{22}$ , as already discussed in Eq. (5.65b).

### 5.4.1 The Seebeck Effect and Thermoelectric Power

If there is a temperature gradient, Eq. (5.74) suggests that there will be a current flow in the absence of an external field. On the other hand, if the current flow is set to zero (open circuit), there will be a voltage across the rod whose ends are held at different temperatures. The *Seebeck effect*, as it was first noticed by T. J. Seebeck in 1821, can be used to produce an electrical power directly from a temperature difference. The *Seebeck coefficient*, also called *thermopower* or *thermoelectric power*, is defined as the induced thermoelectric voltage across the material of unit length per unit temperature difference. Therefore,

$$\Gamma_S = \frac{-\nabla\Phi}{\nabla T} = \frac{L_{12}}{L_{11}} \quad (5.81)$$

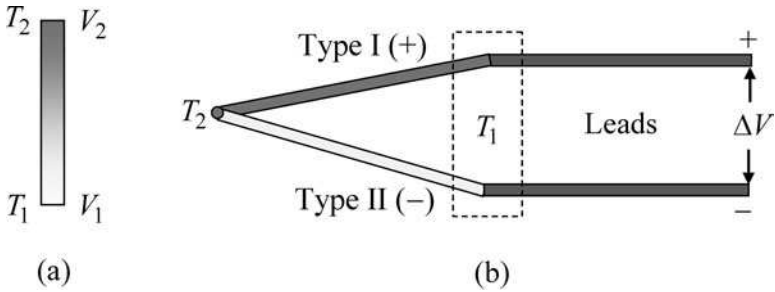
which has units of W/K. To calculate  $L_{12}$  for a metal, we can use Eq. (B.79) to evaluate  $\Psi_1$  in Eq. (5.77). The simplest approach is to assume that  $\tau$  does not change much near the Fermi surface. The result gives (see Problem 5.22)

$$\Gamma_S \approx -\frac{\pi^2 k_B k_B T}{2e \mu_F} \quad (5.82)$$

For metals, the Seebeck coefficient is negative and its magnitude will increase as temperature goes up. From Table 5.2,  $\mu_F = 7$  eV for copper. We have from Eq. (5.82) that  $\Gamma_S = -1.6 \mu\text{V/K}$  at 300 K and  $-3.2 \mu\text{V/K}$  at 600 K. However, the experimental values are positive with  $1.83 \mu\text{V/K}$  at 300 K and  $3.33 \mu\text{V/K}$  at 600 K.<sup>26,27</sup> This sign error is due to the simplification used to evaluate  $\Psi_1$ , and it is an indication that the nearly free electron model may not capture all the fundamental physics of metals. A proper quantum mechanical evaluation based on the actual band structure is rather complicated but has been carried out in some studies.<sup>5,28</sup> Higher values of the Seebeck coefficient can exist in some alloys and semiconductors. Generally speaking, the Seebeck coefficient is positive for *p*-type semiconductors whose majority carriers are holes and negative for *n*-type semiconductors whose majority carriers are electrons.

For a wire whose ends are at different temperatures:  $T_1$  and  $T_2$  in the open circuit shown in Fig. 5.14a, there will be a voltage difference between 1 and 2 according to the relation  $V_2 - V_1 = -\int_{T_1}^{T_2} \Gamma_S(T)dT$ . For *n*-type semiconductors,  $\Gamma_S(T)$  is negative and electrons at the higher-temperature end tend to diffuse toward the lower-temperature end. An electrostatic





**FIGURE 5.14** Illustration of the Seebeck effect. (a) Single wire with a temperature difference between the two ends. (b) A thermocouple made of two different materials.

potential will be built up to balance the diffusion process. Hence, the voltage is higher at the higher-temperature end. Thermoelectric voltage cannot be measured with the same type of wires because the electrostatic potentials would cancel each other. To measure the thermoelectric power, a junction is formed with two types of wires having different Seebeck coefficients, type I (+) and type II (-), as shown in Fig. 5.14b. The leads can be a third type of wire or the same as one of the thermocouple wires. This is of course the familiar thermocouple arrangement for temperature measurement. A reference temperature ( $T_1$ ) is needed because a thermocouple can measure only the temperature difference. The voltage output can be expressed as

$$\Delta V = \int_{T_1}^{T_2} [\Gamma_{S,I}(T) - \Gamma_{S,II}(T)] dT = \Gamma_{I,II} \Delta T \quad (5.83)$$

In thermocouple practice, the difference  $\Gamma_{I,II}$  is called the Seebeck coefficient or thermopower, and the potential difference  $\Delta V$  is called the electromotive force (emf). Because the Seebeck coefficient is zero when a material becomes superconducting ( $\sigma \rightarrow \infty$ ), superconductors have been used to establish an absolute scale of thermoelectric power.<sup>27</sup> In thermometry, a wire with a positive Seebeck coefficient and another with a negative Seebeck coefficient are combined to form a thermocouple junction. For example, a type E thermocouple is made of a nickel-chromium alloy (chromel) and a copper-nickel alloy (constantan); on the other hand, a type J thermocouple is made of copper and constantan. Historically, galvanometer was used to accurately measure the electric current in a potentiometer. The DC voltage can now be measured quickly and very accurately with a digital voltmeter/multimeter (DVM). Detailed discussions about the fundamentals and practice of thermoelectric thermometry based on metallic and alloy wires can be found in Bentley.<sup>26</sup>

## 5.4.2 The Peltier Effect and the Thomson Effect

Equations (5.74 and 5.75) can be combined to eliminate the potential term so that

$$\mathbf{q}'' = \frac{L_{21}}{\sigma} \mathbf{J}_e - \kappa \nabla T \quad (5.84)$$

This equation suggests that there will be a heat flux in the material due to an external electric current, even without any temperature difference. This phenomenon, first discovered by Jean Peltier in 1834, is called the *Peltier effect*, which can be used for refrigeration

(known as *thermoelectric cooling*) by passing through an electric current. The coefficient  $L_{21}/\sigma$  is called the *Peltier coefficient*. It can be seen from Eq. (5.76) and Eq. (5.81) that

$$\Pi = L_{21}/\sigma = T\Gamma_S \tag{5.85}$$

This quantitative relationship between the Seebeck coefficient and the Peltier coefficient was revealed by William Thomson (Lord Kelvin) in the 1850s. Thomson’s thermodynamic derivation led him to discover a third thermoelectric effect, known as the *Thomson effect*, which states that heat can be *released* or *absorbed* when current flows in a material with a temperature gradient. The energy received by a volume element for prescribed  $\mathbf{J}_e$  and  $\nabla T$  can be expressed as follows:

$$\mathbf{J}_e \cdot (-\nabla\Phi) - \nabla \cdot \mathbf{q}'' = \frac{J_e^2}{\sigma} + \nabla \cdot (\kappa\nabla T) - \left( T \frac{d\Gamma_S}{dT} \right) \mathbf{J}_e \cdot \nabla T \tag{5.86}$$

Notice that the common term  $\Gamma_S \mathbf{J}_e \cdot \nabla T$  in both  $\mathbf{J}_e \cdot (-\nabla\Phi)$  and  $\nabla \cdot \mathbf{q}''$  cancels out. In Eq. (5.86), the first term is the heat generated by the Joule heating, the second term is the heat transferred into the control volume due to the temperature gradient, and the third term is caused by the Thomson effect. The last term on the right-hand side is nonzero when there is a current flow with a temperature gradient, unless the Seebeck coefficient is independent of temperature. It should be noted that, like the Seebeck effect and the Peltier effect, the Thomson effect is also a reversible process *per se*. The *Thomson coefficient*  $K$  is defined as the rate of the absorbed heat divided by the product of the current density and the temperature gradient. Thus,

$$K = T \frac{d\Gamma_S}{dT} \tag{5.87}$$

Equation (5.86) has provided a way to determine  $d\Gamma_S/dT$ , after  $\sigma$  and  $\kappa$  are measured at different temperatures. This allows the absolute thermopower to be determined for certain materials at higher temperatures since superconductivity can occur only at very low temperatures. A systematic study has resulted in the determination of absolute thermoelectric power for lead and platinum, which can then be used as reference materials to determine the absolute thermoelectric power for other materials.<sup>27</sup> It should be noted that, before the discovery of high-temperature superconductors, the highest temperature that a material could be made superconducting was 23 K in an alloy. Superconductivity at temperatures above 35 K was discovered in a ceramic material in 1986 and, shortly afterward, superconductivity above the boiling temperature of liquid nitrogen (78 K) has been made possible.

**Example 5-7.** Consider a *p*-type semiconductor rod of diameter  $d = 1$  mm and length  $l = 2$  mm. One end of the rod is in contact with a heat sink at  $T_L = 300$  K, and the other end is in contact with a heat source at  $T_H = 350$  K. What is the open-circuit voltage? If a current  $I = 0.8$  A is allowed to flow from the cold end to the hot end, what is the heat transfer rate to the heat sink? Neglect the temperature dependence of the thermal conductivity, the electrical resistivity, and the Seebeck coefficient by using  $\kappa = 1.1$  W/(m · K),  $r_e = 19 \mu\Omega \cdot \text{m}$ , and  $\Gamma_S = 220 \mu\text{V/K}$ , respectively.

**Solution.** Assume there is no heat transfer via the side of the rod. For an open circuit, the electric potential is higher at the cold end, and the voltage across the rod is  $V_{\text{open}} = \Gamma_S(T_H - T_L) = 11$  mV. The rate of heat transfer to the heat sink by conduction from the heat source is  $q_{L,C} = (\pi d^2/4)\kappa(T_2 - T_1)/L = 21.6$  mW.

When an electric current is running from the cold end to the hot end, the Joule heating is generated uniformly inside the rod. The dissipated heat must reach both ends equally by conduction. The additional heat transfer to the heat sink is  $q_{L,J} = I^2R/2 = 15.5$  mW, where  $R = 48.4$  m $\Omega$  is the resistance of the rod. On the other hand, the Peltier effect results in cooling, or heat removal from the heat sink. From Eq. (5.84) we have  $q_{L,P} = -T_L\Gamma_S I = -52.8$  mW. The combination of the three terms gives the heat transfer rate as  $q_L = q_{L,C} + q_{L,J} + q_{L,P} = -15.7$  mW. The negative sign indicates that heat is removed from the heat sink.

This example demonstrates the Peltier effect for thermoelectric refrigeration. It can be seen that a smaller thermal conductivity will decrease the heat transfer between the two ends, a smaller electrical resistivity will reduce the Joule heating, and a larger Seebeck or Peltier coefficient will enhance the heat removal. For most metals, the thermal conductivity is too high and the Seebeck coefficient is too small for refrigeration applications. Some insulators can have a large Seebeck coefficient but their electrical resistivity is too high for them to be used in thermoelectric devices.

### 5.4.3 Thermoelectric Generation and Refrigeration

With the understanding of the Seebeck effect, the Peltier effect, and the Thomson effect, we are ready to perform a thermodynamic analysis of the thermoelectric generator or refrigerator as illustrated in Fig. 5.15. There are  $N$  pairs of junctions that are connected electrically

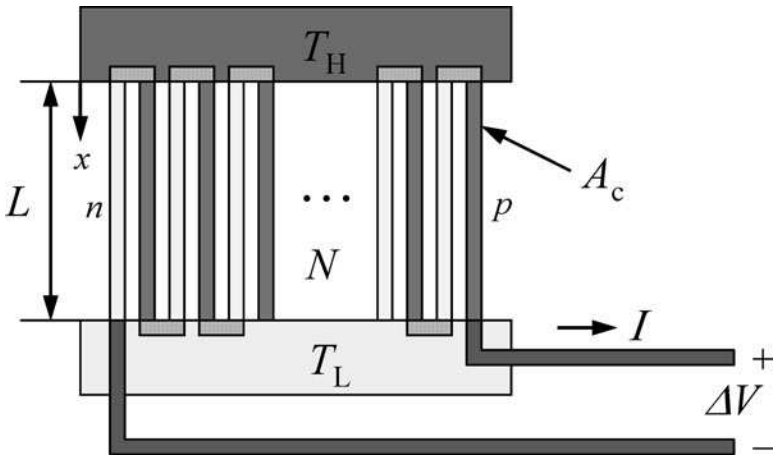


FIGURE 5.15 Illustration of a thermoelectric generator or refrigerator.

in series by metallic interconnects and thermally in parallel between the two heat sinks. The study of thermoelectric generation and refrigeration has become an active research area since the 1950s, along with the development of semiconductor materials or  $p$ - $n$  junctions. Heavily doped semiconductors exhibit large Seebeck coefficients. Alternative  $n$ -type or  $p$ -type semiconductors (or semimetals) are used as thermoelectric materials or *thermoelectric elements*. These include antimony-tellurium (Sb-Te), bismuth-tellurium (Bi-Te), and silicon-germanium (Si-Ge) compounds. More recently, nanostructured materials are investigated as candidates to increase the performance of thermoelectric devices. To simplify the analysis, contact resistances are neglected, and it is assumed that all the thermoelectric elements have the same length  $L$  and the same cross-sectional area  $A_c$ . Furthermore, heat transfer by other modes is neglected except conduction by thermoelectric elements. Because contact electrical resistance is neglected, heat generation by the Joule heating happens due to resistance of the thermoelectric elements only. A load resistance  $R_L$  is used to evaluate the output electric power in the generator. A further assumption is that the thermal and electrical conductivities, as well as the Seebeck coefficient, are independent of temperature. This assumption is reasonable when the temperature difference between the two heat reservoirs is very small.

Consider a thermoelectric generator. In this case, heat is taken from the high-temperature reservoir at  $T_H$  at the rate  $q_H$ , and some heat is released to the low-temperature reservoir at  $T_L$  at the rate  $q_L$ . The generated thermoelectric power is

$$P = I\Delta V = q_H - q_L \quad (5.88)$$

The temperature distribution along the thermoelectric element is not linear, i.e., the temperature gradient is not constant. The steady-state temperature distribution along a single thermoelectric element can be solved by setting Eq. (5.86) to zero. Because of the assumption of constant values of  $I$ ,  $\kappa$ ,  $\sigma$ , and  $\Gamma_S$ , the Thomson coefficient also becomes zero. Therefore, we obtain

$$T(x) = \frac{J_c^2}{2\sigma\kappa}(L-x)x - \frac{x}{L}(T_H - T_L) + T_H \quad (5.89)$$

The resulting heat transfer rates due to temperature gradient are

$$-\kappa A_c \left. \frac{dT}{dx} \right|_{x=0} = \kappa A_c \frac{T_H - T_L}{L} - \frac{I^2 L}{2\sigma A_c} \quad (5.90a)$$

and

$$-\kappa A_c \left. \frac{dT}{dx} \right|_{x=L} = \kappa A_c \frac{T_H - T_L}{L} + \frac{I^2 L}{2\sigma A_c} \quad (5.90b)$$

Clearly, half of the Joule heating goes to the heat source, and the other half goes to the heat sink, as noticed in Example 5-7. Substituting Eq. (5.90) into Eq. (5.84) and using the subscripts  $n$  and  $p$  for different thermoelectric elements, we have

$$q_H = NI\Gamma_{np}T_H + NA_c\kappa_{np}\frac{\Delta T}{L} - N\frac{I^2L}{2A_c\sigma_{np}} \quad (5.91a)$$

$$q_L = NI\Gamma_{np}T_L + NA_c\kappa_{np}\frac{\Delta T}{L} + N\frac{I^2L}{2A_c\sigma_{np}} \quad (5.91b)$$

where  $\Gamma_{np} = \Gamma_{S,p} - \Gamma_{S,n}$ ,  $\kappa_{np} = \kappa_n + \kappa_p$ ,  $\Delta T = T_H - T_L$ , and  $\sigma_{np} = (1/\sigma_n + 1/\sigma_p)^{-1}$ . The output power is therefore

$$P = I\Delta V = q_H - q_L = NI\Gamma_{np}\Delta T - I^2R_0 \quad (5.92)$$

where  $R_0 = NL/(A_c\sigma_{np})$  is the resistance of all thermoelectric elements. The voltage is solely caused by the Seebeck effect, i.e.,  $\Delta V = M\Gamma_{np}\Delta T$ . Assuming the load resistance is  $R_L$ , we have

$$I = \frac{\Delta V}{R_0 + R_L} = \frac{M\Gamma_{np}\Delta T}{R_0 + R_L} \quad (5.93)$$

Substituting Eq. (5.93) into Eq. (5.92), we see that the electric power is indeed  $P = I^2R_L$ . The thermal efficiency can be calculated as follows:

$$\eta = \frac{P}{q_H} = \frac{\frac{R_L}{R_0} \frac{\Delta T}{T_H}}{\frac{1}{Z^*T_H} \left(1 + \frac{R_L}{R_0}\right)^2 + \left(1 + \frac{R_L}{R_0}\right) - \frac{\Delta T}{2T_H}} \quad (5.94)$$

where

$$Z^* = \frac{NL}{A_c\kappa_{np}} \frac{\Gamma_{np}^2}{R_0} = \frac{\sigma_{np}\Gamma_{np}^2}{\kappa_{np}} \quad (5.95)$$

is independent of the geometry.<sup>29</sup> When  $1/Z^*T_H \ll 1$  and  $R_0/R_L \ll 1$ , we have  $\eta \rightarrow 1 - T_L/T_H$ , which is exactly the Carnot efficiency. Increasing  $Z^*$  will improve the efficiency. Hence, minimizing the thermal conduction, reducing the electrical resistance, and increasing the Seebeck coefficient of the thermoelectric elements are essential for improving the performance. A similar analysis can be done for thermoelectric cooling, which is left as an exercise (see Problem 5.25). In general, the *figure of merit* of thermoelectricity is defined as

$$Z = \frac{\sigma \Gamma_S^2}{\kappa} \quad (5.96)$$

which has units of  $1/K$ , and can be nondimensionalized by multiplying it by temperature  $T$ . The resulting dimensionless parameter  $ZT$  (zee-tee) is often quoted as the figure of merit for thermoelectric materials or devices. This applies to both thermoelectric generation and refrigeration (see Problems 5.23 and 5.25).

Because of the compromise between a large electrical conductivity and a small thermal conductivity and the requirement of a large Seebeck coefficient, it has turned out that semiconductors are the best choice for thermoelectric applications. After extensive pursuit in the 1950s, materials with  $ZT$  values between 0.5 and 1 near room temperature have been developed using  $\text{Bi}_x\text{Sb}_{2-x}\text{Te}_3$  and  $\text{Bi}_2\text{Se}_y\text{Te}_{3-y}$ . These materials are essentially doped V-VI semiconductors  $\text{Sb}_2\text{Te}_3$  or  $\text{Bi}_2\text{Te}_3$ . In the past 15 years, intensive theoretical and experimental research has been conducted to increase the thermoelectric device performance by using nanostructured materials. Mildred Dresselhaus and coworkers predicted that multiple quantum wells or superlattices may enhance  $ZT$  values due to quantum confinement as well as a reduction in the phonon thermal conductivity. The idea has been extended to  $\text{PbTe/PbSe}$  superlattice nanowires.<sup>30</sup> Superlattices made of  $\text{SiGe/Si}$  and  $\text{GaAs/AlAs}$  have also been considered. Since 2001, several groups have demonstrated  $ZT$  values exceeding 2.<sup>31–33</sup> Chen's group performed an extensive investigation on the phonon and electron transport in nanostructured materials.<sup>34</sup> The reduction in thermal conductivity may come from a combination of a number of factors including the mean-free-path reduction by boundary scattering, thermal resistance associated with acoustic mismatch or phonon wave scattering at the interface of dissimilar materials, as well as quantum confinement of the phonon density of states. Before moving to the discussion of size effects on thermal conductivity, let us give an overview of irreversible thermodynamics and a brief introduction to nonequilibrium thermodynamics.

#### 5.4.4 Onsager's Theorem and Irreversible Thermodynamics

The set of coupling equations given in Eq. (5.74) and Eq. (5.75) is an example of *irreversible thermodynamics*, pioneered by Lars Onsager in the 1930s. Alternatively, it is also known as the thermodynamics of irreversible processes or Onsager's theorem. Onsager described the phenomenological relations of interrelated or coupled transport processes using the following equation:<sup>35</sup>

$$\mathbf{J}_i = \sum_j \alpha_{ij} \mathbf{F}_j \quad (5.97)$$

where  $\mathbf{J}_i$  is the flux of a physical quantity  $X_i$  with  $J_i = dX_i/dt$ ,  $\alpha_{ij}$  is called the *Onsager kinetic coefficient*, and  $\mathbf{F}_i$  is the  $i$ th generalized driving force or *affinity*. In an equilibrium state, all  $\mathbf{F}_i$ 's are zero. Furthermore, the entropy of a system can be expressed as<sup>36</sup>

$$ds = \sum_i f_i dX_i \quad (5.98)$$

where  $f_i$  is a property that is related to  $\mathbf{F}_i$  such that  $\mathbf{F}_i$  is proportional to the gradient of  $f_i$ . The entropy flux is thus

$$\mathbf{s}'' = \sum_i f_i \mathbf{J}_i \quad (5.99)$$

If an infinitesimal control volume is chosen, the continuity equation can be written as

$$\frac{\partial X_i}{\partial t} + \nabla \cdot \mathbf{J}_i = 0 \quad (5.100)$$

The entropy balance becomes

$$\frac{\partial s}{\partial t} = \dot{s}_{\text{gen}} - \nabla \cdot \mathbf{s}'' \quad (5.101)$$

where  $\partial s / \partial t = \sum_i f_i (\partial X_i / \partial t)$  and  $\nabla \cdot \mathbf{s}'' = \sum_i \nabla f_i \cdot \mathbf{J}_i + \sum_i f_i \nabla \cdot \mathbf{J}_i$ . Using the continuity equation, we obtain the volumetric entropy generation rate:

$$\dot{s}_{\text{gen}} = \sum_i \nabla f_i \cdot \mathbf{J}_i \quad (5.102)$$

Furthermore, the *Onsager reciprocity* is expressed as follows:<sup>35,36</sup>

$$\alpha_{ij} = \alpha_{ji} \quad (5.103)$$

Lars Onsager (1903–1976) received the Nobel Prize in Chemistry in 1968 “for the discovery of the reciprocal relations bearing his name, which are fundamental for the thermodynamics of irreversible processes.” The Onsager reciprocity was even considered by some researchers as the *fourth law of thermodynamics*.

**Example 5-8.** Determine the Onsager kinetic coefficients and the volumetric entropy generation rate for a conductor with a constant current and temperature gradients.

**Solution.** It should be noted that in thermoelectricity,  $\mathbf{J}_1 = \mathbf{J}_e$ ,  $\mathbf{J}_2 = \mathbf{q}''$ ,  $\mathbf{F}_1 = -(1/T)\nabla\Phi$ , and  $\mathbf{F}_2 = \nabla(1/T) = -(1/T^2)\nabla T$ . Thus, the Onsager relations are expressed as

$$\mathbf{J}_e = \alpha_{11} \frac{-\nabla\Phi}{T} - \alpha_{12} \frac{\nabla T}{T^2} \quad (5.104)$$

$$\mathbf{q}'' = \alpha_{21} \frac{-\nabla\Phi}{T} - \alpha_{22} \frac{\nabla T}{T^2} \quad (5.105)$$

Comparing the above expressions with Eq. (5.74) and Eq. (5.75), we find that

$$\alpha_{11} = TL_{11} \quad \alpha_{12} = \alpha_{21} = T^2 L_{12} \quad \text{and} \quad \alpha_{22} = T^2 L_{22} \quad (5.106)$$

The entropy generation rate can be calculated by using Eq. (5.99). Note that

$$ds = \frac{\delta Q - \mu dN}{TV} = \mathbf{q}'' \cdot \nabla \left( \frac{1}{T} \right) + \mathbf{J}_e \cdot \left( -\frac{\nabla\Phi}{T} \right) \quad (5.107)$$

In the steady state, the energy equation, Eq. (5.86), becomes

$$\mathbf{J}_e \cdot (-\nabla\Phi) - \nabla \cdot \mathbf{q}'' = 0 \quad (5.108)$$

Therefore, the volumetric entropy generation rate for 3-D and 1-D cases, respectively, are

$$\dot{s}_{\text{gen}} = \mathbf{q}'' \cdot \nabla \left( \frac{1}{T} \right) + \frac{1}{T} \nabla \cdot \mathbf{q}'' \quad \text{and} \quad \dot{s}_{\text{gen}} = \frac{q''}{T^2} \frac{dT}{dx} + \frac{1}{T} \frac{dq''}{dx} \quad (5.109)$$

These results are consistent with the analysis in Chap. 2 (see Example 2.5 and Problem 2.29). Furthermore, Eq. (5.109) suggests that the Thomson effect is a reversible process that does not cause any entropy generation. The same can be said for both the Seebeck effect and the Peltier effect, which are reversible thermoelectric effects. In addition to thermoelectricity, irreversible thermodynamics has found applications in multicomponent diffusion, nonisothermal diffusion (when both temperature gradient and concentration gradient exist), and some magnetic processes.<sup>36</sup> A further advancement in nonequilibrium thermodynamics was made by Ilya Prigogine (1917–2003) who was awarded the Nobel Prize in Chemistry in 1977. Prigogine's study extended irreversible thermodynamics to systems that are far away from equilibrium and allowed to exchange energy, mass, and entropy with their surroundings. Prigogine and colleagues demonstrated that ordered dissipative systems can be formed from disordered systems, when the systems are far from equilibrium, and dubbed this theory *dissipative structure*, which led to pioneering research in self-organization or self-assembly. The formation of ordered structures from disordered structures has diverse applications in chemical, biological, and social systems.<sup>37</sup> It is beyond the scope of this book to go into details of this theory further.

## 5.5 CLASSICAL SIZE EFFECT ON CONDUCTIVITIES AND QUANTUM CONDUCTANCE

---

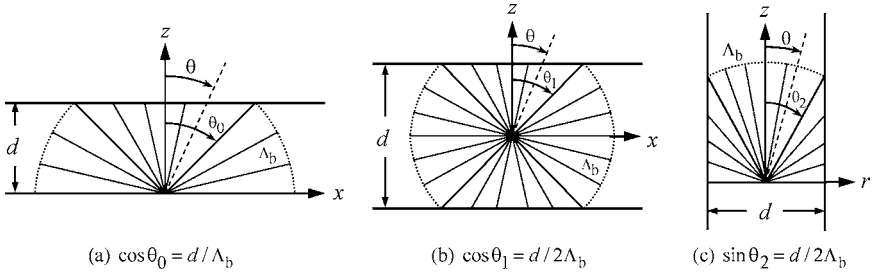
When the characteristic length, such as the thickness of a film or the diameter of a wire and the size of the grains (for polycrystalline solids), is comparable with the mechanistic length, i.e., the mean free path, boundary or interface scattering becomes important. Subsequently, the thermal conductivity (as well as other transport coefficients) becomes size dependent, and can also be anisotropic.<sup>38,39</sup> Because the mean free paths of electrons and phonons tend to increase as temperature goes down, size effects are usually more important at low temperatures. The criteria are also different for different materials. In the following section, we will study the effect of boundary scattering on electrical and thermal conductivities, along with some discussion about the quantum limit of conductance in nanostructures.

### 5.5.1 Classical Size Effect Based on Geometric Consideration

The simple expression of thermal conductivity based on the kinetic theory is  $\kappa = \frac{1}{3}(\rho c_v)v\Lambda_b$  for either electrons or phonons. Here,  $\Lambda_b$  is called the *bulk mean free path*, which is the mean free path when the material is infinitely extended. While the specific heat and the velocity are also size dependent, especially for phonons, let us now focus on the size dependence of the mean free path. The main objective is to illustrate how boundary scattering reduces the thermal conductivity. The argument is also applicable to the electrical conductivity since it is also proportional to the mean free path. Shown in Fig. 5.16 are two geometric configurations to be considered here: (a) and (b) are for a thin film, and (c) is for a thin wire or rod.

In the ballistic transport limit when  $d \ll \Lambda_b$ , if we assume that the mean free path in the film is the same as the thickness  $d$ , i.e.,  $\Lambda_f = d$ , then the conductivity ratio can be obtained as

$$\frac{\kappa_f}{\kappa_b} = \frac{\Lambda_f}{\Lambda_b} = \frac{1}{Kn} \quad (5.110)$$



**FIGURE 5.16** Illustration of free-path reduction due to boundary scattering. (a) A thin film for paths originated from the surface. (b) A thin film for paths originated from the center. (c) A thin wire for paths originated from the center.

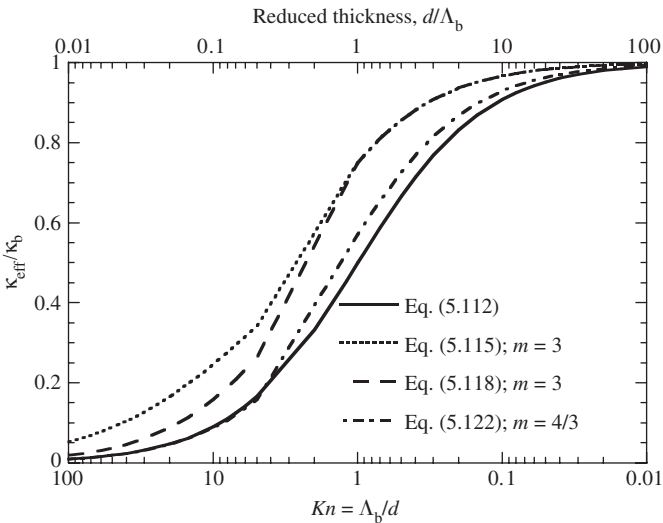
where  $Kn = \Lambda_b/d$  is the Knudsen number, adopted from the theory of rarefied gas dynamics, for electrons or phonons. In the intermediate region, we can apply Matthiessen’s rule as suggested in Eq. (5.52) and Eq. (5.70) such that

$$\frac{1}{\Lambda_{\text{eff}}} = \frac{1}{\Lambda_b} + \frac{1}{\Lambda_r} \tag{5.111}$$

Accordingly,

$$\frac{\kappa_{\text{eff}}}{\kappa_b} = \frac{\Lambda_{\text{eff}}}{\Lambda_b} = \frac{1}{1 + Kn} \tag{5.112}$$

The result calculated from Eq. (5.112) is plotted in Fig. 5.17 to illustrate the size dependence of the effective thermal conductivity. It appears that this simple formula overpredicts the reduction in thermal conductivity, as compared with more realistic models to be discussed next.



**FIGURE 5.17** Reduction in thermal conductivity due to boundary scattering. Note that Eq. (5.116) was used with different  $m$  values for small  $Kn$  numbers.



J. J. Thomson (*Proc. Cambridge Phil. Soc.*, **11**, 120, 1901) was the first to consider the size effect on electrical conductivity of thin films. His argument was extended by Fuchs (*Proc. Cambridge Phil. Soc.*, **34**, 100, 1938) based on the Boltzmann transport equation. The geometric argument assumes boundary scattering is diffuse and inelastic, i.e., the electrons are fully accommodated after scattering by the boundary. The concept of accommodation is the same as that used for ideal gas particles in the free molecule flow regime, discussed in Sec. 4.4. However, the distribution of free paths is not taken into consideration for simplicity. In other words, all paths are assumed to be the same as the mean free path in the bulk. When  $d \ll \Lambda_b$ , we may assume that all energy carriers originate from the boundary. From Fig. 5.16a, we see that

$$\Lambda(\theta) = \begin{cases} dl \cos \theta, & 0 < \theta < \theta_0 \\ \Lambda_b, & \theta_0 < \theta < \pi/2 \end{cases} \quad (5.113)$$

The free paths should be averaged over the hemisphere, and the weighted average can be written and evaluated as follows:

$$\frac{\Lambda_f}{\Lambda_b} = \frac{\int_0^{2\pi} \int_0^{\pi/2} \Lambda(\theta) \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi/2} \Lambda_b \sin \theta d\theta d\phi} = \frac{\ln(Kn) + 1}{Kn} \quad (5.114)$$

Applying Matthiessen's rule again, we have

$$\frac{\kappa_{\text{eff}}}{\kappa_b} = \frac{\Lambda_{\text{eff}}}{\Lambda_b} = \left( 1 + \frac{Kn}{\ln(Kn) + 1} \right)^{-1} \quad (5.115)$$

This equation, however, cannot be applied for small values of  $Kn$  since  $\ln(Kn)$  becomes negative. Let us assume Eq. (5.115) is applicable for  $Kn > 5$ . When  $Kn < 1$ , we may use

$$\frac{\kappa_{\text{eff}}}{\kappa_b} = \frac{\Lambda_{\text{eff}}}{\Lambda_b} = \left( 1 + \frac{Kn}{m} \right)^{-1} \quad (5.116)$$

where  $m \approx 3$  for thin films.<sup>38,39</sup> Equation (5.112) can be considered as a special case of Eq. (5.116) with  $m = 1$ . The results based on Eq. (5.115) and Eq. (5.116) are plotted in Fig. 5.17 for comparison. Interpolation has been taken in the intermediate region when the  $Kn$  value is between 1 and 5. Equations (5.114) and (5.115) did not consider the direction of transport and cannot capture the anisotropic feature due to size effect. Flik and Tien employed a weighted average of the free-path components in the parallel and normal directions of thin films.<sup>39</sup> Their work was extended to different geometries by Richardson and Nori.<sup>40</sup> For the  $z$  direction, the projected mean free path is  $\Lambda_z = \Lambda(\theta) \cos \theta$ ; hence, the weighted average becomes

$$\frac{\Lambda_z}{\Lambda_{b,z}} = \frac{\int_0^{2\pi} \int_0^{\pi/2} \Lambda(\theta) \cos \theta \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi/2} \Lambda_b \cos \theta \sin \theta d\theta d\phi} = \frac{2}{Kn} - \frac{1}{Kn^2} \quad (5.117)$$

The use of Matthiessen's rule allows us to obtain

$$\frac{\kappa_{\text{eff},z}}{\kappa_b} = \left( 1 + \frac{Kn}{2 - Kn^{-1}} \right)^{-1} \quad \text{for } Kn > 5 \quad (5.118)$$

For  $Kn < 1$ , Eq. (5.116) should be used with  $m = 3$ , which can be obtained by integrating over the film when  $Kn \ll 1$ .<sup>39</sup> The result from Eq. (5.118) is also shown in Fig. 5.17. For the  $x$  direction, one may assume that all the electrons originate from the center of the film for simplicity. The component of the free path is  $\Lambda_x = \Lambda(\theta)\sin\theta \cos\phi$ , where  $\phi$  is the azimuthal angle. Due to symmetry, the integration can be carried out in an octant. It can be seen from Fig. 5.16b that  $\Lambda(\theta) = d/(2 \cos\theta)$  for  $0 \leq \theta < \theta_1$ , and  $\Lambda(\theta) = \Lambda_b$  for  $\theta_1 \leq \theta < \pi/2$ , where  $\theta_1 = \cos^{-1}(d/2\Lambda_b)$ . Subsequently,

$$\frac{\Lambda_x}{\Lambda_{b,x}} = \frac{\int_0^{\pi/2} \int_0^{\pi/2} \Lambda(\theta)\sin^2\theta \cos\phi \, d\theta \, d\phi}{\int_0^{\pi/2} \int_0^{\pi/2} \Lambda_b\sin^2\theta \cos\phi \, d\theta \, d\phi} \quad (5.119)$$

After evaluation of the above integrals, we obtain

$$\frac{\kappa_{\text{eff},x}}{\kappa_b} = \frac{2}{\pi Kn} \ln[2Kn(1 + \sin\theta_1)] + 1 - \frac{2\theta_1}{\pi} - \frac{\sin\theta_1}{\pi Kn} \quad (5.120)$$

In the ballistic limit, i.e.,  $Kn \gg 1$ , Eq. (5.120) reduces to  $\kappa_{\text{eff},x}/\kappa_b = (2/\pi Kn)\ln(4Kn)$ . If the free paths were to originate from the boundary, the result could be obtained by replacing  $Kn$  with  $Kn/2$  in Eq. (5.120). While it is perfectly logical to assume that all the carriers originate from the surface for the  $z$  component in the ballistic limit. For thermal transport along the film with a temperature gradient in the  $x$  direction, the carriers must originate from a cross section, i.e., the  $y$ - $z$  plane, inside the film. The transport process along the film is essentially diffusion-like with significant boundary scattering contributions. Anisotropy may arise between  $\kappa_x$  and  $\kappa_z$  due to boundary scattering. A simple argument is that paths with large polar angles are more important for parallel conduction, whereas paths with smaller polar angles are more important for normal conduction. Based on the geometry, it can be seen that paths with smaller polar angles are more likely to be scattered by the boundary. Another reason that causes  $\kappa_x$  to be greater than  $\kappa_z$  is that scattering tends to be more specular for larger incidence polar angles. Specular reflection or elastic scattering does not reduce the conductivity because the incident particles change only the direction without any exchange of energy with the surface. Crystal anisotropy is another major reason for anisotropic conduction, sometimes the dominant reason, as in high-temperature superconducting  $\text{YBa}_2\text{Cu}_3\text{O}_7$  films.<sup>39</sup> Grain boundaries can strongly influence the thermal conductivity in polycrystalline films.<sup>38</sup> For chemical-vapor-deposited polycrystalline diamond films,  $\kappa_x$  may be greater or smaller than  $\kappa_z$  depending on the crystal orientation; see Graebner et al. (*J. Appl. Phys.*, **71**, 5353, 1992).

For circular wires, considering the conduction along a thin wire, as shown in Fig. 5.16c, we have  $\Lambda_z(\theta) = \Lambda_b \cos\theta$  for  $0 < \theta < \theta_2$ , and  $\Lambda_z(\theta) = d \cot\theta/2$  for  $\theta_2 < \theta < \pi/2$ , where  $\theta_2 = \sin^{-1}(d/2\Lambda_b)$ . Thus,

$$\frac{\Lambda_{w,z}}{\Lambda_{b,z}} = \frac{\int_0^{2\pi} \int_0^{\pi/2} \Lambda_z(\theta)\sin\theta \, d\theta \, d\phi}{\int_0^{2\pi} \int_0^{\pi/2} \Lambda_b \cos\theta \sin\theta \, d\theta \, d\phi} = \frac{1}{Kn} - \frac{1}{4Kn^2} \quad (5.121)$$

Applying Matthiessen's rule yields

$$\frac{\kappa_{\text{eff,w}}}{\kappa_b} = \left( 1 + \frac{Kn}{1 - (4Kn)^{-1}} \right)^{-1} \quad (5.122)$$

which can be applied for  $Kn > 5$ . For  $Kn < 1$ , studies have shown that Eq. (5.116) is a good approximation with  $m = 4/3$ .<sup>41,42</sup> The reduction in thermal conductivity for thin wires is also indicated in Fig. 5.17, where the values for  $1 < Kn < 5$  are based on a simple interpolation between the two expressions. Due to the geometric confinement, the reduction in mean free path is more severe for thin wires than for thin films. The geometric argument is easy to understand and may help gain a physical intuition of the size effect due to boundary scattering. In consideration of the classical size effect, it is assumed that Fourier's law is still applicable with a modified thermal conductivity. Derivations based on the BTE are presented next for the size effect on the electron and phonon transport properties along a thin film or a wire.

## 5.5.2 Classical Size Effect Based on the BTE

In Sec. 5.3.3, we derived electrical and thermal conductivities based on the BTE for bulk materials. The relaxation time approximation was adopted, and the distribution function was assumed to be not too far away from equilibrium, i.e., under the local-equilibrium conditions. To determine the size effect on the conductivities along thin films, the same assumptions will be applied. Consider the geometry shown in Fig. 5.16a, with a temperature gradient and an electric field in the  $x$  direction only. Because of the finite thickness in the  $z$  direction, the distribution function should also be an explicit function of  $z$ , viz.,

$$f_1(\varepsilon, T, z) \approx f_0(\varepsilon, T) + \tau(\varepsilon) \left( \frac{eE}{m_e} \frac{\partial f_0}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial v_x} - v_x \frac{\partial f_0}{\partial T} \frac{dT}{dx} - v_z \frac{\partial f_1}{\partial z} \right) \quad (5.123a)$$

Compared with Eq. (5.59), the last term was added because  $f_1$  depends also on  $z$ . Here, the electric field and the temperature gradient are along the  $x$  direction rather than the  $z$  direction as in previous sections on the conductivities. In Eq. (5.123a), we have already replaced  $\partial f_1 / \partial \varepsilon$  with  $\partial f_0 / \partial \varepsilon$  and  $\partial f_1 / \partial T$  with  $\partial f_0 / \partial T$ . Hence, Eq. (5.123a) can also be written as

$$-\frac{eE}{m_e} \frac{\partial f_0}{\partial \varepsilon} \frac{\partial \varepsilon}{\partial v_x} + v_x \frac{\partial f_0}{\partial T} \frac{dT}{dx} + v_z \frac{\partial f_1}{\partial z} = -\frac{f_1 - f_0}{\tau(\varepsilon)} \quad (5.123b)$$

which is nothing but the steady-state BTE under the relaxation time approximation. The general solution can be expressed as

$$f_1 = f_0 + \tau v_x \left( eE \frac{\partial f_0}{\partial \varepsilon} - \frac{\partial f_0}{\partial T} \frac{dT}{dx} \right) \left[ 1 - \Psi(\mathbf{v}) \exp \left( -\frac{z}{\tau v_z} \right) \right], \quad v_z > 0 \quad (5.124a)$$

$$\text{and } f_1 = f_0 + \tau v_x \left( eE \frac{\partial f_0}{\partial \varepsilon} - \frac{\partial f_0}{\partial T} \frac{dT}{dx} \right) \left[ 1 - \Psi(\mathbf{v}) \exp \left( -\frac{d-z}{\tau v_z} \right) \right], \quad v_z < 0 \quad (5.124b)$$

where  $\Psi(\mathbf{v})$  is an arbitrary function that accounts for the accommodation and scattering characteristics. If perfect accommodation is assumed with inelastic and diffuse scattering, then  $\Psi(\mathbf{v}) = 1$ . Let us consider electrical conduction without any temperature gradient. For diffuse scattering, only with  $\Psi(\mathbf{v}) = 1$ , it can be shown that

$$f_1 = f_0 + \tau v_x eE \frac{\partial f_0}{\partial \varepsilon} \left[ 1 - \exp \left( -\frac{z}{\tau v_z} \right) \right], \quad v_z > 0 \quad (5.125a)$$

$$\text{and } f_1 = f_0 + \tau v_x eE \frac{\partial f_0}{\partial \varepsilon} \left[ 1 - \exp \left( -\frac{d-z}{\tau v_z} \right) \right], \quad v_z < 0 \quad (5.125b)$$

We must substitute the distribution function to Eq. (5.61a) and integrate over  $(v_x, v_y, v_z)$ , or over  $(v, \theta, \phi)$  or  $(\varepsilon, \theta, \phi)$  in the spherical coordinate, to obtain  $J_e(z) = -eJ_N(z)$  along the film. Therefore,

$$\begin{aligned}
 J_e(z) = & -e^2 E \int_0^\infty \tau \frac{\partial f_{FD}}{\partial \varepsilon} d\varepsilon \int_0^{2\pi} d\phi \left\{ \int_0^{\pi/2} v_x^2 \left[ 1 - \exp\left(-\frac{z}{\tau v \cos \theta}\right) \right] v^2 \sin \theta d\theta \right. \\
 & \left. + \int_{\pi/2}^\pi v_x^2 \left[ 1 - \exp\left(-\frac{d-z}{\tau v \cos \theta}\right) \right] v^2 \sin \theta d\theta \right\}
 \end{aligned}
 \tag{5.126}$$

Putting  $v_x = v \sin \theta \cos \phi$ , the integration over  $\phi$  can be carried out independently. The average current flux  $\bar{J}_e = (1/d) \int_0^d J_e(z) dz$  can also be obtained. The properties of the Fermi integral allow the integration over  $\varepsilon$  to be carried out and expressed in terms of the properties at the Fermi surface, i.e.,  $\tau(\mu_F)$  and  $v_F$ . Notice that  $\Lambda_b = \tau(\mu_F)v_F$ , and let  $\bar{J}_e = \sigma_f E$ , where  $\sigma_f$  is the effective electrical conductivity of the film. After normalization of the electrical current density based on Eq. (5.61a) and Eq. (5.62), we obtain the following relation:

$$\frac{\sigma_f}{\sigma_b} = F(Kn)
 \tag{5.127}$$

where

$$\begin{aligned}
 F(Kn) = & \frac{3}{4d} \int_0^{\pi/2} \sin^3 \theta \int_0^d \left[ 1 - \exp\left(-\frac{z}{\Lambda_b \cos \theta}\right) \right] dz d\theta \\
 & + \frac{3}{4d} \int_{\pi/2}^\pi \sin^3 \theta \int_0^d \left[ 1 - \exp\left(-\frac{d-z}{\Lambda_b \cos \theta}\right) \right] dz d\theta \\
 = & \frac{3}{2d} \int_0^{\pi/2} \sin^3 \theta \left\{ d - \Lambda_b \cos \theta \left[ 1 - \exp\left(-\frac{d}{\Lambda_b \cos \theta}\right) \right] \right\} d\theta \\
 = & 1 - \frac{3Kn}{8} + \frac{3Kn}{2} \int_1^\infty \left( \frac{1}{t^3} - \frac{1}{t^5} \right) \exp\left(-\frac{t}{Kn}\right) dt
 \end{aligned}
 \tag{5.127a}$$

Note that  $t = 1/\cos \theta$  in the substitution, and the  $m$ th exponential integral is defined as  $E_m(x) = \int_1^\infty e^{-xt} t^{-m} dt$  or  $E_m(x) = \int_0^1 \mu^{m-2} e^{-x/\mu} d\mu$ , which has the relation  $E_{m+1}(x) = e^{-x}/m - (x/m)E_m(x)$ . Equation (5.127a) can also be expressed as

$$F(Kn) = 1 - \frac{3Kn}{8} + \frac{3Kn}{2} \left[ E_3\left(\frac{1}{Kn}\right) - E_5\left(\frac{1}{Kn}\right) \right]
 \tag{5.127b}$$

The asymptotic relations are

$$\frac{\sigma_f}{\sigma_b} \approx 1 - \frac{3Kn}{8} \quad \text{for } Kn \ll 1
 \tag{5.128a}$$

and

$$\frac{\sigma_f}{\sigma_b} \approx \frac{3 \ln(Kn)}{4Kn} \quad \text{for } Kn \gg 1
 \tag{5.128b}$$

which is close to Eq. (5.120) for  $Kn \gg 1$ . The derivation using the BTE presented earlier inherently assumed that the electrons are originated from the film rather than from the

boundaries. Kumar and Vradis performed an extensive comparison between different expressions and relied on a different method to derive the size effect on conductivities.<sup>43</sup>

For thermal conductivity, we can substitute Eq. (5.124) with  $\Psi(\mathbf{v}) = 1$  into Eq. (5.65a) and follow the similar procedure to obtain  $\kappa_t/\kappa_b = F(Kn)$ , where  $F(Kn)$  is given in Eq. (5.127a) or (5.127b). At very low temperatures or near room temperature, the Wiedemann-Franz law is applicable and the reduction in electrical and thermal conductivities are essentially the same. In the intermediate region, one could use different scattering rates or mean free paths for the bulk thermal and electrical conductivities to determine the size effect individually based on Eq. (5.127).

According to the discussion of thermoelectricity in Sec. 5.4, we could in principle quantify the size effect on other coefficients. If the same assumptions are used, to the first-order approximation,  $L_{12}$  and  $L_{21}$  are subject to boundary scattering and will also be reduced according to Eq. (5.127). Because the thermoelectric power is the ratio of the two coefficients, the Seebeck coefficient along the film should be expected to remain the same regardless of boundary scattering. One should be cautious about this conclusion because the assumption of a spherical Fermi surface and the free-electron model are questionable when modeling the thermoelectricity as mentioned previously. The above discussion can be extended to scattering with a specular component. Let the parameter  $p$ , which is called *specularity*, represent the probability of scattering being elastic and specular. For specular and elastic scattering, the carriers will continue to exchange energy and momentum inside the film after the reflection by the boundary. Therefore, these scattering events do not cause any reduction in the effective mean free path or conductivities along the film. If  $p$  is assumed to be independent of the incident direction, the function  $\Psi(\mathbf{v})$  in Eq. (5.124) becomes

$$\Psi(\mathbf{v}) = \frac{1 - p}{1 - p \exp(-d/\tau v_z)} \quad (5.129)$$

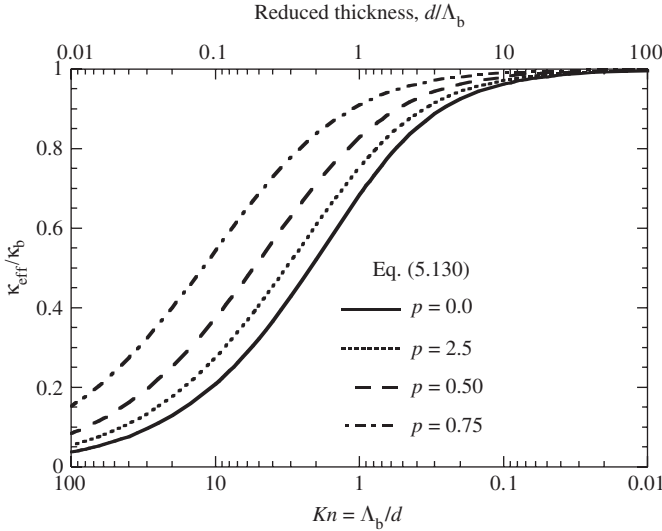
The function given in Eq. (5.127a) may be modified after some tedious derivations as follows:

$$F(Kn, p) = 1 - \frac{3(1 - p)Kn}{2} \int_1^\infty \left( \frac{1}{t^3} - \frac{1}{t^5} \right) \frac{1 - \exp(-t/Kn)}{1 - p \exp(-t/Kn)} dt \quad (5.130)$$

The effects of  $p$  and  $Kn$  on the effective conductivity are shown in Fig. 5.18. The trends with respect to  $Kn$  are very similar to those in Fig. 5.17 obtained from simple geometric considerations. For electronic transport, since the wavelength of the electron is less than 1 nm, usually the boundary scattering can be considered as diffuse, i.e.,  $p = 0$ . For phonons, the wavelength may vary from the atomistic scale up to the size of the crystal. Therefore, the size effect needs to be considered for different phonon frequencies. The parameter  $p$  can be estimated based on the rms surface roughness  $\sigma_{\text{rms}}$  and wavelength  $\lambda$  of the carrier by

$$p = \frac{1}{1 + \delta^2} \quad (5.131)$$

where  $\delta = 4\pi\sigma_{\text{rms}}/\lambda$  at normal incidence. This equation can be derived from the wave scattering theory as detailed by Elson (*Appl. Opt.*, **22**, 3207, 1982). Generally speaking,  $p \ll 1$  when  $\lambda \leq \sigma_{\text{rms}}$ . When  $\lambda > 10\sigma_{\text{rms}}$ , the specular reflection cannot be neglected. In reality, the specularity  $p$  depends on the angle of incidence and tends to increase for carriers with large incidence polar angle  $\theta_i$ , since  $\delta$  should be multiplied by  $\cos \theta_i$ . Furthermore, the actual scattering distribution often consists of a broad specular lobe, and the nonspecular component is not perfectly diffuse. This is similar to light scattering by rough surfaces for which an in-depth discussion will be given in Chap. 9.



**FIGURE 5.18** Size effect on thermal conductivity along a film of thickness  $d$ , as predicted by the BTE with different specularities.

As can be seen from Figs. 5.17 and 5.18, when  $Kn = \Lambda_b/d > 0.1$ , i.e., when  $d < 10\Lambda_b$ , the size effect may be significant, and boundary scattering dominates when  $d < 0.1\Lambda_b$ . Note that Examples 5-5 and 5-6 provide typical numerical values of the bulk mean free paths of electrons in a noble metal and of phonons in silicon. At room temperature, the electron mean free path of a metal is on the order of tens of nanometers; one would expect some size effect when  $d$  is less than 300 nm. However, for a highly pure metal at very low temperatures, the electron mean free path could be on the order of millimeters. In thin films when  $d$  is on the order of micrometers, boundary scattering would dominate the scattering process. For semiconductors, such as silicon, the phonon mean free path is also on the order of tens of nanometers at room temperature. Therefore, the size effect can be neglected for a 1- $\mu\text{m}$ -thick silicon film above room temperature. As temperature is lowered, size effect becomes more and more significant. Numerical calculations dealing with the conductivity reduction are left as exercises.

The above discussion can be extended to conduction along a thin wire. For wires with circular cross sections, the effective conductivity can be expressed as<sup>41,42</sup>

$$\frac{\kappa_w}{\kappa_b} \text{ or } \frac{\sigma_w}{\sigma_b} = 1 - \frac{12}{\pi} \int_0^1 \sqrt{1 - \xi^2} \int_1^\infty \exp\left(-\frac{\xi t}{Kn}\right) \frac{\sqrt{t^2 - 1}}{t^4} dt d\xi \quad (5.132)$$

In particular, the asymptotic approximations with about 1% accuracy are

$$\frac{\kappa_w}{\kappa_b} \text{ or } \frac{\sigma_w}{\sigma_b} \approx 1 - \frac{3}{4}Kn + \frac{3}{8}Kn^3 \text{ for } Kn < 0.6 \quad (5.133a)$$

and 
$$\frac{\kappa_w}{\kappa_b} \text{ or } \frac{\sigma_w}{\sigma_b} \approx \frac{1}{Kn} - \frac{3(\ln Kn + 1)}{8Kn^2} - \frac{2}{15Kn^3} \text{ for } Kn > 1 \quad (5.133b)$$

If the scattering is not completely diffuse, by introducing a specularity parameter  $p$  similar to that for thin films, the expression becomes

$$\frac{\kappa_w}{\kappa_b} \text{ or } \frac{\sigma_w}{\sigma_b} = 1 - \frac{12(1-p)^2}{\pi} \sum_{m=1}^{\infty} m p^{m-1} G(Kn, m) \quad (5.134)$$

where

$$G(Kn, m) = \int_0^1 \sqrt{1-\xi^2} \int_1^{\infty} \exp\left(-\frac{m\xi t}{Kn}\right) \frac{\sqrt{t^2-1}}{t^4} dt d\xi$$

Again, different mean free paths and  $Kn$  numbers should be used for thermal and electrical conductivities in the region where the Wiedemann-Franz law is not applicable.

For phonons, the distribution function depends on the frequency or the wavevector, which can be expressed in the spherical coordinate. The group velocity depends on the dispersion relation for a given phonon mode or branch. The scattering rate is also frequency dependent. Nevertheless, the BTE under the relaxation time approximation can be expressed as follows for a given frequency:

$$v_x \frac{\partial f_{BE}}{\partial T} \frac{dT}{dx} + v_z \frac{\partial f_1}{\partial z} = -\frac{f_1 - f_{BE}}{\tau(\omega)} \quad (5.135)$$

where  $v_x$  and  $v_z$  are the components of the group velocity that depend on the frequency. The solution is similar to Eq. (5.124a) and Eq. (5.124b), especially for the  $z$  dependence. Following the discussions in Sec. 5.3.4 on phonon thermal conductivity, in conjunction with the average heat flux along the film, we can rewrite Eq. (5.73) as follows:

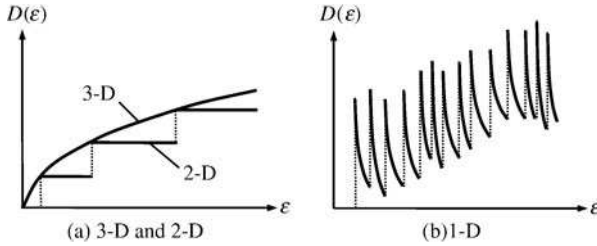
$$\kappa_f = \frac{k_B^4 T^3}{6\pi^2 \hbar^3} \sum_P \int_0^{x_m} \frac{\tau(x) v_g(x)}{v_p^2(x)} \frac{x^4 e^x}{(e^x - 1)^2} F(Kn_x, p) dx \quad (5.136)$$

where  $x = \hbar\omega/k_B T$  is a reduced frequency and  $Kn_x = \tau(x) v_g(x)/d$  is a function of  $x$  or  $\omega$ . In this equation, the summation index  $P$  accounts for all phonon polarizations, the integration is within the cutoff frequencies, and the function  $F(\xi, p)$  can be calculated from Eq. (5.130). If an average  $Kn$  that is independent of the frequency can be used, Eq. (5.136) can be simplified as  $\kappa_f/\kappa_b \approx F(Kn, p)$ . A similar equation can be developed for thin wires. Size-dependent thermal conductivities for single-crystal silicon films have been experimentally observed for film thicknesses from a few micrometers down to 20 nm.<sup>44</sup>

In the earlier discussion, the Fourier result was assumed to hold under the local-equilibrium approximation, with reduced thermal conductivities to include the effect of boundary scattering. More recent works have employed equilibrium and nonequilibrium molecular dynamics to study thermal transport at nanoscales.<sup>45,46</sup> In heterogeneous structures, such as superlattices, the thermal transport is across the multiple layers, and the local-equilibrium assumption breaks down in the ballistic regime. Further discussion of non-Fourier conduction especially for transient processes will be deferred to Chap. 7. For superlattice nanowires, both lateral confinement and longitudinal confinement exist. Each element is like a quantum dot that is confined in all three dimensions. When the quantum confinement becomes significant, the relaxation time approximation used to solve the BTE is not applicable. Next, we will introduce the quantum size effect on electrical and thermal transport processes, with an emphasis on the concept of quantum conductance and its implications.

### 5.5.3 Quantum Conductance

Quantum size effect on the lattice specific heat was discussed in Sec. 5.2. Here, attention is paid to the electrical conductance of metallic materials and thermal conductance of dielectric



**FIGURE 5.19** Electron density of states due to quantum confinement. (a) 2-D quantum wells versus 3-D bulk solids. (b) 1-D quantum wires.

materials. For bulk solids, the density of states for electrons  $D(\varepsilon)$  is proportional to  $\sqrt{\varepsilon}$ , as given in Eq. (5.18) and illustrated in Fig. 5.19a. Note that for phonons or photons, the energy  $\varepsilon = \hbar\omega$  is proportional to the frequency and the density of states  $D(\omega)$  is proportional to  $\omega^2$  when the dispersion is linear [see Eq. (5.35) and Fig. 5.4]. For electrons or holes,  $\varepsilon = p^2/2m^* = \hbar^2k^2/2m^*$ , where  $k$  is the wavevector and  $m^*$  is an effective mass. For electron gas in a 2-D solid, the density of states  $D(\varepsilon) = 2(k/2\pi)(dk/d\varepsilon) = m^*/\pi\hbar^2$ , which can be derived from Eq. (5.37) considering the spin degeneracy. In a quantum well of thickness  $L$ , the energy levels are quantized in the normal or  $z$  direction according to Eq. (3.80), i.e.,  $n^2\hbar^2/(8m^*L^2)$ , where  $n$  is a positive integer. The combined energies can be expressed as

$$\varepsilon_n(k) = \frac{n^2\hbar^2}{2m^*L} + \frac{\hbar^2(k_x^2 + k_y^2)}{2m^*} \quad (5.137)$$

and the resulting density of states is given by

$$D(\varepsilon) = \frac{nm^*}{\pi\hbar^2}, \quad \text{for } \varepsilon_n \leq \varepsilon < \varepsilon_{n+1} \text{ and } n = 1, 2, \dots \quad (5.138)$$

which is a staircase function as depicted in Fig. 5.19a. The reason that the density of states for the  $n$ th subband is multiplied by  $n$  is because  $k_{z,n} = n\pi/L$ , where  $k_{\min} = \pi/L$ . Before applying Eq. (5.36) and Eq. (5.37), we must multiply the total number of modes  $N$  by  $k_{z,n}/(\pi/L)$ . For 1-D quantum wires confined in both  $y$  and  $z$  directions (assuming a rectangular shape of  $L_y \times L_z$ ), the energy levels are given by

$$\varepsilon_{l,n} = \frac{l^2\hbar^2}{2m^*L_y} + \frac{n^2\hbar^2}{2m^*L_z} \quad (5.139)$$

For each subband ( $l, n$ ), the density of states becomes

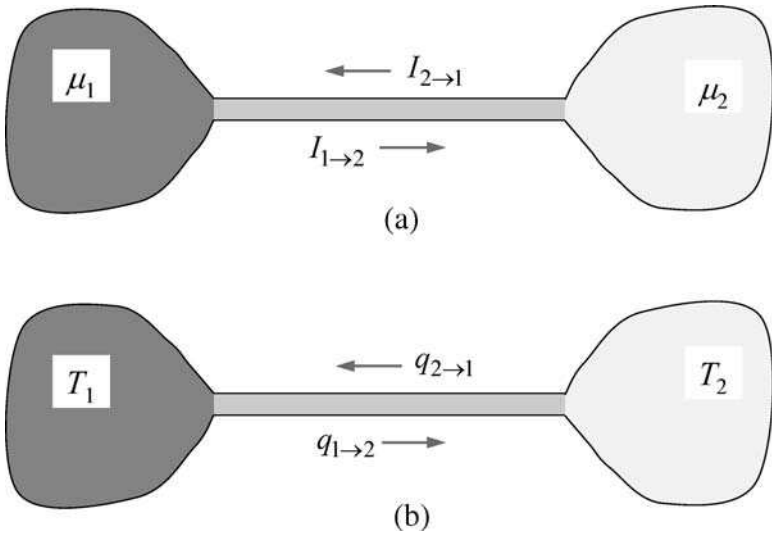
$$D(\varepsilon) = \frac{nl}{\pi\hbar} \sqrt{\frac{2m^*}{\varepsilon - \varepsilon_{l,n}}} \quad (5.140)$$

which has an inverse square-root dependence upon energy and a singularity at  $\varepsilon_{l,n}$ , as shown in Fig. 5.19b. For 3-D confined quantum dots, the energy levels are completely discrete; subsequently, the density of states becomes isolated delta functions.

The quantization of energy levels of electrons or frequencies for phonons in small structures suggests that the resulting transport properties may also be quantized. For example, the electrical conductance may depend on the applied current or force for the nanocontact in a stepwise manner. The thermal conductance of insulators can also be quantized due to limited available phonon modes in small structures and at low temperatures. In this section,



we use conductance rather than conductivity for reasons to be explained soon. Long before the quantization of conductance was experimentally observed, physicists have formulated different theories to understand the transport phenomena in the quantum or ballistic regimes. Landauer and collaborators have developed a formula to treat electrical current flow as transmission probability when carriers are scattered coherently and the resulting ballistic transport behaves quantum mechanically.<sup>47</sup> Landauer's formulation can be easily applied to the 1-D case for conductance through a narrow channel, as illustrated in Fig. 5.20a.



**FIGURE 5.20** Illustration of quantum conductance. (a) Electrical current flow through a narrow metallic channel due to different electrochemical potentials. (b) Heat transfer between two heat reservoirs through a narrow dielectric channel.

Suppose ballistic transmission exists in the channel connecting two reservoirs of different electrochemical potentials; there will be a current flow from 1 to 2 and reversely from 2 to 1. In the absence of losses by scattering and reflection, the net current flow can be expressed as

$$J_e = J_{1 \rightarrow 2} - J_{2 \rightarrow 1} = -e v_F (\mu_1 - \mu_2) D(\epsilon) \quad (5.141)$$

where  $\mu$  is the chemical potential. The derivation can be easily generalized to include the electrostatic potential. Note that the density of states in the 1-D case is  $D(\epsilon) = (\pi \hbar v_F)^{-1}$ , considering the electronic spin degeneracy. Because the voltage drop  $V_1 - V_2 = -(\mu_1 - \mu_2)/e$ , the electrical conductance for complete transmission becomes

$$g_{e0} = \frac{J_e}{V_1 - V_2} = \frac{e^2}{\pi \hbar} \quad \text{or} \quad \frac{2e^2}{h} \quad (5.142)$$

which gives a universal constant with a value of  $7.75 \times 10^{-5} \Omega^{-1}$  or a resistance value of 12.91 k $\Omega$ . This is the quantum conductance for an ideal 1-D conductor, in which there is no resistance or voltage drop associated with the channel itself. The voltage drops are associated with the perturbation at each end of the channel as it interacts with the reservoir.<sup>47</sup> In the above derivation, we also assumed that the Fermi function or the distribution function

can be approximated as a step function (i.e., at absolute zero temperature). By introducing a transmission coefficient  $\xi_{12}$  and the actual distribution function, Eq. (5.141) can be modified as follows:<sup>47,48</sup>

$$J_e = \int_0^\infty (-ev_F)\xi_{12}(\varepsilon)[f_{FD}(\varepsilon,\mu_1) - f_{FD}(\varepsilon,\mu_2)]D(\varepsilon) d\varepsilon \quad (5.143)$$

For small potential differences, using the following approximation:

$$\frac{f_{FD}(\varepsilon,\mu_1) - f_{FD}(\varepsilon,\mu_2)}{\mu_2 - \mu_1} = -\frac{\partial f_{FD}(\varepsilon,\mu)}{\partial \mu} = \frac{\partial f_{FD}(\varepsilon,\mu)}{\partial \varepsilon}$$

we obtain the expression of the electrical conductance:

$$g_e = -\frac{2e^2}{h} \int_0^\infty \xi_{12}(\varepsilon) \frac{\partial f_{FD}}{\partial \varepsilon} d\varepsilon \quad (5.144a)$$

which reduces to Eq. (5.142), at absolute zero temperature, when  $\xi_{12}(0)$  is taken to 1. The transmission coefficient is given by a scattering matrix (the S-matrix) based on the solution of Schrödinger's equation. The solution is in the form of eigenvalues called eigenchannels, each with a transmission coefficient  $\tau_i$  between 0 and 1. Thus, the expression of conductance is reduced to

$$g_e = \frac{2e^2}{h} \sum_i \tau_i \quad (5.144b)$$

Depending on how many propagation modes at the Fermi level are excited, the conductance varies in a discontinuous manner. Conductance quantization has been realized in metallic nanocontacts, nanowires, and carbon nanotubes,<sup>48–51</sup> even at room temperature, and has also been predicted by molecular dynamics simulations.<sup>49</sup> These discoveries are very important for the development of single-electron transistors, nanoelectromechanical systems, nanotribology, and quantum computing.

The ballistic thermal transport process resembles electromagnetic radiation between two blackbodies separated by a vacuum. For a 1-D photon gas, the Stefan-Boltzmann law reads  $q'' \propto T^2$  rather than  $q'' \propto T^4$ . In a solid nanostructure (channel) that links two heat reservoirs, as illustrated in Fig. 5.20b, the ballistic heat conduction can be treated in a similar way so that

$$q_{1 \rightarrow 2} = \frac{1}{2\pi} \sum_P \int_{\omega_p}^{\omega_D} \xi_P(\omega) \hbar \omega f_{BE}(\omega, T_1) d\omega \quad (5.145a)$$

$$q_{2 \rightarrow 1} = \frac{1}{2\pi} \sum_P \int_{\omega_p}^{\omega_D} \xi_P(\omega) \hbar \omega f_{BE}(\omega, T_2) d\omega \quad (5.145b)$$

and

where  $\xi_P(\omega)$  is the transmission coefficient of the polarization branch  $P$ , and it accounts for both scattering in the channel and reflection from the junctions. Here, the upper bound  $\omega_D$  approaches infinity at very low temperatures, and the lower bound is a cutoff frequency for the phonon mode  $P$ . This cutoff frequency is determined by the width of the channel and the order of the propagating phonon modes, like in a waveguide to be discussed in Chap. 10.

More specifically, if a rectangular cross section is considered whose dimensions are  $L_x$  and  $L_y$ , the cutoff frequency for the  $(m,n)$  mode is given by

$$k_{mn} = \frac{\omega_{mn}}{v_a} = \sqrt{\left(\frac{m\pi}{L_x}\right)^2 + \left(\frac{n\pi}{L_y}\right)^2} \quad (5.146)$$

Apparently, a narrow channel enables a large-cutoff wavenumber. Note that the zeroth-order mode always exists because it has a zero cutoff frequency. If the integration in Eq. (5.145) is expressed in terms of wavevector, there will be a group velocity  $v_g$  term. In writing Eq. (5.145), we have assumed that  $v_g = v_p$  or a linear dispersion relation. The net heat transfer is calculated by  $q_{12} = q_{1 \rightarrow 2} - q_{2 \rightarrow 1}$ , which is commonly done in radiation heat transfer. Assuming that the temperature difference is small, we obtain the thermal conductance as

$$g_T = \frac{q_{12}}{T_1 - T_2} = \frac{1}{2\pi} \sum_p \int_{\omega_p}^{\omega_D} \xi_p(\omega) \hbar \omega \frac{\partial f_{BE}(\omega, T)}{\partial T} d\omega \quad (5.147a)$$

or

$$g_T = \frac{k_B^2 \bar{T}}{h} \sum_p \int_{x_p}^{x_D} \xi_p(x) \frac{x^2 e^x}{(e^x - 1)^2} dx \quad (5.147b)$$

Note that  $\bar{T}$  represents the average temperature. At sufficiently low temperatures, only the lowest phonon branches with zero cutoff frequency may contribute to the conductance. If the transmission coefficient is assumed to be unity, each of the lowest phonon modes will contribute to the thermal conductance by

$$g_{T0} = \frac{\pi k_B^2 T}{6\hbar} \quad \text{or} \quad \frac{\pi^2 k_B^2 T}{3h} \quad (5.148)$$

which has a value  $g_{T0}/T = 0.947 \text{ pW/K}^2$  and is another universal constant that can be viewed as the Stefan-Boltzmann constant in a 1-D space for each mode. If the preceding derivation is repeated to obtain electron thermal conductance, we will end up with  $2g_{T0}$  due to the electronic spin degeneracy. Therefore, the Lorentz number  $Lz = \kappa/\sigma T = g_T/g_e T$  in the ballistic regime remains the same as given in Eq. (5.56) for the diffusive regime.<sup>52</sup> Roukes Group has demonstrated experimentally quantum thermal conductance using a 60-nm-thick silicon nitride membrane.<sup>53</sup> They reported a  $16g_{T0}$  behavior at temperatures below 0.6 K since the structure was suspended by four narrow bridges (channels). Each bridge or channel acts like a wire with four phonon modes (two transversal, one longitudinal, and one torsional).

Carbon nanotubes, with very large thermal conductivities, have been known for a while.<sup>54–58</sup> Single-walled carbon nanotubes can be made essentially free from defect scattering and boundary scattering due to atomistic smoothness. The diameter can be made as small as a few nanometers with a length of several micrometers. Thermal conductivities of single-walled and multi-walled nanotubes have been measured with suspended MEMS bridges and are found to exceed that of diamond at room temperature.<sup>57</sup> The thermal conductivity can be calculated from the measured thermal conductance based on an effective cross-sectional area. Above the room temperature, phonon-phonon anharmonic interactions may provide a mean for diffusive-conduction behavior. Nanotube bundles on the other hand are subject to various scattering mechanisms and thus possess a lower thermal conductivity. Furthermore, the contact may be attributed to the reduction in conductance. The contact and interface scattering needs to be further addressed in order to realize the potential of nanotubes for use in heat transfer enhancement. Mingo and Broido calculated the thermal conductance of carbon nanotubes in the ballistic limit.<sup>58</sup> For semiconductor nanotubes at sufficiently low temperatures, the thermal conductance becomes  $16g_{T0}T$  due to the four lowest phonon modes regardless of the length and the cross-sectional area. The

thermal conductivity of carbon nanotubes depends on the length and the cross-sectional area, and will increase with temperature. On the other hand, as the length or temperature increases, scattering becomes important and the conductance reduces. In the diffusion limit, the conductivity is independent of the length and diminishes as temperature further increases. For nanotubes whose band structures are metal-like, such as with (6,0) and (18,0) chiral numbers, electronic ballistic transport may be important; however, electron-phonon scattering will dominate at high enough temperatures.

## 5.6 SUMMARY

---

This chapter began with lattice vibrations (i.e., phonons) in solids and discussed the dimensionality and the quantum size effect on the lattice specific heat. The free-electron theory was applied, assuming a spherical Fermi surface, to predict the electronic specific heat, as well as electrical and thermal conductivities of solids. The Boltzmann transport equation, under the relaxation time approximation and the local-equilibrium assumption, was used to derive the conductivities and thermoelectric coefficients under the framework of irreversible thermodynamics. A brief discussion of the efficiency of thermoelectric power and refrigeration systems was then provided. The classical size effect on electrical and thermal conductivities was presented, followed by the introduction to the concept of conductance quantization for both electrical current and heat flow. The properties were discussed with examples of representative materials, such as noble metals, semiconductors, quantum wells, superlattices, nanowires, and carbon nanotubes. The band theory for electrons and phonons will be introduced in the next chapter as an advanced topic of the transport theory of solids.

## REFERENCES

---

1. M. I. Flik, B. I. Choi, and K. E. Goodson, "Heat transfer regimes in microstructures," *J. Heat Transfer*, **114**, 666–674, 1992.
2. C. L. Tien and G. Chen, "Challenges in microscale conductive and radiative heat transfer," *J. Heat Transfer*, **116**, 799–807, 1994.
3. D. G. Cahill, K. Goodson, and A. Majumdar, "Thermometry and thermal transport in micro/nanoscale solid-state devices and structures," *J. Heat Transfer*, **124**, 223–241, 2002.
4. C. Kittel, *Introduction to Solid State Physics*, 7th ed., Wiley, New York, 1996.
5. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Harcourt College Publishers, Fort Worth, TX, 1976.
6. Y. S. Touloukian and E. H. Buyco (eds.), *Thermophysical Properties of Matter*, Vol. 4: Specific Heat—Metallic Elements and Alloys; Vol. 5: Specific Heat—Nonmetallic Solids, IFI/Plenum, New York, 1970.
7. G. Nilsson and S. Rolandson, "Lattice dynamics of copper at 80 K," *Phys. Rev. B*, **7**, 2393–2400, 1973.
8. A. J. E. Foreman, "Anharmonic specific heat of solids," *Proc. Phys. Soc. (London)*, **79**, 1124–1141, 1962.
9. R. A. MacDonald and W. M. MacDonald, "Thermodynamic properties of fcc metals at high temperatures," *Phys. Rev. B*, **24**, 1715–1724, 1981.
10. V. Novotny, P. P. M. Meincke, and J. H. P. Watson, "Effect of size and surface on the specific heat of small lead particles," *Phys. Rev. Lett.*, **28**, 901–903, 1972; V. Novotny and P. P. M. Meincke, "Thermodynamic lattice and electric properties of small particles," *Phys. Rev. B*, **8**, 4186–4199, 1973.

11. W. Yi, L. Lu, D.-L. Zhang, Z. W. Pan, and S. S. Xie, "Linear specific heat of carbon nanotubes," *Phys. Rev. B*, **59**, R9015–R9018, 1999.
12. C. Dames, B. Poudel, W. Z. Wang et al., "Low-dimensional phonon specific heat of titanium dioxide nanotubes," *Appl. Phys. Lett.*, **87**, 031901/1–3, 2005.
13. A. A. Valladares, "The Debye specific heat in  $n$  dimensions," *Am. J. Phys.*, **43**, 308–311, 1975.
14. W. DeSorbo and W. W. Tyler, "The specific heat of graphite from 13 to 300 K," *J. Chem. Phys.*, **21**, 1660–1663, 1953.
15. R. S. Prasher and P. E. Phelan, "Size effect on the thermodynamic properties of thin solid films," *J. Heat Transfer*, **120**, 1078–1081, 1998; R. S. Prasher and P. E. Phelan, "Non-dimensional size effects on the thermodynamic properties of solids," *Int. J. Heat Mass Transfer*, **42**, 1991–2001, 1999.
16. R. Nicklow, N. Wakabayashi, and H. G. Smith, "Lattice dynamics of pyrolytic graphite," *Phys. Rev. B*, **5**, 4951–4962, 1972.
17. B. S. Tosic, J. P., Setrajic, D. Lj. Mirjanic, and Z. V. Bundalo, "Low-temperature properties of thin films," *Physica A*, **184**, 354–366, 1992.
18. H. P. Baltes and E. R. Hilf, "Specific heat of lead grains," *Solid State Commun.*, **12**, 369–373, 1973; Th. F. Nonnenmacher, "Quantum size effect on the specific heat of small particles," *Phys. Lett.*, **51A**, 213–214, 1975; R. Lautenschlager, "Improved theory of the vibrational specific heat of lead grains," *Solid State Commun.*, **16**, 1331–1334, 1975.
19. M. S. Dresselhaus and P. C. Eklund, "Phonons in carbon nanotubes," *Adv. Phys.*, **49**, 705–814, 2000.
20. J. Hone, B. Batlogg, Z. Benes, A. T. Johnson, and J. E. Fischer, "Quantized phonon spectrum of single-wall carbon nanotubes," *Science*, **289**, 1730–1733, 2000; W. A. de Heer, "A question of dimensions," *Science*, **289**, 1702–1703, 2000.
21. R. Denton, B. Muhlschlegel, and D. J. Scalapino, "Thermodynamic properties of electrons in small metal particles," *Phys. Rev. B*, **7**, 3589–3607, 1973.
22. W. P. Halperin, "Quantum size effects in metal particles," *Rev. Mod. Phys.*, **58**, 533–606, 1986.
23. J. M. Ziman, *Electrons and Phonons*, Oxford University Press, Oxford, UK, 1960; also in the Oxford Classics Series, 2001.
24. R. A. Matula, "Electrical resistivity of copper, gold, palladium, and silver," *J. Phys. Chem. Ref. Data*, **8**, 1147–1298, 1979.
25. Y. S. Touloukian, R. W. Powell, C. Y. Ho, and P. G. Klemens (eds.), *Thermophysical Properties of Matter*, Vol. 1: Thermal Conductivity—Metallic Elements and Alloys; Vol. 2: Thermal Conductivity—Nonmetallic Solids, IFI/Plenum, New York, 1970.
26. R. E. Bentley, *Theory and Practice of Thermoelectric Thermometry*, Springer-Verlag, Singapore, 1998.
27. R. B. Roberts, "The absolute scale of thermoelectricity II," *Phil. Mag. B*, **43**, 1125–1135, 1981.
28. O. Dreirach, "The electrical resistivity and thermopower of solid noble metals," *J. Phys. F: Met. Phys.*, **3**, 577–584, 1973.
29. S. L. Soo, *Direct Energy Conversion*, Prentice-Hall, Englewood Cliffs, NJ, 1968.
30. L. D. Hicks and M. S. Dresselhaus, "Effect of quantum-well structures on the thermoelectric figure of merit," *Phys. Rev. B*, **47**, 12727–12731, 1993; M. S. Dresselhaus, G. Dresselhaus, X. Sun et al., "The promise of low-dimensional thermoelectric materials," *Microscale Thermophys. Eng.*, **3**, 89–100, 1999; T. Koga, O. Rabin, and M. S. Dresselhaus, "Thermoelectric figure of merit of Bi/Pb<sub>1-x</sub>Eu<sub>x</sub>Te superlattices," *Phys. Rev. B*, **62**, 16703–16706, 2000; M. S. Dresselhaus, Y.-M. Lin, O. Rabin, and G. Dresselhaus, "Bismuth nanowires for thermoelectric applications," *Microscale Thermophys. Eng.*, **7**, 207–219, 2003; Y.-M. Lin and M. S. Dresselhaus, "Thermoelectric properties of superlattice nanowires," *Phys. Rev. B*, **68**, 075304, 2003.
31. R. Venkatasubramanian, E. Siivola, T. Colpitts, and B. O'Quinn, "Thin-film thermoelectric devices with high room-temperature figures of merit," *Nature*, **413**, 597–602, 2001.
32. J. P. Heremans, C. M. Thrush, D. T. Morelli, and M.-C. Wu, "Thermoelectric power of bismuth nanocomposites," *Phys. Rev. Lett.*, **88**, 216801, 2002; J. P. Heremans, C. M. Thrush, and D. T. Morelli, "Thermopower enhancement in lead telluride nanostructures," *Phys. Rev. B*, **70**, 115334, 2004.
33. D. Bile, S. D. Mahanti, E. Quarez, K.-F. Hsu, R. Pcionek, and M. G. Kanatzidis, "Resonant states in the electronic structure of the high performance thermoelectrics AgPb<sub>m</sub>SbTe<sub>2+m</sub>: The role of

- Ag-Sb microstructures," *Phys. Rev. Lett.*, **93**, 146403, 2004; K.-F. Hsu, S. Loo, F. Guo et al., "Cubic AgPb<sub>m</sub>SbTe<sub>2+m</sub>: Bulk thermoelectric materials with high figure of merit," *Science*, **303**, 818–821, 2004.
34. G. Chen, "Size and interface effects on thermal conductivity of superlattices and periodic thin-film structures," *J. Heat Transfer*, **119**, 220–229, 1997; G. Chen, "Phonon wave effects on heat conduction in thin films," *J. Heat Transfer*, **121**, 945–953, 1999; G. Chen and A. Shakouri, "Heat transfer in nanostructures for solid-state energy conversion," *J. Heat Transfer*, **124**, 242–252, 2002; T. Zeng and G. Chen, "Interplay between thermoelectric and thermionic effects in heterostructures," *J. Appl. Phys.*, **92**, 3152–3161, 2002; T. Zeng and G. Chen, "Nonequilibrium electron and phonon transport and energy conversion in heterostructures," *Microelectronics Journal*, **34**, 201–206, 2003; R. G. Yang and G. Chen, "Thermal conductivity modeling of periodic two-dimensional nanocomposites," *Phys. Rev. B*, **69**, 195316, 2004.
  35. L. Onsager, "Reciprocal relations in irreversible processes. I & II," *Phys. Rev.*, **37**, 405–426; **38**, 2265–2279, 1931.
  36. H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed., Wiley, New York, 1985.
  37. D. Kondepudi and I. Prigogine, *Modern Thermodynamics: From Heat Engines to Dissipative Structures*, Wiley, New York, 1998.
  38. C. R. Tellier and A. J. Tosser, *Size Effects in Thin Films*, Elsevier, Amsterdam, 1982.
  39. M. I. Flik and C. L. Tien, "Size effect on the thermal conductivity of high- $T_c$  thin-film superconductors," *J. Heat Transfer*, **112**, 872–881, 1990.
  40. R. A. Richardson and F. Nori, "Transport and boundary scattering in confined geometrics: Analytical results," *Phys. Rev. B*, **48**, 15209–15217, 1993.
  41. D. Stewart and P. M. Norris, "Size effect on the thermal conductivity of thin metallic wires: Microscale implications," *Microscale Thermophys. Eng.*, **4**, 89–101, 2000.
  42. S. G. Walkauskas, D. A. Broido, K. Kempa, and T. L. Reinecke, "Lattice thermal conductivity of wires," *J. Appl. Phys.*, **85**, 2579–2582, 1999.
  43. S. Kumar and G. C. Vradis, "Thermal conductivity of thin metallic films," *J. Heat Transfer*, **116**, 28–34, 1994.
  44. M. Asheghi, M. N. Touzelbaev, K. E. Goodson, Y. K. Leung, and S. S. Wong, "Temperature-dependent thermal conductivity of single-crystal silicon layers in SOI substrates," *J. Heat Transfer*, **120**, 30–36, 1998; M. Asheghi, K. Kurabayashi, R. Kasnavi, and K. E. Goodson, "Thermal conduction in doped single-crystal silicon films," *J. Appl. Phys.*, **91**, 5079–5088, 2002; W. Liu and M. Asheghi, "Thermal conductivity measurements of ultra-thin single crystal silicon layers," *J. Heat Transfer*, **128**, 75–83, 2006.
  45. P. K. Schelling, S. R. Phillpot, and P. Keblinski, "Comparison of atomic-level simulation methods for computing thermal conductivity," *Phys. Rev. B*, **65**, 144306, 1999.
  46. A. J. Kulkarni and M. Zhou, "Size-dependent thermal conductivity of zinc oxide nanobelts," *Appl. Phys. Lett.*, **88**, 141921, 2006.
  47. R. Landauer, "Spatial variation of currents and fields due to localized scatters in metallic conduction," *IBM J. Res. Develop.*, **1**, 223–231, 1957; M. Büttiker, Y. Imry, R. Landauer, and S. Pinhas, "Generalized many-channel conductance formula with application to small rings," *Phys. Rev. B*, **31**, 6207–6215, 1985; R. Landauer, "Conductance determined by transmission: probes and quantized constriction resistance," *J. Phys.: Condens. Matter*, **1**, 8099–8110, 1989; Y. Imry and R. Landauer, "Conductance viewed as transmission," *Rev. Mod. Phys.*, **71**, S306–S312, 1999.
  48. G. Rubio, N. Agrait, and S. Vieira, "Atomic-sized metallic contacts: Mechanical properties and electronic transport," *Phys. Rev. Lett.*, **76**, 2302–2305, 1996; N. Agrait, A. L. Yeyati, J. M. van Ruitenbeek, "Quantum properties of atomic-sized conductors," *Phys. Rep.*, **377**, 81–279, 2003.
  49. U. Landman, W. D. Luedtke, N. A. Burnham, R. J. Colton, "Atomistic mechanisms and dynamics of adhesion, nanoindentation, and friction," *Science*, **248**, 454–461, 1990; U. Landman, W. D. Luedtke, B. E. Salisbury, and R. L. Whetten, "Reversible manipulations of room temperature mechanical and quantum transport properties in nanowire junctions," *Phys. Rev. Lett.*, **77**, 1362–1365, 1996.
  50. L. Chico, L. X. Benedict, S. G. Louie, and M. L. Cohen, "Quantum conductance of carbon nanotubes with defects," *Phys. Rev. B*, **54**, 2600–2606, 1996.
  51. S. Frank, P. Poncharal, Z. L. Wang, and W. A. de Heer, "Carbon nanotube quantum resistors," *Science*, **280**, 1744–1746, 1998.

52. K. Schwab, E. A. Henriksen, J. M. Worlock, and M. L. Roukes, "Measurement of the quantum of thermal conductance," *Nature*, **404**, 974–977, 2000; K. Schwab, J. L. Arlett, J. M. Worlock, and M. L. Roukes, "Thermal conductance through discrete quantum channels," *Physica E*, **9**, 60–68, 2001.
53. A. Greiner, L. Reggiani, T. Kuhn, and L. Varani, "Thermal conductivity and Lorenz number for one-dimensional ballistic transport," *Phys. Rev. Lett.*, **78**, 1114–1117, 1997.
54. J. Hone, M. Whitney, C. Piskoti, and A. Zettl, "Thermal conductivity of single-walled carbon nanotubes," *Phys. Rev. B*, **59**, S2514–S2516, 1999; J. Hone, M. C. Llaguno, N. M. Nemes et al., "Electrical and thermal transport properties of magnetically aligned single wall carbon nanotube films," *Appl. Phys. Lett.*, **77**, 666–668, 2000.
55. S. Berber, Y.-K. Kwon, and D. Tománek, "Unusually high thermal conductivity of carbon nanotubes," *Phys. Rev. Lett.*, **84**, 4613–4616, 2000.
56. S. Maruyama, "A molecular dynamics simulation of heat conduction of finite length SWNTs," *Physica B*, **323**, 193–195, 2002; S. Maruyama, "A molecular dynamics simulation of heat conduction of a finite length single-walled carbon nanotube," *Microscale Thermophys. Eng.*, **7**, 41–50, 2003.
57. P. Kim, L. Shi, A. Majumdar, and P. L. McEuen, "Thermal transport measurements of individual multiwalled nanotubes," *Phys. Rev. Lett.*, **87**, 215502, 2001; C. Yu, L. Shi, Z. Yao, D. Li, and A. Majumdar, "Thermal conductance and thermopower of an individual single-wall carbon nanotube," *Nano Lett.*, **5**, 1842–1846, 2005.
58. N. Mingo and D. A. Broido, "Carbon nanotube ballistic thermal conductance and its limits," *Phys. Rev. Lett.*, **95**, 096105, 2005; N. Mingo and D. A. Broido, "Length dependence of carbon nanotube thermal conductivity and the problem of long waves," *Nano Lett.*, **5**, 1221–1225, 2005.

## PROBLEMS

---

- 5.1. Calculate the specific heat of lead, using both the Einstein model and the Debye model, for temperatures equal to 2, 10, 20, 50, 100, 200, 300, 600, and 800 K. Use  $\Theta_D = 88$  K and  $\Theta_E = 65$  K since the specific heats calculated with these values agree with the data well for the whole temperature range. Compare your answer with the values from Touloukian and Buyco.<sup>6</sup> Explain the low-temperature and high-temperature behavior.
- 5.2. In the first stage of designing a refrigeration system that will cool 1 kg of Pb from 300 to 2 K. Assume the Debye model can be used to calculate the temperature-dependent specific heat of lead (with  $\Theta_D = 88$  K). Answer the following questions:
  - (a) How much energy must be removed from Pb?
  - (b) How much entropy must be transferred out from Pb?
  - (c) Assuming that the environment is at 300 K, what is the least amount of work necessary to perform this refrigeration task?
  - (d) Consider the refrigeration in three temperature ranges: (1) from 300 to 100 K; (2) from 100 to 20 K; and (3) from 20 to 2. What is the least amount of work needed in each temperature range?
- 5.3. Plot the Fermi function  $f_{FD}$  versus  $\epsilon$  for  $T = 0, 500,$  and  $5000$  K. Plot the distribution function of free electrons in metal  $f(\epsilon)$  as a function of  $\epsilon$ . Discuss the main features of these plots. (Use eV as the unit for energy.)
- 5.4. The Fermi energy (at 0 K) of copper is  $\mu_F = 7.07$  eV. What is  $\mu(T)$  of Cu at 1000 and 10,000 K? Determine the maximum and root-mean-square free-electron speeds in copper at 0 K. Plot the electron distribution functions in terms of the speed and the kinetic energy for  $T = 0, 300,$  and  $4000$  K.
- 5.5. The Fermi energy of silver is  $\mu_F = 5.51$  eV. Calculate  $\mu(T)$  of Ag at  $T = 400$  and  $4000$  K. What is the rms speed of electrons at 0 K? What is the Fermi velocity? Plot the Fermi function at 0 and 4000 K in one graph and discuss the differences.
- 5.6. For  $k_B T \ll \mu_F$ , the specific heat of free electron gas in metal may be expressed as  $\bar{c}_{v,e} = (\bar{R}/n_c k_B) \int_0^\infty (\partial f_{FD} / \partial T) D(\epsilon) \epsilon d\epsilon$ . Evaluate this integration to obtain Eq. (5.25) by referring to Appendix B.8.
- 5.7. Calculate the Fermi energy of silver using the molecular weight and density. Estimate the spacing between the adjacent atoms of Ag. Calculate and plot the electron specific heat and the lattice specific

heat of Ag at temperatures from 0 to 1000 K. Show in a separate graph the low-temperature behavior. How do your calculated values agree with experimental data found in a heat transfer text?

**5.8.** Calculate the Fermi energy  $\mu_F$  for copper based on the molecular weight and density. What is the rms speed of free electrons in Cu at 0 and 300 K? Find the electronic specific heat and the lattice specific heat in  $J/(kg \cdot K)$  of Cu at 0.1, 1, 10, 30, and 500 K. When can you apply the  $T^3$  law, and when can you use the DeLong-Petit law?

**5.9.** Calculate the electronic specific heat and the lattice specific heat of gold at 1, 10, 100, 300 and 1000 K. Sketch their temperature dependence. At what temperature will the electronic and lattice contributions be the same? How does your calculated result compare with the value given in a heat transfer text?

**5.10.** The Mayer relation for the specific heat can be written as  $c_p - c_v = T\beta_p^2/\rho\kappa_T$ , where  $\beta_p = (1/v)(\partial v/\partial T)_p$  is the isobaric volume expansion coefficient,  $\kappa_T = -(1/v)(\partial v/\partial P)_T$  is the isothermal compressibility, and  $\rho$  is the density. Noting that the sound speed  $v_a$  is defined according to  $v_a^2 = (\partial P/\partial \rho)_s = c_p/(c_p - c_v)/c_p$ , we can write  $(c_p - c_v)/c_p = T\beta_p^2 v_a^2/c_p$ . A simple estimate of the relative difference between the specific heats is readily obtained by assuming that  $v_a$  is independent of temperature,  $c_p$  on the right-hand side is approximately  $3R$ , and  $\beta_p = 3\alpha$ , where  $\alpha$  is the linear thermal expansion coefficient. For silicon,  $\alpha \approx 4.6 \times 10^{-6} K^{-1}$  at 1000 K and  $v_a \approx 5000$  m/s. For copper,  $\alpha \approx 2.2 \times 10^{-5} K^{-1}$  and  $v_a \approx 2500$  m/s. Estimate the relative difference between  $c_p$  and  $c_v$  at 1000 K for silicon and copper.

**5.11.** Graphene is a single sheet of carbon atoms that forms carbon nanotubes by rolling and connecting the ends to form a seamless cylinder. The phonon mode with the lowest speed is the out-of-plane transverse acoustic mode, when the atoms vibrate perpendicular to the plane. It has a dispersion relation  $\omega(k) = ak^2$ , with  $a = 6 \times 10^{-7} m^2 \cdot s$ . It is expected that this mode is the dominate mode for the lattice specific heat at low temperatures (below 100 K). Using the 2-D solid model with the quadratic dispersion to show that  $c_l(T) \propto T$  at low temperatures, i.e.,  $T \ll \Theta_D$ .

**5.12.** Evaluate the specific heat of a thin GaAs film of two different thicknesses:  $L = 2$  and 10 nm. Plot the calculated specific heat with and without planar modes. Compare your results with that predicted by the Debye model for the bulk GaAs at  $T \ll \Theta_D$ .

**5.13.** Develop a computer program to calculate the lattice specific heat of CdS or ZnO<sub>2</sub> cubic nanocrystals with different sizes:  $L = 2, 10,$  and 20 nm. Discuss the low-temperature behavior in terms of Eq. (5.44a) and Eq. (5.44b).

**5.14.** For a nanowire of diameter  $d = 5$  nm, show that  $c_l(T) \propto T$  at low temperatures for a linear dispersion. If the length of the nanowire is  $L = 10d$ , what is the lowest temperature asymptote of the specific heat due to the second quantum effect?

**5.15.** Calculate the electron scattering rate and the mean free path of copper at 295 K. Use the linear relations for the electrical resistivity and the Wiedemann-Franz law to calculate the thermal conductivity at 200, 400, 600, and 800 K. Compare the calculated results with data from a heat transfer textbook.

**5.16.** Calculate the electron scattering rate  $1/\tau$ , the mean free path  $\Lambda$ , the electrical conductivity  $\sigma$ , and the thermal conductivity  $\kappa$  of aluminum near room temperature. If the temperature is increased by 5%, how will  $1/\tau$ ,  $\Lambda$ ,  $\sigma$ , and  $\kappa$  change? Express the scattering rate in both rad/s and Hz. Discuss why one should multiply it by  $2\pi$  to express  $1/\tau$  in Hz.

**5.17.** Sketch the thermal conductivity versus temperature from 0 to 1000 K for silver. What is the dependence of  $\kappa$  on  $T$ , as the temperature approaches absolute zero? How does the thermal conductivity change above 300 K?

**5.18.** Find the data for the electrical and thermal conductivities of a good conductor in a large temperature range, and evaluate when the Wiedemann-Franz law is valid. Show the low-temperature and high-temperature asymptotes for both  $\sigma$  and  $\kappa$ .

**5.19.** In the text, we stated that  $\partial f_{FD}/\partial \epsilon$  is a Dirac delta function and used it to obtain the electrical conductivity in Eq. (5.63). Prove that when  $k_B T \ll \mu_F$ , the integral  $\int_0^\infty G(\epsilon)(\partial f_{FD}/\partial \epsilon)d\epsilon \approx -G(\mu_F)$ , where  $G(x)$  is an analytical function of  $x$ . Then, derive Eq. (5.49) from Eq. (5.63).

**5.20.** Sketch the thermal conductivity of germanium (relatively pure) as a function of temperature.<sup>24</sup> Explain the trend of thermal conductivity at very low temperatures and at above room temperature. Can you assume that the thermal conductivity is independent of temperature near room temperature?

**5.21.** Derive Eq. (5.74) through Eq. (5.80). Show that in Eq. (5.80), the second term is much smaller than the first term for metals.



**5.22.** Prove Eq. (5.82), and calculate the Seebeck coefficient for Ag at 300 and 600 K. The measured Seebeck coefficient of Ag is  $1.51 \mu\text{V/K}$  at 300 K and  $3.72 \mu\text{V/K}$  at 600 K. On the other hand, the Seebeck coefficient for Pt is  $-5.28 \mu\text{V/K}$  at 300 K and  $-11.66 \mu\text{V/K}$  at 600 K. If an Ag-Pt thermocouple is formed with a junction temperature  $T_2 = 600 \text{ K}$  and a reference temperature  $T_1 = 300 \text{ K}$ , find the output voltage (see Fig. 5.14b).

**5.23.** For given values of  $T_L$ ,  $T_H$  and  $Z^*$ , there exists an optimal ratio  $R_L/R_0$  for achieving the maximum efficiency of the thermoelectric generator given in Eq. (5.94). Show that

$$\eta_{\max} = \frac{\Delta T}{T_H} \frac{\sqrt{1 + Z^* T_M} - 1}{\sqrt{1 + Z^* T_M} + T_L/T_H}$$

where  $T_M = (T_H + T_L)/2$ . Calculate the maximum efficiency, normalized to the Carnot efficiency, for  $T_L = 300 \text{ K}$  and  $T_H = 800 \text{ K}$  as a function of the dimensionless parameter  $Z^* T_M$ . Plot it for  $Z^* T_M$  from 0.3 to 3. Discuss the significance of  $ZT$  in thermoelectric devices.

**5.24.** Consider a thermoelectric generator made of two semiconductors working between  $T_L = 300 \text{ K}$  and  $T_H = 600 \text{ K}$ . The  $p$ -type material is made of  $\text{Bi}_{0.5}\text{Sb}_{1.5}\text{Te}_3$ , and the  $n$ -type material is made of  $\text{Bi}_2\text{Se}_{0.75}\text{Te}_{2.25}$ , with the following average properties:  $\kappa_p = 1.2 \text{ W/m} \cdot \text{K}$ ,  $\kappa_n = 1.3 \text{ W/m} \cdot \text{K}$ ,  $r_{e,p} = 15 \mu\Omega \cdot \text{m}$ ,  $r_{e,n} = 13 \mu\Omega \cdot \text{m}$ ,  $\Gamma_p = 210 \mu\text{V/K}$ , and  $\Gamma_n = -190 \mu\text{V/K}$ . Assume that the length  $L = 0.8 \text{ cm}$  and the cross section  $A_c = 0.3 \text{ cm}^2$  for both materials. A generator with a diameter of 10 cm contains 100 pairs ( $N = 100$ ). Find the power output at the maximum efficiency (see Problem 5.23).

**5.25.** Perform a thermodynamic analysis of the thermoelectric cooling using the same configuration as in Fig. 5.15. By noting that no load resistance is needed and the voltage supplied  $\Delta V = N\Gamma_{np}\Delta T + IR_0$ , show that the coefficient of performance of a thermoelectric refrigeration is

$$C.O.P. = \frac{|q_L|}{P} = \frac{I\Gamma_{np}A_c\sigma_{np}T_L - I^2L/2 - A_c^2\sigma_{np}\kappa_{np}\Delta T/L}{I\Gamma_{np}A_c\sigma_{np}\Delta T + I^2L}$$

The maximum C.O.P. can be obtained by setting the derivative with respect to  $I$  equal to zero. Show that

$$(C.O.P.)_{\max} = \frac{T_L}{\Delta T} \frac{\sqrt{1 + Z^* T_M} - T_H/T_L}{\sqrt{1 + Z^* T_M} + 1}$$

where  $T_M = (T_H + T_L)/2$ .

**5.26.** Estimate the thermal conductivity along a copper film with various thicknesses:  $d = 400, 100,$  and  $50 \text{ nm}$  at  $300 \text{ K}$ . What if the temperature is reduced to  $1 \text{ K}$ ?

**5.27.** Estimate the thermal conductivity along a copper wire with various diameters:  $d = 400, 100,$  and  $50 \text{ nm}$  at  $1$  and  $300 \text{ K}$ , respectively. Compare simple geometric averaging of free paths with the BTE. What are the electron de Broglie wavelengths at these temperatures? If the surface roughness parameter  $\sigma_{\text{rms}} = 2 \text{ nm}$ , will the scattering be mostly diffuse or specular at each temperature?

**5.28.** At  $5 \text{ K}$ , calculate the thermal conductivity, perpendicular ( $\kappa_{\text{eff},z}$ ) and parallel ( $\kappa_{\text{eff},x}$ ) to the plane, for a  $200\text{-nm}$ -thick gold film. Calculate the effective thermal conductivity  $\kappa_{\text{eff},w}$  of a gold wire of  $5\text{-}\mu\text{m}$  thickness. Hint: use the bulk resistivity value from Fig. 5.11.

**5.29.** In Example 5-6, we have calculated the properties of a single-crystal silicon at various temperatures. Use simple relations with  $p = 0$  to estimate the thermal conductivities of silicon at temperatures ranging from  $5$  to  $1000 \text{ K}$  along a  $50\text{-nm}$ -thick thin film and a  $100\text{-nm}$ -thick thin wire. Assume the surface roughness  $\sigma_{\text{rms}} = 2 \text{ nm}$ . Will the diffuse model be a good assumption? For the thin film, redo the calculation using the specularity  $p$  estimated based on the thermal phonon wavelength  $\lambda_{\text{th}}$ .

**5.30.** The diameter of a carbon nanotube is determined by its chiral numbers ( $m, n$ ) according to  $d = 0.07834\sqrt{m^2 + mn + n^2}$ . What is the diameter of (10,10) single-walled nanotubes? Assume that the wall thickness (unit atomic layer) is  $0.34 \text{ nm}$ . What is the cross-sectional area? Calculate the phonon thermal conductivity  $\kappa$  in the ballistic limit considering the four phonon modes at  $100 \text{ K}$  for (10,10) nanotubes with length  $L = 100 \text{ nm}, 1 \mu\text{m},$  and  $10 \mu\text{m}$ . Will the ballistic limit of thermal conduction hold at room temperature and above?

---

# CHAPTER 6

---

## ELECTRON AND PHONON TRANSPORT

---

In the preceding chapter on solid properties, we relied on the Drude-Sommerfeld model, which assumes that electrons are completely free and the Fermi surface is spherical and isotropic in all directions of the wavevector. While the concepts of electronic band structures and phonon dispersion in real solids were often mentioned, we have deliberately avoided any details. It is hoped that the free-electron model will help readers gain an intuitive picture of electrons without a deep knowledge of solid state physics. Note that the free-electron model is applicable only for metals, usually good conductors, and cannot be applied to semiconductors. The Sommerfeld theory, albeit successful in quantitatively describing electronic transport for certain metals, does not touch on the fundamental mechanisms of electron scattering and the shape of the Fermi surface. The free-electron model also fails to explain certain phenomena including thermoelectricity. The Hall effect and magnetoresistance, to be discussed in the following section, provide further evidence of the inadequacy of the free-electron model.

This chapter introduces electronic band theory after a brief discussion of electronic structures in atoms, binding in crystals, and crystal lattices. The phonon dispersion relations are presented subsequently and explained in terms of different branches of acoustic and optical phonons. Subsequently, the electron and phonon scattering mechanisms are explored. The next section addresses electronic emission and tunneling phenomena, including photoelectric effect, thermionic emission, field emission, as well as electron tunneling through a potential. A significant portion of this chapter is then devoted to semiconductor materials and devices, with an emphasis on optoelectronic applications such as solar cells, thermophotovoltaic systems, light-emitting diodes (LEDs), and semiconductor lasers including quantum well lasers.

### 6.1 THE HALL EFFECT

---

When a conductor carrying electric current is placed in a magnetic field perpendicular to the current flow, there is a Lorentz force acting on the conductor according to  $\mathbf{F} = \sum q\mathbf{u}_d \times \mathbf{B} = I\mathbf{l} \times \mathbf{B}$ , where  $q$  is the charge of each carrier,  $\mathbf{u}_d$  is the drift velocity of the carrier,  $\mathbf{B}$  is the magnetic induction,  $I$  is the current in the conductor, and  $l$  is the length of the conductor. This principle was used in the electromagnetic motor invented by Michael Faraday in 1821. Because electric current is always defined in the direction of the applied electric field  $\mathbf{E} = -\nabla V$ , the force acting on the conductor is independent of the nature of the carriers (electrons or holes). Microscopically, however, there is a subtle difference that can be distinguished by the experiment first performed by Edwin Hall in 1878 when he was a graduate student at Johns Hopkins University. As shown in Fig. 6.1, an electric current

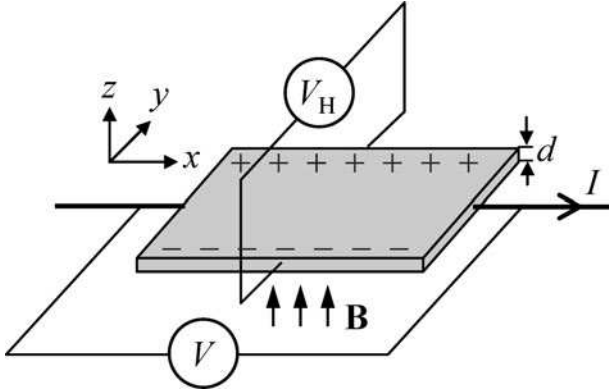


FIGURE 6.1 Illustration of the Hall effect experiment.

passes through a metal foil in the  $x$  direction, while the electrons are drifted opposite to the  $x$  direction. When a uniform magnetic field  $\mathbf{B}$  is applied in the  $z$  direction, the electrons are subjected to a force toward the negative  $y$  direction. Gradually, an electric field is built up across the foil as manifested by a nonzero voltage  $V_H$ , which is called the Hall voltage. The electric potential in the  $y$  direction eventually balances the magnetic force such that the electrons drift in the  $x$  direction only. This effect is called the *Hall effect*. By setting the  $y$  component of the Lorentz force  $\mathbf{F} = q(\mathbf{E} + \mathbf{u}_d \times \mathbf{B})$  to zero, one obtains

$$V_H = \frac{IB}{nqd} \quad (6.1)$$

where  $n$  is the number density of the carrier and  $d$  is the thickness of the conductor.<sup>1,2</sup> The *Hall coefficient* is defined as follows:

$$\eta_H = \frac{V_H d}{IB} = \frac{1}{nq} \quad (6.2)$$

The *Hall resistance* can be defined as  $R_H = V_H/I = B\eta_H/d$ , and its inverse is called the *Hall conductance*. Similarly, the *Hall resistivity* is given by  $r_H = B\eta_H = E_y/J_x$ , where  $E_y$  is the electric field in the  $y$  direction and  $J_x$  is the current density. For metals,  $q = -e$  and  $n = n_e$ , the number density of free electrons, and one would expect a negative Hall resistance.

**Example 6-1.** Find the Hall coefficient and the Hall voltage for a copper foil of  $2 \times 2 \text{ cm}^2$  area, with a thickness of  $10 \mu\text{m}$ . Given the electrical current  $I = 0.5 \text{ A}$  and the magnetic induction  $B = 1.0 \text{ T}$  (tesla)  $= 1.0 \text{ Wb/m}^2$ , what is the voltage drop along the current flow direction?

**Solution.** Based on the previous chapter, the number density of electrons in copper is  $n_e = 8.45 \times 10^{28} \text{ m}^{-3}$ . From Eq. (6.2), we obtain  $\eta_H = -7.4 \times 10^{-5} \text{ cm}^3/\text{C}$ , and from Eq. (6.1) we find  $V_H = 23.7 \mu\text{V}$ , which is a very small voltage but can be measured accurately. Using the resistivity of copper  $r_c = 1.7 \times 10^{-8} \Omega \cdot \text{m}$ , we see that  $V = 850 \mu\text{V}$ , which is much larger than the Hall voltage. The Hall coefficient is much larger for semiconductors because of their usually much lower carrier densities.

Before the discovery of the Hall effect, many people, including James Clerk Maxwell, believed that the force acted only on the conductor but not on the current carriers.<sup>3</sup> Measurement of the Hall coefficient allows the determination of the sign of the charge carriers as well as the carrier concentration. This is important especially for semiconductor materials. The Hall coefficient is positive for  $p$ -type semiconductors, but negative for  $n$ -type

semiconductors. In reality, the Hall coefficient depends also on the applied magnetic field although such a dependence cannot be predicted by the Drude free-electron model. For some common metals like Al, Be, Cd, In, and Zn, the Hall coefficient can even become positive. Therefore, the Hall effect cannot be fully accounted by the free-electron model. It is necessary to understand the electronic structures.

Magnetoresistance is the change in resistance of a material under an applied magnetic field. The magnetoresistance may be transverse, when the applied magnetic field is perpendicular to the current flow, and longitudinal, when the applied magnetic field is parallel to the current flow. In the free-electron theory, resistance is expected to be independent of the strength of the applied transverse magnetic field. In reality, most materials exhibit transverse magnetoresistance that depends on the magnetic field strength. In the late 1980s, researchers observed a giant magnetoresistive (GMR) effect, also called giant magnetoresistance, with extremely thin films of ferromagnetic and metallic layers. The GMR effect has been applied to read heads for magnetic hard disk drives.<sup>4</sup>

Klaus von Klitzing and coworkers in 1980 measured the Hall voltage of a 2-D electron gas using a metal-oxide-semiconductor field-effect transistor (MOSFET), at very low temperatures ( $T \approx 1.5$  K) with a high magnetic field ( $B > 15$  T), at the Grenoble High Magnetic Field Laboratory in France.<sup>5</sup> They found that the Hall conductance is quantized and increases with the applied magnetic field by steps in a staircase sequence. The Hall conductance is a multiple of a fundamental constant,  $1/R_K$ , where

$$R_K = h/e^2 = 25,812.807449 \pm 0.000086 \Omega \quad (6.3)$$

is called the von Klitzing constant. Note that  $e^2/h$  is proportional to the fine-structure constant, which is related to the strength of light-matter interaction in quantum electrodynamics. For this work, von Klitzing was awarded the Nobel Prize in Physics in 1985. The remarkable precision and gauge invariance of quantized conductance allowed the definition of a resistance standard used worldwide since 1990.<sup>6</sup> As discussed in Chap. 5, quantized conductance has also been observed between nanocontacts and nanostructures with an increment of  $2/R_K$ . The discovery of the fractional quantum Hall effect in 1982, on the other hand, rendered three physicists (Robert Laughlin, Horst Störmer, and Daniel Tsui) the 1998 Nobel Prize. This has led to a breakthrough in our fundamental understanding of the physical world. For example, in a 2-D system, electrons may switch between Fermi-Dirac statistics and Bose-Einstein statistics, continuously.<sup>7</sup> More recently, Strohm et al. reported the Hall effect for phonons by applying a magnetic field perpendicular to the heat flow in a paramagnetic dielectric material at low temperatures.<sup>8</sup> A transverse temperature difference was measured, which reverses sign when the magnetic field is inverted.

## 6.2 GENERAL CLASSIFICATIONS OF SOLIDS

---

There are several ways to classify solids. Based on their electrical conductivities, solids may be classified as insulators, semiconductors, or conductors. They may exist in different forms, such as amorphous or crystalline phases, depending on how the atoms in the solids are arranged. A general introduction is given in this section considering chemical bonds and electrical properties of solids. Let us first take a look at the electron configuration in atoms because it is directly related to physical and chemical properties.

### 6.2.1 Electrons in Atoms

The periodic table of elements is arranged sequentially according to atomic number, which is determined by the number of protons inside the nucleus and equal to the number of

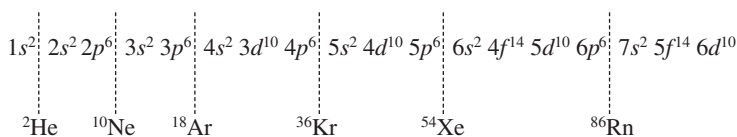
electrons orbiting the nucleus, since an atom itself is charge neutral. The electrons occupy different quantum states, which are fully described by the Schrödinger wave equation as discussed in Chap. 3. By solving the wave equation in spherical coordinates, the number of quantum states can be determined and identified using indices  $n$ ,  $l$ , and  $m$ .<sup>9,10</sup> The first or principal quantum number  $n = 1, 2, 3, 4, \dots$  corresponds to different shells, denoted as K, L, M, N, O,  $\dots$ . In each shell, there are  $n$  subshells defined by the orbital number  $l = 0, 1, 2, \dots, (n - 1)$ . The corresponding symbols are  $s, p, d, f, g, h$ , and so forth. For each  $l$ , the magnetic quantum number  $m = 0, \pm 1, \pm 2, \dots, \pm l$ , which gives a total of  $2l + 1$  orbitals for each subshell. Hence, there are a total of  $n^2$  orbitals in the  $n$ th shell. When spin degeneracy is considered, the total allowable quantum states are  $2n^2$  in the  $n$ th shell. In other words, there are 2, 8, 18, and 32 quantum states in the first (K), second (L), third (M), and fourth (N) shells, respectively. On the other hand, there are  $2(2l + 1)$  quantum states in the  $l$ th ( $l < n$ ) subshell. For example, the  $s, p, d$ , or  $f$  subshell contains correspondingly 2, 6, 10, or 14 quantum states. According to Pauli's exclusion principle, each quantum state can have no more than one electron; i.e., at most only two electrons (one with  $+\frac{1}{2}$  and the other with  $-\frac{1}{2}$  spin) can share the same orbit. According to the Aufbau principle, electrons will fill the lowest energy states first. The electron configuration of an atom is expressed by the numbers in each subshell. For example, we can write for aluminum and calcium, respectively,



Note that the  $4s$  orbitals are filled before the  $3d$  orbitals because the associated energy level of a  $3d$  orbital is higher than that of a  $4s$  orbital. However, the electron configuration for  ${}^{29}\text{Cu}$  is



This is due to the fact that a half-filled or filled  $d$  subshell is more stable than the  $s$  shell of the next level.<sup>10</sup> Similarly, the outermost shells for chromium ( ${}^{24}\text{Cr}$ ) are  $4s^1 3d^5$  not  $4s^2 3d^4$ , and those for gold ( ${}^{79}\text{Au}$ ) are  $6s^1 4f^{14} 5d^{10}$  not  $6s^2 4f^{14} 5d^9$ . The properties of an element depend largely on the filled state of the outermost orbitals. Alkali metals, such as  ${}^3\text{Li}$ ,  ${}^{11}\text{Na}$ , and  ${}^{19}\text{K}$ , have one electron in the outermost orbit and can easily lose it, especially when interacting with halogens whose outermost orbitals can be filled by adding only one electron each. The result is the formation of chemically stable compounds such as  $\text{NaCl}$  and  $\text{CsF}$ . The outermost electrons are called *valence electrons*. The  $4s^1$  electron in copper is largely responsible for its high electrical conductivity because it can leave the atom relatively easily. When the outermost orbitals are completely filled, as in noble gases like He and Ne, the atoms are very stable and reluctant to react with others. Noble gases are also called inert gases since they are monatomic gases at ambient conditions. At the atmospheric pressure, helium must be cooled to 4.2 K for it to condense into liquid. The general sequence of electron configuration in order of increasing energy is

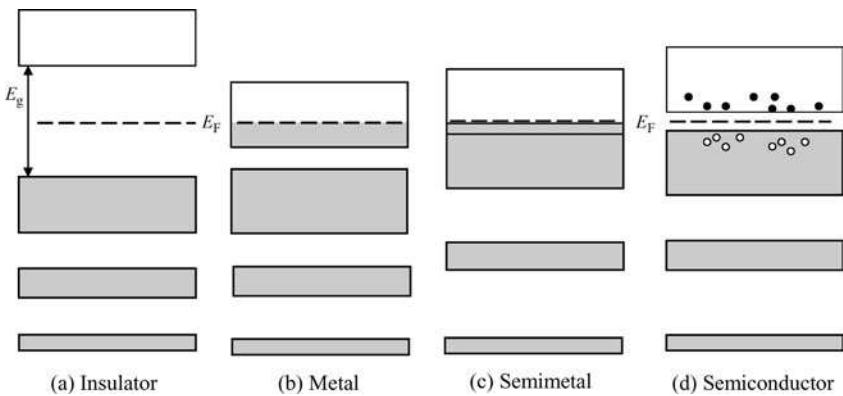


For convenience, each dashed line indicates the electron configuration of an inert gas listed underneath that line. Each noble gas contains a completely filled  $p$  subshell (with the exception of He which has a filled K shell) before the next  $s$  subshell. In atomic physics, *ionization energy* is the energy required to separate an electron from the atomic nucleus. The ionization energy varies periodically according to the atomic number: Alkali metals have the lowest ionization energy because of the single electron in the outermost  $s$  orbitals. On the other hand, inert gases have the highest ionization energy. Helium is the most stable element

with an ionization energy of 24.6 eV. The ionization energy of lithium is only 5.4 eV. For a hydrogen atom, the ionization energy is 13.6 eV as discussed in Chap. 3.

## 6.2.2 Insulators, Conductors, and Semiconductors

The picture of free electron gas depicted in Chap. 5 is an oversimplified version in which the electron energies are limited to a nearly continuous band from the zero energy level up to the Fermi energy or Fermi level. Only those near the Fermi surface contribute to electronic transport properties. Electrons in a single atom are in various discrete energy levels, which are well predicted by quantum mechanics. In real solids, atoms are arranged in close proximity; hence, electrons interact strongly with one another as well as with the crystal lattices, resulting in complex wavefunctions as manifested by their band structures. There exists a large number of *allowable bands* that may be occupied by electrons. Between two consecutive allowable bands, there exists a *forbidden band* that cannot be occupied by any electron. Electrons occupy broad bands with allowable energy states up to the Fermi level. The distinction between insulators and metals can be understood by looking at the electronic states near the Fermi surface as illustrated in Fig. 6.2. A brief



**FIGURE 6.2** Schematic of the energy band for different materials, where  $E_g$  is the bandgap energy and  $E_F$  is the Fermi energy. (a) An insulator has a completely filled valence band and a completely empty conduction band, with a wide bandgap between the two. (b) A metal has a partially filled conduction band and the Fermi level lies in this band. (c) A semimetal, also called a metal, has a conduction band that overlaps the filled valence band. (d) A semiconductor is like an insulator but with a much smaller bandgap and may conduct electricity at elevated temperatures due to thermally excited electrons and holes. Doping or impurities in a semiconductor can result in a large electrical conductance.

qualitative description is given here, whereas more detailed theories are deferred to subsequent sections.

For insulators, the highest occupied band is completely filled as shown in Fig. 6.2a. This is called a *valence band* due to the contribution of valence electrons. The next higher band is a *conduction band* which is completely empty. There exists a large energy gap between the valence band and the conduction band, usually between 5 and 15 eV. Examples are  $E_g \approx 8$  eV for fused silica ( $\text{SiO}_2$ ) and  $E_g \approx 14$  eV for LiF. The Fermi level lies in the middle of the forbidden band. Because the valence band is completely filled, electrons are not free to move around (i.e., change from one quantum state to another) under the influence

of an electric field. An electrical insulator is also called a dielectric. Pure crystalline dielectrics are transparent to visible light because their valence electrons cannot be excited unless the incoming radiation frequency is high enough that the photon energy exceeds the bandgap energy. Note that a photon energy of  $h\nu = 2$  eV corresponds to a visible wavelength  $\lambda = 620$  nm, and that of 10 eV corresponds to  $\lambda = 124$  nm, which lies in the deep ultraviolet. On the other hand, lattice vibrations or phonons in dielectric materials often yield absorption of radiation in the mid-infrared.

A metal has a partially filled conduction band, which is the highest occupied band, as shown in Fig. 6.2*b*. The Fermi level lies inside this allowable band. For some metals like Bi and Sn, the conduction band overlaps the valence band as illustrated in Fig. 6.2*c*. These metals are sometimes called semimetals since their electrical conductivities are not as high as the alkali or noble metals. Because the energy states within the conduction band are continuous, the uppermost electrons in the partially filled conduction band or the top of the valence band can be excited to a higher unoccupied energy level by an arbitrary applied field. Over 80% of the elements in the periodic table are metals (or semimetals). All group Ia (alkali, excluding hydrogen), group IIa (alkaline earth), group IIIa (except boron), and transition (all b groups from columns 3 to 12 of the periodic table) elements are metals. The interaction between electromagnetic radiation and a material is much like applying an electric field to the material, except that the frequency of the applied field is very high. Note that the frequency of red light at  $\lambda = 632$  nm is  $\nu = c/\lambda = 475$  THz. Because of their relatively free electrons, metals interact with electromagnetic radiation strongly. This is manifested by the strong absorption by thin metallic films and the high reflection from polished bulk metals. The strong interaction of metals with microwaves can easily be demonstrated by placing a piece of aluminum foil in a microwave oven and then observing the noises and sparkles as the oven is turned on. At shorter wavelengths in the visible spectrum and in the ultraviolet, additional absorption mechanisms emerge that may be better explained by the particle nature of light.

Semiconductors have band structures similar to those of insulators, except that the energy bandgap  $E_g$  is much narrower, i.e., on the order of 1 eV. For example, diamond has a bandgap of 5.5 eV and is usually classified as an insulator, whereas silicon has a bandgap of 1.1 eV at room temperature and is a semiconductor. Some semiconductors can have a relatively large bandgap and hence are called *wideband semiconductors*. Examples are the III-V semiconductor GaN (3.4 eV) and the II-VI semiconductors CdS (2.4 eV) and ZnS (3.7 eV). Diamond may be considered as a wideband semiconductor because of its crystal structure similar to those of Si and Ge. Pure or *intrinsic* semiconductors are insulators at low temperatures. At higher temperatures, as illustrated in Fig. 6.2*d*, some electrons (dots) can be *thermally excited* from the valence band to the conduction band, leaving holes (circles) in the valence band. Subsequently, electrical current may flow through, although with a large resistance as compared to metals. Bandgap absorption is essential for the interaction of semiconductors with optical radiation. When the photon energy exceeds the energy gap, strong absorption occurs. This is why a silicon wafer looks dark and is opaque to visible light.

By doping the semiconductor with impurities, the charge distribution can be significantly changed; while at the same time, the bandgap and the Fermi level are slightly modified. The semiconductor becomes *extrinsic*, meaning that the number of electrons is no longer the same as that of holes. A group V element, such as phosphorous with five valence electrons, may substitute a small fraction of silicon atoms. The extra valence electrons can be thermally excited to the conduction band via ionization of the impurities. The phosphorus atom is said to be a *donor*, and the doped semiconductor becomes *n*-type since the majority of its carriers are electrons. The electron concentration can be significantly increased to enhance the electrical conductivity. From the band structure point of view, the donated electrons form a filled impurity band right below the conduction band. The difference in energy between the conduction band and the impurity band is called *ionization energy*, which is on the order of 0.05 eV. The ionization energy of a semiconductor has a different meaning from the ionization energy required to separate an electron from the

atomic nucleus discussed earlier. Likewise, when impurities from a group III element such as boron with three valance electrons are introduced, additional holes are created such that the silicon semiconductor becomes a *p*-type semiconductor because of the additional positive charge carriers. The boron atoms are called *acceptors*, which form an empty impurity band right above the valance band.<sup>11</sup> The energy difference between these two bands is also called the ionization energy. Doping can strongly affect the infrared properties of semiconductors because of free carrier absorption. Furthermore, impurities and defects tend to increase phonon scattering and reduce thermal conductivity since thermal transport in semiconductors is mainly by lattice vibration.

### 6.2.3 Atomic Binding in Solids

Two or more atoms can combine to form a molecule, mainly through the electrons in the outermost orbits (i.e., valance electrons), since the electrons in the inner shells remain tightly bonded to their nuclei. The wavefunctions of the valance electrons are significantly modified as compared with those of the individual atoms. There are five major kinds of chemical bonds: the ionic, covalent, molecular, and hydrogen bonds for insulators and the metallic bond for conductors. Solids with identical chemical composition can have different stable forms or phases, which exhibit distinct differences in their appearances as well as electrical, mechanical, and thermal properties. A notable example is carbon, which may exist in the form of diamond, graphite, carbon black (amorphous carbon), or the fullerene family. A crystal contains periodic and densely packed atoms or lattices, whereas an amorphous solid does not have well-organized lattice structures. The atoms in an amorphous solid are disordered and irregular, like those in a liquid, except that they are firmly bonded together. Therefore, a crystal is usually denser and harder than the amorphous phase of the same composition. A crystal usually exhibits distinct facets along the crystalline planes and has a sharp transition between solid and liquid at a fixed melting point. An amorphous solid does not have clear facets when broken. When heated up, an amorphous solid is first softened and then gradually it melts over a wide temperature range. An example is quartz versus fused silica (glass), both made of  $\text{SiO}_2$ . For a given composition, the thermal conductivity is usually much higher in the crystal form because of lattice vibrations.

Alkali metals and alkaline earth metals have one and two valance electrons, respectively, that are loosely bonded. A metal atom can lose its outermost electrons to become a positive ion. On the other hand, the elements in groups VIIa and VIa tend to gain additional electrons to fill the outermost orbits and become negative ions. The positive and negative ions attract each other by electrostatic force and form an *ionic bond*, which is quite strong. *Ionic crystals*, such as NaCl, CsCl, KBr,  $\text{CaF}_2$  and MgO, are hard and usually have high melting points (above 1000 K). They are insulators because the ions cannot move around freely and are transparent in the visible spectrum because of the large bandgap. Nevertheless, some of these crystals are soluble and can be dissolved in water. The solution becomes conductive because of the ions. The positive and negative ions form an electrical dipole and can absorb infrared radiation through lattice vibrations. These solids belong to the group of *polar materials*, in terms of polarizability. Note that the elements in groups Ib (noble metals) and IIb (Zn, Cd, and Hg) resemble those in groups Ia and IIa because of the outermost *s*-orbital electrons. The difference is that groups Ib and IIb also have filled *d* subshells. Therefore, II-VI semiconductors such as ZnSe and CdTe are largely ionic bonded.

The main contribution to the binding energy is the electrostatic or Madelung energy.<sup>2</sup> The long-range electrostatic force between two ions with charges  $q_1$  and  $q_2$  is  $\pm q_1 q_2 / r^2$ , where  $r$  is the separation distance measured from the center of the ion cores. Depending on the sign of the charges, either attractive or repulsive force may occur. The ions arrange themselves in a way that gives the strongest attractive interaction, which is balanced by the short-range repulsive force between atoms. The contribution of the Coulomb attraction to



the total energy of the system is roughly proportional to  $-1/r$ . As atoms are brought very close to each other, the charge distributions or the electron orbits begin to overlap with each other. Pauli's exclusion principle requires some of the electrons move to higher quantum states, resulting in an increased total energy of the system. Associated with the increased energy is a repulsive force between the atoms. The magnitude of this repulsive force varies with  $1/r^{m+1}$  (where  $m$  is between 6 and 10 for alkali halides with NaCl structure<sup>1</sup>), and thus, is negligible at large distances but increases rapidly when the distance is less than 0.5 nm. The repulsive force contributes to the energy of the system by  $1/r^m$ . There exists a minimum energy or equilibrium position of the system when all the repulsive and attractive forces balance each other. Readers are reminded about the similar discussion in Sec. 4.2.4 on the intermolecular force and potential [see Eq. (4.48) and Fig. 4.8]. Understanding the binding energy or the interatomic potential is very important for atomic scale simulations, e.g., those using molecular dynamics.

*Covalent bonds* are formed between nonmetallic elements when the electrons in the outermost orbits are shared by more than one atom. Covalent binding is important for gaseous molecules like  $\text{Cl}_2$ ,  $\text{N}_2$ , and  $\text{CO}_2$ . When the atoms are brought close enough, the electron orbits overlap, allowing them to share one or more electrons. Covalent interactions result in attractive forces, and the binding of atoms is associated with a reduced total energy. *Covalent crystals* consist of an infinite network of atoms joined together by covalent bonds. Examples are diamond, silicon, SiC, and quartz ( $\text{SiO}_2$ ). The whole crystal is better viewed as a large molecule or supermolecule. In diamond structure, each atom is bonded to four neighboring atoms, which form a tetrahedron. In a SiC crystal, each silicon atom is bonded to four carbon atoms and vice versa. In a  $\text{SiO}_2$  crystal, while each silicon atom is bonded to four oxygen atoms at tetrahedral angles, each oxygen atom is bonded only to two silicon atoms. Covalent solids are usually very hard with a high melting point and thermal conductivity. The melting points of quartz and silicon are 1920 K and 1690 K, respectively. Diamond has the highest melting point among all known materials, i.e., 3820 K. At room temperature, the thermal conductivity of diamond is 2300 W/(m · K), which is the highest of all known bulk materials. Pure diamond and intrinsic silicon do not absorb radiation at frequencies lower than that of the corresponding bandgap energy. Because of its wide bandgap, diamond is clear in the visible region and transparent throughout the whole infrared and microwave regions.

Some solids have both ionic and covalent characteristics. Examples are the III-V semiconductors such as GaN, GaAs, and InSb. II-VI materials such as ZnO and CdS have a large proportion (30%) of covalent bond characteristics. Even SiC has some ionic bond characteristics because of the dipoles formed due to different attractive forces by different atoms. Therefore, SiC is also a polar material that can absorb and emit infrared radiation through lattice vibrations.

Inert gases can be solidified at very low temperatures via *molecular bonds*. At atmospheric pressure, argon becomes liquid at temperatures between 84 and 87 K. At temperatures below 84 K, it crystallizes into a dense solid, called a *molecular crystal*. Van der Waals' force caused by induced dipole moments between atoms is responsible for the attraction and binding of atoms. The van der Waals weak interaction gives a long-range potential that is proportional to  $-1/r^6$ , as discussed in Sec. 4.2.4. The repulsive potential for inert gas is proportional to  $1/r^{12}$ . Molecular bonds are also important for many organic molecules.

Hydrogen has only one electron per atom and can form a covalent bond with another to form  $\text{H}_2$  molecule. When interacting with other atoms, a hydrogen atom may be attracted to form a *hydrogen bond*. The hydrogen bond and the resulting electrostatic attraction are important for  $\text{H}_2\text{O}$  molecules, with many striking physical properties in its vapor, water, and ice phases. Hydrogen bonds and molecular bonds are essential to organic molecules and polymers.

*Metallic bonds* are responsible for the binding energy in metals. Pure metals can form densely packed periodic lattices or crystals. Metals often exist in *polycrystalline* form in

which small grains of crystals are joined together randomly, or in *alloy* form in which the atoms are arranged irregularly like an amorphous insulator. Unlike in a covalent crystal where atoms share a few electrons, in a metallic crystal, some valence electrons leave the ion cores completely and are shared by all the ions in the crystal. This is consistent with the picture of free electron gas and describes well the behavior of alkali metals. Transition metals, including the noble metals, contain electrons in the *d* subshell. The metallic bonds are supplemented by covalent and molecular bonds. Due to the relatively free electron gas, metals have high thermal and electrical conductivities. Metallic crystals are also more flexible than nonmetallic crystals, which are usually brittle. The melting points of metals vary significantly. Examples are Hg (234 K), Ga (303 K), Au (1338 K), and W (3695 K). As mentioned in previous chapters, the physical properties would change significantly as the structure is reduced down to hundreds, tens, or even a few atomic layers in one-, two-, or three-dimensions. Examples are carbon nanotubes, silicon nanowires, ZnO nanobelts, and CdSe-ZnS quantum dots. In order to further understand the properties of solids, let us examine the crystal structures more closely in the following section.

### 6.3 CRYSTAL STRUCTURES

---

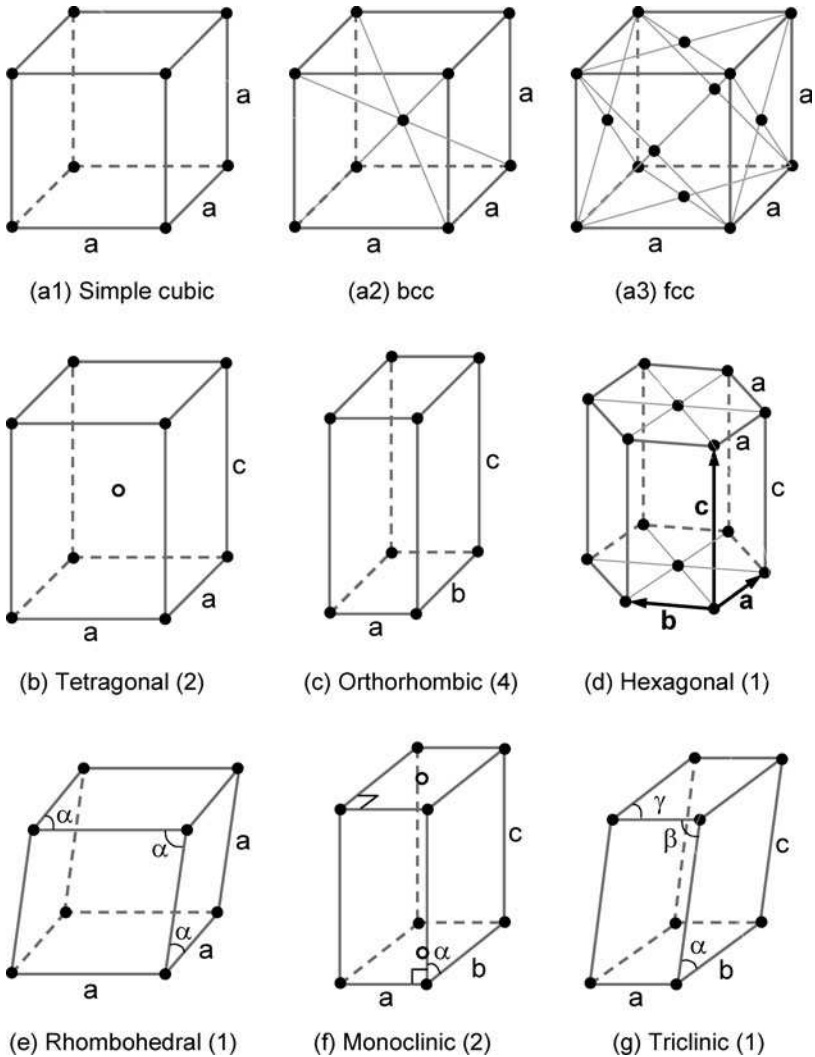
A crystal is constructed by the continuous repetition in space of an identical structural unit. Geometrically speaking, a crystal is a 3-D periodic array, or network, of lattices. All lattice points are identical to one another. For a crystal made of only one type of element, each lattice point may be treated as a single atom or ion. However, this is not necessary as will be illustrated later. In general, each lattice point represents a set of atoms, ions, or molecules, located in its neighborhood. This set of atoms, ions, or molecules is called a *basis*. A *unit cell* of a crystal structure contains both the lattice and the basis, and can be repeated by translations to cover the whole crystal.

It has long been hypothesized that crystalline materials must have some periodicity in their microstructures. In 1913, W. L. Bragg and his father W. H. Bragg used x rays to provide microscopic evidence of the existence of periodic lattice structures. This was a giant step because the distances between atoms are on the order of 0.1 nm. X-ray crystallography provided a powerful tool for the determination of the microscopic structure of solids. The Braggs received the Nobel Prize in Physics in 1915, when Lawrence Bragg was only 25 years old. It was not until 1983 that atomic images were obtained in real space using a scanning tunneling microscope (STM) as discussed in Chap. 1. The physical properties of crystalline solids are largely determined by the arrangement of atoms in a unit cell, in addition to the chemical bonds between atoms. It is of great importance to know the structure of a crystal first in order to understand its electrical, thermal, mechanical, and optical properties.

#### 6.3.1 The Bravais Lattices

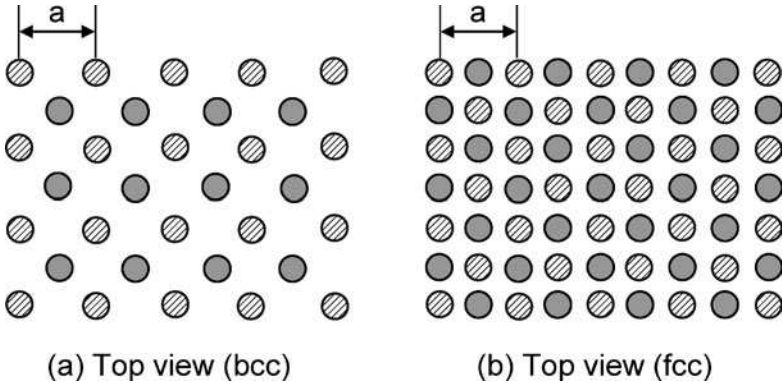
In three dimensions, crystal lattices can be grouped into 14 different types as required by translational symmetry. These are called Bravais lattices, named after French physicist Auguste Bravais (1811–1863), who showed that there are only 14 unique Bravais lattices from the point of view of symmetry. Bravais lattices are then categorized into seven crystal systems, resulting in seven types of *conventional unit cells*, namely, cubic, tetragonal, orthorhombic, hexagonal, rhombohedral, monoclinic, and triclinic, as illustrated in Fig. 6.3.

There are three cubic lattices: the simple cubic with lattice points only on its apexes, the body-centered cubic (bcc) with one additional lattice at the center, and the face-centered cubic (fcc) with one additional lattice at each face, as shown in Fig. 6.3a1, Fig. 6.3a2, and Fig. 6.3a3. To illustrate the difference between bcc and fcc lattices clearly, Fig. 6.4 displays



**FIGURE 6.3** The seven crystal systems with a total of fourteen Bravais lattices, where each point is called a lattice point. The number in parentheses refers to the number of Bravais lattices in the crystal system. (a) Three types of the cubic: simple cubic, body-centered cubic (bcc), and face-centered cubic (fcc). (b) Tetragonal: either simple or body-centered as represented by the empty circle at the center. (c) Orthorhombic: simple, body-centered, face-centered, or based-centered. (d) Hexagonal. (e) Rhombohedral (also trigonal). (f) Monoclinic: simple or base-centered as represented by the empty circles on the opposite faces. (g) Triclinic.

the top views of these two structures with the same  $a$ , which is called the *lattice constant*. Some practical examples will be given soon. If one looks at Fig. 6.4b along the diagonals, the face-centered structure becomes body-centered. However, the lattice constant would become  $a/\sqrt{2}$  along the lateral directions but remains  $a$  in the vertical direction. Such a structure is a special case of the tetragonal, because one side is not the same as the other two.



**FIGURE 6.4** Top views of (a) body-centered cubic and (b) face-centered cubic Bravais lattices. The two different filling patterns (hatched and shaded) represent lattice points on alternative layers as in Fig. 6.3a2 and Fig. 6.3a3.

There are two tetragonal Bravais lattices, the simple and the body-centered, because a face-centered tetragonal lattice can simply be rotated by  $45^\circ$  to become a body-centered one. A tetragonal lattice can be thought of as a cubic lattice stretched in one direction.

In the orthorhombic lattices shown in Fig. 6.3c, the three lattice constants:  $a$ ,  $b$ , and  $c$ , are not equal to each other. Besides the simple, body-centered, and face-centered orthorhombic lattices, there exists a base-centered lattice structure, in which two additional lattices are placed at the center of the top and bottom faces. An orthorhombic lattice can be thought of as a corresponding tetragonal lattice stretched along one side of its square. To produce the additional two, one can simply rotate the tetragonal by  $45^\circ$  and then stretch it.

A hexagonal lattice contains equal triangular or honeycomb layered structures (see Fig. 6.3d). The next three types of Bravais lattices have inclined faces (see Fig. 6.3e, f, and g). The rhombohedral (or trigonal) has equal sides, whereas the triclinic has three different sides and angles. Both contain six parallelogram faces. The monoclinic, on the other hand, has four rectangular faces and two parallelogram faces.

**Example 6-2.** Copper is an fcc lattice. Estimate the lattice constant and the distance between nearest copper atoms (or ion cores to be exact) from the density and the molecular weight of copper.

**Solution.** From Table 5.1, we have for Cu that  $\rho = 8930 \text{ kg/m}^3$  and  $M = 63.5 \text{ kg/kmol}$ . The number density of Cu atoms is  $n = \rho N_A / M = 8.47 \times 10^{28} \text{ m}^{-3}$ . If the structure were simple cubic, we would easily find that  $a = n^{-1/3} = 0.228 \text{ nm}$ , which would also be the closest distance between atoms. For an fcc lattice, there are eight corner points and six face points. If each lattice point is made of one atom, each corner point contains one-eighth of an atom and each face point contains half of an atom inside the cube. Therefore, there are four atoms inside each fcc unit cell. The number of unit cells per unit volume becomes  $n/4$  and the calculated lattice constant is  $a = 0.361 \text{ nm}$  for Cu. The closest distance between atoms is  $a/\sqrt{2} = 0.256 \text{ nm}$ . If we assume that all the atoms are rigid spheres that are closely packed (touching one another), then we can calculate the packing density or the fraction of occupied space. Assume that the diameter of an atom is  $d$ . For a simple cubic lattice,  $a = d$  and there is only one atom per lattice. Hence,  $f = (1/6)\pi d^3/a^3 = 0.52$ . For an fcc lattice,  $a = d\sqrt{2}$  and  $f = 4(1/6)\pi d^3/a^3 = 0.74$ . What is the packing density for a bcc lattice then?

Some solids with bcc or fcc lattices are listed in Table 6.1, along with others that form a hexagonal close-packed (hcp) lattice. An hcp lattice can be considered as two Bravais hexagonal lattices that are interlocked at  $c/2$ . Each lattice point is surrounded by, at equal

**TABLE 6.1** Crystal Structures and Lattice Constants of Common Elements.<sup>1,2</sup> Room Temperature Values Unless Otherwise Indicated. Note that  $1 \text{ \AA} = 0.1 \text{ nm}$ 

fcc		bcc		hcp		
Element	$a$ (Å)	Element	$a$ (Å)	Element	$a$ (Å)	$c$ (Å)
Ar (4.2 K)	5.26	Ba	5.02	H (4 K)	3.75	6.12
Ag	4.09	Cr	2.88	Be	2.27	3.59
Al	4.05	Cs (78 K)	6.05	Cd	2.98	5.62
Au	4.08	Fe	2.87	Er	3.56	5.59
Ca	5.58	K (5 K)	5.23	Gd	3.64	5.78
Ce	5.16	Li (78 K)	3.49	Mg	3.21	5.21
Cu	3.61	Mo	3.15	Ti	2.95	4.69
Pb	4.95	Na (5 K)	4.23	Tl	3.46	5.53
Pd	3.89	Nb	3.30	Y	3.65	5.73
Pt	3.92	V	3.03	Zn	2.66	4.95
Yb	5.49	W	3.16	Zr	3.23	5.15

distances, 12 neighboring points: 3 above, 3 below, and 6 at the same height. Imagine that atoms are rigid spheres with a diameter  $d$ ; we can show that  $a = d$  and  $c = d\sqrt{8/3}$  for an hcp lattice. Each sphere is in contact with 12 others. It can be seen from Table 6.1 that these hcp crystals follow the ratio  $c/a = \sqrt{8/3} \approx 1.633$  within  $\pm 16\%$ .

### 6.3.2 Primitive Vectors and the Primitive Unit Cell

A set of *primitive vectors* can be defined for Bravais lattices  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  so that the vector between any two lattice points can be expressed by

$$\mathbf{R} = m\mathbf{a} + n\mathbf{b} + l\mathbf{c} \quad (6.4)$$

where  $m$ ,  $n$ , and  $l$  are integers. For a simple cubic lattice, we can simply assign  $\mathbf{a} = a\hat{\mathbf{x}}$ ,  $\mathbf{b} = a\hat{\mathbf{y}}$ ,  $\mathbf{c} = a\hat{\mathbf{z}}$ , as can be seen from Fig. 6.3a1. However, the assignment of primitive vectors is not unique. The parallelepiped formed by the three vectors is called a *primitive unit cell*, whose volume  $V_{\text{uc}} = \mathbf{a} \times \mathbf{b} \cdot \mathbf{c}$  remains the same no matter how the primitive vectors are chosen. Taking the bcc lattice as an example, we may choose the primitive vectors as either

$$\mathbf{a} = a\hat{\mathbf{x}}, \mathbf{b} = a\hat{\mathbf{y}}, \mathbf{c} = 0.5a(\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}) \quad (6.5a)$$

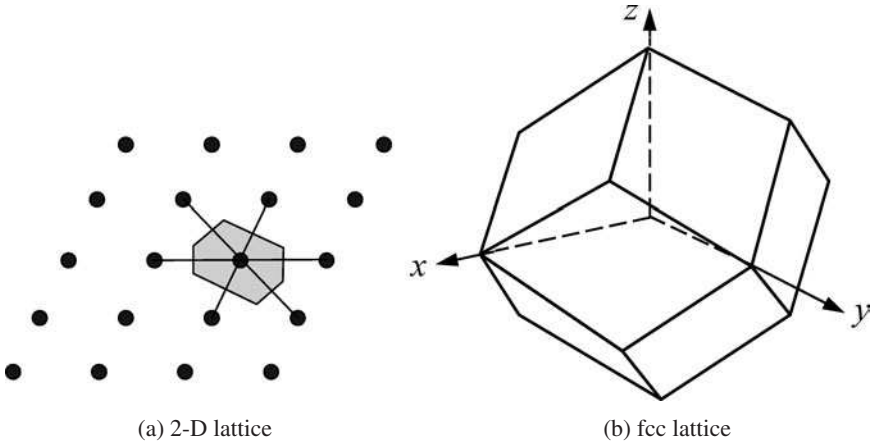
$$\text{or} \quad \mathbf{a} = 0.5a(-\hat{\mathbf{x}} + \hat{\mathbf{y}} + \hat{\mathbf{z}}), \mathbf{b} = 0.5a(\hat{\mathbf{x}} - \hat{\mathbf{y}} + \hat{\mathbf{z}}), \mathbf{c} = 0.5a(\hat{\mathbf{x}} + \hat{\mathbf{y}} - \hat{\mathbf{z}}) \quad (6.5b)$$

From Eq. (6.5b), we see that  $\mathbf{a} + \mathbf{b} + \mathbf{c}$  points to the center point and  $\mathbf{a} + \mathbf{b} = a\hat{\mathbf{z}}$ . Either way, we end up with  $V_{\text{uc}} = 0.5a^3$ , suggesting that the Bravais bcc lattice is not a primitive cell. In fact, only the simple Bravais lattices are primitive unit cells. Of course, there are other ways of choosing the primitive vectors. For a Bravais fcc lattice, we can write

$$\mathbf{a} = 0.5a(\hat{\mathbf{y}} + \hat{\mathbf{z}}), \mathbf{b} = 0.5a(\hat{\mathbf{x}} + \hat{\mathbf{z}}), \mathbf{c} = 0.5a(\hat{\mathbf{x}} + \hat{\mathbf{y}}) \quad (6.6)$$

Each vector conveniently ends at the three face-centered points. The total volume of the primitive cell becomes  $V_{\text{uc}} = 0.25a^3$ , as expected.

Another way of choosing a unit cell is to follow the two steps: (1) Draw lines to connect a given lattice point to all nearby lattice points. (2) At the midpoint and normal to these lines, draw new lines or planes. The smallest volume enclosed in this way is called the *Wigner-Seitz primitive cell*, as illustrated in Fig. 6.5. The Wigner-Seitz cell for a 2-D lattice



**FIGURE 6.5** The Wigner-Seitz cells: (a) for a 2-D lattice as shown by the shaded region and (b) for an fcc lattice as shown by the rhombic dodecahedron.

becomes a hexagon whose opposite sides are parallel, and that for an fcc lattice is a rhombic dodecahedron. The longer diagonal of each rhombic face is  $\sqrt{2}$  times that of the shorter diagonal. There are six apexes where four surfaces meet and eight apexes where three surfaces meet. The distance between opposite axes joined by four faces is exactly the Bravais lattice constant  $a$ . The axes:  $x$ ,  $y$ , and  $z$ , pass through these six apexes as well as the center. Each Wigner-Seitz cell contains only one lattice point, and it has been proven to be a primitive cell.

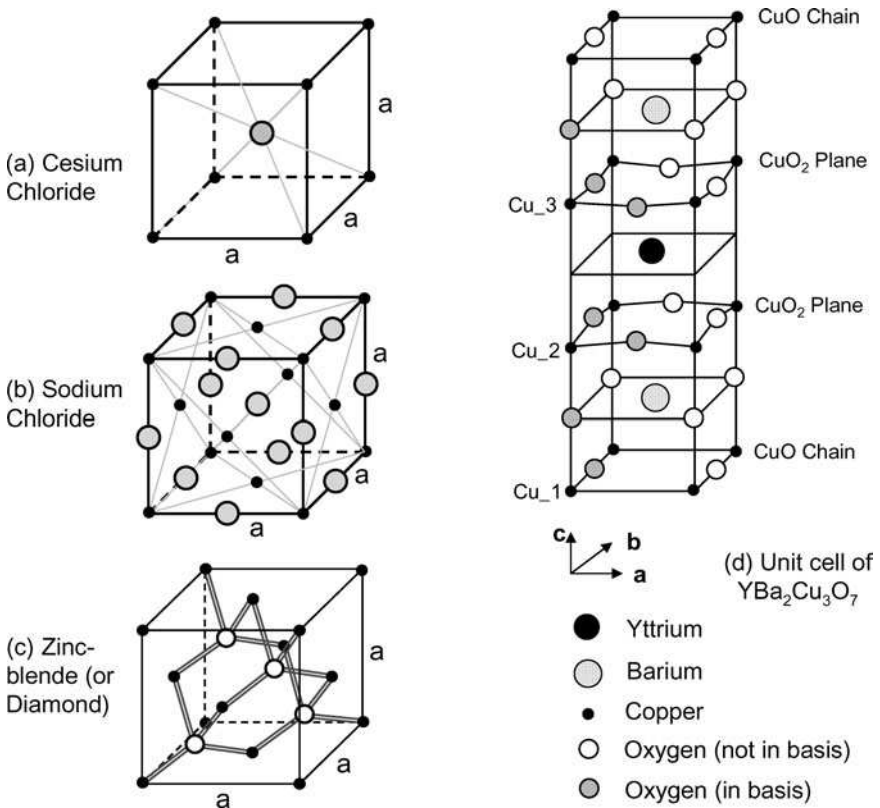
It is convenient to describe the orientation of the crystal plane by the Miller indices, which are three integers  $h$ ,  $k$ , and  $l$ , without common factors, and denoted by  $(hkl)$ . These numbers give a vector  $h\mathbf{a} + k\mathbf{b} + l\mathbf{c}$  that is perpendicular to the plane. For example, if  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  are along the  $x$ ,  $y$  and  $z$  axes, respectively, the six surfaces of the cubic unit cell are represented by  $(001)$ ,  $(00\bar{1})$ ,  $(010)$ ,  $(0\bar{1}0)$ ,  $(100)$ , and  $(\bar{1}00)$ , where a negative sign is denoted by a bar on top of the number. The whole set of surfaces can be denoted by  $\{100\}$  due to symmetry. Most commercial semiconductor wafers are  $(100)$  oriented and some  $(111)$ . The way to find the smallest  $h, k, l$  of any specified crystal facet is first to extend the plane so that it intersects the axes formed by the lattice vectors. Find the intercepts on each axis in terms of multiples of the unit cell vector, e.g.,  $(2, 4, -6)$ ; the numbers must be integers for any specified crystal plane. Take the reciprocals of these numbers, which are  $(\frac{1}{2}, \frac{1}{4}, -\frac{1}{6})$ . Multiply them by the least common multiple, which is 12 in this example. Put into the Miller indices, i.e.,  $(6, 3, \bar{2})$ . All parallel planes are characterized by the same set of Miller indices.

**Example 6-3.** Find all angles between the  $(100)$ ,  $(111)$ , and  $(311)$  surfaces in a cubic lattice.

**Solution.** For two vectors  $\mathbf{a}$  and  $\mathbf{b}$ ,  $\mathbf{a} \cdot \mathbf{b} = ab \cos \alpha = x_a x_b + y_a y_b + z_a z_b$ . Thus, the angle between  $(100)$  and  $(111)$  planes is  $\alpha = \cos^{-1}(1/\sqrt{3}) = 54.7^\circ$ ; that between  $(100)$  and  $(311)$  planes is  $\alpha = \cos^{-1}(3/\sqrt{11}) = 25.2^\circ$ ; and that between  $(111)$  and  $(311)$  planes is  $\alpha = \cos^{-1}(\frac{3+1+1}{\sqrt{11} \times 3}) = 29.5^\circ$ .

### 6.3.3 Basis Made of Two or More Atoms

With respect to the primitive vector and basis, a bcc lattice can be thought of as a simple cubic with a basis made of two atoms, one at  $(0,0,0)$  and the other at  $a(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ . Each of the eight lattice points contains the same basis by translation, according to Eq. (6.4), and the unit vectors along the three orthogonal sides of the cubic. The simple cubic lattice having a basis of two atoms, however, breaks some of the symmetry of the Bravais cubic lattice and is called a non-Bravais lattice. Lattices with a basis consisting of more than one atom have important practical applications as discussed in the following. The cesium chloride structure is made of two types of elements, each forming a simple Bravais lattice, as shown in Fig. 6.6a. The two Bravais lattices can be thought of as being placed in identical



**FIGURE 6.6** Crystalline structures. (a) Cesium chloride; (b) Sodium chloride. (c) Zincblende, which becomes a diamond structure when the atoms in the empty circles are the same as the filled ones. (d)  $\text{YBa}_2\text{Cu}_3\text{O}_7$  superconductor whose lattice constants are approximately  $a = 0.38$ ,  $b = 0.39$ , and  $c = 1.17$  nm.

positions first, and then one is moved by  $a(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$  so that the point at the origin is translated to the center of the other. It is not a body-centered cubic lattice. Rather, the crystal structure can be viewed as a simple cubic with a base of two ions, Cs at  $(0,0,0)$  and Cl at  $a(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ . The

sodium chloride structure is more common. In this case, it can be considered as two fcc lattices made of different ions. The two fcc lattices are then translated exactly the same way as in the CsCl structure. The resulting structure is shown in Fig. 6.6*b*, where each ion is surrounded by six ions of the other type. The lattice constants of some common crystals are listed in Table 6.2. It can be seen that most ionic crystals form NaCl or CsCl structures.

The crystal structures of diamond and zincblende semiconductors are also derivatives of the cubic structure. The zincblende structure is formed from two fcc lattices with different types of atoms, displaced along the body diagonal by one-quarter the length of the diagonal. Specifically, the basis is made of one atom at (0,0,0) and the other atom at  $a(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ , as shown in Fig. 6.6*c*. A total of four atoms are moved completely inside the cube, and each atom has a covalent bond with each of the four adjacent atoms, which together form a tetrahedron. Examples of zincblende structure are GaAs, SiC, and so forth. A diamond structure can be viewed as a special case of a zincblende structure for which there is only one type of element, such as C, Si, or Ge. The outermost subshell of Si is  $3s^23p^2$ , and the  $s$  subshell is filled. By promoting an  $s$ -electron to a  $p$ -state to form  $sp^3$  hybrids, four covalent bonds can be formed. This is also true for C and Ge. In essence, the diamond lattice can be thought of as an fcc lattice with a basis containing two identical atoms: one is on the corner

**TABLE 6.2** Crystal Properties of Some Compounds and Semiconductors at Room Temperature.<sup>1,2</sup> For Semiconductors, the Bandgap Energy is Indicated, and “i” in Parentheses Denotes an Indirect Bandgap

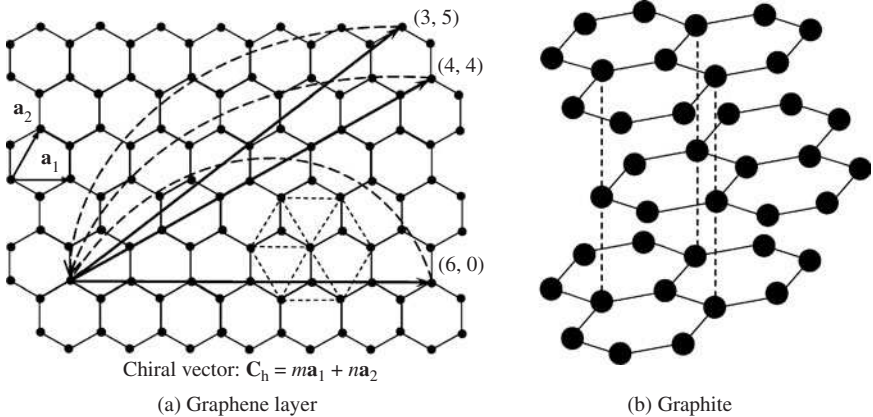
Compound		Semiconductors		
Composition	$a$ (Å)	Composition	$a$ (Å)	$E_g$ (eV)
Sodium chloride structure		Diamond structure		
LiF	4.02	C	3.57	5.47 (i)
LiCl	5.13	Si	5.43	1.11 (i)
NaBr	5.97	Ge	5.66	0.66 (i)
NaCl	5.64	Zincblende structure		
KBr	6.60	BN	3.62	7.5 (i)
KCl	6.29	CdS	5.82	2.42
CsF	6.01	CdSe	6.05	1.70
AgCl	5.55	CdTe	6.48	1.56
AgBr	5.77	GaAs	5.65	1.42
MgO	4.21	GaN (w)	5.45	3.36
MgS	5.20	GaP	5.45	2.26 (i)
CaO	4.81	GaSb	6.43	0.72
CaS	5.69	HgTe	6.04	< 0
CaSe	5.91	InAs	5.87	0.36
BaTe	6.99	InP	6.48	1.35
Cesium chloride structure		InSb	4.35	0.17
CsCl	4.12	SiC	4.63	2.36
CsBr	4.29	ZnO	5.41	3.35
CsI	4.57	ZnS	5.67	3.68
TlBr	3.97	ZnSe	6.09	2.58



and the other on the body diagonal at a distance of one-quarter diagonal. Table 6.2 also presents commonly used diamond and zincblende semiconductors with associated lattice constants and bandgap energies. Notice that GaN crystal is wurtzite in its stable form with a hexagonal symmetry. This is also the case for AlN and InN, which are not shown in the table. The III-nitride materials have a wide band, and thus are important for UV-blue-green LEDs and lasers. On the other hand, ZnS, ZnO, CdS, and CdSe can also be wurtzite. HgTe is a semimetal with a negative bandgap and can be mixed with the wideband semiconductor CdTe to form the ternary compound of  $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ , which can be used as infrared detectors, namely, mercury-cadmium-telluride (MCT) detectors.

Yttrium-barium-copper oxide ( $\text{YBa}_2\text{Cu}_3\text{O}_7$ ) is a high-temperature superconductor, which becomes superconducting at temperatures below 91 K.<sup>12</sup> It belongs to the cuprate-perovskite family and is a ceramic material when one oxygen atom is removed from the unit cell to form  $\text{YBa}_2\text{Cu}_3\text{O}_6$ .<sup>13</sup> The crystal structure of  $\text{YBa}_2\text{Cu}_3\text{O}_7$  is a simple orthorhombic lattice, whose basis contains 13 atoms, as shown in Fig. 6.6d. The structure is very close to a tetragonal one since  $a \approx b$ . The properties of  $\text{YBa}_2\text{Cu}_3\text{O}_7$  are highly anisotropic in the  $c$ -axis direction. Superconductivity is found in the  $a$ - $b$  plane, which is presumed due to the  $\text{CuO}_2$  planes above and below the yttrium atom. Other examples of Bravais lattices include As, Sb, and Bi with rhombohedral lattices; In and Sn with tetragonal lattices; and Ga, Cl, Br, and S (rhombic) with orthorhombic lattices.<sup>1</sup>

Graphite is a form of carbon made of layered structures as shown in Fig. 6.7. When separated from others, each individual layer or sheet is called a graphene. In the graphite structure,



**FIGURE 6.7** Crystal structures of (a) graphene layer and (b) graphite. Carbon nanotubes can be viewed as rolling a graphene sheet in a direction perpendicular to the chiral vector.

each carbon atom is covalently bonded to three others in the plane and loosely bonded between planes. There are relatively free electrons, and hence graphite is a conductor. The layer of graphite or graphene has a honeycomb shape, and at first sight, it may be difficult to link it with the arrays of triangles in the hexagonal lattice. It becomes more obvious, however, if a basis is chosen to contain two atoms so that a hexagon with all diagonals can be seen by the dashed lines in Fig. 6.7a. In this way, graphite can be considered as a hexagonal close-packed structure. If the basis is chosen to contain four atoms, graphite may be thought as a simple hexagonal structure.<sup>1</sup>

The structure of carbon nanotubes (CNTs) can be understood based on the graphene structure and the chiral vector,

$$\mathbf{C}_h = m\mathbf{a}_1 + n\mathbf{a}_2 \quad (6.7)$$

Different CNTs are based on rolling in the chiral vector so that the axis is perpendicular to the chiral vector and the magnitude of the chiral vector becomes the perimeter of the tube. The diameter of the tube becomes

$$d_t = \frac{C_h}{\pi} = \frac{a_{C-C}}{\pi} \sqrt{3(m^2 + mn + n^2)} \quad (6.8)$$

where  $a_{C-C} = 0.142$  nm is the nearest distance between the carbon atoms in graphene.<sup>14</sup> Notice that the chiral vector has a magnitude  $a = a_{C-C}\sqrt{3} = 0.246$  nm. In calculating the cross-sectional surface area of a single-walled CNT, one could use  $a$  as the wall thickness and obtain

$$A_c = \pi D a = 3(a_{C-C})^2 \sqrt{(m^2 + mn + n^2)} \quad (6.9)$$

Take (20,20) SWNT as an example; we have  $d_t = 2.7$  nm and  $A_c = 2.1$  nm<sup>2</sup>. Some researchers suggested using a layer thickness equal to the separation of graphite as 0.335 nm, which gives  $A_c = \pi D \times 0.335 = 2.9$  nm<sup>2</sup>. Note that for a solid wire  $A_c = \pi D^2/4 = 5.8$  nm<sup>2</sup>.

## 6.4 ELECTRONIC BAND STRUCTURES

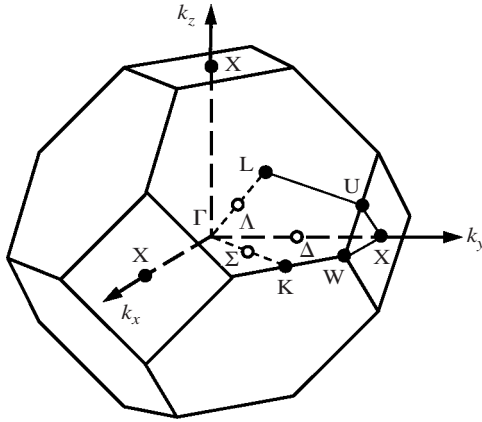
The behavior of electrons in solid is complicated because the solution of wavefunctions involves a rather complicated many-body problem. Electrons in solids can be thought of as in a periodic potential due to the periodic arrays of atoms. Electronic band structures are functions that describe the electron states in the energy versus wavevector space. Let us first look at the reciprocal lattice in three dimensions.

### 6.4.1 Reciprocal Lattices and the First Brillouin Zone

The reciprocal lattice of a crystal structure is defined in the  $\mathbf{k}$ -space (wavevector-space). Since a crystal is a periodic array of lattices in real space, the reciprocal lattice can be obtained by performing a spatial Fourier transform of the crystal. For a simple orthorhombic lattice with the primitive vectors  $\mathbf{a} = a\hat{\mathbf{x}}$ ,  $\mathbf{b} = b\hat{\mathbf{y}}$ , and  $\mathbf{c} = c\hat{\mathbf{z}}$ , the reciprocal lattice can be defined by the three vectors  $\mathbf{A} = (2\pi/a)\hat{\mathbf{x}}$ ,  $\mathbf{B} = (2\pi/b)\hat{\mathbf{y}}$ , and  $\mathbf{C} = (2\pi/c)\hat{\mathbf{z}}$ , which define another orthorhombic. The product of the volumes of the unit lattice and the reciprocal lattice is  $8\pi^3$ . Some of this aspect was discussed in Chap. 5. In general, the reciprocal primitive vectors can be generated by

$$\mathbf{A} = 2\pi \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})}; \quad \mathbf{B} = 2\pi \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})}; \quad \mathbf{C} = 2\pi \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \quad (6.10)$$

In solid state physics, a Brillouin zone is defined as a Wigner-Seitz cell in the reciprocal lattice and the smallest of which is called the first Brillouin zone. The definition of the Brillouin zone gives a vivid geometric interpretation of the Bragg diffraction condition and thus is of importance in the study of electron and phonon states in crystals, as well as their interactions with electromagnetic waves. Figure 6.8 shows the first Brillouin zone of a face-centered cubic lattice. The directions  $k_x$ ,  $k_y$ , and  $k_z$  are called the [100], [010], and [001]



**FIGURE 6.8** The first Brillouin zone of a face-centered cubic structure. The shape is a truncated octahedron with eight hexagons and six squares. This is also the Wigner-Seitz cell for a bcc lattice, whose first Brillouin zone has the same shape as the Wigner-Seitz cell for an fcc lattice shown in Fig. 6.5*b*.

directions, respectively. The center of the Brillouin zone is called the  $\Gamma$ -point, and the intersection of the three axes with the zone edge is called the X-point. The *body diagonal*, or the [111] direction, meets the zone edge at the L-point. Other points of interest such as K, W, and U at the zone edges and  $\Delta$ ,  $\Lambda$ , and  $\Sigma$ , located halfway between the zone center and an edge, can also be defined.

### 6.4.2 Bloch's Theorem

The total potential in a crystal includes the core-core, electron-electron, and electron-core Coulomb interactions. For solving electron wavefunctions subjected to such a potential, one would have to deal with a many-body problem, which turned out to be very difficult in mathematics. However, this problem can be simplified using the so-called nearly free electron model, in which each electron moves in the average field created by the other electrons and ions. This is also called the one-electron model. The Hamiltonian operator  $H$  for the one-electron model is given as

$$H = \frac{p_e^2}{2m_e} + U(\mathbf{r}) \quad (6.11)$$

where  $p_e$  and  $m_e$  are the momentum and the mass of the electron, respectively, and  $U(\mathbf{r})$  is a periodic potential function resulted from both the electron-electron and electron-core interactions. The one-electron Schrödinger equation is therefore (see Sec. 3.5.1)

$$\left[ -\frac{\hbar^2}{2m_e} \nabla^2 + U(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}) \quad (6.12)$$

where  $E$  is the electron energy and  $\psi(\mathbf{r})$  is the electron wavefunction. The periodicity of the lattice structure yields the boundary condition,

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \quad (6.13)$$

where  $\mathbf{R}$  is the vector between two lattice points. Note that  $U(\mathbf{r})$  is called the periodic potential and can be expanded as a Fourier series in terms of the reciprocal lattice vector  $\mathbf{G}$  as follows:

$$U(\mathbf{r}) = \sum_{\mathbf{G}} U_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} \quad (6.14)$$

The reciprocal lattice vector can be expressed as  $\mathbf{G} = l_1\mathbf{A} + l_2\mathbf{B} + l_3\mathbf{C}$ , where  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  are primitive vectors of the reciprocal lattice as given in Eq. (6.10), and the integers  $l_1$ ,  $l_2$ , and  $l_3$  are indices. In Eq. (6.14),  $U_{\mathbf{G}}$ 's are complex Fourier expansion coefficients for a given set of  $l_1$ ,  $l_2$ , and  $l_3$ .

According to the Bloch theorem, the wavefunction of an electron in a periodic potential must have the form:

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \quad (6.15)$$

where  $u_{\mathbf{k}}(\mathbf{r})$  is a periodic function with the periodicity of the lattice, similar to Eq. (6.13), and thus  $\psi(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \psi(\mathbf{r})$ . The wavefunction  $\psi(\mathbf{r})$  can also be expressed as a Fourier series summed over all values of the permitted wavevector such that

$$\psi(\mathbf{r}) = \sum_{\mathbf{k}} C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (6.16)$$

The summation is over all wavevectors  $\mathbf{k}$ 's. From Eq. (6.16), we have

$$\nabla^2 \psi(\mathbf{r}) = \sum_{\mathbf{k}} C_{\mathbf{k}} (i\mathbf{k})^2 e^{i\mathbf{k}\cdot\mathbf{r}} = - \sum_{\mathbf{k}} k^2 C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (6.17)$$

The combination of Eq. (6.14) and Eq. (6.16) gives

$$U(\mathbf{r})\psi(\mathbf{r}) = \sum_{\mathbf{k}} \sum_{\mathbf{G}} U_{\mathbf{G}} C_{\mathbf{k}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (6.18)$$

Using Eq. (6.16) through Eq. (6.18), we can rewrite the Schrödinger equation as follows:

$$\sum_{\mathbf{k}} \frac{\hbar^2 k^2}{2m_e} C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} + \sum_{\mathbf{k}} \sum_{\mathbf{G}} U_{\mathbf{G}} C_{\mathbf{k}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} = \sum_{\mathbf{k}} E C_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \quad (6.19)$$

The coefficients of each Fourier component must be equal on both sides of the equation. Thus,

$$\left( \frac{\hbar^2 k^2}{2m_e} - E \right) C_{\mathbf{k}} + \sum_{\mathbf{G}} U_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} = 0 \quad (6.20)$$

where  $C_{\mathbf{k}-\mathbf{G}}$  is the coefficient for the term with  $(\mathbf{k} - \mathbf{G})$  in the exponent, i.e.,  $e^{i(\mathbf{k}-\mathbf{G})\cdot\mathbf{r}}$  in Eq. (6.16). Equation (6.20) is paramount in the electronic band theory of crystals, and it called the *central equation* according to Kittel.<sup>2</sup> When  $U(\mathbf{r}) \equiv 0$ , Eq. (6.20) reduces to  $E_{\mathbf{k}}^0 = \hbar^2 k^2 / (2m_e)$  by noting that  $\hbar k = p_e$  for free electrons, as used in the Sommerfeld theory. Under the influence of a periodic potential, the relationship becomes more complex because it is a set of linear equations for infinite numbers of coefficients. Because the equation is homogeneous, the determinant of the characteristic matrix must be zero. In some cases, the terms can be significantly reduced to yield simple solutions with insightful physics.

Consider the 1-D case when the Fourier components are relatively small compared with the kinetic energy of electrons at the zone boundary. This is the weak-potential assumption. At the first Brillouin zone boundaries, we have

$$k = G/2 = \pi/a \quad (6.21)$$

Because there are only two values of  $k$  and  $G$ , Eq. (6.20) reduces to the following two equations due to symmetry:

$$(E_{\mu}^0 - E)C_{\mu} + UC_{-\mu} = 0 \quad (6.22a)$$

and

$$(E_{-\mu}^0 - E)C_{-\mu} + UC_{\mu} = 0 \quad (6.22b)$$

where  $\mu = \frac{1}{2}G$  is introduced merely for the convenience of notation. These equations have solutions only when the determinant becomes zero, i.e.,

$$\begin{vmatrix} E_{\mu}^0 - E & U \\ U & E_{-\mu}^0 - E \end{vmatrix} = 0 \quad (6.23)$$

Because  $E_{\mu}^0 = E_{-\mu}^0 = \hbar^2\mu^2/(2m_e)$ , the two roots are then obtained as

$$E = E_{\mu}^0 \pm U = \frac{(\hbar\pi/a)^2}{2m_e} \pm U \quad (6.24)$$

The two solutions at the zone edge, i.e.,  $k = \pi/a$ , are actually on two  $E(k)$  curves. When  $k$  is near the zone edge, we can express the central equation, Eq.(6.20), as the following two equations:<sup>1,2</sup>

$$(E_k^0 - E)C_k + UC_{k-G} = 0 \quad (6.25a)$$

and

$$(E_{k-G}^0 - E)C_{k-G} + UC_k = 0 \quad (6.25b)$$

By setting its determinant to zero, we obtain

$$E(k) = \frac{1}{2}(E_k^0 + E_{k-G}^0) \pm \left[ \frac{1}{4}(E_k^0 - E_{k-G}^0)^2 + U^2 \right]^{1/2} \quad (6.26)$$

which gives two branches near the zone edge, as shown in Fig. 6.9. A bandgap of  $2U$  is formed at the first Brillouin zone edge. The corresponding wavefunctions at the zone edge are

$$\psi_{1,2}(x) = \frac{1}{\sqrt{2L}}(e^{i\pi x/a} \pm e^{-i\pi x/a}) \quad (6.27a)$$

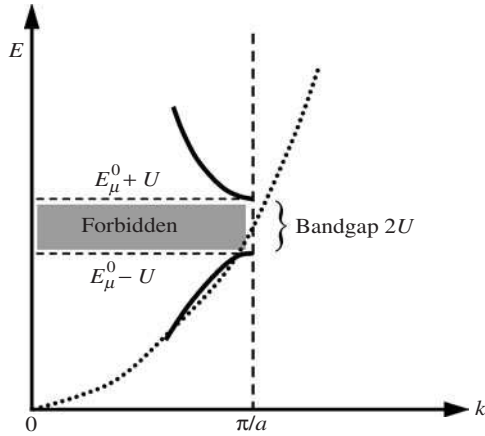
where  $L$  is the length of the crystal. This forms two standing waves:

$$\psi_1(x) = \sqrt{2/L}\cos(\pi x/a) \quad \text{and} \quad \psi_2(x) = i\sqrt{2/L}\sin(\pi x/a) \quad (6.27b)$$

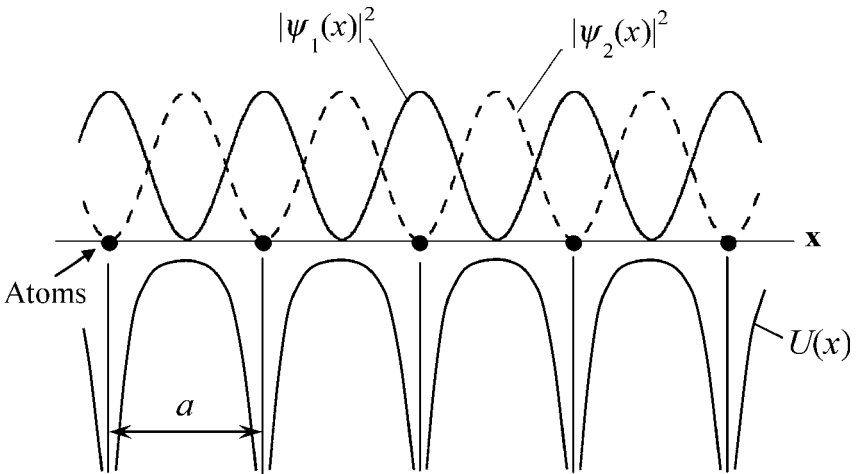
The lower energy state  $E_{\mu}^0 - U$  corresponds to  $\psi_1$  with a probability density  $|\psi_1|^2$  peaked at core sites, as shown in Fig. 6.10. The probability density function describes electrons that are piled up close to the core site. The upper energy state  $E_{\mu}^0 + U$  corresponds to  $\psi_2$  with a probability density  $|\psi_2|^2$  that distributes electrons between the cores. The energy difference between these two states is the origin of formation of the gap at the Brillouin zone edge. On the other hand, away from the zone edge, the electron wavefunction can be expressed as

$$\psi(x) \approx L^{-1/2}e^{\pm ikx} \quad (6.28)$$

which are propagating waves that characterize the wavelike behavior of free electrons.<sup>2,15</sup>

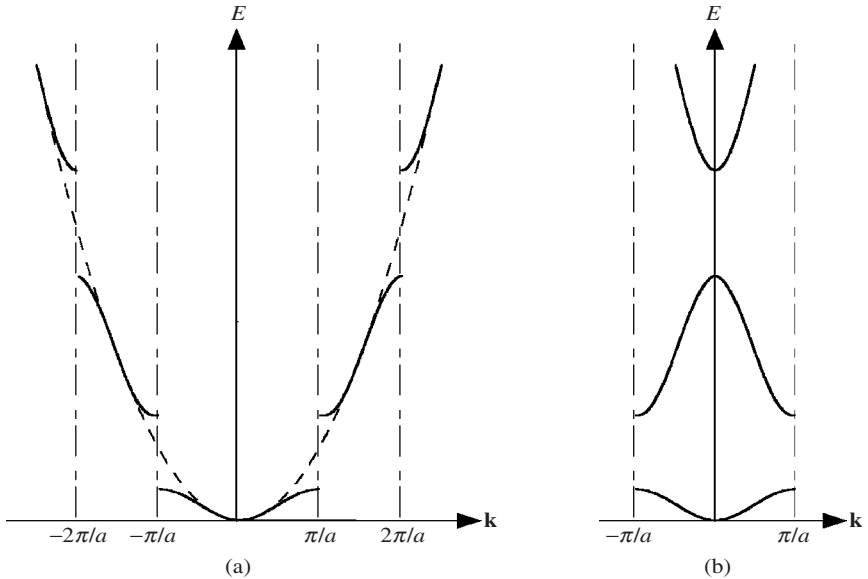


**FIGURE 6.9** Plot of the energy bands, where the solid curves are based on Eq. (6.26). The lower and upper bands correspond to the choice of the minus and plus signs, respectively. When  $k = G/2 = \pi/a$ , the two bands are separated by a gap of magnitude  $2U$ . The dotted line, on the other hand, represents free-electron behavior  $E \propto k^2$ .



**FIGURE 6.10** The upper part of the figure plots the probability density  $|\psi|^2$  in a 1-D weak potential at the edge of the first Brillouin zone; the lower part of the figure illustrates the actual potential  $U(x)$  of electrons.

When all the Brillouin zones and their associated Fourier components are included, the result is a set of curves, as those shown in Fig. 6.11a. An easier way to show this is to use the Kronig-Penney model, first formulated in 1931, in which the potential is assumed to be a square-well array.<sup>2,9</sup> The details are left as an exercise (see Problem 6.12). The allowable bands are illustrated by the solid curves in Fig. 6.11. If the electron were completely



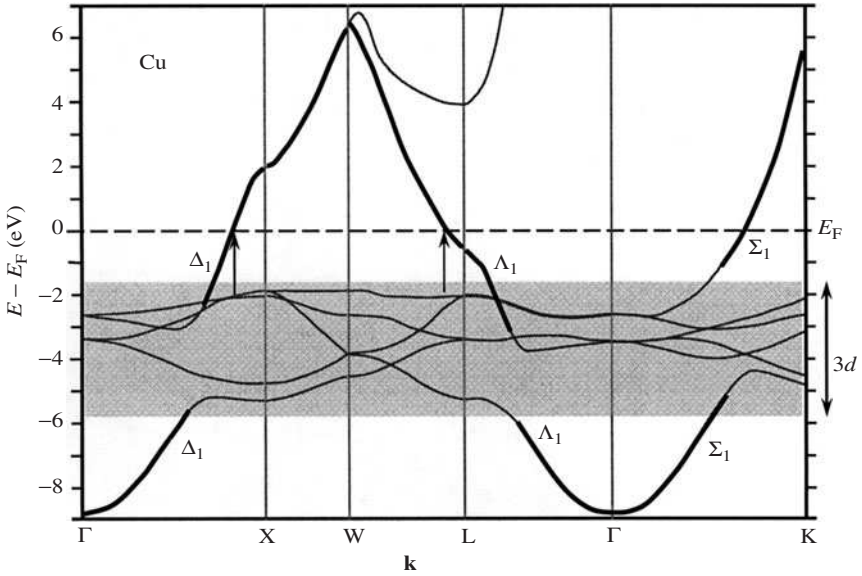
**FIGURE 6.11** Representation of the electronic band structure. (a) The extended-zone scheme. (b) The reduced-zone scheme.

free, then  $E(k) = E_k^0 = \hbar^2 k^2 / (2m_e)$  would be a parabola, as illustrated by the dashed curve in Fig. 6.11a, without any bandgap. It is useful to plot all the energy levels in the first Brillouin zone. This can be done by folding the branches in Fig. 6.11a, which is known as the *extended-zone scheme*, using the reciprocal lattice vector. The result is shown in Fig. 6.11b, which is called the *reduced-zone scheme* for the representation of the electronic bands.

### 6.4.3 Band Structures of Metals and Semiconductors

The nearly free electron model described in the previous section assumes a weak potential and cannot predict the behavior of electrons in the inner orbits or near the nuclei. A simple way to calculate the electronic structure of inner electrons, such as those in the  $d$  subshells, is the *tight-binding method*, which assumes that the potential is so large that electrons can hardly move out of the ion core. Due to the complicated 3-D structure and the multiple number of outermost electrons in each atom, the actual electronic band structures are rather complicated. More advanced methods include the augmented plane-wave (APW) method, the Korringa-Kohn-Rostoker (KKR) Green function method, and the pseudopotential method. More details can be found from Ashcroft and Mermin,<sup>1</sup> Kittel,<sup>2</sup> and Omar,<sup>15</sup> and references therein.

It can be shown that the number of orbits in a band in the first Brillouin zone is the same as the number of unit cells in the crystal,  $N$ . According to the Pauli exclusion principle, the number of electrons that can occupy a band is  $2N$ . For copper, the outermost electron configuration is  $4s^1 3d^{10}$ . The  $s$ - and  $d$ -subshell electrons result in six bands (with some overlap), as can be seen from Fig. 6.12, along the direction according to the first Brillouin zone depicted in Fig. 6.8.<sup>16-18</sup> The  $d$  bands are from 2 to 5.5 eV below the Fermi level and are completely filled. The  $s$  band, illustrated by the thicker line segments, is interrupted by the  $d$  bands. The



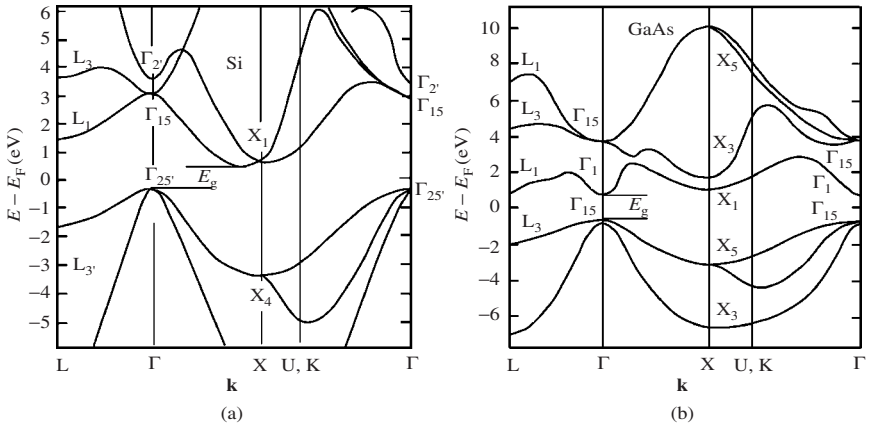
**FIGURE 6.12** Calculated energy band structure of copper, adapted from Segall,<sup>16</sup> Burdick,<sup>17</sup> and Hummel.<sup>18</sup>

$s$  band is only half filled and half empty. For alkali metals, there is only one valence electron and the  $s$  band is continuous. Electrons in the  $s$  band can be easily excited from below the Fermi level to above the Fermi level within the same band. This explains why copper is a conductor. When radiation is incident on a copper surface, because of the relatively high frequency, free electrons have an inductive characteristic and tend to reflect the radiation. The absorption of photons will cause the electrons in the  $s$  band to reach a higher level within the same band. If the phonon energy exceeds 2 eV, transition from the top  $d$  band to the  $s$  band right above the Fermi level is possible, as indicated by the two arrows in Fig. 6.12. The *interband transitions* result in strong absorption as well as a reduction in reflection of copper at wavelengths shorter than about 0.6  $\mu\text{m}$ . Pure copper has a red-brown color because it does not reflect blue and violet colors. Gold has a similar interband transition that absorbs short-wavelength visible light. On the other hand, for silver, the interband transition occurs at a much shorter wavelength. Thus, silver can reflect light in the whole visible spectrum.

The Fermi surface is anisotropic and not spherical for real crystals. For alkali metals with bcc lattices, such as Na and K, the Fermi surface is nearly spherical lying inside the first Brillouin zone.<sup>1</sup> The Fermi surface of Al is close to the free electron surface for an fcc lattice with three conduction electrons per atom. For noble metals, due to the effect of  $d$  bands, the Fermi surface is characterized by a sphere that bulges out in the eight  $\langle 111 \rangle$  directions.

The electronic band structures of Si and GaAs in the first Brillouin zone are shown in Fig. 6.13, along reciprocal lattice directions.<sup>19–21</sup> Si and GaAs are chosen here because these two types of semiconductors have distinct energy gap features that can represent a wide range of semiconductor materials. Degeneracy causes additional subbands within the conduction and valence bands. *Intraband transitions* refer to the excitation or relaxation of electrons between subbands. For intrinsic semiconductors, the Fermi level lies right in the





**FIGURE 6.13** Calculated energy band structure of (a) silicon and (b) gallium arsenides, adapted from Cohen and Bergstresser,<sup>19</sup> Herman and Spicer,<sup>20</sup> and Cohen and Chelikowsky.<sup>21</sup>

middle between the bottom of the conduction band and the top of the valence band. The valence bands are formed by the bonded valence electrons, and they are completely filled at low temperatures. The electrons in the conduction band are dissociated from the atom and hence become free charges. The bandgap energy, or energy gap,  $E_g$  is the difference between the energies at the top of the valence band ( $E_V$ ) and the bottom of the conduction band ( $E_C$ ). The values of  $E_g$  for some semiconductors are included in Table 6.2. For Si, as shown in Fig. 6.13a, the bottom of the conduction band and the top of the valence band do not occur at the same  $k$ . This type of semiconductor is called an *indirect gap* semiconductor. For a *direct gap* semiconductor, such as GaAs, the bottom of the conduction band and the top of the valence band occur at the same value of  $k$  at the  $\Gamma$ -point, as shown in Fig. 6.13b. The mechanism for electron transition between the valence band and the conduction band in a direct gap semiconductor is completely different from that in an indirect gap semiconductor. Additional discussion about radiation absorption will be given in Chap. 8.

At absolute zero temperature, there are no electrons in the conduction band and the valence band is completely filled. When the temperature increases or there exist optical excitations, electrons in the valence band can transit to the conduction band, leaving behind some vacancies in the valence band. The vacancies left in the valence band are called *holes*, which have the same mass but opposite charge as electrons. Usually the electrons are found almost exclusively in levels near the conduction band minima, while the holes are found in the neighborhood of the valence band maxima. Therefore, the energy versus wavevector relations for the carriers can generally be approximated by quadratic forms in the neighborhood of such extrema, i.e.,

$$E_e(k) = E_C + \frac{\hbar^2 k^2}{2m_e^*} \quad \text{and} \quad E_h(k) = E_V - \frac{\hbar^2 k^2}{2m_h^*} \quad (6.29)$$

where subscript e and h are for electrons and holes, respectively,  $E_C$  is the energy at the bottom of the conduction band, and  $E_V$  is the energy at the top of the valence band. In the 1-D case, the effective mass  $m^*$  for electrons and holes are defined as

$$\frac{1}{m_e^*} = \frac{1}{\hbar^2} \frac{d^2 E_e}{dk^2} \quad \text{and} \quad \frac{1}{m_h^*} = -\frac{1}{\hbar^2} \frac{d^2 E_h}{dk^2} \quad (6.30)$$

where the negative sign is assigned to make the effective mass of the hole positive at the top of the valance band. Effective mass is defined based on the quantum mechanical description of the group velocity and the acceleration of charge carriers, respectively, as

$$v_g = \frac{1}{\hbar} \frac{\partial E}{\partial k} \quad \text{and} \quad a = \frac{dv_g}{dt} = \frac{1}{\hbar} \frac{\partial^2 E}{\partial k^2} \frac{dk}{dt} = \frac{1}{\hbar^2} \frac{\partial^2 E}{\partial k^2} F \quad (6.31)$$

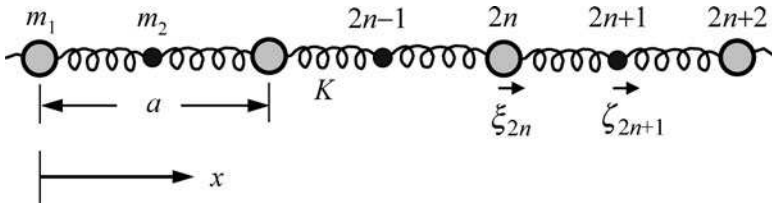
where  $F = dp_g/dt = \hbar dk/dt$  is the force exerted on the charge carrier due to an electric field. In 3-D case, the effective mass depends also on the direction and is a  $3 \times 3$  tensor.<sup>15</sup>

## 6.5 PHONON DISPERSION AND SCATTERING

In the above discussion of electronic band structures, it is assumed that the cores of atoms are fixed. In a real crystal, however, the cores of atoms are vibrating about their equilibrium positions and the vibration of atoms has an important influence on energy storage and transport in crystals. Lattice vibration causes elastic waves to propagate in crystalline solids. Phonons are the energy quanta of lattice waves. For a given vibration frequency  $\omega$ , the energy of a phonon  $\hbar\omega$  is the smallest discrete value of energy. Thermal vibrations in crystals are thermally excited phonons, like the thermally excited photons in a blackbody cavity.

### 6.5.1 The 1-D Diatomic Chain

Phonon dispersion describes the relationship between the vibration frequency and the phonon wavevector. A simple example is given first for a diatomic chain of linear spring-mass arrays, as shown in Fig. 6.14. It is assumed that the spring constant  $K$  is the same



**FIGURE 6.14** A chain of two atoms with different masses  $m_1$  and  $m_2$  linked by springs of the same spring constant  $K$ , where  $\xi$  and  $\zeta$  denote the displacements of individual atoms from their equilibrium positions.

between the nearest-neighbor atoms. The spring is a conceptual representation of the combined attractive and repulsive forces, which can be assumed linear if the displacement is sufficiently small. Anharmonic vibrations may become significant at high temperatures. Another assumption of the nearest-neighbor model is that the forces on an atom come from the nearest neighbors only.<sup>22</sup> The equation of motion of the atoms can be written as follows:

$$m_1 \frac{d^2 \xi_{2n}}{dt^2} = K(\zeta_{2n+1} + \zeta_{2n-1} - 2\xi_{2n}) \quad (6.32a)$$

and

$$m_2 \frac{d^2 \zeta_{2n+1}}{dt^2} = K(\xi_{2n+2} + \xi_{2n} - 2\zeta_{2n+1}) \quad (6.32b)$$

where  $\xi_{2n}$  is the displacement of the atom with mass  $m_1$  indexed by an even number and  $\xi_{2n+1}$  is the displacement of the atom with mass  $m_2$  indexed by an odd number.<sup>23</sup> To solve these equations, let  $\xi_{2n} = A_1 \exp[i(nka - \omega t)]$  and  $\xi_{2n+1} = A_2 \exp[i(n + 1/2)ka - i\omega t]$ , and substitute them into Eq. (6.32). After some manipulations, we obtain the following equations:

$$(2K - m_1\omega^2)A_1 - 2K \cos(ka/2)A_2 = 0 \quad (6.33a)$$

$$(2K - m_2\omega^2)A_2 - 2K \cos(ka/2)A_1 = 0 \quad (6.33b)$$

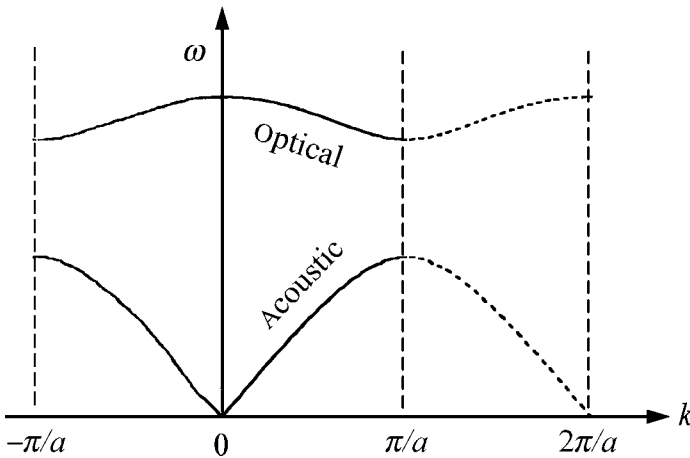
The determinant must be zero, viz.,

$$\begin{vmatrix} 2K - m_1\omega^2 & -2K \cos(ka/2) \\ -2K \cos(ka/2) & 2K - m_2\omega^2 \end{vmatrix} = 0 \quad (6.34)$$

which gives  $m_1 m_2 \omega^4 - 2K(m_1 + m_2)\omega^2 + 4K^2[1 - \cos^2(ka/2)] = 0$ , and its two roots for  $\omega^2$  are

$$\omega^2 = K \left( \frac{1}{m_1} + \frac{1}{m_2} \right) \pm K \left[ \left( \frac{1}{m_1} + \frac{1}{m_2} \right)^2 - \frac{4 \sin^2(ka/2)}{m_1 m_2} \right]^{1/2} \quad (6.35)$$

The resulting  $\omega - k$  curves are the dispersion relations, as shown in Fig. 6.15. Two branches are formed when  $m_1 \neq m_2$ . The upper branch that corresponds to the plus sign is



**FIGURE 6.15** Phonon dispersion of the linear diatomic chain, calculated by the nearest-neighbor model. The first Brillouin zone is between  $-\pi/a$  and  $\pi/a$ .

called the *optical phonon branch*, or simply *optical branch*, because it is important for infrared activities in ionic solids. The lower branch that corresponds to the minus sign is called the *acoustic branch*. At very low frequencies, the atoms in the unit cell move in phase with each other. Such a behavior is characteristic for a sound wave.

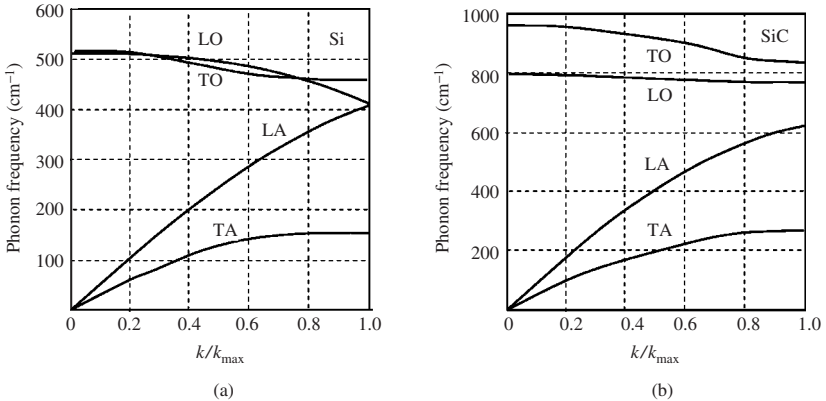
It can be seen that the dispersion curves vary periodically with  $k$ . The results outside the first Brillouin zone merely reproduce lattice dynamics that can be fully described by the dispersion curves in the first Brillouin zone. Due to the periodicity of the solution in terms of  $k$ , we may treat a value of  $k$  outside the first Brillouin zone by subtracting an appropriate integer times the reciprocal lattice constant  $2\pi/a$  to give a value of  $k$  within the limits of the first Brillouin zone. Given that  $|k| \leq \pi/a$ , the phonon wavelength is specified by

$$\lambda = \frac{2\pi}{k}, 2a \leq \lambda < \infty \quad (6.36)$$

This makes perfect sense as the wavelength should not be smaller than the lattice constants, as explained in previous chapter (see Fig. 5.3). For solids with small dimensions, there is also a limit of the maximum wavelength  $2L$ . For  $k \ll \pi/a$ , the acoustic branch gives  $\omega \propto k$ , which is a linear dispersion relation. At  $k = \pi/a$ ,  $\omega = \sqrt{2K/m_1}$  and  $\sqrt{2K/m_2}$ , and the two branches are separated when  $m_1 \neq m_2$ . In this case, it should be noticed that the group velocity  $v_g \equiv d\omega/dk = 0$ . Only standing waves exist. If  $m_1 = m_2$ , then the upper and lower branches will be continuous at  $k = \pi/a$  and the slope is not zero. However, the lattice constant needs to be modified to  $a/2$  in Fig. 6.14, and thus, the range of the first Brillouin zone is between  $-2\pi/a$  and  $2\pi/a$ . The upper branch should be unfolded at  $k = \pi/a$  to connect smoothly with the lower branch. The result is a single branch covering the first Brillouin zone. The proof is left as an exercise.

## 6.5.2 Dispersion Relations for Real Crystals

The above discussion can be extended to 3-D systems, in which lattice vibrations allow both transverse and longitudinal modes. For the case of two atoms per primitive cell, there are one longitudinal and two transverse branches for both acoustic and optical vibration modes. The phonon dispersion relations for silicon and silicon carbide are shown in Fig. 6.16.<sup>25–28</sup>



**FIGURE 6.16** Optical and acoustical branches of phonon dispersion. (a) Si [100] direction, adapted from Brockhouse,<sup>25</sup> Dolling,<sup>26</sup> and Tubino et al.<sup>27</sup> (b) SiC, adapted from Feldman et al.<sup>28</sup>

Experimental determination of the phonon dispersion curves were made with neutron scattering<sup>25,26</sup> for Si and Raman scattering for SiC.<sup>28</sup> Because  $m_1 = m_2$  for Si, the longitudinal optical (LO) and longitudinal acoustic (LA) branches meet at the zone edge and thus the

group velocity is not equal to zero there. For SiC on the other hand, the two roots in Eq. (6.35) are different because  $m_1 \neq m_2$ . There exists a frequency gap between the LO and LA branches at the zone edge. The frequency gap is forbidden for propagating waves, i.e., no phonons can propagate at frequencies within the gap, similar to the bandgap for electrons. The group velocities of LO and LA phonon modes are zero at the zone edge; this can be seen by the flat dispersion curves. One should not worry about the negative or positive sign of the group velocity as it is merely a result of folding the dispersion curves. The group velocity is always in the direction of energy transfer. It should be mentioned that the speed of sound and the phonon propagation speed refer to the group velocity, not to the phase velocity.

According to the wave-particle duality, a phonon with energy  $\hbar\omega$  should also have an associated momentum, given by

$$\mathbf{p} = \hbar\mathbf{k} \quad (6.37)$$

where  $\mathbf{k}$  is the wavevector of the phonon. There is a distinction between phonons and photons. Photons do not carry any physical momentum because the physical momentum associated with lattice vibration is zero, except when all lattices are in phase. On the other hand, when interacting with other elementary particles, such as electrons or photons, the wavevector must follow the selection rule such that it looks as if a phonon has a real momentum given by Eq. (6.37). This momentum is often called the *crystal momentum*.<sup>1,2</sup>

The group velocity of phonons in the optical branches is usually small, and subsequently, optical phonons contribute little to the thermal conduction in solids. On the other hand, optical phonons can interact or scatter with acoustic phonons, especially at elevated temperatures, to reduce the thermal conductivity.<sup>23</sup> Although LA phonons have higher group velocities than TA phonons, one must consider also the frequency distribution of phonons since phonons obey Bose-Einstein statistics [see Eq. (5.71) and discussions in Chap. 5]. At low temperatures, TA phonons are dominant contributors to the heat conduction as well as the specific heat of insulators and semiconductors. As the temperature goes up, LA phonons become important. While optical phonons contribute little to the heat conduction, they contribute about half of the heat capacity above room temperature. This is because group velocity does not enter the equation for specific heat [see Eq. (5.30)]. In general, if there are  $q$  atoms in the primitive cell or basis, there will be one longitudinal and two transverse acoustic branches, and  $q - 1$  longitudinal and  $2(q - 1)$  transverse optical branches. However, degeneracy of the transverse branches may occur due to symmetry.<sup>22-24</sup> An example of complex materials is the family of zeolites, which are hydrated aluminosilicate minerals that exhibit nanoporous crystalline structures. Zeolites have important applications as filters, catalysts, solar collector, and adsorption refrigeration. Greenstein et al. studied the thermal properties of MFI zeolite films considering the phonon dispersion.<sup>29</sup> MFI is a special type of zeolite that has ordered channel directions and an average pore size of 0.6 nm. The calculation of specific heat and thermal conductivity involved summation over 864 polarizations (phonon branches) over all wavevectors in the first Brillouin zone. The modeling results were in reasonable agreement with experiments.<sup>29</sup>

Another important aspect of phonon transport is scattering. The mean free path of phonons is often small compared with the size of crystals. For nanostructures, on the contrary, the mean free path can be larger than the characteristic length, resulting in boundary scattering. Some qualitative discussions have been given in the previous chapter. A summary of the characteristics of phonon and photon is given in Table 6.3. In most situations, phonons are treated as particles, especially in dealing with interactions among phonons themselves as well as with electrons, photons, and defects. For long-wavelength phonons, lattice vibration can also be described by a sound wave or an acoustic wave of three polarizations. To analyze the acoustic wave behavior, the crystal is viewed as a continuous medium because the individual vibration of atoms is not of interest. Acoustic mismatch theory is based on the reflection,

**TABLE 6.3** Comparison of the Characteristics of Phonon and Photon

Phonon	Photon
Bose-Einstein statistics	Bose-Einstein statistics
Massless	Massless
Energy: $\varepsilon = h\nu$	Energy: $\varepsilon = h\nu$
Phase speed: $v_p = \omega/k$	Phase speed: $v_p = \lambda\nu$
Mechanical vibration (existence in solids and some liquids, such as liquid helium)	Electromagnetic waves (existence in any medium as well as in vacuum)
Both transverse and longitudinal	Transverse only
Crystal momentum: $\mathbf{p} = \hbar\mathbf{k}$	Physical momentum: $\mathbf{p} = \hbar\mathbf{k}$
Frequency: less than $\approx 50$ THz	Frequency: no limit
Group velocity: less than $\approx 2 \times 10^4$ m/s	Group velocity: order of $10^8$ m/s
Mean free path: $\approx 10$ to $100$ nm (except at very low temperatures and in nanotubes)	Mean free path: no limit (largely dependent on the medium)

transmission, and emission of acoustic waves to predict the thermal boundary resistance, to be discussed in Chap. 7. A brief discussion of the microscopic conservation or selection rules during scattering events involving phonons is presented next.

### 6.5.3 Phonon Scattering

Phonon scattering governs the thermal transport properties of dielectric and semiconductor materials. Proper modeling of phonon scattering is important for the application of the Boltzmann transport equation (BTE), considering the frequency-dependent scattering rate. The anharmonic nature of the interatomic potential offers a coupling mechanism for phonon-phonon interactions, which was not included in the linear oscillator model. The phonon-phonon scattering is inelastic because the phonon frequency before the scattering event is different from that after the event. The energy conservation requires the scattering to involve at least three phonons. A three-phonon process is mostly common since the probability is usually much larger than the values for processes involving four or more phonons. In a three-phonon process, either two phonons interact to form a third one, or one phonon breaks into two others. The phonon energy and crystal momentum are conserved as given by<sup>1,2</sup>

$$\hbar\omega_1 + \hbar\omega_2 = \hbar\omega_3 \quad \text{or} \quad \hbar\omega_1 = \hbar\omega_2 + \hbar\omega_3 \quad (6.38)$$

$$\hbar\mathbf{k}_1 + \hbar\mathbf{k}_2 = \hbar\mathbf{k}_3 \quad \text{or} \quad \hbar\mathbf{k}_1 = \hbar\mathbf{k}_2 + \hbar\mathbf{k}_3 \quad (6.39)$$

In Eq. (6.38) and Eq. (6.39), the left-hand-side terms are for phonon(s) before scattering and the right-hand-side terms are for phonon(s) after scattering. The processes just described are called *normal (or N) processes*, in which the wavevectors of phonons are inside the first Brillouin zone. Since both the energy and the momentum are conserved, *N* processes do not alter the direction of energy flow. Hence, *N* processes make no contribution to the thermal resistance and do not affect the thermal conductivity.

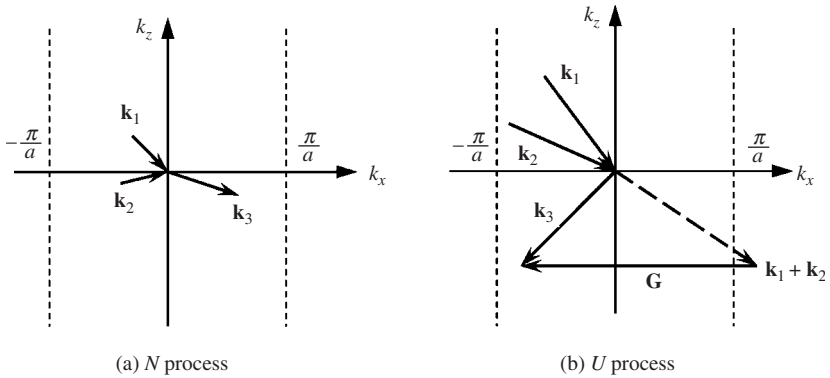
Scattering is also permitted when two phonons interact to form a third one, whose wavevector is outside the Brillouin zone. This can be understood by the equivalence of phonons with the same energy but with different wavevectors  $\mathbf{k}'$  and  $\mathbf{k}$  that follow the relationship:

$$\mathbf{k}' = \mathbf{k} + \mathbf{G} \quad (6.40)$$

where  $\mathbf{G}$  is a reciprocal lattice vector. The reverse process is also possible with the assistance of  $\mathbf{G}$  so that one phonon is annihilated to create two others. The momentum relations given in Eq. (6.39) need to be modified as follows after dropping  $\hbar$  in all terms:

$$\mathbf{k}_1 + \mathbf{k}_2 = \mathbf{k}_3 + \mathbf{G} \quad \text{or} \quad \mathbf{k}_1 + \mathbf{G} = \mathbf{k}_2 + \mathbf{k}_3 \quad (6.41)$$

These equations, combined with the energy conservation described by Eq. (6.38), describe the *umklapp* (or *U*) processes. The net momentum is not conserved in the *U* processes, which introduce thermal resistance and thus reduce the thermal conductivity. Figure 6.17



**FIGURE 6.17** Schematic illustrations of phonon-phonon scattering processes.

schematically shows the relationship between the wavevectors for an *N* process and a *U* process. An *N* process can be viewed as the general case of a *U* process when  $\mathbf{G} = 0$ .

Above room temperature, *U* processes dominate and the thermal conductivity decreases linearly as temperature increases. This is because the scattering rate  $\gamma = 1/\tau$  between acoustic phonons due to a *U* process can be described by<sup>23</sup>

$$\gamma_U = (A\omega + B\omega^2)T \quad (6.42)$$

where  $A$  and  $B$  are positive constants. When the temperature is reduced, the *U* process becomes weaker because of the shift in phonon distribution function toward longer wavelengths. As shown in Fig. 5.13, as the temperature is decreased below room temperature, the thermal conductivity increases to a maximum and then decreases due to the reduction in the specific heat. Four-phonon processes are also possible. Four-phonon scattering includes the annihilation of two phonons to create two others, the annihilation of one phonon to create three others, and the annihilation of three phonons to create another. The calculation of the probability of scattering is more involved.<sup>23</sup> Ecsedy and Klemens estimated the scattering rate due to four-phonon processes to be<sup>30</sup>

$$\gamma_{\text{Four}} \propto \omega^2 T^2 \quad (6.43)$$

In the temperature range from 300 to 1000 K, the probability of four-phonon processes is two to three orders of magnitude less than that of the three-phonon *U* processes.

In addition to the phonon-phonon interactions, phonons may also have interactions with defects (such as impurities, vacancies, or dislocations) and boundaries. These scattering processes may also influence the mean free path of phonons. Scattering of phonons by

defects is elastic since the phonon energy remains the same. At temperatures near the Debye temperature, phonon-phonon interactions are dominant. As the temperature drops, the wavelengths of phonons become comparable to the size of defects, and the scattering of phonons by defects is important. The scattering rate for phonon-defect scattering is independent of temperature but dependent on the phonon wavelength. This can be modeled using the Rayleigh scattering theory for small particles such that the scattering rate due to defects is inversely proportional to the fourth power of the phonon wavelength  $\lambda$ , viz.,

$$\gamma_{\text{ph-d}} \propto \lambda^{-4} \quad (6.44)$$

When the bulk mean free path is comparable or greater than the characteristic dimension, such as the thickness of the film or the diameter of the wire, scattering of phonons by boundaries becomes important. Boundary scattering is important for nanostructure materials and at low temperatures when the phonon mean free path is large, as discussed extensively in the previous chapter.

In metals and semiconductors, electronic transport becomes important. The scattering of charge carriers controls the electric conduction in solids and dominates the thermal conduction in metals. Carrier-carrier inelastic scattering is negligible except for highly conductive materials, such as a high-temperature superconductor. Since lattice vibrations are enhanced with increasing temperature, electron-phonon scattering usually dominates the scattering process at high temperatures; while at low temperatures, lattice vibrations are weak and defect scattering becomes important. The vibration of lattice ions causes deviations from the perfect periodic lattice and distorts the carrier wavefunction. This is more easily visualized as the scattering of electrons by phonons. Both the acoustic branch and the optical branch can scatter electrons. Usually, the energy of acoustic phonons can be neglected compared with the electron energy. Therefore, scattering by acoustic phonons is essentially elastic. Scattering by optical phonons is inelastic because the exchange of energy between the carriers and the phonons can be significant. This process facilitates the energy transfer between electrons and phonons, which is associated with Joule heating. For materials with two different atoms per primitive cell, the asymmetric charge distribution in the chemical bond forms a dipole. Scattering by optical phonons in these materials is called *polar scattering*, which can effectively scatter electrons or holes. The energy and momentum conservations for carrier-phonon scattering can be written as

$$E_f = E_i \pm \hbar\omega_{\text{phonon}} \quad (6.45a)$$

and

$$\mathbf{k}_f + \mathbf{G} = \mathbf{k}_i \pm \mathbf{k}_{\text{phonon}} \quad (6.45b)$$

where subscripts  $i$  and  $f$  indicate the initial and final states of the carrier, the minus sign corresponds to phonon emission, and the plus sign corresponds to phonon absorption. The momentum of an electron is similar to that of a phonon and is also referred to as the *crystal momentum*. If  $\mathbf{G}$  is set to zero, the process is an  $N$  process; otherwise, it is a  $U$  process as in phonon-phonon scattering. In semiconductors at low temperatures, only  $N$  processes are energized. In metals and semiconductors, the electron-phonon scattering rate typically ranges from  $10^{12}$  to  $10^{14}$  Hz at room temperature. Near or above the Debye temperature, the specific heat is almost a constant and the number of phonons increases linearly with temperature. Hence, the electron-phonon scattering rate is proportional to temperature in metals, resulting in nearly temperature-independent thermal conductivity, while the electrical resistance is proportional to temperature.

An electron or hole in a periodic lattice does not really collide with ions. The transport of free carriers can be viewed as the propagation of a wave in a periodic potential created by the ions. In addition to lattice vibrations, defects or impurities may break the periodicity of the potential or alter its amplitude. Kinetic theory gives the defect scattering rate  $\gamma_{\text{e-d}}$  as

$$\gamma_{\text{e-d}} = n_d \sigma_d v_e \quad (6.46)$$



where  $n_d$  and  $\sigma_d$  are the defect number density and cross-sectional area, respectively, and  $v_e$  is the average carrier velocity. For metals, the electron velocity is the Fermi velocity  $v_F$ , which is on the order of  $10^6$  m/s. For semiconductors, the random velocity of electrons or holes can be calculated by

$$v_{th} = (3k_B T/m^*)^{1/2} \tag{6.47}$$

which is called the *thermal velocity* and is on the order of  $10^5$  m/s at room temperature.

In semiconductors, the interband transition requires the conservation of both energy and momentum. This can occur by electronic transitions when interacting with the incident radiation. For indirect gap semiconductors, however, the photon itself cannot provide a large enough change in momentum. Therefore, a phonon is either emitted or absorbed for momentum conservation. The energy and momentum conservation equations are, respectively,

$$E_f - E_i = \hbar\omega_{\text{photon}} \pm \hbar\omega_{\text{phonon}} \tag{6.48a}$$

and

$$\mathbf{k}_f - \mathbf{k}_i = \mathbf{k}_{\text{photon}} \pm \mathbf{k}_{\text{phonon}} \tag{6.48b}$$

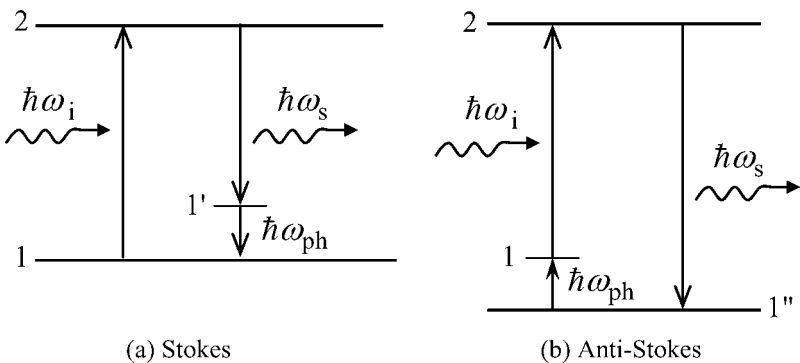
where the plus and minus signs correspond to phonon absorption and emission, respectively. This kind of transition is called the indirect interband transition. For a direct interband transition, there is no need to emit or absorb a phonon and, thus, the last term in either of Eq. (6.48a) or Eq. (6.48b) should be dropped out. The interaction of photons with solids will be left to Chap. 8 (Sec. 8.4) for a more detailed discussion about the absorption and emission processes.

In addition to the absorption and the emission, photons may be scattered by phonons, causing a nonlinear effect. There exists inelastic scattering when photons are scattered by phonons, resulting in x-ray scattering, neutron scattering, Raman scattering, and Brillouin scattering. In Raman scattering, the creation (emission) or annihilation (absorption) of a phonon causes a shift in the frequency of the radiation, namely, the Stokes or anti-Stokes shifts, as shown in Fig. 6.18. The energy conservation equations are

$$\hbar\omega_s = \hbar\omega_i - \hbar\omega_{ph}, \text{ for a Stokes shift} \tag{6.49a}$$

and

$$\hbar\omega_s = \hbar\omega_i + \hbar\omega_{ph}, \text{ for an anti-Stokes shift} \tag{6.49b}$$



**FIGURE 6.18** Illustration of Raman scattering: (a) the Stokes and (b) the anti-Stokes processes.

where subscripts i, s, and ph are for incident photon, scattered photon, and phonon, respectively. Because the interaction involved two photons and one phonon, the momentum of the phonon is restricted to small values. The Raman effect, or the Raman scattering, was named

after Indian physicist C. V. Raman (1888–1970), who won the Nobel Prize in Physics in 1930 for the discovery. The intensity of the anti-Stokes shift is usually much weaker than that of the Stokes shift. In certain cases, however, the phonons generated by the Stokes process can subsequently participate in the anti-Stokes process, causing a strong excitation to the anti-Stokes component. It is interesting to note that the anti-Stokes component actually pumps energy out from the material, resulting in a radiative cooling effect.

Note that the resulting photon can interact with the phonon again, creating a cascade process that emits  $m$  phonons. The photon energy is reduced by  $m$  times the energy per phonon. The probability decreases as the order increases. Raman spectroscopy has become a major analytical instrument for the study of solids. High-intensity lasers, high-resolution spectrometers, and sensitive detectors such as photomultiplier tubes (PMTs) are often employed to measure narrow Raman lines. The Raman intensity and intensity ratio depend upon temperature, as illustrated in Fig. 6.19. The ratio of the Raman intensities can be expressed by

$$\frac{I_{\text{anti-Stokes}}}{I_{\text{Stokes}}} = \left( \frac{\omega_i - \omega_{\text{ph}}}{\omega_i + \omega_{\text{ph}}} \right)^2 \exp\left(-\frac{\hbar\omega_{\text{ph}}}{k_B T}\right) \quad (6.50)$$

which can be used for surface temperature measurements.<sup>31</sup>

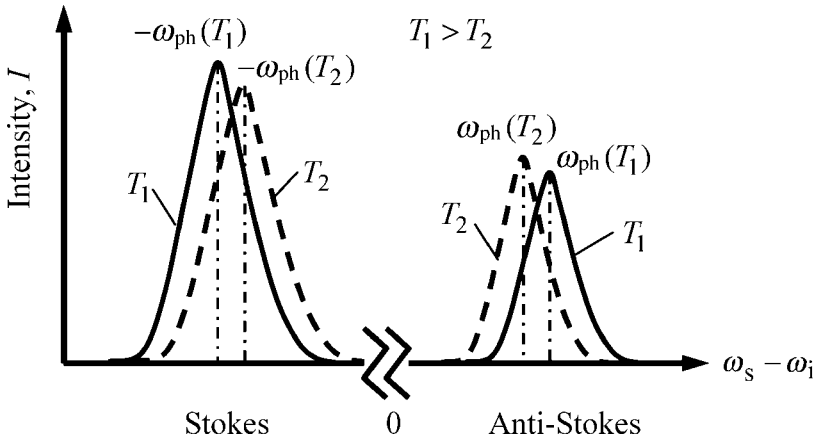


FIGURE 6.19 Raman intensity for the Stokes and anti-Stokes scattering at two different temperatures.

**Example 6-4.** Neutron scattering by phonons is important for measuring the dispersion relations. Express the energy conservation and the momentum conservation during the neutron-phonon scattering in terms of the wavevector and the mass of the neutron, and the wavevector and the frequency of the phonon. Assume the process involves one phonon only.

**Solution.** A neutron has a mass  $m_n = 1.673 \times 10^{-27}$  kg that is 1834 times that of an electron. Based on the wave-particle duality, the kinetic energy of a neutron can be expressed as  $E_n = p^2/2m_n = \hbar^2 k^2/2m_n$ ; thus, the energy conservation becomes

$$\frac{\hbar^2 k_s^2}{2m_n} = \frac{\hbar^2 k_i^2}{2m_n} \pm \hbar\omega_{\text{ph}} \quad (6.51)$$

where  $k_i$  and  $k_s$  are the magnitude of wavevector of the incident and scattered neutrons. The wavevector selection rule gives

$$\mathbf{k}_s + \mathbf{G} = \mathbf{k}_i \pm \mathbf{k}_{\text{ph}} \quad (6.52)$$

These relations characterize the inelastic scattering of neutrons by phonons. The plus and minus signs refer to the process that absorbs or releases phonons, respectively.

## 6.6 ELECTRON EMISSION AND TUNNELING

In all the discussions given so far, electrons are confined to the solid. Emission or discharge of electrons from a solid surface to vacuum or through a barrier (such as in a metal-insulator-metal multilayer structure) is possible, under the influence of an incident electromagnetic wave, an electric field, or a heating effect. Because of the importance of electron emission and tunneling to fundamental physics and device applications, the basic concepts are described in this section.

### 6.6.1 Photoelectric Effect

In 1887, Heinrich Hertz observed the *photoelectric effect* or *photoemission*. Shortly afterward, the phenomenon was experimentally studied by several others, including J. J. Thomson, who discovered electron as a subatomic particle. When radiation is incident on a metal plate, the electrons in the metal can be excited by absorbing the energy of the electromagnetic wave to escape the surface, as illustrated in Fig. 6.20a. The actual apparatus

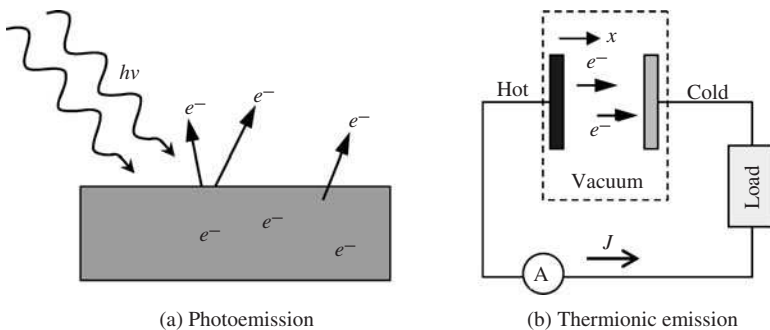


FIGURE 6.20 Illustration of (a) the photoelectric effect and (b) the thermionic emission.

used for measuring the ejected photoelectrons was to use another electrode and measure the current flow via a closed circuit. This is similar to the arrangement shown in Fig. 6.20b for thermionic emission, but with photons incident on the left plate without heating up any of the plates. If the frequency of the incident radiation is not high enough, no electrical current can be measured no matter how intense the incident radiation is. Saying in other words, there appears to be a threshold frequency for photoemission to occur in a given material. The photoelectric effect was explained in 1905 by Albert Einstein with the concept of light quanta, postulated by Max Planck a few years earlier. Although Einstein also made seminal contributions to the theory of relativity and Brownian motion, he was awarded the Nobel Prize in Physics in 1921 mainly for his discovery of the law that governs the photoelectric effect.

From the Fermi-Dirac distribution function of free electron gas, we can see that at low temperatures, electrons fill all energy levels up to the Fermi energy  $E_F$ . Note that we use  $E$  and  $E_F$  as the relative electron energy and, thus, they can be either positive or negative. Because of the binding of the electron with the rest of the solid, an additional energy, called

the *work function*  $\psi$ , must be provided to the electron for it to escape from the solid. For Ag, Al, Au, Cu, Fe, Pb, and W, the work function ranges from 4 to 5 eV, which corresponds to a wavelength in the ultraviolet region from 250 to 300 nm. For Na, K, Cs, and Ca, the work function ranges from 2 to 3 eV, which falls in the visible spectrum. Because a photon can interact with only one electron at a time, the photon energy  $h\nu$  must exceed the work function in order for the incident radiation to eject electrons from the surface. If  $h\nu > \psi$ , the photon energy may be absorbed by an electron right at the Fermi level. Subsequently, the electron will have a kinetic energy of

$$\frac{1}{2}m_e v_{e,\max}^2 = h\nu - \psi \quad (6.53)$$

after leaving the surface. If an electron is below the Fermi level, the kinetic energy of the ejected electron will be smaller than that given by Eq. (6.53). Therefore, Eq. (6.53) predicts the *maximum kinetic energy* of an electron for the prescribed photon frequency. A direct method for the determination of the work function is to measure the kinetic energy distribution of the photoelectrons, for a given frequency of the incident radiation.

One of the applications of photoemission is to measure the electron binding energy using the x-ray photoelectron spectroscopy (XPS), which is also called the electron spectroscopy for chemical analysis (ESCA). The basic principle for XPS is

$$E_{\text{bd}} = h\nu - \frac{1}{2}m_e v_e^2 - \psi \quad (6.54)$$

where  $E_{\text{bd}}$  stands for the binding energy with respect to the Fermi energy. The high-energy photons from an x-ray source (200 to 2000 eV) can interact with the inner electrons and eject them out of the surface. The photoelectron intensity can be plotted as a function of the electron kinetic energy using an electron energy analyzer. The intensity peaks are associated with the binding energies of the particular atomic structures. Comparing with the recorded photoelectron spectra, XPS allows the determination of the chemical composition of the substance near the surface. Swedish physicist Kai Siegbahn shared the Nobel Prize in Physics in 1981 for his contribution leading to the practical application of XPS. Furthermore, ultraviolet photoemission spectroscopy (UPS) with photon energies ranging from 5 to 100 eV, often from a synchrotron radiation source, has been used to study the electronic band structures.

## 6.6.2 Thermionic Emission

The charge emission from hot bodies was independently discovered by British scientist Frederick Guthrie in 1873, with a heated iron ball, and Thomas Edison in 1880, while working on his incandescent bulbs. *Thermionic emission* was extensively studied in the early 1900s by Robert Millikan, Nobel Laureate in Physics in 1923; Owen Richardson, Nobel Laureate in Physics in 1928; and Irving Langmuir, Nobel Laureate in Chemistry in 1932, among others.

With the understanding of the work function as the threshold energy that an electron must gain to escape the solid, it becomes straightforward to explain the emission of electrons from a heated metal. We use metal here to illustrate thermionic emission because good conductors can be better approximated by the Sommerfeld theory. The distribution function of a free electron gas has been extensively discussed in Chap. 5 (Sec. 5.1.3). At absolute zero temperature, all states below the Fermi level are filled by electrons and all states above the Fermi level are empty. Note that this picture is consistent with the electronic band theory. At elevated temperatures, the distribution function is modified as illustrated in Fig. 5.5. Some electrons will have energies above  $E_F$  (or  $\mu_F$ , as used in Chap. 5). Because the distribution function becomes zero only when  $E \rightarrow \infty$ , a small fraction of electrons must occupy energy levels exceeding  $E_F + \psi$ . We wish to quantitatively evaluate the current density or the charge flux from the hot plate to the cold plate, as illustrated in Fig. 6.20*b*. Let the electron flow be along the  $x$  direction.

From Eq. (5.16), the number of electrons per unit volume between  $\mathbf{v}$  and  $\mathbf{v} + d\mathbf{v}$  is

$$n(\mathbf{v})dv_x dv_y dv_z = 2 \left( \frac{m_e}{h} \right)^3 \frac{dv_x dv_y dv_z}{e^{(E-E_F)/k_B T} + 1} \quad (6.55)$$

where  $E = \frac{1}{2}m_e(v_x^2 + v_y^2 + v_z^2)$  is the kinetic energy of an electron. The current density in the  $x$  direction is given by

$$J_x = (1 - r') \iiint (-e)v_x n(\mathbf{v}) dv_x dv_y dv_z \quad (6.56)$$

where  $r'$  is the electron reflection coefficient or the fraction of electron reflected by the receiver. The integration is from  $-\infty$  to  $\infty$  in both the  $y$  and  $z$  directions. In order for an electron to escape in the  $x$  direction, the following criterion must be satisfied:

$$v_x > v_{x,0} = \sqrt{2(E_F + \psi)/m_e} \quad (6.57)$$

This equation suggests that the integration is carried out only in the tail of the distribution function, where the  $x$  velocity is positive and the kinetic energy is sufficiently large, i.e.,  $E - E_F > \psi$ , which is on the order of several electron volts. Note that  $k_B T = 0.086$  eV at 1000 K and 0.026 eV at 300 K. When  $\psi/k_B T = 4$ , dropping the unity term in the denominator of Eq. (6.55) causes less than 2% error. The error becomes even smaller at a larger  $v_x$  so that its impact on the integration is negligibly small. For this reason, it appears safe to substitute the Fermi-Dirac distribution by the Maxwell-Boltzmann distribution, viz.,

$$J_x = -2e(1 - r') \left( \frac{m_e}{h} \right)^3 e^{E_F/k_B T} \int_{v_{x,0}}^{\infty} v_x \exp\left(-\frac{1}{2}m_e v_x^2\right) dv_x \quad (6.58)$$

$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}m_e v_y^2 - \frac{1}{2}m_e v_z^2\right) dv_y dv_z$$

The result is the famous *Richardson-Dushman equation* for the current density:

$$J = A_{RD}(1 - r')T^2 e^{-\psi/k_B T} \quad (6.59)$$

where  $A_{RD} = 4\pi m_e e k_B^2 / h^3 = 1.202 \times 10^6$  A/(m<sup>2</sup> · K<sup>2</sup>) is called the *Richardson constant*, and the direction of  $J$  is as shown in Fig. 6.20b. The heat transfer associated with the electron flow can be evaluated by considering the kinetic energy associated with each electron,  $\frac{1}{2}m_e v^2 \approx \frac{1}{2}m_e v_x^2$ , i.e.,

$$q_x'' = (1 - r') \iiint v_x \left( \frac{1}{2}m_e v_x^2 \right) n(\mathbf{v}) dv_x dv_y dv_z = (\psi + E_F + k_B T) J_x / e \quad (6.60)$$

This equation suggests that the average energy of the “hot electron” is  $\psi + E_F + k_B T$ , as expected. Vacuum tubes operate based on the principle of thermionic emission. Vacuum tubes had wide applications in the mid twentieth century in radio, TV, and computer systems, but have largely been replaced by transistors nowadays. Thermionic generators produce electricity without any moving parts and belong to the category of direct energy converters. Extensive discussion of the thermodynamics and efficiency of thermionic converters can be found from Hatsopoulos and Gyftopoulos.<sup>32</sup>

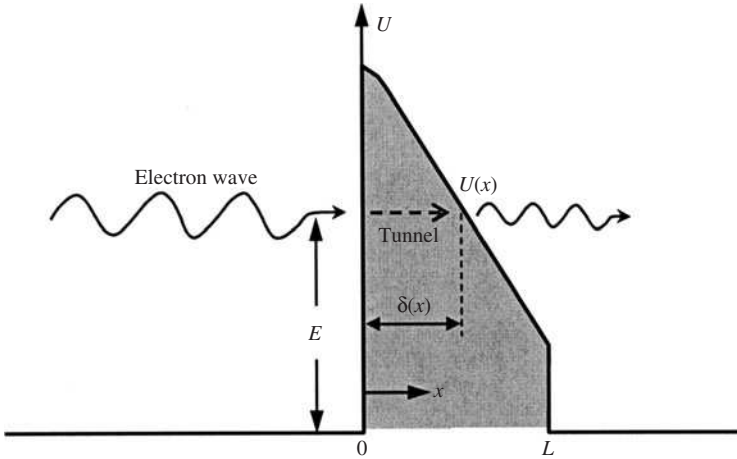
In some applications, a voltage can be applied between the electrodes. Furthermore, a semiconductor can be used to form a Schottky barrier between a metal and a semiconductor.<sup>11</sup> The applied voltage changes the potential distribution so that it gradually decreases inside the barrier. Furthermore, the work function can be significantly reduced. Assuming the transmission coefficient is unity, Eq. (6.59) can be modified to the following for the net charge transfer:

$$J_{\text{net}} = A^* T^2 e^{-\psi^*/k_b T} (e^{e\Delta V/k_b T} - 1) \quad (6.61)$$

where  $A^*$  should be calculated according to the effective mass,  $\psi^*$  is the effective work function, and  $\Delta V$  is the applied voltage.<sup>11</sup> In deriving Eq. (6.61), we assumed that hot electrons from the cathode will go through the barrier through ballistic processes. This means that the electron mean free path must be larger than the thickness of the semiconductor film. Otherwise, the electron transport is governed by diffusion because of collisions with phonons or impurities. When diffusion occurs, the electron transport under the influence of a temperature difference is described by the thermoelectric effect, based on irreversible thermodynamics, as discussed in Chap. 5. When the barrier thickness is extremely small, another phenomenon called quantum tunneling may occur such that an electron whose energy is lower than the potential barrier has a chance to transmit through the barrier. Tunneling effect will be discussed in the next subsection. Mahan and coworkers pointed out that, for the thermionic phenomenon to be the dominant transport mechanism, the electron mean free path in the barrier must be greater than the thickness of the barrier.<sup>33</sup> Furthermore, the latter must exceed the characteristic length, below which tunneling becomes significant. Thermionic emission in semiconductor heterogeneous structures has been extensively studied in the last decade, for both refrigeration and power generation.<sup>33-34</sup> The refrigeration process is a reversed thermionic power generation process. In *thermionic refrigeration*, the cold cathode emits electrons to the room-temperature anode as a result of the applied voltage. In order to achieve any cooling effect, energy that is carried through by the electrical current must be greater than that by heat conduction via lattice vibration from the hot electrode to the cold electrode. The nonequilibrium electron and phonon transport phenomena have also been investigated. In some cases, both thermionic and thermoelectric effects may show up.<sup>35,36</sup> In other cases, thermionic and tunneling effects can work together or against each other.<sup>37,38</sup>

### 6.6.3 Field Emission and Electron Tunneling

From the above discussion, we have noticed that thermionic emission may be enhanced or even reversed (from a colder cathode to a hotter anode) by an applied electric field. Some thermal excitation is necessary for part of the electrons to occupy energy levels above the Fermi level by a finite amount, prescribed by the work function. This is commonly referred to as a *potential barrier* or a *potential hill*. An electron must acquire sufficient energy for it to surmount the barrier. When the field strength is very high, however, electrons at energy levels lower than the height of the barrier can tunnel through the potential hill. The word *tunneling* gives a vivid (but inaccurate) picture of the tunneling phenomenon as if a hole were drilled for the electrons without sufficient energy to pass through a potential hill, without climbing to its top first. This phenomenon of electron emission at high applied field is called *field emission*, which can occur at very low temperatures. The applied electric field can exceed several billion volts per meter. Because of the high field, field emission can occur only in ultrahigh vacuum (UHV); otherwise, ionization of the gas molecules would occur that can cause *discharge glow*. In essence, field emission is a form of *quantum tunneling*, which cannot be understood within the framework of classical mechanics. The electron motion is governed by Schrödinger's wave equation, and the transmission can be predicted by the probability of finding an electron on the other side of the potential hill, as illustrated in Fig. 6.21.



**FIGURE 6.21** Illustration of quantum tunneling through a potential barrier by an electron wave.

In 1928, Fowler and Nordheim provided the first quantum mechanical derivation of the field emission current density  $J$  as follows:

$$J = C \left( \frac{\Delta V}{L} \right)^2 \exp \left( - \frac{\alpha \psi^{3/2}}{\Delta V/L} \right) \quad (6.62)$$

This is called the Fowler-Nordheim equation, in which  $\Delta V/L$  is the electric field,  $C$  and  $\alpha$  are two positive constants, and  $\psi$  is the work function defined previously.

The WKB approximation is commonly used to find the transmission probability  $\tau'$  of tunneling. WKB (also KWB or BWK) stands for Wentzel, Kramers, and Brillouin, although a fourth person Jeffreys was also included in some literature—so, the abbreviation appeared as JWKB. The main assumption in the WKB approximation is that the potential  $U(x)$  is a slow function of  $x$ .<sup>40</sup> In the region where the electron energy  $E$  is greater than  $U(x)$ , the wavefunction is of the form

$$\Psi(x,t) = A \exp(-i\omega t) \exp \left[ \pm \frac{i}{\hbar} \sqrt{2m_e(E - U)} \right] \quad (6.63a)$$

where  $A$  is the amplitude of the electron wave. In the region where  $E < U(x)$ , the wavefunction is of the form

$$\Psi(x,t) = A \exp(-i\omega t) \exp \left[ \pm \frac{1}{\hbar} \sqrt{2m_e(U - E)} \right] \quad (6.63b)$$

The transmission probability or transmission coefficient can be approximated as

$$\tau'(E) = \exp \left[ - \frac{2}{\hbar} \int_0^{\delta} dx \sqrt{2m_e(U - E)} \right] \quad (6.64)$$

where  $\delta$  is the width of the potential at  $E$ .<sup>40</sup>

**Example 6-5.** Assume that  $U(x) = \psi - (x/L)e\Delta V$  and  $\delta(E) = L(\psi - E)/e\Delta V$ , i.e., linearly varying barrier whose highest potential is  $\psi$  at  $x = 0$ . Find the transmission coefficient.

**Solution.** For the triangular barrier shown in Fig. 6.21, we note that

$$\int_0^\delta dx \sqrt{U - E} = \int_0^\delta dx \sqrt{\psi - E - xe\Delta V/L} = \frac{(\psi - E)^{3/2}}{e\Delta V/L} \int_1^0 \sqrt{1 - u} du = \frac{2(\psi - E)^{3/2}}{3e\Delta V/L}$$

Substituting this equation into Eq. (6.64), we obtain

$$\tau'(E) = \exp\left[-\frac{4(\psi - E)^{3/2}}{3\hbar\Delta V/L} \sqrt{2m_e}\right] \quad (6.65)$$

When  $E \ll \psi$ , we see that  $\tau' \approx \exp(-\alpha\psi^{3/2}L/\Delta V)$ , where  $\alpha = (4/3\hbar)\sqrt{2m_e}$ . At elevated temperatures, however, we need to consider the energy distribution of electrons. Esaki and coworkers demonstrated that the resonant tunneling of electron waves may allow the transmission coefficient to approach unity in superlattice and double-barrier structures.<sup>41</sup> Electron tunneling is similar to photon tunneling of electromagnetic waves, to be discussed in Chap. 10.

The tunneling current density can be calculated by

$$J_t = \int_{E_{\min}}^{E_{\max}} e\tau'(E)n(E)dE \quad (6.66)$$

where  $E$  is the kinetic energy in the  $x$  direction,  $E_{\max}$  corresponds to the energy at the top of the potential barrier,  $E_{\min}$  is a reference energy, and  $n(E)dE$  is the number of available electrons, with energy between  $E$  and  $E + dE$ , per unit area per unit time, given as

$$n(E) = \frac{m_e k_B T}{2\pi^2 \hbar^3} \ln\left[1 + \exp\left(-\frac{E - E_F}{k_B T}\right)\right] \quad (6.67)$$

Some analytical expressions similar to Eq. (6.62) have been presented to approximate the integration of Eq. (6.66).<sup>42,43</sup>

The energy transfer during field emission or electron tunneling can also be evaluated.<sup>37,44,45</sup> A salient difference between thermionic emission and field emission is that thermionic emission always gives out energy as the electrons are emitted and transfer the energy to the other side of the barrier. This is because the emitted electrons are in the high-energy tail of the distribution function, called hot electrons, with a much higher effective temperature than the equilibrium cathode temperature. On the other hand, field emission allows electrons with energies much lower than that corresponding to the equilibrium temperature to escape the surface. Since the replacement electrons have a higher average energy than the emitted electrons, a heating effect occurs that increases the cathode temperature. Depending on the geometry, temperature, transmission coefficient, and energy distribution, both heating and cooling of the cathode are made possible by field emission. This is known as the Nottingham effect originally published in 1941.

Some applications of quantum tunneling in semiconductors and superconductors were discussed in Chap. 1. One of the applications of electron tunneling was the invention of scanning tunneling microscope (STM). Xu et al. developed a model for the energy exchange by the tunneling electrons and made a comparison with STM measurements.<sup>44</sup> They considered the Nottingham effect on both electrodes, as well as resistive heating. At short distances, thermionic emission, field emission, and photon tunneling could occur simultaneously. Photon tunneling will be studied in Chap. 10. Fisher and Walker analyzed the energy transport in nanoscale field emission processes by considering the geometry of the emission tip.<sup>45</sup> Quantum size effect may play a role in modifying some of the critical parameters. Field emission by nanotubes has been proposed for nanoscale manufacturing and thermal writing.<sup>46</sup> Wong et al. performed a detailed thermal analysis during electron beam heating and laser processing.<sup>47</sup> Carbon nanotube field emission display (CNT-FED) has been demonstrated at the Samsung SDI Company by Chung et al. (*Appl. Phys. Lett.*, **80**, 4045, 2002) and is being commercialized as a flat-panel display technology. While



CNT-FEDs resemble the cathode-ray tubes (CRTs) in many ways, it can be made thin and flat with a much lower applied voltage. Carbon nanotube field emission has also been demonstrated for the generation of x ray by Yue et al. (*Appl. Phys. Lett.*, **81**, 355, 2002) and luminescent tubes by Bonard et al. (*Appl. Phys. Lett.*, **78**, 2775, 2001).

## 6.7 ELECTRICAL TRANSPORT IN SEMICONDUCTOR DEVICES

Semiconductors are the most important materials for microelectronics, MEMS, and optoelectronics. Much of the discussions in Chap. 5 and the previous sections of this chapter are applicable to semiconductors, especially for the energy storage and transport by phonons. This section focuses on the basics of electrical transport and properties for some common semiconductor devices used in optoelectronics.

### 6.7.1 Number Density, Mobility, and the Hall Effect

The calculation of the number density of electrons and holes at any given temperature  $T$  is very important for the determination of the electrical, optical, and thermal properties of semiconductor materials and devices. The free electron gas model can be modified to describe the electron and hole distributions and the transport in semiconductors. The Fermi-Dirac distribution function is applicable to electrons and holes according to

$$f_e(E) = \frac{1}{e^{(E-E_F)/k_B T} + 1} \quad \text{and} \quad f_h(E) = \frac{1}{e^{(E_F-E)/k_B T} + 1} \quad (6.68)$$

Note that  $f_e(E) + f_h(E) \equiv 1$ . The number density of electrons or holes is given by

$$n_e = \int_{E_C}^{\infty} \frac{D_e(E)dE}{e^{(E-E_F)/k_B T} + 1} \quad \text{and} \quad n_h = \int_{-\infty}^{E_V} \frac{D_h(E)dE}{e^{(E_F-E)/k_B T} + 1} \quad (6.69)$$

where  $D_e(E)$  and  $D_h(E)$  are the densities of states in the conduction and valence bands, respectively. With the approximated quadratic forms of the conduction and valence bands, Eq. (6.29), the densities of states can be written as

$$D_e(E) = M_C \left. \frac{dk}{dE} \right|_C = \frac{M_C}{2\pi^2} \left( \frac{2m_e^*}{\hbar^2} \right)^{3/2} (E - E_C)^{1/2} \quad (6.70a)$$

and

$$D_h(E) = \left. \frac{dk}{dE} \right|_V = \frac{1}{2\pi^2} \left( \frac{2m_h^*}{\hbar^2} \right)^{3/2} (E_V - E)^{1/2} \quad (6.70b)$$

where  $M_C$  is the number of equivalent minima in the conduction band. Equation (6.70a) and Eq. (6.70b) are derived based on the parabolic shape near the bottom of the conduction band for electrons or the top of the valence band for holes. The effective mass of electrons is a geometric average over the three major axes because the effective mass of silicon depends on the crystal direction. The effective mass of holes is an average of heavy holes and light holes because there exist different subbands.<sup>11</sup> At moderate temperatures,  $E_C - E_F \gg k_B T$  and  $E_F - E_V \gg k_B T$  are satisfied; subsequently,  $f_e$  and  $f_h$  can be approximated with the classical Maxwell-Boltzmann distribution:

$$f_e(E) \approx e^{(E_F-E)/k_B T} \quad \text{and} \quad f_h(E) \approx e^{(E-E_F)/k_B T} \quad (6.71)$$

We can carry out the integrations in Eq. (6.69) and thus obtain

$$n_e = N_C e^{-(E_C - E_F)/k_B T} \quad (6.72a)$$

and

$$n_h = N_V e^{-(E_F - E_V)/k_B T} \quad (6.72b)$$

where  $N_C = 2M_C(2\pi m_e^* k_B T/h^2)^{3/2}$  and  $N_V = 2(2\pi m_h^* k_B T/h^2)^{3/2}$  are called the *effective density of states* in the conduction band and in the valance band, respectively. The combination of Eq. (6.72a) and Eq. (6.72b) gives, in terms of  $E_g = E_C - E_V$ ,

$$n_e n_h = N_{th}^2 = N_C N_V e^{-E_g/k_B T} \propto T^3 e^{-E_g/k_B T} \quad (6.73)$$

This expression does not involve the Fermi energy. Therefore, it holds for both intrinsic and doped semiconductors. The number density  $N_{th}$  can be viewed as thermally excited electron-hole pairs per unit volume. It is also referred to as the number density of intrinsic carriers because  $n_e = n_h = N_{th}$ , in an intrinsic semiconductor. It can be seen that the number densities increase with temperature so that the electrical conductivity of an intrinsic semiconductor increases with temperature. The Fermi energy for an intrinsic semiconductor can be obtained by setting  $n_e = n_h$  in Eq. (6.72a) and Eq. (6.72b), yielding

$$E_F = \frac{E_C + E_V}{2} + \frac{k_B T}{2} \ln\left(\frac{N_V}{N_C}\right) \approx \frac{E_C + E_V}{2} \quad (6.74)$$

The Fermi energy for an intrinsic semiconductor is expected to lie in the middle of the forbidden band or the bandgap. The requirement for the approximate distributions given in Eq. (6.71) to hold with less than 2% error is  $E_g/k_B T > 8$ , such that  $\exp[-E_g/(2k_B T)] < 0.02$ . For  $E_g > 0.8$  eV, we have  $T < 1150$  K. One should keep in mind that  $E_g$  reduces as temperature increases. For silicon,  $E_g \approx 1.11$  eV at 300 K and  $\approx 0.91$  eV at 900 K.

When impurities of either donors or acceptors or both are involved, the calculation of Fermi energy and number densities becomes more involved.<sup>11,15</sup> Let  $N_D$  and  $N_A$  stand respectively for the number densities (i.e., doping concentrations) of donors (e.g., P and As) and acceptors (e.g., B and Ga). In brief, the energy level of donors  $E_D$  is usually lower but very close to  $E_C$ . As a result, the Fermi energy  $E_F$  goes up but is always below  $E_D$ . The difference  $E_C - E_D$  is called the ionization energy of donors, which is required for the donors to become ionized. The ionization of donors increases the number of free electrons, and the semiconductor is said to be of *n*-type. For the semiconductor Si, the ionization energy for P is 45 meV and that of As is 54 meV. Likewise, the energy level of acceptors  $E_A$  is slightly above  $E_V$ , and  $E_A - E_V$  is called the ionization energy of acceptors. The ionization of acceptors increases the number of holes, and the semiconductor is said to be of *p*-type. For the semiconductor Si, the ionization energy for B is 45 meV and that of Ga is 72 meV. Note that there are  $5.0 \times 10^{22}$  cm<sup>-3</sup> (atoms per cubic centimeters) for silicon. For *n*-type silicon with an arsenic doping concentration of  $N_D = 5.0 \times 10^{16}$  cm<sup>-3</sup>, the impurities occupy one atomic site per million. Because of the change in Fermi energy, most of the impurities are ionized at room temperature, when the doping concentration is less than  $5.0 \times 10^{17}$  cm<sup>-3</sup>. For fully ionized impurities, the charge neutrality requires that

$$n_e + N_A = n_h + N_D \quad (6.75)$$

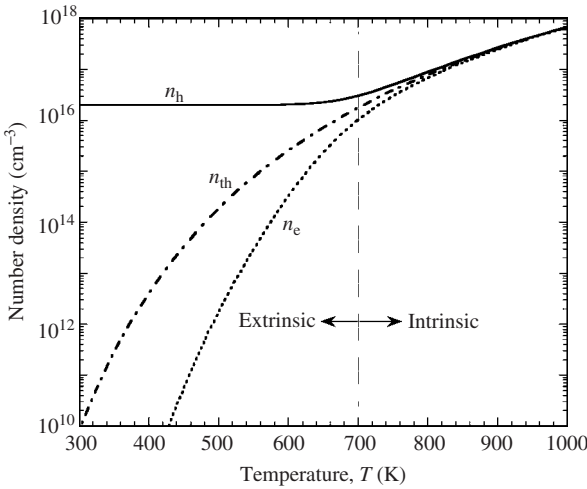
If the impurities are partially ionized,  $N_D$  and  $N_A$  in Eq. (6.75) should be replaced by the ionized donor and acceptor concentrations, respectively.

**Example 6-6.** For boron-doped Si with  $N_A = 2.0 \times 10^{16} \text{ cm}^{-3}$ , find  $n_{\text{th}}$ ,  $n_e$ , and  $n_h$  at temperatures from 300 to 1000 K; compare your answers with the values for intrinsic silicon. Assume  $m_e^* = 0.3m_e$  and  $m_h^* = 0.6m_e$ . Use  $M_C = 6$  and  $E_g(T) = 1.155 - 0.000473T^{3/4}(T + 636) \text{ eV}$ .

**Solution.** This is a  $p$ -type semiconductor with  $N_D = 0$ , and from Eq. (6.75), we have  $n_h = n_e + N_A$ . Substituting it into Eq. (6.72), we have  $n_e^2 + N_A n_e = N_{\text{th}}^2 = N_C N_V e^{-E_g/k_B T}$ , i.e.,

$$n_e = \frac{1}{2} \left( \sqrt{N_A^2 + 4N_{\text{th}}^2} - N_A \right) \quad \text{and} \quad n_h = \frac{1}{2} \left( \sqrt{N_A^2 + 4N_{\text{th}}^2} + N_A \right) \quad (6.76)$$

The calculated values of  $N_C = 2.47 \times 10^{19} \text{ cm}^{-3}$  and  $N_V = 1.17 \times 10^{19} \text{ cm}^{-3}$  at 300 K are somewhat lower than the recommended values of  $N_C = 2.86 \times 10^{19} \text{ cm}^{-3}$  and  $N_V = 2.66 \times 10^{19} \text{ cm}^{-3}$ .<sup>11</sup> The results are plotted in Fig. 6.22 for comparison. In the extrinsic region when  $T < 700 \text{ K}$ , the



**FIGURE 6.22** Calculated number densities for Example 6-6.

majority carriers are holes, and  $n_h \approx N_A$  depends little on temperature. In the intrinsic region when  $T > 800 \text{ K}$ ,  $n_e \approx n_h \approx N_{\text{th}}$ , due to thermal excitation. It should be mentioned that at very low temperatures, i.e.,  $T < 100 \text{ K}$ , ionization is not very effective and  $n_h \ll N_A$ . Therefore, the low-temperature region is called freeze-out zone, which is not shown in the plot.

The Drude free-electron model predicts  $\sigma = \tau n_e e^2 / m_e$ , as given in Eq. (5.49), which can be applied for both electrons and holes, using proper effective masses and relaxation times. In semiconductor physics, the term *mobility* is often used, and defined as

$$\mu_e = \frac{e\tau_e}{m_e^*} \quad \text{and} \quad \mu_h = \frac{e\tau_h}{m_h^*} \quad (6.77)$$

The physical significance is that mobility is the drift velocity per unit applied field, i.e.,

$$\mathbf{u}_{\text{d,e}} = -\mu_e \mathbf{E} \quad \text{and} \quad \mathbf{u}_{\text{d,h}} = \mu_h \mathbf{E} \quad (6.78)$$

The electrical conductivity of a semiconductor is thus

$$\sigma = en_e \mu_e + en_h \mu_h \quad (6.79)$$

Depending on the impurity and temperature range, one term may be dominant, or both the terms may be comparable. It is crucial to understand the scattering mechanism in semiconductors. In metals, all the conducting electrons are near the Fermi surface, and their average energy cannot be described by the classical statistics because  $\bar{\epsilon} = \frac{1}{2}m_e\bar{v}_e^2 \approx \frac{3}{5}\mu_F \neq \frac{3}{2}k_B T$  (see Example 5-2). For semiconductors, on the other hand, Eq. (6.71) tells us that  $\bar{\epsilon} = \frac{1}{2}m_e^*v_e^2 \approx \frac{3}{2}k_B T = \frac{1}{2}m_e^*v_{th}^2$ , and the classical statistics is applicable to a large temperature range. Thermal velocity is the velocity of electrons or holes at the equilibrium temperature and was given in Eq. (6.47). At sufficiently high temperatures when phonon scattering dominates, the electron mean free path  $\Lambda_e \propto 1/T$ . Based on the relation  $\tau_e = \Lambda_e/\bar{v}_e$ , we have

$$\mu_{ph} \propto T^{-3/2} \quad (6.80a)$$

where  $\mu_{ph}$  is the contribution of carrier-phonon scattering. Equation (6.80a) describes intrinsic semiconductor without defects. The scattering by impurities results in a mobility given by

$$\mu_d \propto T^{3/2}/N_d \quad (6.80b)$$

where  $N_d$  stands for the concentration of the ionized impurities. The combination gives the mobility for either electron or hole as follows:

$$\frac{1}{\mu} = \frac{1}{\mu_{ph}} + \frac{1}{\mu_d} \quad (6.81)$$

For intrinsic semiconductor, the electrical conductivity is very small and proportional to  $\exp[-E_g/(2k_B T)]$  so that the electrical conductivity increases with temperature. For intermediately doped semiconductors, there exists a maximum value of the mobility below room temperature due to the opposite temperature dependence of  $\mu_{ph}$  and  $\mu_d$ . At that temperature, the electrical conductivity is maximum. As the temperature goes up beyond room temperature, the conductivity decreases due to the increased phonon scattering. When the semiconductor reaches the intrinsic region, the number density suddenly increases and the conductivity increases again with temperature.

The Hall effect is very useful in measuring the mobility of semiconductors. In the extrinsic region, the Hall effect allows measurement of the type and concentration of the carriers. The measurements are usually carried out with the van der Pauw method, which is a four-probe technique for determining the electrical resistance and the Hall coefficient. The data of electrical resistivity and number density allow the extraction of the mobility, based on the effective mass determined using cyclotron resonance technique.

When both the carriers are significant to the transport properties, the situation is rather interesting. Referring to Fig. 6.1, when current flows to the positive  $x$  direction, we have  $v_{e,x} < 0$  and  $v_{h,x} > 0$ . The magnetic force drives both the electrons and the holes toward the negative  $y$  direction, such that  $v_{e,y} < 0$  and  $v_{h,y} < 0$  if  $E_y = 0$ . At steady state, a finite  $E_y$ , known as the Hall field, may exist. Since there is no net current flow in the  $y$  direction, we must have

$$J_x = -ev_{e,x}n_e + ev_{h,x}n_h \quad (6.82a)$$

$$J_y = -ev_{e,y}n_e + ev_{h,y}n_h = 0 \quad (6.82b)$$

In general, both  $v_{e,y}$  and  $v_{h,y}$  are not zero. The Lorentz force  $\mathbf{F} = q(\mathbf{E} + \mathbf{u}_d \times \mathbf{B})$  in the  $y$  direction is related to the drift velocities for electrons or holes by

$$-eE_y + ev_{e,x}B = ev_{e,y}\mu_e \quad (6.83a)$$

and

$$eE_y - ev_{h,x}B = ev_{h,y}\mu_h \quad (6.83b)$$

Rewrite Eq. (6.83a) and Eq. (6.83b) as  $n_e \mu_e (E_y - v_{e,x} B) = -n_e v_{e,y}$  and  $n_h \mu_h (E_y - v_{h,x} B) = n_h v_{h,y}$ , respectively. Compared with Eq. (6.82b), we notice that  $n_e \mu_e (E_y - v_{e,x} B) + n_h \mu_h (E_y - v_{h,x} B) = 0$ , or

$$\frac{E_y}{B} = \frac{n_e \mu_e v_{e,x} + n_h \mu_h v_{h,x}}{n_e \mu_e + n_h \mu_h}$$

Combining it with Eq. (6.82a), we obtain the Hall coefficient as follows:

$$\eta_H = \frac{E_y}{J_x B} = \frac{n_e \mu_e v_{e,x} + n_h \mu_h v_{h,x}}{e(n_e \mu_e + n_h \mu_h)(-n_e v_{e,x} + n_h v_{h,x})}$$

Substituting  $v_{e,x} = -\mu_e E_x$  and  $v_{h,x} = \mu_h E_x$  into the previous equation, we obtain

$$\eta_H = \frac{n_h \mu_h^2 - n_e \mu_e^2}{e(n_h \mu_h + n_e \mu_e)} \quad (6.84)$$

after canceling  $E_x$ . The Hall coefficient for semiconductors may be positive or negative, and becomes zero when  $n_h \mu_h^2 = n_e \mu_e^2$ . The drift velocities in the y direction, however, cannot be zero unless  $B = 0$  or  $J_x = 0$ .

## 6.7.2 Generation and Recombination

The generation, recombination, and diffusion processes are directly related to the charge transport in semiconductors and optoelectronic devices. This section takes photoconductivity as an example to illustrate the generation and recombination processes, followed by a brief discussion of luminescence.

Much has been said previously about absorption of light that causes a transition in the electronic states in solids. The bandgap absorption of Si, Ge, and GaAs corresponds to the wavelengths in the visible and near-infrared spectral regions. The excitation of electrons from the valence band to the conduction band by the absorption of radiation increases the conductivity of the semiconductor dramatically. This is known as *photoconductivity* and can be used for sensitive radiation detectors. For some semiconductors, the bandgap is very narrow so that transitions can happen at longer wavelengths. For example, the bandgap energy of  $\text{Hg}_{0.8}\text{Cd}_{0.2}\text{Te}$  is 0.1 eV at 77 K (liquid nitrogen temperature), and the material can be used as infrared detectors, which are commonly referred to as MCT detectors. At very low temperatures, impurities cannot be ionized thermally even though the ionization energy is very small. For boron-doped germanium, the ionization energy  $E_A - E_V \approx 10$  meV corresponds to a wavelength of about 120  $\mu\text{m}$ .<sup>48</sup> Therefore, Ge:B can be used as far-infrared radiation detectors. There are two groups of radiation detectors. The first group is called thermal or bolometric detectors, which rely on the temperature change of the detector as a result of the absorbed radiation. The temperature change can be monitored by a temperature-dependent property, such as the electrical resistance. An example is the superconductive bolometer, which relies on the drastic change in resistance with temperature, near the superconducting-to-normal-state transition or the critical temperature  $T_c$ . The second group is called nonthermal, nonbolometric, or nonequilibrium detectors. An example is the photoconductive detector in which the conductivity changes as a result of the direct interaction of electrons with photons.

Before the radiation is incident on the photoconductive detector, the conductivity can be expressed as  $\sigma_0 = en_{e,0}\mu_e + en_{h,0}\mu_h$  at thermal equilibrium. Under the influence of an

incident radiation with photon energies greater than the bandgap, additional electron-hole pairs are created so that the concentration is increased by  $\Delta n$  for both types of carriers. The relative change in the electrical conductance  $\Delta\sigma/\sigma_0$  can be expressed as

$$\frac{\Delta\sigma}{\sigma_0} = \frac{\Delta n(\mu_e + \mu_h)}{n_{e,0}\mu_e + n_{h,0}\mu_h} \quad (6.85)$$

Here,  $\Delta n$  is the net increase in carrier concentration as a result of both generation and recombination. The *generation* is associated with the absorbed radiation and depends on the intensity of the incident light and the *quantum efficiency*, which is wavelength dependent. The quantum efficiency is the percentage of the incoming photons that generate an electron-hole pair. The *recombination* is a relaxation process because the excess charges are not at thermal equilibrium. If the incident radiation is blocked off, the semiconductor will quickly reach an equilibrium with the conductivity  $\sigma_0$ . The characteristic time of the recombination process is called the *recombination lifetime* or *recombination time*  $\tau_{rc}$ . While it is also related to electron scattering, lattice scattering, and/or defect scattering, the recombination time is usually much longer than the relaxation time used in charge transport processes. The net rate of change can be expressed as the rate of generation (creation) minus the rate of recombination (annihilation), viz.,

$$\frac{dn}{dt} = n_g - \frac{n - n_0}{\tau_{rc}} \quad (6.86)$$

Under a steady-state incident radiation, we can set  $dn/dt = 0$  so that  $\Delta n = n - n_0 = \tau_{rc}n_g$ . Suppose that the incoming photon is of frequency  $\nu$  in Hz with a spectral irradiance  $I_\nu$  in  $W/(m^2 \cdot Hz)$ , and the detector has an effective area  $A$ , thickness  $d$ , and absorptance  $\alpha_\nu$ . We have

$$n_g = \frac{\alpha_\nu I_\nu A}{h\nu Ad} = \frac{\alpha_\nu I_\nu}{h\nu d} \quad (6.87)$$

Substituting into Eq. (6.85), we obtain the *sensitivity* of a photoconductive detector as follows:

$$\frac{1}{I_\nu} \frac{\Delta\sigma}{\sigma_0} = \frac{\alpha_\nu \tau_{rc}(\mu_e + \mu_h)}{h\nu(n_{e,0}\mu_e + n_{h,0}\mu_h)d} \quad (6.88)$$

Increasing the recombination time  $\tau_{rc}$  improves the sensitivity but decreases the speed or response time of the detector. Photoconductivity requires that  $h\nu > E_g$  for bandgap absorption to occur. However, the sensitivity decreases toward higher frequencies, or shorter wavelengths, because there are fewer photons per unit radiant power. Consequently, the sensitivity of a photoconductive detector increases with wavelength first and then suddenly drops to zero close to the band edge. The absorptance depends on the thickness  $d$ , which should be 2 to 3 times the radiation penetration depth.

In photoconductivity, the recombination is not associated with the emission of radiation, and therefore, it is said to be nonradiative. The Auger effect and multiphonon emission are two common processes of nonradiative recombination. In the Auger effect, the energy released by a recombining electron-hole pair is absorbed by another electron in the conduction band, which subsequently relaxes to the equilibrium condition by the emission of phonons. In a multiphonon emission process, the recombination of an electron-hole pair is associated with the release of a cascade of phonons, each having a much lower energy. More details on the recombination process and how to calculate the associated lifetime can be found from Sze.<sup>11</sup>

Radiative recombination can also occur and is very important for light-emitting applications, such as *luminescence*, which is essentially the inverse process of absorption. The excitation of electrons may be accomplished by passing through an electrical current. An example is the semiconductor light-emitting diode, in which the electronic transition from

the conduction band to the valence band can result in optical radiation. Photoluminescence is often referred to as *fluorescence*, when the emission occurs at the same time as the absorption, or *phosphorescence*, when the emission continues for a while after the excitation.

### 6.7.3 The $p$ - $n$ Junction

The  $p$ - $n$  junction is familiar to every reader although many of us are unfamiliar with the underlying physics. Let us first take a look at the charge transport by diffusion, which is a very important process in semiconductor applications. Diffusion takes place when there is a spatial nonuniformity in the carrier concentration. The principle is the same as the diffusion of ideal gas molecules described in Sec. 4.2.3. Using Fick's law, we can write the current densities resulting from the diffusion of electrons and holes as follows:

$$J_e = eD_e \frac{dn_e}{dx} \quad \text{and} \quad J_h = -eD_h \frac{dn_h}{dx} \quad (6.89)$$

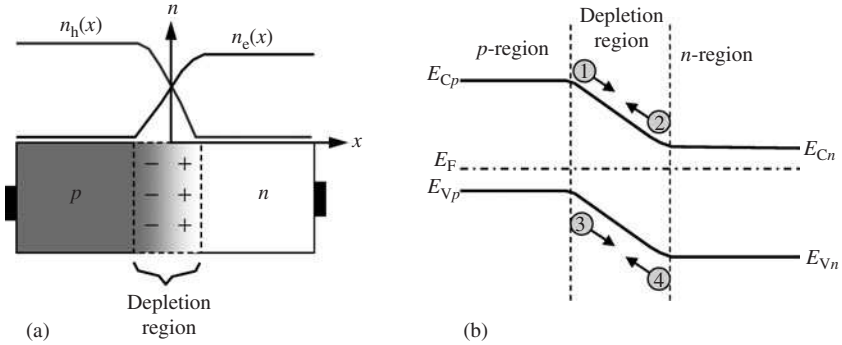
where the diffusion coefficient for electrons and holes can be related to the mean free path and the average velocity by  $D_e = \frac{1}{3}\Lambda_e\bar{v}_e$  and  $D_h = \frac{1}{3}\Lambda_h\bar{v}_h$ , according to Eq. (4.42). Assuming  $\bar{v}_e \approx \bar{v}_h \approx v_{th}$ , we have  $D_e = \frac{1}{3}\Lambda_e v_{th} = \frac{1}{3}\tau_e v_{th}^2$ . Combined with  $\frac{1}{2}m_e^* v_{th}^2 = \frac{3}{2}k_B T$ , we obtain

$$D_e = \frac{\tau_e k_B T}{m_e^*} = \frac{\mu_e k_B T}{e} \quad (6.90)$$

which is known as the *Einstein relation*. A similar equation holds also for the holes. In transient heat conduction, the *thermal diffusion length* is usually calculated by  $l_{th} = \sqrt{\alpha t}$ , where  $\alpha = \kappa/\rho c_p$  is the thermal diffusivity and  $t$  is a characteristic time. The *diffusion length* for electrons is defined as  $l_d = \sqrt{D_e \tau_e}$ , which is proportional to  $\tau_e$  and  $v_{th}$ . The diffusion velocity is sometimes defined as  $v_d = l_d/\tau_e = v_{th}/\sqrt{3}$ . The factor of  $\sqrt{3}$  reduction arises because the diffusion velocity is the average thermal velocity along one direction only. In semiconductors, charge transfer is a combined effect of the carrier drift and diffusion. Electron diffusion is not important for metals because of the large drift velocity given by the Fermi velocity, which changes little with temperature at moderate temperatures.

Through oxidation, lithography, diffusion and ion implantation, and metallization, semiconductor  $p$ - $n$  junctions can be fabricated with microelectronics manufacturing technology.<sup>11</sup> A  $p$ - $n$  junction consists of a  $p$ -type semiconductor, with a high hole concentration, joined with an  $n$ -type semiconductor, with a high electron concentration, as shown in Fig. 6.23.

If one compares Fig. 6.23a with Fig. 4.7b, the process looks similar to a binary diffusion. Because of the concentration gradient, holes will diffuse right and electrons will diffuse left. Diffusion causes the region near the interface to be depleted, so that there are fewer free holes on the left side and fewer free electrons on the right side of the depletion region. Keep in mind that electrons and holes are charged particles. As they leave the host material, ions of opposite charges are left behind. This results in a charge accumulation, as shown in Fig. 6.23a, that leads to a built-in potential in the depletion region that will inhibit further diffusion. As a consequence of this built-in potential, the energy in the  $p$ -doped region is raised relative to that in the  $n$ -doped region, as shown in Fig. 6.23b. The Fermi level is the same everywhere, and it is closer to the conduction band for the  $n$ -type and the valence band for the  $p$ -type.



**FIGURE 6.23** Schematic of a  $p$ - $n$  junction at thermal equilibrium. (a) The device and carrier concentrations, including the charge distribution in the depletion region. The width of the depletion region is exaggerated for clarity. (b) The energy band diagram for the  $p$ - $n$  junction near the depletion region. The dash-dotted line is the Fermi level. The four processes are (1) electron drift, (2) electron diffusion, (3) hole diffusion, and (4) hole drift.

**Example 6-7.** Prove that the Fermi energy in a  $p$ - $n$  junction is independent of  $x$  at thermal equilibrium, as shown in Fig. 6.23b

**Solution.** Without any externally applied voltage, the current densities become

$$J_e = J_{e,\text{drif}} + J_{e,\text{diff}} = -en_c\mu_c \frac{dV_0}{dx} + eD_c \frac{dn_c}{dx} \quad (6.91a)$$

$$J_h = J_{h,\text{drif}} + J_{h,\text{diff}} = -en_h\mu_h \frac{dV_0}{dx} - eD_c \frac{dn_h}{dx} \quad (6.91b)$$

where  $V_0$  is the built-in potential and the electric field is  $-dV_0/dx$ . Because a high potential  $V_0$  means a smaller electron kinetic energy, we have

$$\frac{dV_0}{dx} = -\frac{1}{e} \frac{dE_C}{dx} = -\frac{1}{e} \frac{dE_V}{dx} \quad (6.92)$$

From Fig. 6.23, we see that the built-in electric field in the depletion region points toward the negative  $x$  direction and thus  $dV_0/dx > 0$ ; consequently,  $dE_C/dx < 0$  and  $dE_V/dx < 0$ . Employing Eq. (6.72), we notice that

$$\frac{1}{n_c} \frac{dn_c}{dx} = \frac{1}{k_B T} \left( \frac{dE_F}{dx} - \frac{dE_C}{dx} \right) \quad \text{and} \quad \frac{1}{n_h} \frac{dn_h}{dx} = \frac{1}{k_B T} \left( \frac{dE_V}{dx} - \frac{dE_F}{dx} \right) \quad (6.93)$$

Substituting Eq. (6.90), Eq. (6.92) and Eq. (6.93) into Eq. (6.91a) and setting  $J_e = 0$ , we end up with

$$\frac{dE_F}{dx} = 0 \quad (6.94)$$

This equation can also be derived using Eq. (6.91b); hence, the Fermi energy  $E_F$  is independent of  $x$ .

A popular application of  $p$ - $n$  junction is as a diode rectifier, which allows current to flow easily with a forward bias but becomes highly resistive when the bias is reversed. For the configuration shown in Fig. 6.23, a forward bias means that the electrical field is in the positive  $x$  direction, opposite to the built-in field. Qualitatively, this can be understood as a forward bias removes the barrier for holes to diffuse right and for electrons to diffuse left. On



the other hand, a reverse bias creates an even stronger barrier for these diffusion processes. Quantitatively, it can be shown that for an externally applied voltage  $V$  (positive for forward bias and negative for reverse bias), the current density can be expressed as

$$J = J_s \left[ \exp\left(\frac{eV}{k_B T}\right) - 1 \right] \quad (6.95)$$

where  $J_s$  is the saturation current density, which depends on the diffusion coefficient, scattering time, number density, and other factors. Since  $dJ/dV = (J_s k_B T/e) \exp(eV/k_B T)$ , the electrical conductance increases with  $V$  for forward bias, and decreases to zero as  $V \rightarrow -\infty$ . It should be noted that in practice, the width of the depletion region is often less than  $0.5 \mu\text{m}$  and the built-in potential may be around 1 V through the depletion region. There is actually a very large built-in field.

Heterojunction is a junction of dissimilar semiconductors with different bandgap energies. The energy band diagram can be very different from that shown in Fig. 6.23b. The Fermi energies can be different on each side. Bipolar transistors were invented in 1947 at Bell Labs. It is based on two  $p$ - $n$  junctions arranged in a  $p$ - $n$ - $p$  or  $n$ - $p$ - $n$  configuration. Field-effect transistors (FETs) work on a different principle. As shown in Fig. 1.3, the free electrons cannot move from the source to the drain because of the lack of free carriers in the  $p$ -type wafer. If a negative voltage is applied to the gate, electrons below the gate will be pushed even further, and there is still little chance for the electron to flow from the source to the drain. However, as soon as a positive voltage is applied to the gate, electrons will be attracted to the region below it and form a path for electricity to flow from the source to the drain. Furthermore, a transistor can amplify the signal since only a weak signal is necessary to the gate. Metal-oxide-semiconductor field-effect transistors (MOSFETs) have become the most important device in contemporary integrated circuits. Thermal management is important for such devices because of the local heating or hot spots where Fourier's law often fails to predict the temperature history. More discussion on nonequilibrium heat conduction will be given in Chap. 7. A brief discussion on photovoltaic devices will be given next.

## 6.7.4 Optoelectronic Applications

The photovoltaic effect is a direct energy conversion process in which electromagnetic radiation, incident upon a  $p$ - $n$  junction, generates electron-hole pairs. The built-in electric field in the  $p$ - $n$  junction tends to push the generated holes to the  $p$ -region and the generated electrons to the  $n$ -region, resulting in a reverse photocurrent. Solar cells and photovoltaic detectors have been developed and applied for over half a century. Thermophotovoltaic (TPV) devices have also been considered as energy conversion systems that allow recycling of the waste heat.<sup>49</sup> Figure 6.24 shows a typical TPV cell and the associated electrical circuit. When the incident radiation with a photon energy greater than the bandgap energy  $E_g$  of the cell material strikes the  $p$ - $n$  junction, an electron-hole pair is generated at the location as each photon is absorbed. Carriers generated in the depletion region are swept by the built-in electric field and then collected by the electrodes at the ends of the cell, resulting in a drift current. For radiation absorbed near the depletion region, the minority carriers (electrons in the  $p$ -region, and holes in the  $n$ -region) tend to diffuse toward the depletion region, yielding a diffusion current. If the load resistance  $R_L$  is zero, i.e., in the case of a short circuit, there is a photocurrent  $I_{ph}$  flowing in the circuit due to the combination of the diffusion and drift of charge carriers. The direction of this current is indicated on Fig. 6.24. If the circuit is open, or the load resistance  $R_L$  approaches infinite, a positive open-circuit voltage is built up due to irradiation. This gives the maximum voltage

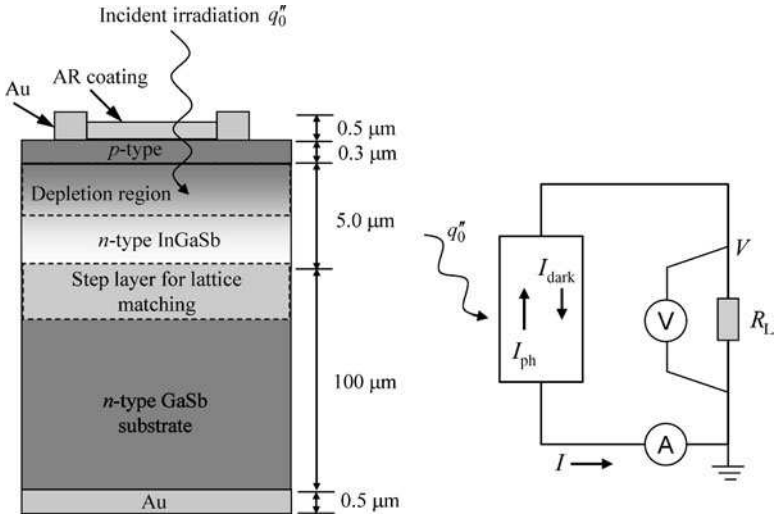


FIGURE 6.24 Schematic of a typical TPV cell with its circuit diagram on the right.<sup>49</sup>

$V = V_{\max}$ , when no current flows through the load, i.e.,  $I = 0$ . When the load has a finite resistance  $R_L$ , a voltage  $V$  is developed, not only across the load but also across the photovoltaic cell. This voltage reduces the built-in potential of the cell as if a forward bias is applied to the  $p$ - $n$  junction. Subsequently, the diffusion of minority carriers produces a forward current, which is called the *dark current* in photovoltaic devices. The current  $I$  flowing through the load resistor becomes

$$I = -I_{\text{ph}} + I_s \left[ \exp\left(\frac{eV}{k_B T}\right) - 1 \right] \quad (6.96)$$

The first term on the right is the photocurrent, or the short-circuit current, which depends on the incident photon flux, quantum efficiency, as well as the transport properties. The second term on the right is the dark current  $I_{\text{dark}}$ , and  $I_s$  is the saturation current, as defined in Eq. (6.95) based on the current density. The dark current is zero, when  $V = 0$ . For the photovoltaic cell, shown in Fig. 6.24, if the incident radiation flux  $q_0'' = 0$ , then  $I$ ,  $I_{\text{ph}}$ , and  $V$  become zero at thermal equilibrium. Basu et al. recently provided an extensive review of the operation principle and the state of the art in TPV technology, as well as the potential application of microscale radiative heat transfer for performance improvement.<sup>49</sup>

Light-emitting diodes (LEDs) are based on  $p$ - $n$  junctions as well but with direct gap semiconductors. At low forward bias voltages, the recombination processes are essentially nonradiative. At high forward bias voltages, however, radiative recombination results in the emission of photons. The emission is a spontaneous process and is incoherent. Depending on the materials used and their bandgaps, LEDs can emit in the ultraviolet, visible, and infrared regions.

Semiconductor lasers are based on the stimulated emission process, as discussed in Chap. 3, and have numerous important applications due to their small size, portability, and ease of operation. Semiconductor lasers have been used in laser printers, optical fiber communication, CD reading/writing, and so forth. The key is to create population inversion so that lasing can occur. Quantum well lasers, based on quantum confinement, offer significant advantages over conventional semiconductor lasers, such as low threshold current,

high output power, high speed, and so forth. Further explanation of the optical and electronic characteristics of semiconductor lasers can be found from Sze<sup>11</sup> and Zory,<sup>50</sup> for example.

## 6.8 SUMMARY

---

This chapter began with an introduction to the atomic structures, chemical bonds, and crystal lattices. Emphasis was given on electronic band structures and phonon dispersion relations, allowing one to gain a deeper knowledge of solid state physics, beyond the previous chapter. Photoelectric effect, thermionic emission, and field emission were described in subsequent sections to stress the interrelation between these phenomena. The basic electrical transport processes in semiconductors, such as number density, mobility, electrical conductivity, charge diffusion, and photoconductivity were explained. The  $p$ - $n$  junction was discussed along with applications, such as photovoltaic cells, LEDs, and semiconductor lasers.

## REFERENCES

---

1. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Harcourt College Publishers, Fort Worth, TX, 1976.
2. C. Kittel, *Introduction to Solid State Physics*, 7th ed., Wiley, New York, 1996.
3. J. E. Avron, D. Osadchy, and R. Seiler, "A topological look at the quantum Hall effect," *Physics Today*, 38–42, 2003.
4. G. A. Prinz, "Magnetoelectronics," *Science*, **282**, 1660–1663, 1998.
5. K. von Klitzing, G. Dorda, and M. Pepper, "New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance," *Phys. Rev. Lett.*, **45**, 494–497, 1980; K. von Klitzing, "The quantized Hall effect," *Rev. Mod. Phys.*, **58**, 519–530, 1986.
6. A. Hartland, "The quantum Hall effect and resistance standards," *Metrologia*, **29**, 175–190, 1992.
7. J. P. Eisenstein and H. L. Strörmer, "The fractional quantum Hall effect," *Science*, **248**, 1510–1516, 1990.
8. C. Strohm, G. L. J. A. Rikken, and P. Wyder, "Phenomenological evidence for the phonon Hall effect," *Phys. Rev. Lett.*, **95**, 155901, 2005.
9. J. E. Lay, *Statistical Mechanics and Thermodynamics of Matter*, Harper Collins Publishers, New York, 1990.
10. F. Yang and J. H. Hamilton, *Modern Atomic and Nuclear Physics*, McGraw-Hill, New York, 1996.
11. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., Wiley, New York, 1981; S. M. Sze, *Semiconductor Devices: Physics and Technology*, 2nd ed., Wiley, New York, 2002.
12. R. J. Cava, "Structure chemistry and the local charge picture of copper oxide superconductors," *Science*, **247**, 656–662, 1990.
13. B. I. Choi, Z. M. Zhang, M. I. Flik, and T. Siegrist, "Radiative properties of Y-Ba-Cu-O films with variable oxygen content," *J. Heat Transfer*, **114**, 958–964, 1992.
14. M. S. Dresselhaus and P. C. Eklund, "Phonons in carbon nanotubes," *Adv. Phys.*, **49**, 705–814, 2000.
15. M. A. Omar, *Elementary Solid State Physics: Principles and Applications*, Addison-Wesley, New York, 1975.
16. B. Segall, "Fermi surface and energy band of copper," *Phys. Rev.*, **125**, 109–122, 1962.
17. G. A. Burdick, "Energy band structure of copper," *Phys. Rev.*, **129**, 138–150, 1963.
18. R. E. Hummel, *Electronic Properties of Materials*, Springer-Verlag, Berlin, 1993.

19. M. L. Cohen and T. K. Bergstresser, "Band structure and pseudopotential from factors for fourteen semiconductors of the diamond and zinc-blende structures," *Phys. Rev.*, **141**, 789–796, 1966.
20. F. Herman and W. E. Spicer, "Spectral analysis of photoemission yields in GaAs and related crystals," *Phys. Rev.*, **174**, 906–908, 1968.
21. M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors*, Springer-Verlag, Berlin, 1988.
22. M. Born and K. Huang, *Dynamic Theory of Crystal Lattices*, Oxford University Press, London, 1954.
23. G. P. Srivastava, *The Physics of Phonons*, Adam Hilger, Bristol, 1990.
24. J. M. Ziman, *Electrons and Phonons*, Oxford University Press, Oxford, 1960.
25. B. N. Brockhouse, "Lattice vibration in silicon and germanium," *Phys. Rev. Lett.*, **2**, 256–258, 1959.
26. G. Dolling, in *Second Symposium on Inelastic Scattering of Neutrons in Solids and Liquids*, Chalk River, Canada (IAEA, Vienna, 1963, Vol. II, p. 37).
27. R. Tubino, L. Piseri, and G. Zerbi, "Lattice dynamics and spectroscopic properties by a valence force potential of diamondlike crystals: C, Si, Ge, and Sn," *J. Chem. Phys.*, **56**, 1022–1039, 1972.
28. D. W. Feldman, J. H. Parker, Jr., W. J. Choyke, and L. Patrick, "Phonon dispersion curves by Raman scattering in SiC, polytypes 3C, 4H, 6H, 15R, and 21R," *Phys. Rev.*, **173**, 787–793, 1968.
29. A. M. Greenstein, S. Graham, Y. C. Hudiono, and S. Nair, "Thermal properties and lattice dynamics of polycrystalline MFI zeolite films," *Nanoscale & Microscale Thermophys. Eng.*, **10**, 321–331, 2006.
30. D. J. Ecsedy and P. G. Klemens, "Thermal conductivity of dielectric crystals due to four-phonon processes and optical modes," *Phys. Rev. B*, **15**, 5957–5962, 1977.
31. Z. M. Zhang, "Surface temperature measurement using optical techniques," *Annu. Rev. Heat Transfer*, **11**, 351–411, 2000.
32. G. N. Hatsopoulos and E. P. Gyftopoulos, *Thermionic Energy Conversion*, Vol. 1 (1973); Vol. 2 (1979), MIT Press, Cambridge, MA.
33. G. D. Mahan, "Thermionic refrigeration," *J. Appl. Phys.*, **76**, 4362–4366, 1994; G. D. Mahan and L. M. Woods, "Multilayer thermionic refrigeration," *Phys. Rev. Lett.*, **80**, 4016–4019, 1998; G. D. Mahan, J. O. Sofo, and M. Barkowiak, "Multilayer thermionic refrigerator and generator," *J. Appl. Phys.*, **83**, 4683–4689, 1998.
34. A. Shakouri and J. E. Bowers, "Heterostructure integrated thermionic coolers," *Appl. Phys. Lett.*, **71**, 1234–1236, 1997; A. Shakouri, C. LaBounty, J. Piprek, P. Abraham, and J. E. Bowers, "Thermionic emission cooling in single barrier heterostructures," *Appl. Phys. Lett.*, **74**, 88–89, 1999.
35. D. Vashaee and A. Shakouri, "Nonequilibrium electrons and phonons in thin film thermionic coolers," *Microscale Thermophys. Eng.*, **8**, 91–100, 2004; D. Vashaee and A. Shakouri, "Electronic and thermoelectric transport in semiconductor and metallic superlattices," *J. Appl. Phys.*, **95**, 1233–1245, 2004.
36. T. Zeng and G. Chen, "Interplay between thermoelectric and thermionic effects in heterostructures," *J. Appl. Phys.*, **92**, 3152–3161, 2002; T. Zeng and G. Chen, "Nonequilibrium electron and phonon transport and energy conversion in heterostructures," *Microelectronics J.*, **34**, 201–206, 2003.
37. Y. Hishinuma, T. H. Geballe, B. Y. Mozysh, and T. W. Kenny, "Refrigeration by combined tunneling and thermionic emission in vacuum: Use of nanometer scale design," *Appl. Phys. Lett.*, **78**, 2572–2574, 2001; Y. Hishinuma, T. H. Geballe, B. Y. Mozysh, and T. W. Kenny, "Measurements of cooling by room-temperature thermionic emission across a nanometer gap," *J. Appl. Phys.*, **94**, 4690–4696, 2003.
38. T. Zeng, "Thermionic-tunneling multilayer nanostructures for power generation," *Appl. Phys. Lett.*, **88**, 153104, 2006.
39. R. H. Fowler and L. Nordheim, "Electron emission in intense electric field," *Proc. Royal Soc. Lond. A*, **119**, 173–181, 1928.
40. D. J. Griffiths, *Introduction to Quantum Mechanics*, 2nd ed., Pearson Prentice Hall, Upper Saddle River, NJ, 2005.

41. R. Tsu and L. Esaki, "Tunneling in a finite superlattice," *Appl. Phys. Lett.*, **22**, 562–564, 1973; L. L. Chang, L. Esaki, and R. Tsu, "Resonant tunneling in semiconductor double barriers," *Appl. Phys. Lett.*, **24**, 593–595, 1974.
42. J. W. Gadzuk and E. W. Plummer, "Field emission energy distribution (FEED)," *Rev. Mod. Phys.*, **45**, 487–548, 1973.
43. L. Nilsson, O. Groening, P. Groening, O. Kuettel, and L. Schlapbach, "Characterization of thin film electron emitters by scanning anode field emission spectroscopy," *J. Appl. Phys.*, **90**, 768–780, 2001.
44. J. B. Xu, K. Lauger, R. Moller, K. Dransfeld, and I. H. Wilson, "Energy-exchange processes by tunneling electrons," *Appl. Phys. A*, **59**, 155–161, 1994.
45. T. S. Fisher, "Influence of nanoscale geometry on the thermodynamics of electron field emission," *Appl. Phys. Lett.*, **79**, 3699–3701; T. S. Fisher and D. G. Walker, "Thermal and electrical energy transport and conversion in nanoscale electron field emission process," *J. Heat Transfer*, **124**, 954–962, 2002.
46. C. Trinkle, P. Kichambare, R. R. Vallance et al., "Thermal transport during nanoscale machining by field emission of electrons from carbon nanotubes," *J. Heat Transfer*, **125**, 546, 2003; R. R. Vallance, A. M. Rao, and M. P. Mengüç, "Processes for nanomachining using carbon nanotubes," US Patent No. 6,660,959, Dec. 9, 2003.
47. B. T. Wong, M. P. Mengüç, and R. R. Vallance, "Nano-scale machining via electron beam and laser processing," *J. Heat Transfer*, **126**, 566–576, 2004.
48. R. J. Keyes (ed.), *Optical and Infrared Detectors*, Springer-Verlag, Berlin, 1980.
49. S. Basu, Y.-B. Chen, and Z. M. Zhang, "Microscale radiation in thermophotovoltaic devices—A review," *Int. J. Ener. Res.*, **31**, in press, 2007. (Published online 6 Dec. 2006.)
50. P. S. Zory, Jr. (ed.), *Quantum Well Lasers*, Academic Press, San Diego, 1993.

## PROBLEMS

- 6.1. Consider a phosphorus-doped 250- $\mu\text{m}$ -thick silicon wafer, with a doping concentration of  $10^{17} \text{ cm}^{-3}$ . The applied current is 10 mA, and the magnetic induction is 0.5 T.
  - (a) Determine the Hall coefficient and the Hall voltage, assuming there is only one type of carrier.
  - (b) For a chip of area  $1 \times 1 \text{ cm}^2$  and resistivity  $0.075 \Omega \cdot \text{cm}$ , what is the voltage drop along the direction of current flow?
- 6.2. Consider the Hall experiment arranged in Fig. 6.1, under steady-state operation and with uniform magnetic field. Assume a current is flowing in the  $y$  direction.
  - (a) Show that  $v_x = -(e\tau/m_e)E_x - \omega_c\tau v_y$  and  $v_y = -(e\tau/m_e)E_y + \omega_c\tau v_x$ , when the current is carried by electrons. Here,  $\tau$  is the relaxation time,  $v_x$  and  $v_y$  are the electron drift velocities in the  $x$  and  $y$  directions, respectively, and  $\omega_c = eB/m_e$  is called the *cyclotron frequency*.
  - (b) Prove Eq. (6.1) by setting  $v_y = 0$ .
- 6.3. Express the electron configurations for Ag and Au. Based on the orbital occupation of outer electrons, discuss the similarities in their chemical and electrical properties.
- 6.4. Express the electron configurations for Ca and Zn. Based on the orbital occupation of outer electrons, discuss the similarities in their chemical and electrical properties.
- 6.5. Give a general discussion of insulators, semiconductors, and metals. Explain why glass ( $\text{SiO}_2$ ) is transparent, silicon wafers appear dark, and aluminum foils look bright. What are the types of chemical bonds in  $\text{SiO}_2$ , Si, and Al?
- 6.6. How many billiard balls can you pack in a basket with a volume of  $0.25 \text{ m}^3$ ? Assume that the balls are rigid spheres with a diameter  $d = 43 \text{ mm}$  and mass  $m = 46 \text{ g}$ . Arrange the spheres in a crystal lattice according to the diamond, simple cubic, bcc, fcc, and hcp structures. What is the total weight for each arrangement? [Hint: Show that for close-packed spheres, the fraction of volume occupied by the spheres is  $\sqrt{3}\pi/16 \approx 0.340$  (diamond),  $\pi/6 \approx 0.524$  (simple cubic),  $\sqrt{3}\pi/8 \approx 0.680$  (bcc), and  $\sqrt{2}\pi/6 \approx 0.740$  (fcc or hcp).
- 6.7. (a) Count the number of atoms inside a unit cell of  $\text{YBa}_2\text{Cu}_3\text{O}_7$  as shown in Fig. 6.6d, and confirm that it is the same as that in the basis.

(b) Find the density of  $\text{YBa}_2\text{Cu}_3\text{O}_7$  crystal based on the dimensions of the unit cell, noting that the molecular weight  $M = 88.9$  (Y), 137.3 (Ba), 63.6 (Cu), and 16.0 (O) kg/kmol.

**6.8.** (a) Calculate the diameter and cross-sectional area for CNTs with chiral indices  $(m,n) = (5,5)$ ,  $(8,8)$ ,  $(10,10)$ ,  $(10,20)$ , and  $(20,40)$ .

(b) Take  $(40,40)$  SWNTs of 10- $\mu\text{m}$  length, with a thermal conductivity  $\kappa = 3200 \text{ W}/(\text{m} \cdot \text{K})$  at room temperature. Align sufficient nanotubes to make a bundle with a diameter of 1  $\mu\text{m}$ ; how many wires are needed?

(c) Neglect the effect of interface and defects on the thermal conductivity. What is the heat transfer rate if the temperatures at both ends are 320 and 300 K?

(d) Compare the heat transfer rate if the CNT is replaced by a Si nanowire of 1- $\mu\text{m}$  diameter and 10- $\mu\text{m}$  length.

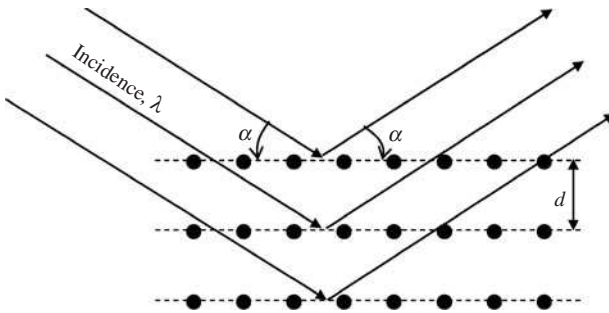
**6.9.** The interatomic potential for a KBr crystal can be expressed as  $\phi(r) = -\alpha_M e^2/(4\pi\epsilon_0 r) + C(alr)^m$ , where  $\alpha_M$  is the Madelung constant, which is 1.748 for crystals with NaCl structure,  $\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2/\text{J} \cdot \text{m}$  is the electric permittivity of vacuum,  $a = 6.60 \times 10^{-10} \text{ m}$  is the lattice constant,  $m = 8.85$ , and  $C = 2.65 \times 10^{-21} \text{ J}$  for KBr. Note that  $r$  is in meter.

(a) Plot the attractive potential, the repulsive potential, and the combined potential in eV as a function of  $r$  in  $\text{\AA}$ .

(b) Find the equilibrium distance, which should be the nearest distance between  $\text{K}^+$  and  $\text{Br}^-$  ions.

(c) At the equilibrium distance, what are the attractive and repulsive forces between each ion pair?

**6.10.** Bragg's x-ray diffraction formula relates the angle  $\alpha$  of diffraction maximum and the x-ray wavelength  $\lambda$  as follows:  $2d\sin\alpha = n\lambda$ , where  $n$  is the refractive index that can be taken as unity in the x-ray region,  $d$  is the spacing between adjacent layers of atoms, and  $\alpha$  is measured between the incidence and the crystal plane, as shown in Fig. P6.10. This formula can be understood by the constructive interference between the two layers.



**FIGURE P6.10** Schematic of Bragg's x-ray diffraction experiment.

(a) To measure a spacing  $d = 3.12 \text{ \AA}$ , what is the maximum wavelength  $\lambda$  that can still be used to perform the experiment successfully?

(b) In an x-ray experiment,  $\lambda = 1.5 \text{ \AA}$ . Assume that the errors in  $\lambda$  and  $n$  are negligible. How accurately must one determine  $\alpha$  in order to measure the spacing with an uncertainty of  $0.01 \text{ \AA}$ ?

**6.11.** Using Eq. (6.10) to show that the reciprocal lattice of a hexagon is also a hexagon, as shown in Fig. 6.3d, calculate the volumes of the direct and reciprocal lattices in terms of  $a$  and  $c$ .

**6.12.** Use the Kronig-Penney model to solve the Schrödinger equation for an electron in a square well array. Referring to Fig. 6.10, assume that the potential function is  $U(x) = 0$  at  $0 \leq x \leq (a-b)/2$  and  $(a+b)/2 \leq x \leq a$ , and  $U(x) = U_0 > 0$  at  $(a-b)/2 \leq x \leq (a+b)/2$ . Note that  $x = 0$  is at the core of atom location and the potential is periodic. Find the conditions for the solutions to exist. For simplicity, you may now assume  $b \rightarrow 0$  and  $U_0 \rightarrow \infty$  to obtain the relation for  $E(k)$ . Plot this function to illustrate the electronic band structure.

**6.13.** Discuss the difference between interband transitions and transitions that occur within a band for copper. Explain why copper appears reddish brown.

**6.14.** What is the difference between a direct bandgap semiconductor and an indirect bandgap semiconductor? Why are Si and GaAs wafers opaque to the visible light?

- 6.15.** Prove Eq. (6.33) and Eq. (6.35) first. Then, plot the phonon dispersion curves for a diatomic chain with mass ratio  $m_1/m_2$  equal to 1, 2, 3, and 4. What happens when  $m_1/m_2 = 1$ ?
- 6.16.** Approximate  $k_{\max} = \pi/a$ , and find the group velocities of LA and TA phonons for Si and SiC at  $k = 0.3k_{\max}$ , based on Fig. 6.16. What is the phase speed at  $k = 0.3k_{\max}$  for LO phonon in SiC? [Hint: Convert the unit of  $\omega$  from  $\text{cm}^{-1}$  to  $\text{rad/s}$  first.]
- 6.17.** Prove Eq. (6.59) and Eq. (6.60). Assume that  $\psi = 0.4 \text{ eV}$  and  $E_F = 3 \text{ eV}$ , estimate the error in Eq. (6.59) caused by approximating the Fermi-Dirac distribution with the Maxwell-Boltzmann distribution in the numerical evaluation.
- 6.18.** Clearly explain the differences between thermionic emission and field emission.
- 6.19.** For a gallium-doped silicon with  $N_A = 5 \times 10^{16} \text{ cm}^{-3}$ , use the information from Example 6-6 to calculate the number density of electrons and holes from 300 to 1000 K. Assume the effect of impurity on the mobility can be neglected so that  $\mu_e = 1450 \text{ cm}^2/(\text{V} \cdot \text{s})$  for electrons, and  $\mu_h = 500 \text{ cm}^2/(\text{V} \cdot \text{s})$  for holes at 300 K. Determine the electrical resistivity of the doped silicon from 300 to 1000 K.
- 6.20.** For a single-type doped silicon with  $\mu_e = 1350 \text{ cm}^2/(\text{V} \cdot \text{s})$  and  $\mu_h = 450 \text{ cm}^2/(\text{V} \cdot \text{s})$  at 400 K, the Hall coefficient is zero. Is this semiconductor  $n$ -type, or  $p$ -type? What is the impurity concentration? [Hint: Use the parameters given in Example 6-6.]
- 6.21.** For a single-type doped silicon with  $\mu_e = 1350 \text{ cm}^2/(\text{V} \cdot \text{s})$ ,  $\mu_h = 450 \text{ cm}^2/(\text{V} \cdot \text{s})$ , and  $N_{\text{th}} = 2 \times 10^{10} \text{ cm}^{-3}$ , calculate and plot the Hall coefficient for  $p$ -type doping, with  $N_A$  ranging from 0 to  $2 \times 10^{12} \text{ cm}^{-3}$ . Discuss, without calculation, the trend with  $n$ -type doping.
- 6.22.** For a phosphorus-doped silicon,  $N_d = 2 \times 10^{15} \text{ cm}^{-3}$ ,  $\mu_e = 1350 \text{ cm}^2/(\text{V} \cdot \text{s})$ , and  $\mu_h = 450 \text{ cm}^2/(\text{V} \cdot \text{s})$ , at 300 K. Use the parameters from Example 6-6 as needed.
- (a) Calculate the thermal velocity and the diffusion length for the electrons and holes at room temperature.
- (b) Find the electrical conductivity at room temperature.
- (c) Plot the thermal velocity and wavelength as a function of temperature.
- 6.23.** Show the  $I$ - $V$  curve of a  $p$ - $n$  junction, based on Eq. (6.95), using dimensionless groups  $J/J_s$  and  $eV/k_B T$ . Discuss the meaning of saturation current density.
- 6.24.** Show the  $I$ - $V$  curve for a photovoltaic cell, and determine the open voltage. Show, on the same diagram, the  $I$ - $V$  curve without irradiation, i.e., with no photocurrent. Discuss the meaning of dark current.

---

# CHAPTER 7

---

## NONEQUILIBRIUM ENERGY TRANSFER IN NANOSTRUCTURES

---

Fourier's law and the associated heat diffusion equation comprise one of the most celebrated models in mathematical physics. Joseph Fourier in 1824 wrote: *Heat, like gravity, penetrates every substance of the universe; its rays occupy all parts of space. . . . The theory of heat will hereafter form one of the most important branches of general physics.* Soon afterward, heat transfer also became an important engineering field, essential to the second industrial revolution and the development of modern technologies.

Recall the discussion of heat interaction and heat transfer in Chap. 2. We have treated heat conduction as a diffusion process based on the concept of local thermal equilibrium. This allows us to define and determine the equilibrium temperature at each location in a body instantaneously, under the continuum assumption described in Chap. 1. The local-equilibrium condition breaks down at the microscale when the characteristic length  $L$  is smaller than a mechanistic length scale, such as the mean free path  $\Lambda$ . For conduction by molecules, consider a rarefied gas between two parallel plates at different temperatures. If the mean free path is much greater than the separation distance, i.e., the Knudsen number  $Kn = \Lambda/L \gg 1$ , the gas is in the free molecule regime and its velocity distribution cannot be described by Maxwell's distribution function. Furthermore, the transport becomes ballistic rather than diffusive. Nonequilibrium energy transfer refers to the situation when the assumption of local equilibrium does not hold. This can occur in solid nanostructures even at the room temperature and in steady state, or in bulk solids under the influence of short pulse heating.

For heat conduction *across* a dielectric thin film, when the thickness is much smaller than the phonon mean free path, which increases as the temperature goes down, the condition of local equilibrium is not satisfied. Hence, the phonon statistics at a given location cannot be described by the equilibrium distribution function at any given temperature. Strictly speaking, temperature cannot be defined inside the medium. However, an *effective* temperature is typically adopted, based on the statistical average of the particle energies. In the case of heat transfer across a thin dielectric film or between two plates separated by a rarefied molecular gas, the effective temperature distribution cannot be described by the heat diffusion theory derived from Fourier's law using the concept of equilibrium temperature without considering the temperature jumps at the boundaries. This has already been demonstrated in Chap. 4 (see Fig. 4.12). Consider a metal or a superconductor that is subjected to ultrafast pulsed-laser heating, in which the pulse duration may range from several femtoseconds to a few nanoseconds. The electrons gain energy quickly to reach a state that is far from equilibrium with the crystal lattice or the phonon system. The transport processes during and immediately after the laser pulse become nonequilibrium both temporally and spatially. Conventional Fourier's law cannot be directly applied.



In Chap. 5, we have considered the size effect on thermal transport in solids. Two approaches have been used under different situations. In the first situation, we apply Matthiessen's rule to account for the reduction in mean free path by assuming that Fourier's law is still applicable but with a size-dependent thermal conductivity. In the second situation, where the transport is completely ballistic, we use the concept of quantum conductance based on the Landauer formulation to solve the problem in a straightforward manner. The definition of an effective thermal conductivity is particularly useful for the study of transport processes *along* a thin film or a thin wire, when the length in the direction of transport is much greater than the mean free path. In this case, a local equilibrium can be established, and thus, the energy transfer is well described by Fourier's law, even though the thickness is less than the mean free path. Here, the only microscale effect is the classical size effect, which arises from boundary scattering of electrons in a metal or phonons in an insulator or a semiconductor. For energy transport across a thin film or in a multilayer structure, on the other hand, the local-equilibrium condition breaks down when the film thickness is much smaller than the mean free path. Furthermore, thermal boundary resistance (TBR) may become significant at the interfaces. Because of the wave-particle duality, the electron wave or phonon wave effect may need to be considered in some cases. For non-metallic crystalline materials, the most commonly used method to study thermal transport is based on the Boltzmann transport equation (BTE) of phonons. Various assumptions and techniques have been developed to solve the phonon BTE. In very small structures, such as nanotubes or nanowires, atomistic simulations may prove more effective.

This chapter first describes the phenomenological theories in which the energy transport processes are represented by a single differential equation or a set of differential equations that can be solved with appropriate initial and boundary conditions. These equations are often called non-Fourier heat equations, which can be considered as extensions of the Fourier heat conduction model, for better or worse. The second section summarizes statistical and atomistic modeling techniques. While the BTE, Monte Carlo method, and molecular dynamics simulations have been presented in Chap. 4, the discussion in the present chapter stresses the application in solid nanostructures, including thermal boundary resistance (TBR) and multilayer structures, with some up-to-date references on solid conduction, multiscale modeling, and thermal metrology.

## 7.1 PHENOMENOLOGICAL THEORIES

A fundamental difficulty of Fourier's heat conduction theory was thought to be that a thermal disturbance in one location of the medium would cause a response at any other location instantaneously, as required by the mathematical solution of the diffusion equation. In theory, the speed of heat propagation appears to be unlimited; this has been viewed by some as a direct violation of the principle of causality. Let us begin with an example of 1-D transient heating of a semi-infinite medium. Assume that the medium is homogeneous, with constant thermal properties, and is initially at a uniform temperature  $T(x,0) = T_i$ . The thermal diffusivity of the medium is  $\alpha = \kappa/(\rho c_p)$ , where  $\kappa$ ,  $\rho$  and  $c_p$  are the thermal conductivity, density, and specific heat of the material, respectively. The wall at  $x = 0$  is heated with a constant heat flux  $q_0''$  at  $0 < t \leq t_p$ , where  $t_p$  is the width of the step heating, and insulated at  $t > t_p$ . The solution of the temperature distribution  $T(x,t)$  can be found from Carslaw and Jaeger<sup>1</sup> and Özişik<sup>2</sup> as follows:

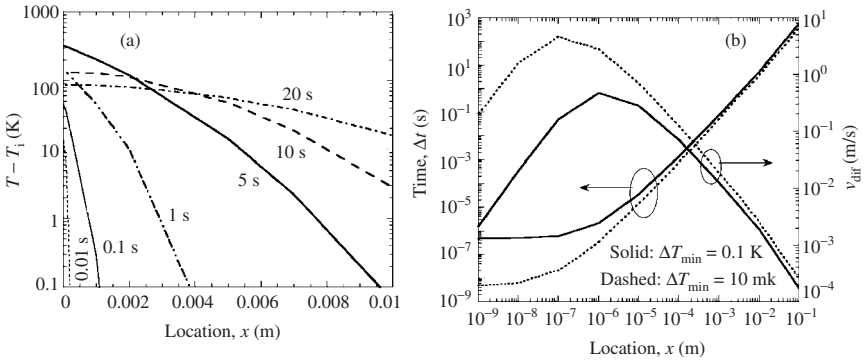
$$T(x,t) - T_i = 2q_0'' \frac{\sqrt{\alpha t}}{\kappa} F(\xi) \quad \text{at } 0 < t \leq t_p \quad (7.1a)$$

$$T(x,t) - T_i = 2q_0'' \frac{\sqrt{\alpha t}}{\kappa} \left[ F(\xi) - \eta F\left(\frac{\xi}{\eta}\right) \right] \quad \text{at } t > t_p \quad (7.1b)$$

where  $\xi = x/\sqrt{4\alpha t}$ ,  $\eta = \sqrt{1 - t_p/t}$ , and  $F(\xi) = \exp(-\xi^2)/\sqrt{\pi} - \xi \operatorname{erfc}(\xi)$  with  $\operatorname{erfc}$  being the complementary error function as given in Appendix B.1.2. While  $F(10) = 1.0 \times 10^{-46}$  and the right-hand sides of both Eq. (7.1a) and Eq. (7.1b) are essentially negligible when  $x > 6\sqrt{\alpha t}$ , the paradox is that a nonzero response must not occur faster than the speed of the thermal energy carriers, such as the Fermi velocity in metals or the speed of sound in dielectrics. In reality, this rarely causes any problem because a signal that is below the noise level cannot be detected by any physical instrument, as will be discussed in the following example:

**Example 7-1.** A thick plate of fused silica  $\text{SiO}_2$ , initially at room temperature, is heated at one surface by a heat flux of  $2.0 \times 10^5 \text{ W/m}^2$  for 5 s and then insulated. Treat the heated surface to be at  $x = 0$ , and assume the other surface is at  $x \rightarrow \infty$ . Plot the temperature distributions at various times. Imagine a temperature sensor is placed at certain locations with instantaneous response and zero additional heat capacity. Estimate the time for the thermometer to sense the temperature rise as a function of the location  $x$ . Assume that the thermophysical properties of the glass are constant,  $\kappa = 1.43 \text{ W/(m} \cdot \text{K)}$ , and  $\alpha = 8.5 \times 10^{-7} \text{ m}^2/\text{s}$ .

**Solution.** The temperature distribution is shown in Fig. 7.1a at  $t = 0.01, 0.1, 1, 5, 10,$  and  $20 \text{ s}$ . During the heating, the temperature monotonically increases with time and the heat flux is always



**FIGURE 7.1** (a) The temperature distributions at various times. (b) The time required for a given location to acquire a minimum temperature rise and the estimated thermal diffusion speed.

positive. After the heat input is stopped when  $t = 5 \text{ s}$ , the temperature near the surface decreases but is still the highest and the temperature decreases toward increasing  $x$ . While the predicted temperature rises everywhere instantaneously, the magnitude may be too small to be observed practically. We can calculate the time  $\Delta t$  required for a minimum temperature rise  $\Delta T_{\text{min}}$ , specified by the thermometer sensitivity. Let us choose  $\Delta T_{\text{min}} = 10 \text{ mK}$  and  $0.1 \text{ K}$  for illustration. The average thermal diffusion speed can be estimated by  $v_{\text{dif}}(x) = x/\Delta t$ , for any given location  $x$ . The results are shown in Fig. 7.1b. In reality, diffusion is often a slow process near room temperature. For the example given here,  $v_{\text{dif}}$  for  $\Delta T_{\text{min}} = 10 \text{ mK}$  is between 1 and 5 m/s, for  $5 \text{ nm} < x < 5 \mu\text{m}$ , and goes down rapidly at  $x > 5 \mu\text{m}$ . At  $x = 10 \text{ mm}$ ,  $v_{\text{dif}}$  is only 2 to 3 mm/s. On the other hand, the speed of sound in glass is on the order of 5 km/s, which is several orders of magnitude greater than the average thermal diffusion speed.

Recall that the uncertainty principle in quantum mechanics states that  $\Delta E \Delta t > \hbar$ , suggesting that we cannot measure time and energy simultaneously with unlimited precision. From statistical mechanics, the distribution function allows a small fraction of particles to have a very high speed or to travel a very large distance without collision, although the probability may be extremely low. Based on the uncertainty principle and statistical mechanics, it seems convincing that Fourier’s law, in its applicable regime, does not violate

the principle of causality. What is physically problematic and practically impossible is to provide a temperature impulse to the surface or at any given location instantaneously. We further conclude that the heat diffusion equation does not produce an infinite speed of thermal energy propagation; rather, it is often a very slow process. Microscopically, Fourier's law fails when a local equilibrium is not established, as explained earlier. At the same time, the concept of an equilibrium temperature cannot be applied. It is critically important for the technological advancement to establish and apply thermal transport theories, both microscopically and macroscopically, under nonequilibrium conditions.

Several phenomenological theories have been developed to describe transient heat transfer processes in solids and micro/nanostructures. Applications of transient and ultrafast heating include laser processing, nanothermal fabrication, and the measurement of thermophysical properties. In the literature, there appears to be controversial experimental evidence on the existence of certain phenomena predicted by the hyperbolic heat conduction. Furthermore, there exists a large division as regards the formulation and the interpretation of the theories of non-Fourier conduction. While the intention is to provide a clear and objective presentation, the discussion will inevitably reflect the author's personal views and limitations at the time the manuscript was prepared. This section should help readers gain a general understanding of the basic concepts and phenomena related to non-Fourier heat conduction. Although relatively few papers out of a large number of publications are cited in the text and the reference section, interested readers can easily trace the relevant literature from the cited sources.

### 7.1.1 Hyperbolic Heat Equation

Several earlier studies have pointed out that the instantaneous response may be an indication of a nonphysical feature of the Fourier heat theory. Carlo Cattaneo in 1948 used kinetic theory of gas to derive a rate equation given by

$$\mathbf{q}''(\mathbf{r}, t) + \tau_q \frac{\partial \mathbf{q}''(\mathbf{r}, t)}{\partial t} = -\kappa \nabla T(\mathbf{r}, t) \quad (7.2)$$

which is called the *modified Fourier equation* or *Cattaneo equation*. The historical contributions by James Clerk Maxwell in 1867 and Pierre Vernotte in 1958 have been extensively reviewed by Joseph and Preziosi and will not be repeated here.<sup>3</sup> In Eq. (7.2),  $\tau_q$  is a kind of relaxation time, originally thought to be the same as  $\tau$ , i.e., the average time between collisions. The energy equation for heat conduction involving an internal source or volumetric heat generation rate  $\dot{q}(\mathbf{r}, t)$  is

$$\dot{q}(\mathbf{r}, t) - \nabla \cdot \mathbf{q}''(\mathbf{r}, t) = \rho c_p \frac{\partial T(\mathbf{r}, t)}{\partial t} \quad (7.3)$$

The divergence of Eq. (7.2) and the time derivative of Eq. (7.3) give two equations, which can be combined with Eq. (7.3) to eliminate the heat flux terms. The resulting differential equation for constant properties can be written as

$$\frac{\dot{q}}{\kappa} + \frac{\tau_q}{\kappa} \frac{\partial \dot{q}}{\partial t} + \nabla^2 T = \frac{1}{\alpha} \frac{\partial T}{\partial t} + \frac{\tau_q}{\alpha} \frac{\partial^2 T}{\partial t^2} \quad (7.4)$$

This is the *hyperbolic heat equation*, in contrast to the heat diffusion equation or parabolic heat equation. Without heat generation, we can rewrite Eq. (7.4) as

$$\nabla^2 T = \frac{1}{\alpha} \frac{\partial T}{\partial t} + \frac{1}{v_{tw}^2} \frac{\partial^2 T}{\partial t^2} \quad (7.5)$$

which is a telegraph equation or a damped wave equation. The solution of the hyperbolic heat equation results in a propagating wave, the amplitude of which decays exponentially as it travels. The speed of this *temperature wave* in the high-frequency limit, or the short-time limit, is given by

$$v_{tw} = \sqrt{\alpha/\tau_q} \quad (7.6)$$

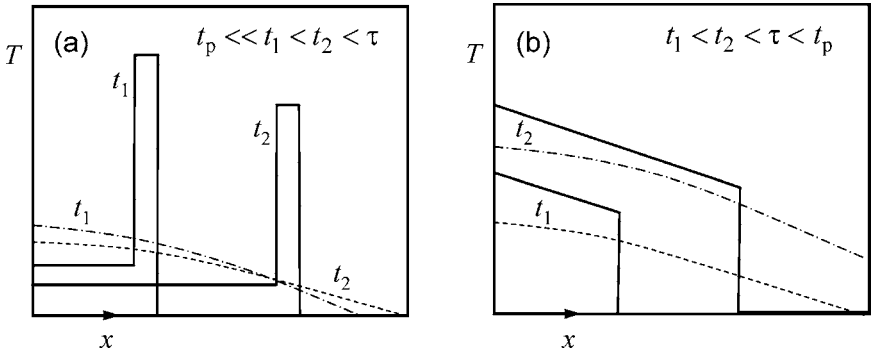
The amplitude of the temperature wave decays according to  $\exp(-t/\tau_q)$  due to the damping caused by the first-order time-derivative term  $(1/\alpha)(\partial T/\partial t)$ , which is also called the diffusion term. For an insulator, from the simple kinetic theory, we have  $\kappa = \frac{1}{3}(\rho c_v)v_g^2\tau$ . Noting that  $c_v = c_p$  for an incompressible solid and assuming  $\tau_q = \tau$ , we get

$$v_{tw} = v_g/\sqrt{3} \quad (7.7)$$

Equation (7.7) relates the speed of the temperature wave to the speed of sound in an insulator. The square root of three can be understood as due to the randomness of thermal fluctuations in a 3-D medium, just like the relation between the velocity and its components,  $\overline{v^2} = \overline{v_x^2} + \overline{v_y^2} + \overline{v_z^2}$ , in kinetic theory. Equation (7.5) indeed sets a limit on the heat propagation speed, which is manifested by a sharp wavefront that travels at  $v_{tw}$  inside the medium for a sudden temperature change at the boundary. As a wave equation, the solution is a temperature field with both an amplitude and a phase. Theoretically, the temperature wave can be reflected by another boundary and can interfere, constructively or destructively, with a forward propagating wave. The interaction between the temperature waves may also result in a resonance effect, a typical wave phenomenon. Numerous analytical and numerical predictions have been made, as referenced in the work of Özişik and Tzou,<sup>4</sup> along with Yeung and Lam,<sup>5</sup> Haji-Sheikh et al.,<sup>6</sup> and Gembarovic and Gembarovic, Jr.<sup>7</sup> It should be noted that the terms *heat wave*<sup>3</sup> and *thermal wave*<sup>4</sup> have also been frequently used in the literature to describe the temperature wave behavior. The term “temperature wave” is used in this chapter for the wavelike behavior associated with the hyperbolic-type heat equations, because “heat wave” might be confused with the calamitous weather phenomenon and “thermal wave” might be confused with the diffusion wave used in photoacoustic techniques. Bennett and Patty (*Appl. Opt.*, **21**, 49, 1982) clarified: *The term thermal wave interference is used to mean the superposition of simple harmonic solutions of the thermal diffusion equation. Although wavelike in nature there are important differences between thermal waves arising from a differential equation that is of the first order in time and waves that are solution to a wave equation that is of the second order in time.* In the heat transfer literature, thermal wave often refers to periodic-heating techniques used widely for thermo-physical property measurements.

Let us consider an example of a semi-infinite solid under a constant heat flux at the surface. Figure 7.2 illustrates the solutions for a small  $t_p$  and a large  $t_p$ , compared with  $\tau$ . Here again, we have assumed  $\tau_q = \tau$ . The propagation speed is equal to  $v_{tw}$ , and the pulse wavefront is given by  $x_1 = v_{tw}t_1$  and  $x_2 = v_{tw}t_2$ . Hence,  $x_1 < x_2 < \Lambda$ , where  $\Lambda = v_g\tau$  is the mean free path. In the case of a short pulse, the temperature pulse propagates and its height decays by dissipating its energy to the medium as it travels. The parabolic heat equation, on the other hand, predicts a continuous temperature distribution without any wavefront (see Fig. 7.2).

As time passes on, the first-order time derivative, or the diffusion term, in Eq. (7.5) dominates. If the relative change of  $\partial T/\partial t$  or  $\mathbf{q}''$  during one  $\tau_q$  is large, then the wave feature is important. This should happen immediately after a sudden thermal disturbance that results in a temporal nonequilibrium, as well as a spatial nonequilibrium near the heat pulse or the wavefront. After a sufficiently long time, usually 5 to 10 times  $\tau_q$ , a local equilibrium will be reestablished, and the thermal field can be described by the parabolic heat equation. At steady state, the hyperbolic and parabolic equations predict the same results. While Eq. (7.4) is mathematically more general than the heat diffusion equation, it should not be taken as



**FIGURE 7.2** (Not to scale) Illustration of the solution of the hyperbolic heat equation at short timescales. (a) A short pulse,  $t_p \ll \tau$ . (b) A long pulse,  $t_p > \tau$ . The solid curves are the solutions of the hyperbolic heat equation, Eq. (7.5), and the dash-dotted and dashed curves are the solutions, calculated from Eq. (7.1), obtained from the heat diffusion equation.

a correction, or a more realistic theory than the Fourier conduction model, because the Cattaneo equation has not been justified on a fundamental basis, nor has it been validated by any plausible experiments.

Many researchers have investigated the hyperbolic heat equation based on the second law of thermodynamics.<sup>8–10</sup> It has been found that the hyperbolic heat equation sometimes predicts a negative entropy generation and even allows energy to be transferred from a lower-temperature region to a higher-temperature region. The entropy generation rate for heat conduction without an internal source can be calculated by<sup>10</sup>

$$\dot{s}_{\text{gen}} = -\frac{1}{T^2} \mathbf{q}'' \cdot \nabla T = \frac{1}{T^2} \mathbf{q}'' \cdot \left( \mathbf{q}'' + \tau_q \frac{\partial \mathbf{q}''}{\partial t} \right) \quad (7.8)$$

The above equation was obtained by setting the energy and entropy balances as follows:

$$\rho \frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{q}'' \quad \text{and} \quad \rho \frac{\partial s}{\partial t} = -\nabla \cdot \left( \frac{\mathbf{q}''}{T} \right) + \dot{s}_{\text{gen}} \quad (7.9)$$

Note that  $du = Tds$ . A negative entropy generation can easily be numerically demonstrated from Eq. (7.5) during the temperature wave propagation. Here, a negative entropy generation does not constitute a violation of the second law of thermodynamics because the concept of “temperature” in the hyperbolic heat equation cannot be interpreted in the conventional sense due to the lack of local thermal equilibrium. Extended irreversible thermodynamics has been proposed by Jou et al. by modifying the definition of entropy such that it is not a property of the system anymore but depends on the heat flux vector.<sup>11</sup> The theory of extended irreversible thermodynamics is self-consistent but has not been experimentally validated; hence, it cannot be taken as a generalized thermodynamic theory. Similarly, the hyperbolic heat equation should not be treated as a more general theory over Fourier’s heat conduction theory.

**Example 7-2.** Derive the modified Fourier equation, or the Cattaneo equation, based on the BTE under the relaxation time approximation.

**Solution.** Tavernier (*C. R. Acad. Sci.*, **254**, 69, 1962) first showed that the Cattaneo equation could be derived for phonons and electrons using the relaxation time approximation of the BTE. Let us first review Sec. 4.3.2, where we have derived Fourier’s law based on the BTE. Again, let us start

by assuming that the temperature gradient is in the  $x$  direction only. The transient 1-D BTE under the relaxation time approximation can be written as follows:

$$\frac{\partial f}{\partial t} + v_x \frac{\partial f}{\partial x} = \frac{f_0 - f}{\tau}$$

A further assumption is made such that  $\partial f/\partial x \approx \partial f_0/\partial x = (\partial f_0/\partial T)(\partial T/\partial x)$ , which is exactly the condition of local equilibrium. Multiplying the earlier equation by  $\tau \epsilon v_x$  and then integrating each term over the momentum space, we obtain by noting  $\int_{\omega} \epsilon v_x f_0 d\omega = 0$  that

$$\frac{\partial}{\partial t} \int_{\omega} \tau v_x f \epsilon d\omega + \int_{\omega} \tau v_x^2 \frac{\partial f_0}{\partial T} \frac{\partial T}{\partial x} \epsilon d\omega = - \int_{\omega} v_x f \epsilon d\omega \quad (7.10a)$$

$$\text{or} \quad \tau \frac{\partial q_x''}{\partial t} - \kappa \frac{\partial T}{\partial x} = -q_x'' \quad (7.10b)$$

which can be generalized to the 3-D case as given in Eq. (7.2), after replacing  $\tau$  with  $\tau_q$ .

The derivation given in this example, however, does *not* provide a microscopic justification of the hyperbolic heat equation, because it is strictly valid only under the local-equilibrium assumption with an averaged relaxation time. The local-equilibrium assumption prohibits application of the derived equation to length scales comparable or smaller than the mean free path.<sup>12</sup> Suppose a thermal disturbance occurs at a certain time and location; after a duration of time that is much longer than the relaxation time, the Fourier law and the parabolic heat equation are well justified because both the spatial and temporal local-equilibrium conditions are met. On the other hand, if we wish to use the modified Fourier equation to study the transient behavior at a timescale less than  $\tau$ , then the disturbance will propagate by a distance shorter than the mean free path, as shown in Fig. 7.2. Therefore, the derivation based on the BTE, under local-equilibrium and relaxation time approximations, is not a microscopic proof of the hyperbolic heat equation, which is meaningful only in a nonequilibrium situation. To this end, it appears that Maxwell made the right choice in dropping terms involving the relaxation time in the paper (*Phil. Trans. R. Soc. London*, **157**, 49, 1867), by assessing that *the rate of conduction will rapidly establish itself*.

While the previous derivation does not support Eq. (7.2), it does not disprove Eq. (7.2) either because the relaxation time approximation is not a very good model in the nonequilibrium regime. The local-equilibrium assumption breaks down completely at extremely short timescales. The basic assumption in the relaxation time approximation is that the distribution function is not too far from equilibrium. For a heat pulse with a duration less than  $\tau$ , the relaxation time approximation should generally be applied when  $t > \tau$ , regardless of whether we are dealing with a thin film or a semi-infinite medium. What may be concluded is that we have failed to prove either by any fundamental theory or by any credible experiments that the Cattaneo equation, originated from the kinetic theory according to the relaxation time approximation, is a physical law that extends Fourier's law to the nonequilibrium regime. Atomistic simulations, based on molecular dynamics and the lattice Boltzmann method, have provided further evidence that the hyperbolic heat equation is not applicable at very short timescales or in the nonequilibrium regime, where the applicability of the relaxation time approximation is also questionable.<sup>13,14</sup> For this reason, we have intentionally avoided phrases like "generalized Fourier's equation" and "modified Fourier's law" in describing Eq. (7.2).

One might argue that when  $\tau_q$  was identified as the average time  $\tau$  between collisions, under the relaxation time approximation, Eq. (7.7) could give the appropriate heat propagation speed, which is one-third of the speed of sound, as observed in liquid helium and some solids at low temperatures. This is a misinterpretation because the phenomenon, related to the second sound with a characteristic speed  $v_{2\text{nd}} = v_g/\sqrt{3}$ , cannot occur by a single relaxation mechanism, as will be shown later. Nevertheless, after some modifications, there

exist special cases when the modified heat equation becomes physically plausible and practically applicable. The modified equation does not produce sharp wavefronts like those illustrated in Fig. 7.2.

### 7.1.2 Dual-Phase-Lag Model

Chester (*Phys. Rev.*, **131**, 2013, 1963) first explained the lagging behavior associated with the Cattaneo equation. He pointed out that the physical significance of the modified Fourier equation lies in that there exists a finite buildup time after a temperature gradient is imposed on the specimen for the onset of a heat flow, which does not start instantaneously but rather grows gradually during the initial period on the order of the relaxation time  $\tau$ . Conversely, if the thermal gradient is suddenly removed, there will be a *lag* in the disappearance of the heat current. Gurtin and Pipkin (*Arch. Ration. Mech. Anal.*, **31**, 113, 1968) introduced the memory effect to account for the delay of the heat flux with respect to the temperature gradient. They expressed the heat flux as an integration of the temperature gradient over time, in analogy with the stress-strain relationship of viscoelastic materials with instantaneous elasticity. The linearized constitutive equation reads

$$\mathbf{q}''(\mathbf{r}, t) = - \int_{-\infty}^t K(t - t') \nabla T(\mathbf{r}, t') dt' \quad (7.11)$$

where  $K(\xi)$  is a kernel function. When  $K(\xi) = \kappa \delta(\xi)$ , Eq. (7.11) reduces to Fourier's law; when  $K(\xi) = (\kappa/\tau_q) e^{-\xi/\tau}$ , Eq. (7.11) reduces to the Cattaneo equation. By assuming

$$K(\xi) = \kappa_0 \delta(\xi) + \frac{\kappa_1}{\tau_q} e^{-\xi/\tau} \quad (7.12)$$

Joseph and Preziosi showed that the heat flux can be separated into two parts:<sup>3</sup>

$$\mathbf{q}''(\mathbf{r}, t) = -\kappa_0 \nabla T - \frac{\kappa_1}{\tau_q} \int_{-\infty}^t \exp\left(-\frac{t-t'}{\tau_q}\right) \nabla T(\mathbf{r}, t') dt' \quad (7.13a)$$

Hence, 
$$\mathbf{q}'' + \tau_q \frac{\partial \mathbf{q}''}{\partial t} = -\kappa \nabla T - \tau_q \kappa_0 \frac{\partial}{\partial t} \nabla T \quad (7.13b)$$

where  $\kappa = \kappa_0 + \kappa_1$  is the steady-state thermal conductivity, as can be seen from Eq. (7.13a). Combined with Eq. (7.3), the heat equation becomes a partial differential equation of the Jeffreys type,

$$\nabla^2 T + \tau_T \frac{\partial}{\partial t} \nabla^2 T = \frac{1}{\alpha} \frac{\partial T}{\partial t} + \frac{\tau_q}{\alpha} \frac{\partial^2 T}{\partial t^2} \quad (7.14)$$

where  $\tau_T = \tau_q \kappa_0 / \kappa$  is known as the *retardation time*.<sup>3</sup> The Jeffreys equation was originally developed in the early twentieth century to relate deformation with stress in the earth's mantle. Unless  $\tau_T = 0$  or  $\kappa_0 = 0$ , Eq. (7.14) maintains the diffusive feature and produces an instantaneous response, albeit small, throughout the medium for an arbitrary thermal disturbance.

In a series of papers published in the early 1990s, Tzou extended the lagging concept to a dual-phase-lag model, as described in his monograph published in 1997.<sup>15</sup> He started with the assumption that

$$\mathbf{q}''(\mathbf{r}, t + \tau_q) = -\kappa \nabla T(\mathbf{r}, t + \tau_T) \quad (7.15)$$

The introduction of a delay time  $\tau_T$  in Eq. (7.15) implies the existence of a lag in the temperature gradient, with respect to the heat flux driven by an internal or external heat source. The rationale of the phenomenological equation given in Eq. (7.15) was that, in some cases, the heat flux might be viewed as the result of a preceding temperature gradient; in other cases, the temperature gradient might be viewed as the result of a preceding heat flux. The heat flux and the temperature gradient can switch roles in the relationship between “cause” and “effect.” Moreover, both lags might occur simultaneously in certain materials under dramatic thermal disturbances, such as during short-pulse laser heating.<sup>4,15</sup> These primitive arguments should not be scrutinized rigorously; rather, they are merely thinking instruments to help us gain an intuitive understanding of the heat flux and temperature gradient relationship. After applying the Taylor expansion to both sides of Eq. (7.15) and using the first-order approximation, one immediately obtains

$$\mathbf{q}'' + \tau_q \frac{\partial \mathbf{q}''}{\partial t} = -\kappa \nabla T - \tau_T \kappa \frac{\partial}{\partial t} \nabla T \quad (7.16)$$

which is mathematically identical to Eq. (7.13b), with the substitution of  $\tau_q \kappa_0 = \tau_T \kappa$ . Applying the first-order approximation of Eq. (7.15), one may end up with  $\mathbf{q}'' + (\tau_q - \tau_T) (\partial \mathbf{q}'' / \partial t) = -\kappa \nabla T$ , or  $\mathbf{q}'' = -\kappa \nabla T - (\tau_T - \tau_q) \partial(\nabla T) / \partial t$ , or even  $\mathbf{q}'' + (\tau_q - \frac{1}{3} \tau_T) (\partial \mathbf{q}'' / \partial t) = -\kappa \nabla T - \frac{2}{3} \tau_T \kappa \partial(\nabla T) / \partial t$ . These equations are merely special cases of Eq. (7.16), after regrouping  $\tau_q$  and  $\tau_T$ . The only requirement for Eq. (7.16) to make logical sense is that both  $\tau_q$  and  $\tau_T$  are nonnegative. The reason that a lag in time has been called a phase lag is perhaps because the temperature field can be viewed as a Fourier transform:  $T(\mathbf{r}, t) = \int_{-\infty}^{\infty} \tilde{T}(\mathbf{r}, \omega) e^{-i\omega t} d\omega$ , where  $\tilde{T}(\mathbf{r}, \omega)$  is the Fourier component at frequency  $\omega$ . The actual phase lag  $\omega \tau_T$  (or  $\omega \tau_q$  for heat flux) depends on the frequency. Equation (7.16) is mathematically more general and has some advantages over the Cattaneo equation. From now on, Eq. (7.14) will be called the *lagging heat equation*. It is straightforward to include the source terms in the lagging heat equation, as well as to treat thermophysical properties as temperature dependent. The solution, however, becomes more and more difficult as the complexity increases. Numerous studies have appeared in the literature on analytical solutions and numerical methods.<sup>4,15–18</sup>

It should be noted that in Eq. (7.12),  $\kappa_0$  and  $\kappa_1$  denote the effective and elastic conductivities, respectively, and are supposed to be nonnegative.<sup>3</sup> Therefore,  $\tau_T$  must not be greater than  $\tau_q$ . In fact, the ratio  $\eta = \kappa_0 / (\kappa_0 + \kappa_1)$  is a direct indication of whether thermal behavior can be described by heat diffusion (when  $\eta = 1$ ) or the hyperbolic heat equation (when  $\eta = 0$ ). In general,  $0 \leq \eta \leq 1$ , and the thermal process lies somewhere between the two extremes prescribed by Fourier’s law and the Cattaneo equation. In other words, there will be wavelike features in the solution, which is superimposed by an instantaneous diffusive response throughout the medium. The diffusive response here, as well as in Fourier’s law, does not correspond to an infinite speed of propagation. Rather, it is well justified by quantum statistics as explained previously.

The dual-phase-lag model relaxes the requirement of  $\tau_T \leq \tau_q$ ; but in the mean time, it produces a negative thermal conductivity component, i.e.,  $\kappa_1 < 0$ , according to Eq. (7.12). This drawback has long been overcome by Tzou, who proposed a new memory function in accordance with Eq. (7.16) as follows:<sup>15</sup>

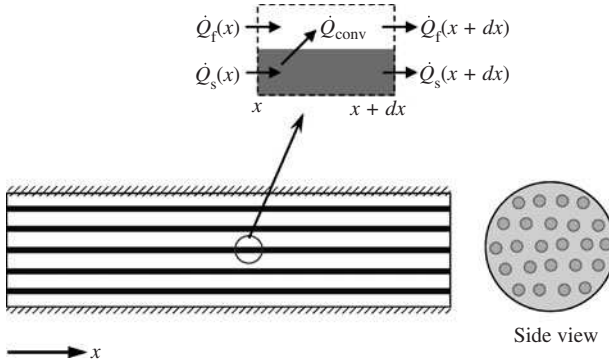
$$\mathbf{q}''(\mathbf{r}, t) = -\frac{\kappa}{\tau_q} \int_{-\infty}^t \exp\left(-\frac{t-t'}{\tau_q}\right) \left[ \nabla T(\mathbf{r}, t') + \tau_T \frac{\partial}{\partial t'} \nabla T(\mathbf{r}, t') \right] dt' \quad (7.17)$$

Equation (7.17) suggests that the heat flux depends not only on the history of the temperature gradient but also on the history of the time derivative of  $\nabla T$ . When  $\tau_T = 0$ , Eq. (7.17) becomes the Cattaneo equation. When  $\tau_T = \tau_q$ , Eq. (7.17) reduces to Fourier’s law. However,  $\tau_T > \tau_q$  is theoretically permitted because Eq. (7.17) does not presume that the



thermal conductivity is composed of an effective conductivity and an elastic conductivity. The inclusion of  $\tau_T > \tau_q$  makes Eq. (7.16) more general than the original Jeffreys-type equation, which is based on Eq. (7.13a). The extension to the region  $\tau_T > \tau_q$  enables the lagging heat equation to describe the behavior of parallel heat conduction, which can occur in a number of engineering situations.

Sometimes, a microscale phenomenon can be understood easily if a macroscale analog can be drawn. For this reason, let us consider the solid-fluid heat exchanger shown in Fig. 7.3.



**FIGURE 7.3** Illustration of heat transfer in a solid-fluid heat exchanger, where long solid rods are immersed in a fluid inside a sealed pipe, which is insulated from the outside.

Assume that a fluid is stationary inside a sealed pipe, filled with long solid rods. The pipe is insulated from the outside. If the rods are sufficiently thin, we may use the average temperature in a cross section and assume that heat transfer takes place along the  $x$  direction only. Let us denote the temperatures of the solid rods and the fluid by  $T_s(x, t)$  and  $T_f(x, t)$ , respectively, and take their properties  $\kappa_s$ ,  $C_s = (\rho c_p)_s$ ,  $\kappa_f$ , and  $C_f = (\rho c_p)_f$  to be constant. Note that  $C_s$  and  $C_f$  are the *volumetric heat capacities*. Given the rod diameter  $d$ , the number of rods  $N$ , and the inner diameter  $D$  of the pipe, the total surface area per unit length is  $P = N\pi D$ , and the total cross-sectional areas of the rods and the fluid are  $A_c = N\pi d^2/4$  and  $A_f = (\pi/4)(D^2 - Nd^2)$ , respectively. Assume the average convection coefficient is  $h$ . The energy balance equations can be obtained using the control volume analysis as follows:

$$C_s \frac{\partial T_s}{\partial t} = \kappa_s \frac{\partial^2 T_s}{\partial x^2} - G(T_s - T_f) \quad (7.18a)$$

and

$$C_f \frac{\partial T_f}{\partial t} = G(T_s - T_f) \quad (7.18b)$$

where  $G = hP/A_c$  and  $C_f' = C_f A_f/A_c$ . In writing Eq. (7.18b), we have assumed that  $\kappa_f \ll \kappa_s$  and dropped the term  $\kappa_f(\partial^2 T_f/\partial x^2)$ . Equations (7.18a) and (7.18b) are coupled equations that can be solved for the prescribed initial and boundary conditions. These are completely macroscopic equations governed by Fourier's law of heat conduction. Nevertheless, we can combine Eq. (7.18a) and Eq. (7.18b) to eliminate  $T_f$  and, consequently, obtain the following differential equation for  $T_s$ :

$$\frac{\partial^2 T_s}{\partial x^2} + \tau_T \frac{\partial}{\partial t} \left( \frac{\partial^2 T_s}{\partial x^2} \right) = \frac{1}{\alpha} \frac{\partial T_s}{\partial t} + \frac{\tau_q}{\alpha} \frac{\partial^2 T_s}{\partial t^2} \quad (7.18c)$$

where  $\alpha = \kappa_s/(C_s + C'_s)$ ,  $\tau_T = C'_s/G$ , and  $\tau_q = C_s\tau_T/(C_s + C'_s) < \tau_T$ . The same equation can also be obtained for the fluid temperature  $T_f$ . Here,  $\tau_q$  does not have the meaning of relaxation time, and the solutions of Eq. (7.18) exhibit diffusion characteristics. Equation (7.18c) is completely physical but should not be viewed as a wave equation; rather, it describes a parallel or coupled heat diffusion process. The concept of dual phase lag can still be applied. It should be noted that, due to the initial temperature difference between the rod and the fluid, a local equilibrium is not established at any  $x$  inside the pipe until after a sufficiently long time.

Although no fundamental physics can be gained from this example, it can help us appreciate that the lagging heat equation may be useful for describing the behavior in inhomogeneous media. Minkowycz et al. studied the heat transfer in porous media by considering the departure from local thermal equilibrium and obtained higher-order differential equations similar to Eq. (7.18c).<sup>19</sup> On the other hand, Kaminski made an experimental attempt to determine  $\tau_q$  in the hyperbolic heat equation, by measuring the time interval between when the heat source was turned on and when a temperature signal was detected.<sup>22</sup> The heat source and the thermometer used were long needles, placed in parallel and separated by a gap of 5 to 20 mm. What the experiment actually measured was the average thermal diffusion speed  $v_{\text{diff}}$  if the cylindrical geometry and the initial conditions were properly taken into consideration in the analysis. The main problem with this frequently cited paper and similar studies in the 1990s was that most researchers did not realize that the hyperbolic heat equation is physically unjustified to be superior to the parabolic heat equation; instead, they thought that the parabolic equation was only a special case of the more general hyperbolic equation. It appears that the Cattaneo equation and the associated hyperbolic heat equation are unlikely to be able to characterize any heat transfer problems successfully without additional terms. Many researchers have already expressed doubt about the applicability of the hyperbolic heat equation, though not so many have realized that an instantaneous response is a legitimate property, rather than a drawback of the diffusion equation. Electron gas and phonon gas in solids are quantum mechanical particles, which do not have memory of any kind. Ideal molecular gases obey classical statistics and do not have memory either, unless the deposited energy is too intense to cause ionization or reaction.

Does the temperature wave exist? What is a temperature wave anyway? In the early 1940s, Russian theoretical physicist Lev Landau (1908–1968) used a two-fluid model to study the behavior of quasiparticles in superfluid helium II and predicted the existence of a second sound, propagating at a speed between  $v_g/\sqrt{3}$  and  $v_{sg}$ , depending on the temperature. Note that the group velocity is the same as the phase velocity for a linear dispersion. Above the  $\lambda$ -point, where superfluidity is lost, the second sound should also disappear. Landau was awarded the Nobel Prize in Physics in 1962 for his pioneering theories of condensed matter at low temperatures. He authored with his students a famous book series in mechanics and physics. Landau's prediction was validated experimentally (*J. Phys. USSR*, **8**, 381, 1944) by Peshkov, who further postulated the existence of a second sound in crystals, when scattering by defects becomes minimized. It was not until the mid 1960s that the second sound associated with heat pulse propagation was observed in solid helium (below 1 K) and other crystals at low temperatures (below 20 K). The second sound can occur only at very low temperatures when the mean free path of phonons in the  $U$  processes, in which the total momentum is not conserved, is longer than the specimen size; while at the same time, the scattering rate of the  $N$  processes, in which the total momentum is conserved, is high enough to dominate other scattering processes. It should be noted that while the  $N$  processes have a much shorter mean free path than the size of the specimen, scattering by  $N$  processes does not dissipate heat (see Sec. 6.5.3). Callaway simplified the BTE for phonon systems by a two-relaxation-time approximation, which should be applicable when  $t > \tau_N$ :

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} = \frac{f_0 - f}{\tau} + \frac{f_1 - f}{\tau_N} \quad (7.19)$$

where  $\tau$  stands for the relaxation time for the  $U$  processes,  $\tau_N$  is the relaxation time for the  $N$  processes, and  $f_0$  and  $f_1$  are the associated equilibrium distribution functions.<sup>21</sup> Guyer and Krumhansl solved the linearized BTE and derived the following equation for the phonon effective temperature:

$$\nabla^2 T + \frac{9\tau_N}{5} \frac{\partial}{\partial t} \nabla^2 T = \frac{3}{\tau v_a^2} \frac{\partial T}{\partial t} + \frac{3}{v_a^2} \frac{\partial^2 T}{\partial t^2} \quad (7.20)$$

where  $v_a$  is the average phonon speed.<sup>22</sup> Assuming a linear dispersion, it can be evaluated using Eq. (5.10). Substituting  $\alpha = \tau v_a^2/3$ ,  $\tau_q = \tau$ , and  $\tau_T = 9\tau_N/5$ , we see that Eq. (7.20) is identical to Eq. (7.14). The condition  $t > \tau_N$  can be satisfied even at  $t < \tau$  since  $\tau_N \ll \tau$ . The significance of Eq. (7.20) lies in that the temperature wave or the second sound is not universal, but rather, requires strict conditions to be met.<sup>22</sup> When the condition  $\tau_N \ll \tau$  is satisfied, we have  $\tau_T \ll \tau_q$  and the energy transfer is dominated by wave propagation. At higher temperatures, the scattering rate for the  $U$  processes is usually very high, and the  $N$  processes contribute little to the heat conduction or thermal resistance, as discussed in Chap. 6. Therefore, the reason why temperature waves have not been observed in insulators at room temperature is not because of the small  $\tau$ , in the range from  $10^{-10}$  to  $10^{-13}$  s, but because of the lack of mechanisms required for a second sound to occur. No experiments have ever shown a second sound in metals, as suggested by the hyperbolic heat equation.

Recently, Shiomi and Maruyama performed molecular dynamics simulations of the heat conduction through (5,5) single-walled carbon nanotubes, 25 nm in length, for several femtoseconds.<sup>23</sup> They found that the wavelike behavior could be fitted by the lagging heat equation, but could not be described by the hyperbolic heat equation due to local diffusion. The ballistic nature of heat propagation in nanotubes has already been explained in Chap. 5. They suspected that optical phonons might play a major role in the non-Fourier conduction process.<sup>23</sup> Tsai and MacDonald studied the strong anharmonic effects at high temperature and pressure using molecular dynamics.<sup>24</sup> Their work predicted a second sound response. The coupling of elastic and thermal effects was thought to be important. Studies on thermomechanical effects such as thermal expansion, thermoelasticity, and shock waves can be found from Tzou<sup>15</sup> and Wang and Xu,<sup>25</sup> and will not be discussed further.

Tang and Araki clearly delineated four regimes in the lagging heat equation, according to the ratio  $\eta = \tau_T/\tau_q$ .<sup>17</sup> (1) When  $\eta = 0$ , it is a damped wave, i.e., hyperbolic heat conduction. (2) When  $0 < \eta < 1$ , it is wavelike diffusion, for which wave features can be clearly seen if  $\eta \ll 1$ . (3) When  $\eta = 1$ , it is pure diffusion or diffusion, i.e., Fourier's conduction. (4) When  $\eta > 1$ , it is called over-diffusion, which makes the dimensionless temperature decay faster than pure diffusion would. In the next section, we will discuss a microscopic theory on short-pulse laser heating of metals, which falls in the regime of over-diffusion, or parallel conduction.

### 7.1.3 Two-Temperature Model

With a short laser pulse, 5 fs to 500 ps, free electrons absorb radiation energy and the absorbed energy excites the electrons to higher energy levels. The "hot electrons" move around randomly and dissipate heat mainly through electron-phonon interactions. Following the work of Kaganov et al. (*Sov. Phys. JETP*, **4**, 173, 1957), Anisimov proposed a *two-temperature model*, which is a pair of coupled nonlinear equations governing the effective temperatures of electrons and phonons.<sup>26</sup> This model was experimentally confirmed later by Fujimoto et al. (*Phys. Rev. Lett.*, **53**, 1837, 1984) and Brorson et al. (*Phys. Rev. Lett.*, **59**, 1962, 1987). The two-temperature model was introduced to the heat transfer community by Qiu and Tien, who also analyzed the size effect due to boundary scattering and performed experiments with thin metallic films.<sup>27</sup> In the two-temperature model, it was

assumed that the electron and phonon systems are each at their own local equilibrium, but not in mutual equilibrium. The electron temperature could be much higher than the lattice (or phonon) temperature due to absorption of pulse heating. Therefore,

$$C_e \frac{\partial T_e}{\partial t} = \nabla \cdot (\kappa \nabla T_e) - G(T_e - T_s) + \dot{q}_a \quad (7.21a)$$

$$C_s \frac{\partial T_s}{\partial t} = G(T_e - T_s) \quad (7.21b)$$

Here, the subscripts e and s are for the electron and phonon systems, respectively,  $C$  is the volumetric heat capacity,  $G$  is the electron-phonon coupling constant, and  $\dot{q}_a$  is the source term that represents the absorbed energy rate per unit volume during the laser pulse and drops to zero after the pulse. Heat conduction by phonons is neglected, and thus, the subscript e is dropped in the thermal conductivity  $\kappa$ . Note that  $\mathbf{q}'' = -\kappa \nabla T_e$ , according to Fourier's law. We have already given a macroscopic example of parallel heat transfer, as shown in Fig. 7.3, which should ease the understanding of the phenomenological relations given in Eq. (7.21). Equation (7.21) originates from microscopic interactions between photons, electrons, and phonons. In order to examine the parameters in Eq. (7.21) and their dependence on  $T_e$  and  $T_s$ , let us assume that the lattice temperature is near or above the Debye temperature, for simplicity. In such a case, electron-electron scattering and electron-defects scattering are insignificant compared with electron-phonon scattering. It is expected that the electron relaxation time is inversely proportional to the lattice temperature, i.e.,  $\tau \approx \tau_{e-ph} \propto T_s^{-1}$ . The meaning of the relaxation time is that the electron system can be assumed to be at internal local equilibrium when  $t > \tau$ , which is the condition for Eq. (7.21) to be applicable. Boundary scattering may play a role for very thin films or in polycrystalline materials. An effective mean free path can be introduced to modify the scattering rate.<sup>27-29</sup> The volumetric heat capacity for the lattice or phonons,  $C_s = \rho c_p$ , is a weak function of the lattice temperature; the volumetric heat capacity of electrons, from Eq. (5.25), becomes

$$C_e = \frac{\pi^2 n_e k_B^2}{2\mu_F} T_e = \gamma_s T_e \quad (7.22)$$

Recall that  $C_e$  is relatively small compared with  $C_s$ , even at several thousand kelvins. From the simple kinetic theory, the thermal conductivity is

$$\kappa = \frac{\pi^2 n_e k_B^2}{3m_e} \tau T_e \approx \frac{\kappa_{eq}}{T_s} T_e \quad (7.23)$$

where  $\kappa_{eq}$  is the thermal conductivity when  $T_e = T_s$ , which can be set as the room temperature value. The term  $T_e$  in Eq. (7.23) comes from the heat capacity. The size effect can be included using an effective relaxation time. Theoretically, the coupling constant can be estimated by

$$G = \frac{\pi^2 m_e n_e v_a^2}{6\tau T_s} \quad \text{or} \quad G = \frac{\pi^4 (n_e v_a k_B)^2}{18\kappa_{eq}} \quad (7.24)$$

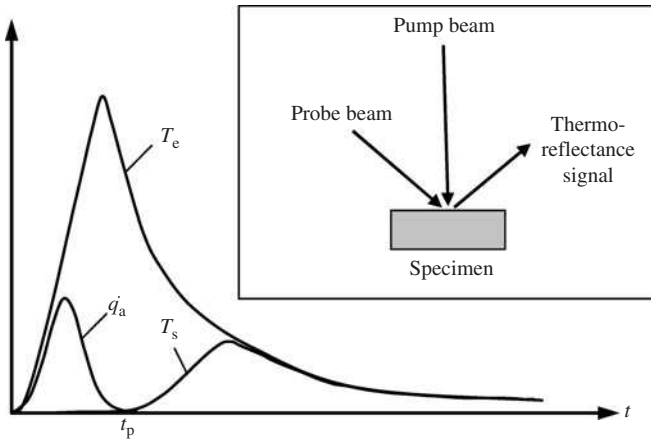
which is independent of temperature, when boundary scattering is not important but proportional to the square of the speed of sound in the metal. With the speed of sound in the low-frequency limit, the dispersion is linear; thus, we do not have to worry about the difference between the phase velocity and the group velocity. From Eq. (5.10), we have

$$v_a = \frac{k_B \Theta_D}{h} \left( \frac{4\pi}{3n_a} \right)^{1/3} \quad (7.25)$$

When boundary scattering is included,  $G$  is expected to increase from the bulk value and depend on the lattice temperature. Using the Debye temperature and for  $n_a = n_e$ , we have

$$G = \frac{\pi^2}{12 \times \sqrt[3]{4}} \frac{n_e k_B^2 \Theta_D^2}{\tau T_s \mu_0} \approx 0.518 \frac{n_e k_B^2 \Theta_D^2}{\tau T_s \mu_0} \quad (7.26)$$

Typical values of  $G$  are on the order of  $10^{16}$  W/(m<sup>3</sup> · K), e.g.,  $G \approx 2.9 \times 10^{16}$  W/(m<sup>3</sup> · K) for gold. The behavior of the electron and phonon temperatures near the surface is shown in Fig. 7.4, for a short pulse. The electron temperature rises quickly during the pulse and



**FIGURE 7.4** Illustration (not to scale) of ultrafast thermoreflectance experiments and the associated electron and phonon temperatures near the surface, during a short pulse.

begins to decrease afterward; in the mean time, the lattice temperature gradually increases until the electron and lattice systems reach a thermal equilibrium. Both the temperatures will go down as heat is carried away from the surface. Note that the electron temperature can rise very high due to its small heat capacity, but the lattice or solid may be just slightly above room temperature. If the temperatures of electron and lattice were assumed the same, Eq. (7.21) reduces to the simple Fourier heat conduction equation, which in turn predicts a much lower temperature rise, because the heat capacity of the lattice is much higher than that of the electrons.

Given such a short timescale and the nonequilibrium nature between electrons and phonons locally, no contact thermometer could possibly measure the effective electron temperature. Experiments are usually performed by the femtosecond or picosecond thermoreflectance technique, also known as the pump-and-probe method, shown in the inset of Fig. 7.4. The reflectance of the surface depends on the electron temperature  $T_e$ . The experimental setup is rather involved and cannot measure the temperature distribution inside the material. The procedure is to send a pump pulse train that is synchronized with a probe pulse train at a fixed delay time. The electron temperature change near the surface is deduced by comparison of the reflectance measurements at different delay times. Electron-phonon coupling, boundary scattering, and thermal boundary resistance can all affect the thermoreflectance signal. Comparing with the model described in Eq. (7.21), along with the

dependence of the reflectance on the electron temperature, the microscopic characteristics can be analyzed. Ultrafast thermoreflectance techniques have become an important thermal metrology tool for the study of electron-phonon interactions, TBR, and thermophysical properties.<sup>30,31</sup> Thermionic emission can also occur from the surface, especially when the electrons are excited to higher energy states.<sup>32</sup>

Similarly to what has been done for Eq. (7.18), Eq. (7.21a) and Eq. (7.21b) can be combined to formulate partial differential equations for either the electron or phonon temperature. Neglecting the temperature dependence of the parameters, one obtains the following differential equations for the electron temperature and the phonon temperature, respectively,

$$\nabla^2 T_e + \tau_T \frac{\partial}{\partial t} \nabla^2 T_e + \frac{\dot{q}_a}{\kappa} + \frac{\tau_T \partial \dot{q}_a}{\kappa \partial t} = \frac{1}{\alpha} \frac{\partial T_e}{\partial t} + \frac{\tau_q \partial^2 T_e}{\alpha \partial t^2} \quad (7.27a)$$

$$\nabla^2 T_s + \tau_T \frac{\partial}{\partial t} \nabla^2 T_s + \frac{\dot{q}_a}{\kappa} = \frac{1}{\alpha} \frac{\partial T_s}{\partial t} + \frac{\tau_q \partial^2 T_s}{\alpha \partial t^2} \quad (7.27b)$$

where  $\alpha = \kappa/(C_e + C_s)$ ,  $\tau_T = C_s/G$ , and  $\tau_q = \tau_T C_e/(C_e + C_s) \approx C_e/G \ll \tau_T$ . These equations are identical to the lagging heat equations and can be solved with appropriate boundary conditions. The results again belong to the regime of over-diffusion, or parallel conduction, without any wavelike features. Cooling caused by thermionic emission is usually neglected, and the surface under illumination can be assumed adiabatic. A 1-D approximation further simplifies the problem. The solution follows the general trends depicted in Fig. 7.4. The situation will be completely changed if a phase change occurs or if the system is driven to exceed the linear harmonic behavior.<sup>15,25</sup>

The term  $\tau_q$  is clearly not the same as the relaxation time  $\tau$  due to collision. The resulting solution is more diffusive than wavelike. In the literature,  $\tau_q$  is commonly referred to as the *thermalization time*. The physical meaning of  $\tau_q$  is a *thermal time constant* for the electron system to reach an equilibrium with the phonon system. For noble metals at room temperature, the relaxation time  $\tau$  is on the order of 30 to 40 fs, the thermalization time  $\tau_q$  is 0.5 to 0.8 ps, and the *retardation time*  $\tau_T$  is 60 to 90 ps. In practice, we need to consider the temperature dependence of the parameters in Eq. (7.21), as mentioned earlier. Some numerical solutions, considering temperature dependence, and comparisons with experiments can be found from Smith et al.<sup>33</sup> and Zhou and Chiu.<sup>34</sup> Given that the two-temperature model cannot be applied to  $t < \tau$ , due to the limitation of Fourier's law, one may prefer to use a pulse width  $t_p$  between 100 and 200 fs and measure the response during several picoseconds until the thermalization process is complete, i.e., the electron and phonon temperatures become the same. This first-stage measurement allows the determination of the coupling constant  $G$ . In the case of a thin film, the TBR sets a barrier for heat conduction between the film and the substrate. The time constant of the film can range from several tens to hundreds of picoseconds. Therefore, the TBR between the film and the substrate can be determined by continuing the observation of thermoreflectance signals for 1 to 2 ns after each pulse. Fitting the curves in the second-stage measurement allows an estimate of the TBR. Of course, one could use a longer pulse width  $t_p$  to determine the TBR. Most advanced femtosecond research laboratories are equipped with Ti:sapphire lasers whose pulse widths range from 50 to 500 fs. Femtosecond lasers with a pulse width of 25 fs have also been used in some studies; see for example Li et al. (*J. Opt. Soc. Am. B*, **15**, 2404, 1998; *Phys. Rev. Lett.*, **82**, 2394, 1999). For  $t_p$  below 50 fs, Eq. (7.21a) is not applicable during the heating, at least for noble metals. The relaxation time for Cr is about 3 fs, and Eq. (7.21) can be safely applied even with  $t_p = 10$  fs. However, the processes below 20 fs may largely involve electron-electron inelastic scattering, thermionic emission, ionization, phase transformation, chemical reaction, and so forth. Other difficult issues associated with the reduced pulse width include widened frequency spectrum, increased pulse intensity, decreased pulse

energy, and so forth. A simple hyperbolic heat flux formulation cannot properly address these issues at  $\tau_p < \tau$ . One must investigate the physical and chemical processes occurring at this timescale in order to develop a physically plausible model, with or without the concept of effective temperatures. Femtosecond laser interactions with dielectric materials have also been extensively studied (see Jiang and Tsai<sup>35</sup> and references therein).

Let us reiterate the major points presented in this section: (a) Fourier's law, which is limited to local equilibrium conditions, does not predict an infinite speed of heat diffusion, nor does it violate the principle of causality. An instantaneous response at a finite distance is permitted by quantum statistics although the probability of such a response sharply approaches zero as the distance increases. An instantaneous temperature change or heat flux at a precise location is not physically possible. Only under the continuum assumption, we can use the concept of sudden change of temperature at the boundary. (b) Heat diffusion is usually a very slow process, compared with the speed of sound. The temperature wave, or the second sound, has been observed only in helium and some very pure dielectric crystals, at low temperatures, where the  $U$  processes are ballistic and the  $N$  processes have a very high scattering rate. However, the simple hyperbolic heat equation has been proved neither theoretically nor experimentally. There is no need to collect previous or future experimental evidence to test the hyperbolic heat equation, which was ill-formulated in the first place. (c) All kinds of non-Fourier equations are based on some sort of effective temperature, which are not measurable using a contact thermometer. The principle of contact thermometry is the zeroth law of thermodynamics, which originates from the theory of thermal equilibrium. The concept of coldness or hotness should be abandoned in reference to nonequilibrium energy transport processes. Noncontact thermometry, on the other hand, relies on certain physical responses to deduce the equilibrium temperature or the effective temperature of the system being measured. (d) The memory hypothesis and the lagging argument are phenomenological models that may be useful in the study of certain non-equilibrium or parallel conduction processes, but are not universally applicable. These and similar equations must be derived and applied on a case-by-case basis. It is important to understand the microscopic processes occurring at the appropriate length scales and timescales in order to develop physically reliable models.

## **7.2 HEAT CONDUCTION ACROSS LAYERED STRUCTURES**

---

In Sec. 5.5.2, we have given a detailed discussion on the heat conduction along a thin film using the BTE, under the local equilibrium assumption. An effective thermal conductivity can be used after taking proper account of boundary scattering. The heat conduction problem can thus be well described by Fourier's law using the effective thermal conductivity. As mentioned earlier, for heat transfer across a film or a superlattice, the condition of local equilibrium breaks down in the acoustically thin limit. The local distribution function cannot be approximated by an equilibrium distribution function at any temperature. Conventional Fourier's law breaks down because it relies on the definition of an equilibrium temperature and the existence of local equilibrium. It is natural to ask the following two questions: (1) Is it possible for us to define an effective temperature? (2) Can Fourier's law still be useful in the nonequilibrium regime, according to the effective temperature? This section presents the equation of phonon radiative transfer (EPRT) and the solution of EPRT for thin films under the relaxation time approximation. A resistance network representation is present to illustrate how Fourier's law of heat conduction may be applied inside the medium, at least approximately, with temperature-jump boundary conditions. Because of the importance of understanding the boundary conditions, this section also discusses models of thermal boundary resistance (TBR) in layered structures.

### 7.2.1 Equation of Phonon Radiative Transfer (EPRT)

The phonon BTE under the relaxation time approximation, in a region with heat generation, may be written as

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \frac{\partial f}{\partial \mathbf{r}} = \frac{f - f_0}{\tau(\omega, T)} + S_0 \quad (7.28)$$

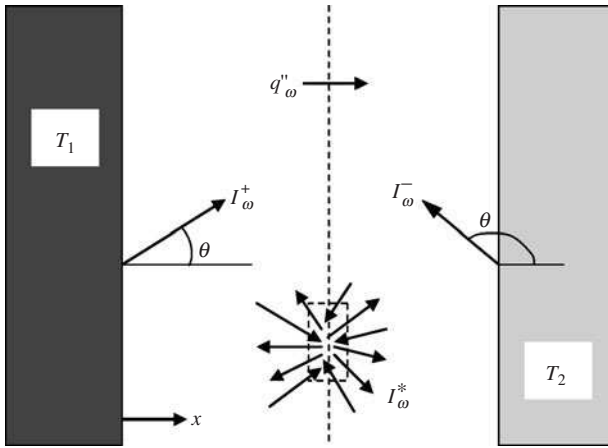
where the second term  $S_0$  on the right-hand side is a source term to model the generation of phonons due to heat dissipation, such as electron-phonon scattering. Phonon-phonon scattering is already included in the first term on the right-hand side. The scattering rate may also include phonon-defect scattering. Many studies have treated phonon transport in analogy to thermal radiative transfer.<sup>12, 36-43</sup> In the following, a simplified case is used to illustrate how to model heat transfer across a thin film as well as multilayer structures. Let us consider a film of thickness  $L$  between two boundaries without any internal source. The phonon BTE becomes

$$\frac{\partial f}{\partial t} + v_x \frac{\partial f}{\partial x} = \frac{f_0 - f}{\tau} \quad (7.29)$$

Realizing the nonequilibrium distribution function may be anisotropic, let us define

$$I_\omega(x, \Omega, t) = \frac{1}{4\pi} \sum_P v_g \hbar \omega f D(\omega) \quad (7.30)$$

where  $P$  is the number of phonon modes or polarizations. Equation (7.30) gives the *phonon intensity*, which is the energy transfer rate in the direction  $\Omega$  from a unit area, per unit frequency and per unit solid angle. The geometry of the problem and illustration of the intensity is given in Fig. 7.5. In this section, we use  $v_g$  for the group velocity and  $v_p$  for the



**FIGURE 7.5** Schematic of phonon radiative transfer inside a dielectric medium between two walls maintained at temperatures  $T_1$  and  $T_2$ . These walls are like heat reservoirs, but their surfaces are not necessarily blackbodies.

phase velocity. Note that  $v_x = v_g \cos \theta$ , where  $\theta$  is the polar angle. Substituting Eq. (7.30) into Eq. (7.29), we obtain

$$\frac{1}{v_g} \frac{\partial I_\omega}{\partial t} + \mu \frac{\partial I_\omega}{\partial x} = \frac{I_\omega^* - I_\omega}{v_g \tau} \quad (7.31)$$



where  $\mu = \cos \theta$  and  $I_{\omega}^*(\omega, T)$  is the intensity for equilibrium distribution that is independent of the direction. Equation (7.31) is called the equation of phonon radiative transfer (EPRT).<sup>12,40</sup> Comparing the EPRT with the ERT given in Eq. (2.52), we see that the scattering terms are neglected in the EPRT, and the emission and the absorption are replaced by the phonon collision terms. The phonon mean free path  $\Lambda = v_g \tau$  is also called the phonon penetration depth (see Example 4-2). The inverse of the penetration depth  $1/\Lambda$  corresponds to the absorption coefficient in the ERT. Conversion to the EPRT allows well-established theories and numerical techniques, developed in radiative transfer, to be applied to solve Eq. (7.31) and to interpret the physical significance of the solutions.<sup>44,45</sup> If  $\tau$  does not depend on frequency, we are dealing with a gray medium.

If the phonon Knudsen number  $Kn = \Lambda/L \ll 1$ , then most phonons will collide with phonons or defects inside the medium. This regime is called the *acoustically thick limit*, in analogy to the *optically thick limit* for photons. This is also known as the macroscale regime or the local equilibrium situation. Unless at a very short timescale, when a sudden local disturbance occurs, we expect that Fourier's law is applicable and the heat conduction is by diffusion. On the other hand, if  $Kn = \Lambda/L \gg 1$ , phonons originated from one boundary will most likely reach the other boundary without colliding with other phonons or defects inside the medium. This is the ballistic regime, corresponding to free molecule flow for molecular gases. This regime is called the *acoustically thin limit*, where the phonon distribution inside the medium cannot be characterized by an equilibrium distribution function if the walls are at different temperatures, even in the steady state. Because we are dealing with the radiative transfer for phonons as we do for photons, from now on, we will refer  $Kn = \Lambda/L \gg 1$  as the *radiative thin limit* and  $Kn = \Lambda/L \ll 1$  as the *radiative thick limit*. Because the BTE is more fundamental than Fourier's law, it works for either limit as well as between the two limits. It would be very useful if a macroscopic model can also be developed to bridge these two limits. Rather than referring readers to more specialized journal papers, in the following, we present some basic formulations that are logically connected with materials presented in earlier chapters.

Note that  $I_{\omega}^0$  is the equilibrium distribution function, which is independent of the direction. Using Bose-Einstein statistics, we have

$$I_{\omega}^*(\omega, T) = \sum_p \frac{v_g \hbar \omega}{e^{\hbar \omega / k_B T} - 1} \frac{k^2 dk}{(2\pi)^3 d\omega} = \sum_p \frac{\hbar \omega^3}{8\pi^3 v_p^2 (e^{\hbar \omega / k_B T} - 1)} \quad (7.32)$$

This equilibrium distribution is also the distribution function for blackbody radiation with  $v_p$  replaced by the speed of light. Integrating Eq. (7.32) over all frequencies gives the total intensity for all three phonon modes as follows:

$$I^*(T) = \int_0^{\infty} I_{\omega}^*(\omega, T) d\omega = \frac{3k_B^4 T^4}{8\pi^3 \hbar^3 v_a^2} \int_0^{\infty} \frac{x^3 dx}{e^x - 1} = \frac{\sigma'_{SB} T^4}{\pi} \quad (7.33)$$

where  $\sigma'_{SB} = \pi^2 k_B^4 / (40 \hbar^3 v_a^2)$  is the phonon Stefan-Boltzmann constant, and  $v_a$  is the average phase velocity of the two translational and one longitudinal phonon modes, defined according to Eq. (5.7). Let us consider a solid at temperatures higher than the Debye temperature. The integration can be carried out to an upper limit  $\omega_m$  with  $x_m = \hbar \omega_m / k_B T \ll 1$ . From the discussion following Eq. (5.13), one can easily show that

$$I^*(T) = \int_0^{\omega_m} I_{\omega}^*(\omega, T) d\omega = \frac{\omega_m^3 k_B T}{8\pi^3 v_p^2} \quad (7.34)$$

This integration is a good approximation, even at temperatures slightly lower than the Debye temperature. When phonons are at equilibrium, the energy flux is  $\pi I^*$ , which is

obtained by integrating  $I^* \cos \theta d\Omega$  over the hemisphere. According to Eq. (4.12), the energy density can be expressed as

$$u(T) = \frac{4\pi}{v_g} I^*(T) \quad (7.35)$$

Note that the volumetric heat capacity  $C = du/dT$ . We therefore obtain the low-temperature relation of the specific heat, i.e., the  $T^3$  law, and the high-temperature relation of the specific heat, i.e., the Dulong-Petit law, as already derived in Sec. 5.1.2. It is important to pay attention to the meaning of  $C$  in the kinetic expression of thermal conductivity:

$$\kappa = \frac{1}{3} C v_g^2 \tau \quad (7.36)$$

At very low temperatures, when  $T \ll \Theta_D$ ,  $C$  is the volumetric heat capacity of all phonon modes combined because only low-frequency modes or acoustic branches contribute to the specific heat. However, at temperatures close to the Debye temperature, phonons in the optical branches contribute little to the thermal conductivity, as already discussed in Chap. 6. The relative contributions of LA and TA branches are also temperature dependent. The Debye temperature for most materials, except diamond, is not much higher than room temperature (see Table 5.2). Therefore, we must treat  $C$  as a fraction of the volumetric specific heat in dealing with Si, GaAs, Ge, ZnS, or GaN, near room temperature. Also, we must use the appropriate upper limit in the integral in calculating the total energy transfer when applying the EPRT. The heat flux per unit frequency interval can thus be expressed as

$$q''_{\omega} = \int_{4\pi} I_{\omega} \cos \theta d\Omega = 2\pi \int_{-1}^1 I_{\omega} \mu d\mu \quad (7.37)$$

Energy balance at any given location requires that the incoming flux be the same as the outgoing flux, for both steady and transient states. This is the criterion for *radiative equilibrium*, which can be expressed as follows:

$$4\pi \int_0^{\omega_m} \frac{1}{\Lambda_{\omega}} I_{\omega}^* d\omega = 2\pi \int_0^{\omega_m} \int_{-1}^1 \frac{1}{\Lambda_{\omega}} I_{\omega} d\mu d\omega \quad (7.38)$$

where  $\Lambda_{\omega}$  is the mean free path at  $\omega$ ,  $4\pi$  on the left-hand side came from the integration over all solid angles in a sphere, and  $2\pi$  on the right-hand side came from integration over the azimuth angles. Equation (7.38) gives a definition of an *effective phonon temperature*  $T^*$  based on  $I_{\omega}^*(T^*, \omega)$ . An equivalent expression can be obtained based on the energy density, viz.,

$$u(T^*) = \sum_P \sum_K \hbar \omega f(\omega, \Omega) \quad (7.39)$$

It follows that the *local equilibrium condition* can be rewritten as

$$I_{\omega}^* = \frac{1}{2} \int_{-1}^1 I_{\omega} d\mu \quad (7.40)$$

Local equilibrium is a sufficient, but not necessary, condition for radiative equilibrium given in Eq. (7.38), regardless whether the medium is gray or not. The physical significance of Eq. (7.40) is that the angular average of the intensity, at a given location and time, can be described by an equilibrium intensity at the effective temperature. Obviously, Eq. (7.40) is not applicable in the radiative thin limit, unless the temperature difference between the two boundaries is negligibly small.

**Example 7-3.** For a dielectric medium of thickness  $L = 0.01 \Lambda$ , where  $\Lambda$  is independent of wavelength. The boundary or wall temperatures are  $T_1 = 100 \text{ K}$  and  $T_2 = 20 \text{ K}$ . Both the temperatures are much lower than the Debye temperature. Assume that reflection at the boundaries is negligible, i.e., the walls can be modeled as blackbodies. Find the steady-state temperature of the medium and the heat flux through the medium.

**Solution.** Because  $Kn = \Lambda/L \gg 1$ , the medium is said to be in the radiative thin limit, in which phonons travel from one wall to another ballistically with little chance of being scattered by other phonons or defects inside the medium. The forward intensity can be expressed as  $I_\omega^+ = I_\omega^+(T_1, \omega)$  for  $\mu > 0$ , and the backward intensity  $I_\omega^- = I_\omega^-(T_2, \omega)$  for  $\mu < 0$ . From Eq. (7.37), we have

$$q_x'' = \int_0^\infty q_\omega'' d\omega = 2\pi \int_0^1 \int_0^\infty (I_\omega^+ - I_\omega^-) \mu d\mu d\omega = \sigma'_{\text{SB}} (T_1^4 - T_2^4) \quad (7.41)$$

For heat conduction, the above equation is called the Casimir limit (*Physica*, **5**, 595, 1938). To numerically evaluate this equation, we need data for  $v_a$ . From Eq. (7.38), we have

$$\sigma'_{\text{SB}} T^4 = \frac{\pi}{2} \int_0^{\omega_m} (I_\omega^+ + I_\omega^-) d\omega = \frac{1}{2} (\sigma'_{\text{SB}} T_1^4 + \sigma'_{\text{SB}} T_2^4) \quad (7.42)$$

where  $T$  is the effective temperature inside the medium  $0 < x < L$ . Since  $T(0) = T_1$  and  $T(L) = T_2$  are the boundary conditions, there is a temperature jump at each boundary. We notice immediately that Eq. (7.40) cannot be satisfied with the temperature defined previously. If we force

$$I_\omega^* = \frac{1}{2} \int_{-1}^1 I_\omega d\mu = \frac{1}{2} (I_\omega^+ + I_\omega^-) \quad (7.43)$$

we would end up with different temperatures at each frequency. In the next chapter (Sec. 8.2.3), we will further discuss the concept of monochromatic temperature. If the walls are not black but diffuse-gray with emissivities  $\varepsilon_1$  and  $\varepsilon_2$ , similar to Eq. (2.51), the heat flux becomes

$$q_x'' = \frac{\sigma'_{\text{SB}} T_1^4 - \sigma'_{\text{SB}} T_2^4}{1/\varepsilon_1 + 1/\varepsilon_2 - 1} \quad (7.44)$$

## 7.2.2 Solution of the EPRT

The two-flux method is very helpful in developing a solution of the EPRT in planar structures. The equations for the forward and backward intensities, denoted respectively by superscripts (+) and (-) can be separated. In the steady state, we have:

$$\mu \frac{\partial I_\omega^+}{\partial x} = \frac{I_\omega^* - I_\omega^+}{\Lambda}, \quad \text{when } 0 < \mu < 1 \quad (7.45a)$$

$$\mu \frac{\partial I_\omega^-}{\partial x} = \frac{I_\omega^* - I_\omega^-}{\Lambda}, \quad \text{when } -1 < \mu < 0 \quad (7.45b)$$

where we have assumed that the medium is gray.<sup>44,45</sup> If we further assume that the walls are diffuse and gray, then the boundary conditions become

$$T(0) = T_1 \quad \text{and} \quad T(L) = T_2 \quad (7.46)$$

Thus, 
$$I_\omega^+(0, \mu) = \varepsilon_1 I_\omega^*(T_1) + (1 - \varepsilon_1) I_\omega^-(0, \mu) \quad (7.47)$$

$$I_\omega^-(L, \mu) = \varepsilon_2 I_\omega^*(T_2) + (1 - \varepsilon_2) I_\omega^+(L, \mu) \quad (7.48)$$

The solutions of Eq. (7.45a) and Eq. (7.45b) can be expressed as follows:

$$I_{\omega}^{+}(x, \mu) = I_{\omega}^{+}(0, \mu) \exp\left(-\frac{x}{\Lambda \mu}\right) + \int_0^x I_{\omega}^{*}(\xi) \exp\left(-\frac{x-\xi}{\Lambda \mu}\right) \frac{d\xi}{\Lambda \mu} \quad \text{for } \mu > 0 \quad (7.49)$$

$$\text{and } I_{\omega}^{-}(x, \mu) = I_{\omega}^{-}(L, \mu) \exp\left(\frac{L-x}{\Lambda \mu}\right) - \int_x^L I_{\omega}^{*}(\xi) \exp\left(-\frac{x-\xi}{\Lambda \mu}\right) \frac{d\xi}{\Lambda \mu} \quad \text{for } \mu < 0 \quad (7.50)$$

In Eq. (7.49), the first term represents intensity originated from the left surface, after being attenuated, and the second term is the contribution of generation that is subject to attenuation as well. Equation (7.50) is viewed reversely for intensity from the right to the left. The spectral heat flux, defined in Eq. (7.37), can be written as

$$\begin{aligned} q_{\omega}'' &= 2\pi \int_0^1 \left[ I_{\omega}^{+}(0, \mu) \exp\left(-\frac{x}{\Lambda \mu}\right) - I_{\omega}^{-}(L, -\mu) \exp\left(-\frac{L-x}{\Lambda \mu}\right) \right] \mu d\mu \\ &+ 2\pi \int_0^x I_{\omega}^{*}(\xi) E_2\left(\frac{x-\xi}{\Lambda}\right) \frac{d\xi}{\Lambda} - 2\pi \int_x^L I_{\omega}^{*}(\xi) E_2\left(\frac{\xi-x}{\Lambda}\right) \frac{d\xi}{\Lambda} \end{aligned} \quad (7.51a)$$

Here again,  $E_m(x) = \int_0^1 \mu^{m-2} e^{-x/\mu} d\mu$  is the  $m$ th exponential integral. If the surface is diffuse, then

$$\begin{aligned} q_{\omega}'' &= 2\pi I_{\omega}^{+}(0) E_3\left(\frac{x}{\Lambda}\right) - 2\pi I_{\omega}^{-}(L) E_3\left(\frac{L-x}{\Lambda}\right) \\ &+ 2\pi \int_0^x I_{\omega}^{*}(\xi) E_2\left(\frac{x-\xi}{\Lambda}\right) \frac{d\xi}{\Lambda} - 2\pi \int_x^L I_{\omega}^{*}(\xi) E_2\left(\frac{\xi-x}{\Lambda}\right) \frac{d\xi}{\Lambda} \end{aligned} \quad (7.51b)$$

Note that energy balance requires that  $\frac{dq_x''}{dx} = \int_0^{\omega_m} \frac{\partial}{\partial x} q_{\omega}''(x, \omega) d\omega = 0$ . Differentiation of Eq. (7.51a) yields

$$\begin{aligned} \frac{\partial q_{\omega}''}{\partial x} &= -\frac{2\pi}{\Lambda} I_{\omega}^{+}(0) E_2\left(\frac{x}{\Lambda}\right) - \frac{2\pi}{\Lambda} I_{\omega}^{-}(L) E_2\left(\frac{L-x}{\Lambda}\right) \\ &- \frac{2\pi}{\Lambda} \int_0^L I_{\omega}^{*}(\xi) E_1\left(\frac{|x-\xi|}{\Lambda}\right) \frac{d\xi}{\Lambda} + \frac{4\pi}{\Lambda} I_{\omega}^{*}(x) \end{aligned} \quad (7.52)$$

In radiative transfer, we call  $J_1 = \int \pi I_{\omega}^{+}(0) d\omega$  and  $J_2 = \int \pi I_{\omega}^{-}(L) d\omega$  the total radiosities at surfaces 1 and 2, respectively, and  $e_b(T) = \int \pi I_{\omega}^{*} d\omega$  the total blackbody emissive power. Therefore,

$$2e_b(T(x)) = J_1 E_2\left(\frac{x}{\Lambda}\right) + J_2 E_2\left(\frac{L-x}{\Lambda}\right) + \int_0^L e_b(T(\xi)) E_1\left(\frac{|x-\xi|}{\Lambda}\right) \frac{d\xi}{\Lambda} \quad (7.53)$$

This is the same as the radiative equilibrium condition, given in Eq. (7.38). We cannot set Eq. (7.52) to zero at all frequencies, when local equilibrium does not exist, even for a gray medium.

**Example 7-4.** Find the temperature distribution, the heat flux, and the thermal conductivity for a gray medium, with diffuse-gray surfaces, in the radiative thick limit, i.e.,  $Kn \ll 1$ , under two extreme conditions: (1)  $T_1, T_2 \ll \Theta_b$  and (2)  $T_1, T_2 > \Theta_b$ .

**Solution.** In the radiative thick limit, the first two terms in Eq. (7.51a) can be dropped as long as  $x$  is not too close to either surface. Applying the first-order Taylor expansion  $I_\omega^*(x) = I_\omega^*(\xi) + (dI_\omega^*/dx)(x - \xi) + \dots$  and letting  $z = (x - \xi)/\Lambda$  in the third and fourth terms, we obtain

$$q_x'' = -4\pi\Lambda \frac{\partial I_\omega^*}{\partial x} \int_0^\infty z E_2(z) dz = -\frac{4\pi}{3} \Lambda \frac{\partial I_\omega^*}{\partial x} \quad (7.54)$$

since  $\int_0^\infty z E_2(z) dz = 1/3$ . In fact, this equation applies to everywhere inside the medium because the spectral heat flux is continuous in the radiative thick limit. Integrating Eq. (7.54) over the frequencies of interest, we see that, under condition (1),

$$q_x'' = -\frac{16\sigma_{\text{SB}}' T^3}{3} \Lambda \frac{dT}{dx}, \quad \text{when } T \ll \Theta_{\text{D}} \quad (7.55a)$$

This is nothing but a heat diffusion equation if we define the thermal conductivity as

$$\kappa(T) = (3/16)\sigma_{\text{SB}}' T^3 \Lambda \quad (7.55b)$$

Comparing with  $\kappa(T) = \frac{1}{3}C_{\text{Vg}}\Lambda$ , we notice from the previous equation that  $C_{\text{Vg}} = \frac{9}{16}\sigma_{\text{SB}}' T^3$ . In the radiative thick limit, the temperature distribution is continuous at the wall, i.e.,  $T(0^+) = T(0) = T_1$  and  $T(L^-) = T(L) = T_2$ . Furthermore, the radiosity at the wall becomes the blackbody emissive power, even though the surface is not black; thus, we can integrate Eq. (7.54) over  $x$  from 0 to  $L$ :

$$\int_0^L q_x'' dx = \frac{4\Lambda}{3} \sigma_{\text{SB}}' \int_{T_1}^{T_2} 4T^3 dT$$

which gives

$$q_x'' = \frac{4}{3} Kn (\sigma_{\text{SB}}' T_1^4 - \sigma_{\text{SB}}' T_2^4) \quad (7.56a)$$

as well as the temperature distribution:

$$T(x) = [T_1^4 - \frac{x}{L}(T_1^4 - T_2^4)]^{1/4} \quad (7.56b)$$

which is linear in terms of the fourth power of temperature. From the definition of thermal resistance,  $q_x'' = (T_1 - T_2)/R_t''$ , we have

$$R_t'' = \frac{3(T_1 + T_2)(T_1^2 + T_2^2)}{4\sigma_{\text{SB}}' Kn} \quad (7.57)$$

Under condition (2), when the temperature is greater than the Debye temperature, we have

$$q_x'' = -\frac{\omega_{\text{m}}^3 k_{\text{B}}}{6\pi^2 v_{\text{p}}^2} \Lambda \frac{dT}{dx} \quad \text{when } T > \Theta_{\text{D}} \quad (7.58)$$

The thermal conductivity becomes  $\kappa(T) = \omega_{\text{m}}^3 k_{\text{B}} \Lambda / (6\pi^2 v_{\text{p}}^2)$ , which implies that  $C_{\text{Vg}} = \omega_{\text{m}}^3 k_{\text{B}} / (3\pi^2 v_{\text{p}}^2)$ . A proper  $\omega_{\text{m}}$  should be chosen so that only propagating phonons are considered. Assuming that the temperature difference is small so that we can approximate the thermal conductivity as a constant, we have

$$q_x'' = \frac{C_{\text{Vg}} Kn}{3} (T_1 - T_2) \quad (7.59)$$

The thermal resistance becomes  $R_t'' = 3/(C_{\text{Vg}} Kn)$ , which increases as  $L$  increases. The temperature distribution is linear. One should realize that the scattering rate increases with temperature, due to phonon-phonon scattering, and depends on the frequency. If we look at the radiative equilibrium condition again, by assuming  $T_1 > T_2$ , we see that  $I_\omega^* > I_\omega^- > I_\omega^+$ . Therefore, local equilibrium is not

a stable-equilibrium state. In the radiative thick limit, the difference between  $I_{\omega}^{+}$  and  $I_{\omega}^{-}$  is caused by the spatial variation of  $I_{\omega}^e$  as can be clearly seen from Eq. (7.49) and Eq. (7.50). Hence, Eq. (7.40) is a good approximation. In the radiative thin limit, according to Eq. (7.34), Eq. (7.59) becomes

$$q_x'' = \frac{1}{4} C_{V_g} (T_1 - T_2) \quad (7.60)$$

Although no closed form exists for the solution of the ERT between the thick and thin limits, a number of approximation techniques and numerical methods can be used to provide satisfactory solutions, such as the discrete ordinates method ( $S_N$  approximation) and the spherical harmonics method ( $P_N$  approximation). It is important to see that except in the radiative thick limit, energy transfer occurs inside the medium in two ways: one is through exchange with the walls, and the other is through diffusion. For this reason, a ballistic-diffusion approximation has been developed to solve the EPRT; see Chen (*Phys. Rev. Lett.*, **86**, 2297, 2001). In general, the temperature distribution looks like that in Fig. 4.12b if  $T_2$  is comparable to the Debye temperature. If  $T_1 \ll \Theta_D$ , then the temperature distribution can be plotted in terms of  $T^4$  so that the distribution looks more or less linear. There exists a temperature jump such that  $T(0^+) \neq T(0)$  and  $T(L^-) \neq T(L)$ , except in the radiative thick limit. Understanding that the temperature is only an effective temperature and given such a temperature distribution, one may assume that there is a thermal resistance at each boundary and an internal thermal resistance, which may be described by Fourier's heat conduction.<sup>41</sup> For thermal radiative transfer in the absence of heat conduction, there exists a radiation slip or radiation jump at the boundary, unless the medium is optically thick. Without a participating medium, photons do not scatter on itself to dissipate heat or transfer heat by diffusion. This is a distinction between photons and phonons. Radiation slip is manifested by a discontinuous change of the intensity at the boundary. The temperature in the medium adjacent to the wall differs from the surface temperature. Such a temperature jump does not exist in classical Fourier's heat conduction theory; however, both velocity slip and temperature jump have already been incorporated in microfluidics research, as discussed in Chap. 4 [see Eq. (4.94)]. The temperature-jump concept was first applied in the study of heat conduction in rarefied gases over 100 years ago. A straightforward approach for phonon transport is to sum up the thermal resistances in the radiative thin and thick limits. The heat flux at very low temperatures can be expressed as

$$q_x'' = \frac{4\Lambda}{3L} \frac{\sigma'_{SB} T_1^4 - \sigma'_{SB} T_2^4}{1 + \left( \frac{1}{\varepsilon_1} - \frac{1}{2} + \frac{1}{\varepsilon_2} - \frac{1}{2} \right) \frac{4Kn}{3}} \quad (7.61)$$

Here, we separately write  $(1/\varepsilon_1 - 1/2)$  and  $(1/\varepsilon_2 - 1/2)$  to emphasize the thermal resistance due to radiation slip at each boundary. In the radiative thick limit, the temperature jump approaches to zero as  $Kn \rightarrow 0$ . Basically, Eq. (7.61) reduces to Eq. (7.44) and Eq. (7.56a), in the extremes. If the walls can be treated as blackbodies, i.e.,  $\varepsilon_1 = \varepsilon_2 = 1$ , and the temperature difference between  $T_1$  and  $T_2$  is small, we can approximate the heat flux as follows:

$$q_x'' = \frac{\kappa_b}{L} \frac{\Delta T}{1 + 4Kn/3} = \kappa_{\text{eff}} \frac{\Delta T}{L} \quad (7.62)$$

where  $\Delta T = T_1 - T_2 \ll T_2 < T_1$ , the bulk thermal conductivity  $\kappa_b(T) = \frac{16}{3} \sigma'_{SB} T^3 \Lambda$ , and the effective conductivity of the film is

$$\kappa_{\text{eff}} = \frac{\kappa_b}{1 + 4Kn/3} \quad (7.63)$$

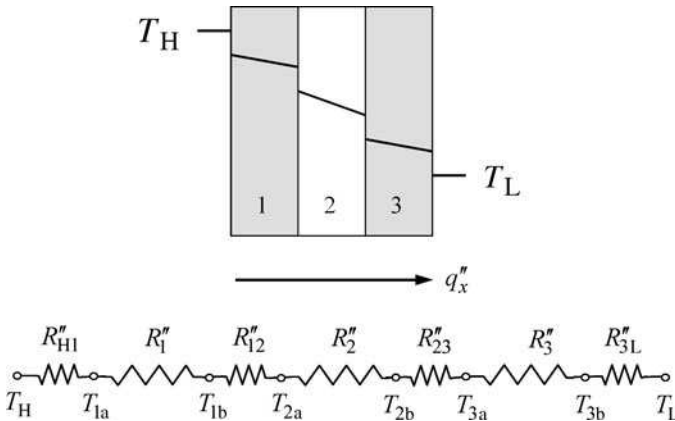
At relatively high temperatures close to the Debye temperature, from Eq. (7.59) and Eq. (7.60), we can write

$$q_x'' = \frac{\kappa_b}{L} \frac{T_1 - T_2}{1 + \left( \frac{1}{\varepsilon_1} + \frac{1}{\varepsilon_2} - 1 \right) \frac{4Kn}{3}} = \kappa_{\text{eff}} \frac{T_1 - T_2}{L} \quad (7.64)$$

where  $\kappa_b(T) = \frac{1}{3}Cv_g\Lambda$ . Equation (7.64) gives the same conductivity ratio  $\kappa_{\text{eff}}/\kappa_b$  as in Eq. (7.63) for blackbody walls. These effective thermal conductivities are on the same order of magnitude as we have derived in Sec. 5.5.5, based on simple geometric arguments and Matthiessen's rule for the mean free path given in Eq. (5.116). In previous chapters, however, we did not elaborate in detail the nature of nonequilibrium and the necessity of defining an effective temperature. It is interesting that different schools of thought can result in rather consistent results. The heat diffusion equation *per se* cannot tell us the cause of a temperature jump or how to evaluate it. The phonon BTE enables us to explore the microscopic phenomena and helps evaluate the parameters and the properties. The microscopic understanding and the macroscopic phenomenological equations can indeed work together to provide an effective thermal analysis tool.

The preceding discussions are consistent with the detailed derivation of the temperature jump or the radiation slip, originally formulated by Deissler (*J. Heat Transfer*, **C86**, 240, 1964), for situations not too far from the radiative thick limit. Nevertheless, the expressions given in Eq. (7.61) and Eq. (7.64) can be approximately applied between the diffusion and ballistic extremes. It should be noted that when the temperature jump is treated as a thermal resistance at the boundary, Fourier's law can be used for the heat conduction inside the medium with bulk thermal conductivity. This is very different from heat conduction along the film.

While there seems to be no problem in understanding the meaning of emissivity for optical radiation, a question still remains as how to interpret the boundary conditions in the case of phonon conduction. If a multilayer structure is considered, we need to better understand the reflection and the transmission of phonons at the interfaces between dissimilar materials. A three-layer structure is shown in Fig. 7.6 to illustrate the temperature distribution



**FIGURE 7.6** Temperature distribution in a multilayer structure, with thermal boundary resistance, and the thermal resistance network representation. Here,  $R_i''$  is the internal resistance in the  $i$ th layer due to heat conduction, and  $R_{ij}''$  is the thermal boundary resistance between the  $i$ th and  $j$ th media. Two temperatures are needed to specify the effective temperature of different media at the interface.

in a multilayer structure. Depending on the temperature range, it seems that we can conveniently determine the internal thermal resistance with Fourier's law, i.e.,  $R_i'' = L_i/\kappa_i$ . For the thermal resistance at the interface inside the layered structures, we could replace the emissivity with the transmissivity  $\Gamma_{ij}$  such that  $R_{ij}'' = \frac{4}{3}(\Lambda_i/\kappa_i)(\Gamma_{ij}^{-1} - \frac{1}{2}) + \frac{4}{3}(\Lambda_j/\kappa_j)(\Gamma_{ji}^{-1} - \frac{1}{2})$ . At the boundaries, we can still use  $R_{\text{H1}}'' = \frac{4}{3}(\Lambda_1/\kappa_1)(\epsilon_1^{-1} - \frac{1}{2})$  and  $R_{\text{3L}}'' = \frac{4}{3}(\Lambda_3/\kappa_3)(\epsilon_3^{-1} - \frac{1}{2})$ . The heat flux can be estimated by  $q_x'' = (T_{\text{H}} - T_{\text{L}})/R_{\text{tot}}''$ , where  $R_{\text{tot}}''$  is the sum of all thermal resistances. The effective thermal conductivity of the whole layered structure becomes  $\kappa_{\text{eff}} = L_{\text{tot}}/R_{\text{tot}}''$ . The details were presented by Chen and Zeng, who further considered non-diffuse surfaces and defined equivalent equilibrium temperatures.<sup>41</sup> The assumption is that the deviation from the radiative thick limit is not significant. If we are dealing with the ballistic regime, we might need to consider phonon wave effects as well as the quantum size effect. Models for thermal boundary resistance will be discussed in the next subsection. It is intriguing to apply the same approach to electron systems for the study of both electrical conductivity and thermal conductivity of metallic solids, as well as metal-dielectric multilayer structures. The thermal resistance network method, however, cannot be easily extended to multidimensional problems or to transient heating by a localized heat source. Statistical models or atomistic simulations are necessary. Therefore, the extension of Fourier's law for 1-D nonequilibrium heat transfer should be considered only as a special case.

### 7.2.3 Thermal Boundary Resistance (TBR)

Thermal resistance at the interface between dissimilar materials is very important for heat transfer in heterostructures. Let us first clarify the difference between thermal contact resistance and thermal boundary resistance (TBR). The former refers to the thermal resistance between two bodies, usually with very rough surfaces whose root-mean-square roughness  $\sigma_{\text{rms}}$  is greater than  $0.5 \mu\text{m}$ , brought or joined together mechanically. For thermal contact resistance, readers are referred to a recent comprehensive review by Yovanovich.<sup>46</sup> Originally, TBR refers to the resistance at the interface between two solids or between a liquid and a dielectric at low temperatures. Even when the materials are in perfect contact with each other, reflections occur when phonons travel toward the boundary, because of the difference in acoustic properties of adjacent materials. In practice, the interface can be atomically smooth, or with a roughness ranging from several tenths of a nanometer to several nanometers. The thermal resistance between a solid material and liquid helium is called the Kapitza resistance, first observed by the Russian physicist and 1978 Nobel Laureate Pyotr Kapitza, in the 1940s. This thermal resistance results in a temperature discontinuity at the boundary and has been modeled, based on the acoustic mismatch model (AMM). Thermal boundary resistance exists between two dielectrics as well as between a metal and a dielectric. In a thin-film structure, an interface is often accompanied by the formation of an intermediate layer of mixed atoms. An extensive review of earlier studies can be found in the work of Swartz and Pohl in 1989;<sup>38</sup> see also Stoner and Maris (*Phys. Rev. B*, **48**, 16373, 1993). Prasher and Phelan (*J. Supercond.*, **10**, 473, 1997) reviewed the studies of TBR of high-temperature superconductors in both the normal and superconducting states, for applications in superconducting electronics and radiation detectors.

Little showed that the heat flux across the boundary of a perfectly joined interface between two solids is proportional to the difference in the fourth power of temperature on each side of the interface.<sup>39</sup> This can be understood based on previous discussions of phonon radiative transfer and blackbody radiation. Consider longitudinal phonon modes that follow the linear dispersion in a Debye crystal, and assume that the interface is perfectly smooth.



At any given frequency, the transmission coefficients can be written as follows (with a small modification for consistency):<sup>39,47</sup>

$$\tau_{12} = \frac{4\rho_1\rho_2v_{l1}^2\cos\theta_1\cos\theta_2}{(\rho_1v_{l1}\cos\theta_2 + \rho_2v_{l2}\cos\theta_1)^2} \quad (7.65a)$$

$$\tau_{21} = \frac{4\rho_1\rho_2v_{l2}^2\cos\theta_1\cos\theta_2}{(\rho_1v_{l1}\cos\theta_2 + \rho_2v_{l2}\cos\theta_1)^2} \quad (7.65b)$$

where subscripts 1 and 2 denote the media 1 and 2, respectively,  $\rho$  is the density,  $v_l$  is the propagation speed of longitudinal phonons, and  $\theta$  is the polar angle, as illustrated in Fig. 7.7. The scattering is assumed to be purely elastic since the phonon frequency is conserved. An analog of Snell's law can be written as follows:

$$\frac{1}{v_{l1}}\sin\theta_1 = \frac{1}{v_{l2}}\sin\theta_2 \quad (7.66)$$

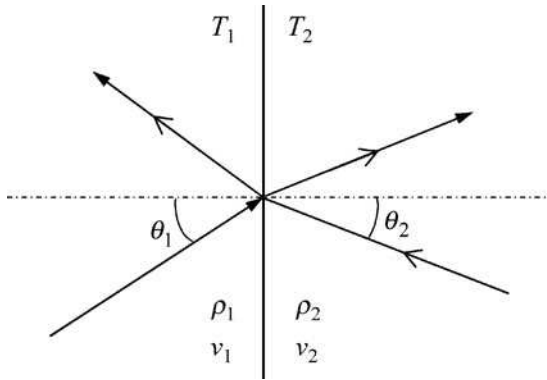


FIGURE 7.7 Schematic of phonon transport across an interface between two semi-infinite media, each at a thermal equilibrium.

If  $v_{l1} > v_{l2}$ , for incidence from medium 2 to 1, there exists a critical angle  $\theta_c = \sin^{-1}(v_{l2}/v_{l1})$ , beyond which all phonons will be reflected. Due to the boundary resistance, there will be a temperature difference across the interface. By assuming that the phonons are at equilibrium on either side, the heat flux from medium 1 to 2 can be expressed as follows:

$$q''_{1 \rightarrow 2} = \frac{1}{4\pi} \int_0^{\omega_m} \int_0^{2\pi} \int_0^{\pi/2} \hbar\omega v_{l1} f_1(\omega, T_1) \tau_{12} D(\omega) \cos\theta_1 \sin\theta_1 d\theta_1 d\phi_1 d\omega \quad (7.67)$$

If the distribution function is isotropic over the hemisphere, we have

$$q''_{1 \rightarrow 2} = \frac{1}{4} \frac{\Gamma_{12}}{v_{l1}^2} \int_0^{\omega_m} \hbar\omega v_{l1}^3 f_1(\omega, T_1) D(\omega) d\omega \quad (7.68)$$

where

$$\Gamma_{12} = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} \tau_{12} \cos\theta_1 \sin\theta_1 d\theta_1 d\phi = 2 \int_0^{\pi/2} \tau_{12} \cos\theta_1 \sin\theta_1 d\theta_1 \quad (7.69)$$

can be viewed as the hemispherical transmissivity. Note that

$$\frac{\Gamma_{21}}{v_{l2}^2} = 2 \int_0^{\theta_c} \frac{\tau_{21}}{v_{l2}^2} \cos \theta_2 \sin \theta_2 d\theta_2 = \frac{\Gamma_{12}}{v_{l1}^2} \quad (7.70)$$

For the Debye density of states, we have

$$\frac{1}{4\pi} v_i \hbar \omega f(\omega, T) D(\omega) d\omega = \frac{\hbar \omega^3}{8\pi^3 v_l^2 (e^{\hbar \omega / k_B T} - 1)}$$

Therefore, the net heat flux across the interface becomes

$$q_x'' = q_{1 \rightarrow 2}'' - q_{2 \rightarrow 1}'' = \frac{1}{4} \frac{\Gamma_{12}}{v_{l1}^2} \int_0^{\omega_m} \hbar \omega [v_{l1}^3 f_1(\omega, T_1) - v_{l2}^3 f_2(\omega, T_1)] D(\omega) d\omega \quad (7.71)$$

or

$$q_x'' = \frac{\Gamma_{12}}{v_{l1}^2} \frac{k_B^4}{8\pi^2 \hbar^3} \left( T_1 \int_0^{x_{m1}} \frac{x^3 dx}{e^x - 1} - T_2^4 \int_0^{x_{m2}} \frac{x^3 dx}{e^x - 1} \right) \quad (7.72)$$

where  $x_{mj} = \hbar \omega / k_B T_j$ . In the low-temperature limit, we obtain

$$q_x'' = \frac{\Gamma_{12}}{v_{l1}^2} \frac{\pi^2 k_B^4}{120 \hbar^3} (T_1^4 - T_2^4) \quad (7.73)$$

After replacing  $v_{j1}^{-2}$  with  $\sum_j v_{j1}^{-2} = v_{l1}^{-2} + 2v_{t1}^{-2}$ , i.e., one longitudinal and two transverse phonon modes, we obtain

$$q_x'' = \frac{\pi^2 k_B^4}{120 \hbar^3} (T_1^4 - T_2^4) \Gamma_{12} \sum_j v_{j1}^{-2} \quad (7.74)$$

The TBR can now be obtained as  $R_b'' = (T_1 - T_2) / q_x''$ . Furthermore, by assuming that the temperature difference is small, we can approximate  $R_b''$  by

$$R_b'' = \frac{30 \hbar^3 T^{-3}}{\pi^2 k_B^4 \Gamma_{12} \sum_j v_{j1}^{-2}} \quad (7.75)$$

which is inversely proportional to  $T^3$ .

The characteristic wavelength is the most probable wavelength in the phonon distribution function. It can be approximated by

$$\lambda_{mp} \approx a \frac{\Theta_D}{T} \quad (7.76)$$

where  $a$  is the lattice constant, on the order of 0.3 to 0.6 nm.<sup>47</sup> Only when  $\lambda_{mp} \gg \sigma_{rms}$ , can we assume that the scattering is completely specular. Even for atomically smooth interfaces, the characteristic wavelength for phonons will be on the same order of magnitude as the rms surface roughness, when the temperature approaches the Debye temperature. The specularity parameter introduced in Eq. (5.131) is often used to approximate the fraction of specular reflection with respect to the total reflection. Another expression of the specularity parameter is

$$p = \exp\left(-\frac{16\pi^2 \sigma_{rms}^2}{\lambda^2}\right) \quad (7.77)$$

This equation has often been wrongly expressed with  $\pi^2$  being mistaken as  $\pi^3$  in the heat conduction literature, following a hidden typo in Ziman's book, *Electrons and Phonons*.<sup>48</sup> In the high-temperature limit, TBR is expected to be small, especially when compared with conduction in the solids. Other considerations are (a) the interface may not be perfectly smooth, (b) there exists an upper limit of the frequency or a lower limit of wavelength, and (c) phonons on either sides of the boundary may not be in a local-equilibrium state. These difficulties post some real challenges in modeling TBR. Nevertheless, we shall present the diffuse mismatch model (DMM) that was introduced by Swartz and Pohl.<sup>38</sup> In the DMM, it is assumed that phonons will be scattered according to a probability, determined by the properties of the two media but independent of where the phonons are originated. For phonons coming from medium 1, the transmission and reflection probabilities are related by  $\Gamma_{12} + R_{12} = 1$ . For phonons originated from medium 2, on the other hand,  $\Gamma_{21} = R_{12}$  and  $R_{21} = \Gamma_{12}$ . Hence, the reciprocity requires that

$$\Gamma_{12} + \Gamma_{21} = 1 \quad (7.78a)$$

We can rewrite Eq. (7.70), considering all three polarizations, as follows:

$$\Gamma_{12} \sum_j v_{j1}^{-2} = \Gamma_{21} \sum_j v_{j2}^{-2} \quad (7.78b)$$

Solving Eq. (7.78a) and Eq. (7.78b), we get

$$\Gamma_{12} = \frac{\sum_j v_{j2}^{-2}}{\sum_j v_{j1}^{-2} + \sum_j v_{j2}^{-2}} \quad (7.79)$$

The heat flux can be calculated according to

$$q_x'' = \frac{k_B^4}{8\pi^2\hbar^3} \left( T_1^4 \int_0^{x_{m1}} \frac{x^3 dx}{e^x - 1} - T_2^4 \int_0^{x_{m2}} \frac{x^3 dx}{e^x - 1} \right) \Gamma_{12} \sum_j v_{j1}^{-2} \quad (7.80)$$

Equation (7.79) and Eq. (7.80) are the only equations needed to calculate TBR with the DMM. In addition to the Debye temperatures and the speeds of longitudinal and transverse waves, one would need to determine the upper limits of the integrals in Eq. (7.80). Alternatively, Eq. (7.80) can be recast using the volumetric heat capacities and the group velocities to obtain

$$q_x'' = \frac{1}{4} (C_{1v_{g1}} T_1 - C_{2v_{g2}} T_2) \Gamma_{12} \quad (7.81)$$

One must be careful in applying the heat capacity in Eq. (7.81) since the heat capacity in the expression of thermal conductivity is different from  $\rho c_p$ , unless at very low temperatures. Both the AMM and the DMM assume that the phonons are in equilibrium on each side of the interface, and do not take into account the nonequilibrium distribution of phonons. In multilayer thin films, especially in quantum wells and superlattices, when the film thickness is comparable with or smaller than the phonon mean free path, thermal transport inside the film cannot be modeled as pure diffusion anymore. A detailed treatment of temperature-jump conditions and boundary resistance in superlattices was performed by Chen and Zeng.<sup>40,41</sup> Majumdar (*J. Heat Transfer*, **113**, 797, 1991) proposed a modified AMM by modeling interface roughness using a fractal structure. In this study, the reflection was approximated by geometric optics, which is applicable when the phonon wavelength is smaller than the autocorrelation length of the rough surface. TBR between highly dissimilar materials, metal-metal interface, and metal-dielectric interface have been the areas of some recent studies; see Majumdar and Reddy (*Appl. Phys. Lett.*, **84**, 4768), Ju et al.

(*J. Heat Transfer*, **128**, 919, 2006), Lyeo and Cahill (*Phys. Rev. B*, **73**, 144301, 2006), and Gundrum et al. (*Phys. Rev. B*, **72**, 245426, 2006).

Let us consider how to model transient heat conduction in thin films. The single relaxation time approximation appears to be limited to the timescale  $t > \tau$ . Joshi and Majumdar (*J. Appl. Phys.*, **74**, 31, 1993) performed a transient analysis of the EPRT. However, the use of Eq. (7.43) implies the presumption of the local-equilibrium condition, which is not valid for large temperature gradients. If worked out properly, the EPRT is applicable for  $t > \tau$ . If the temperature jumps can be properly taken into consideration, it appears that Fourier's law should be applicable when  $t \gg \tau$ . The question is, "How long will it take for the temperature-jump conditions to be justified?" In order to model a timescale less than  $\tau$ , it would be interesting to see if there exists another scattering mechanism that has a much smaller relaxation time than  $\tau$  and that does not transfer or dissipate heat, like the  $N$  processes discussed earlier or phonons in the optical branch. In the two-fluid model, the superfluid moves forward freely, without any viscosity, but conserves kinetic energy as it moves around. This is the principle of superfluidity in liquid helium and superconductivity for electrons. The superfluid does not carry thermal energy, nor does it dissipate heat. Although the  $N$  processes do not carry heat forward, these processes are important for the redistribution of phonons. The two-relaxation-time model developed by Callaway<sup>21</sup> and Guyer and Krumhansl<sup>22</sup>, or the Jeffrey-type equation, might be applicable in extreme cases, e.g., in a nanotube at very small timescales, on the order of femtoseconds. The solution describes a wavelike characteristic that is a combination of a damped wave and a weak diffusion process, which enables an instantaneous response, intrinsic to all heat conduction processes, as justified by statistical mechanics. As mentioned earlier, the wavelike behavior has been demonstrated recently in SWNTs, via molecular dynamics simulation, although a lot of work needs to be done to extend the simulation to multilayer structures. Let us emphasize again that Fourier's heat diffusion appears to be universal for heat conduction, and the hyperbolic heat equation, Eq. (7.5), can be neither physically justified nor practically useful. In the extremely acoustic thin limit, we are dealing with quantum conductance or the Schrödinger wave equation. This wave phenomenon cannot be explained by the hyperbolic heat equation.

### 7.3 HEAT CONDUCTION REGIMES

There has been a continuous effort to delineate the regimes of microscale heat conduction since 1992. A number of references have already been cited in Chap. 5. A recent effort has been made by Escobar et al.<sup>43</sup> Following the previous discussions in this chapter, let us schematically depict the regimes of heat conduction, especially by electrons and phonons in crystalline solids, as in Fig. 7.8. Here,  $\tau_c$  is known as *effective collision interaction time*, or simply *collision time*, since collision does not occur instantaneously but is through intermolecular potential and force interactions. These forces become important only when the particles come very close to each other. Of course, this is the classical picture of atomic or molecular interactions. Electrons and phonons are quantum mechanical particles; thus, the interaction is via the wavefunctions predicted by Schrödinger's equations. For ultrafast pulse heating, the collision time can be the time required for a photon and an electron to interact. Generally speaking, the collision time is much shorter than the relaxation time and neglected in the BTE. The characteristic phonon or electron wavelength  $\lambda$  is assumed to be less than the mean free path  $\Lambda$ .

Region 1 is the macroscale regime where Fourier's law and the heat diffusion equation can be applied, when the timescale is greater than  $\tau$  and the length scale is greater than about  $10\Lambda$ . Region 2 is called the mesoscale or quasi-equilibrium regime, which is characterized by the classical size effect. This region is also known as the first microscale. For

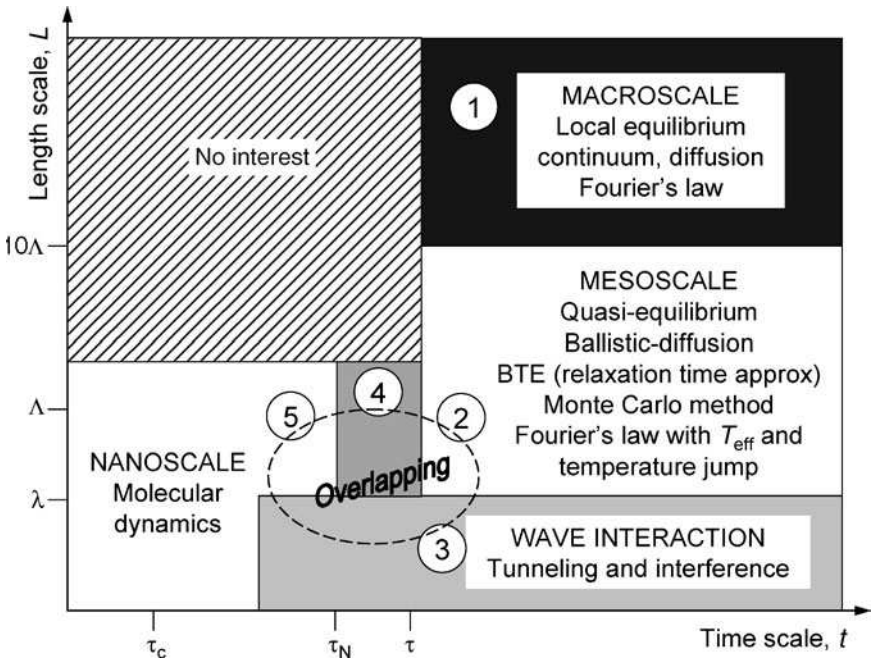


FIGURE 7.8 Heat conduction regimes.

heat transfer along a film or a wire, local-equilibrium assumption is appropriate and boundary scattering reduces the effective mean free path and thermal conductivity. For heat transfer across a film or a multilayer, it is possible to use Fourier's law inside the medium by considering an effective temperature and the temperature-jump boundary condition. It is difficult, if not impossible, to apply Fourier's law to complex geometries or local heating. The two-temperature model for fast laser heating can be in either region 1 or 2, depending on how the length scale is compared with the mean free path. Most of the research on microscale heat transfer between 1990 and 2005 dealt with the microscale phenomena in region 2.

Region 3 is the regime of wave behavior, which is described by Schrödinger's wave equations and where quantum tunneling can occur. Quantum size effect becomes significant on thermal conductivity and specific heat. Quantum conductance is a special case of quantum tunneling, for which the ballistic processes are confined in one dimension through a channel. For very thin layers, wave interference may become important. However, due to the interface roughness, the coherence may be destroyed so that the energy ray method or the particle approach can still be applied at very small length scales. We will give a comprehensive treatment of electromagnetic wave interference and scattering phenomena in subsequent chapters, without discussing the nature of acoustic waves further. The region on the upper left is said to be of no interest at short timescales because a thermal disturbance cannot travel that far and affect the temperature field.

Region 4 is designed to represent the wavelike behavior, described by the Jeffreys-type equation, Eq. (7.14). When we say Jeffreys-type equation, we mean that both  $\kappa_0$  and  $\kappa_1$  in Eq. (7.12a) are positive. As discussed earlier,  $\tau_N$  is the second relaxation time for phonon scattering that does not transfer or dissipate thermal energy, as in the  $N$  processes. In this regime, the BTE based on the two-relaxation-time approximation may be applied.<sup>21,22</sup> This regime includes the heat pulse propagation and the second sound in dielectric crystals, at

low temperatures. It suffices to say that this region, while of great academic interest, has very limited applications. The pure hyperbolic heat equation, however, predicts a nonphysical wavefront and cannot be applied without the additional diffusion term. Nevertheless, theoretical studies of the hyperbolic heat equation have helped in a better understanding of the heat transfer behavior at short timescales and, subsequently, facilitated the development of more realistic models. While the lagging heat equation can mathematically describe both wavelike behavior and parallel heat conduction, it does not provide any new physics. On the other hand, the memory concept may be related to the anharmonic and nonlinear effects that are inherent to the solid and crystal structures. Study of the thermomechanical and thermoelastic effects, and thermal transport in polymers and inhomogeneous materials, such as biological materials, may require empirical and semiempirical models. The lagging heat equation or similar differential equations may be quite helpful in these applications.

Region 5 belongs to the nanoscale regime, where it is necessary to employ quantum or sometimes classical molecular dynamics to study the underlying phenomena. The dashed ellipse indicates the overlapping between different regions, where molecular dynamics simulation may provide rich information as well as a bridge between different timescales and length scales.

Holland (*Phys. Rev.*, **132**, 2461, 1963) analyzed the effect of different polarizations on the thermal conductivity of germanium. Several studies have employed the Monte Carlo method to solve phonon transport equations.<sup>36,37,49</sup> The lattice Boltzmann method has also been employed in a number of publications.<sup>14,43</sup> Molecular dynamics has been applied to the study of TBR, including the interface between SWNTs.<sup>50–53</sup> Chung et al. investigated the effect of different dispersion models on the lattice thermal conductivity.<sup>54</sup> Narumanchi et al. used a finite volume method to solve the 2-D BTE during transient heat transport for a local heating source in silicon.<sup>55</sup> They also demonstrated the feasibility to include phonon dispersion and polarization in the model. Nonequilibrium phonon transport in dimensions less than 100 nm has become an important issue in silicon-on-insulator transistors. Several studies have focused on a multiscale approach to model the thermal transport phenomena at the device level.<sup>42,43,56–58</sup> Sinha and Goodson provided an extensive review on multiscale modeling.<sup>42</sup>

Thermal metrology includes measurements of temperature (thermometry), specific heat (calorimetry), and heat flux. Thermophysical properties, such as thermal conductivity and specific heat, can be measured with steady-state, modulated, or pulsed heating techniques. MEMS and NEMS have enabled the fabrication of miniaturized heaters and sensors. Furthermore, optical techniques such as thermoreflectance, Raman spectroscopy, photothermal radiometry, fluorescence, and laser flash techniques have been widely used in the measurement of thermal properties of nano/microstructured materials. Scanning thermal microscopy and near-field optical microscopy have further improved the spatial resolution. A large number of publications can be found from the bibliography of the present and previous chapters [see, e.g., Cahill et al.<sup>31</sup> and references therein]. Recently, Abel et al. employed micro-Raman spectroscopy to measure the temperature distribution in silicon microstructures with a spatial resolution of 1  $\mu\text{m}$ .<sup>59</sup> Lee et al. performed a steady-state characterization of heated AFM cantilevers over a range pressures for thermal metrology applications.<sup>60</sup> Park et al. analyzed the frequency response of heated AFM cantilevers in the frequency range from 10 Hz to 1 MHz, and observed high-order harmonic responses, such as  $3\omega$ ,  $5\omega$ , and  $7\omega$ , at frequencies below 100 kHz and impedance effect at higher frequencies.<sup>61</sup> Park et al. also investigated thermal behavior of heated cantilevers at cryogenic temperatures, down to 78 K.<sup>62</sup> By measuring the thermal response at various frequencies, this study extracted the specific heat near the cantilever tip and the thermal conductivity along the heavily doped silicon legs, at temperatures ranging from 80 to 200 K. There appears to be a significant reduction in the thermal conductivity for the free-standing silicon cantilever, with a thickness of 0.59  $\mu\text{m}$ , at low temperatures. These studies demonstrate that heated AFM cantilevers have become a promising thermal analysis tool at the micro- and nanoscales.<sup>59–62</sup>

## 7.4 SUMMARY

---

The present chapter, along with Chaps. 5 and 6, provided a comprehensive treatment of thermal properties of and the transport processes in micro/nanostructured solid materials. This chapter focused on the transient and nonequilibrium heat conduction, when the local equilibrium condition is not satisfied to justify the conventional heat diffusion theory, based on Fourier's law. Several modified phenomenological theories were critically reviewed with an emphasis on their application regimes. The phonon BTE was presented using the EPRT, and the solutions were discussed for the nonequilibrium heat transfer across a thin film or a multilayer structure. The basic models of TBR were outlined. Finally, a heat transfer regime was developed to assist readers in choosing an appropriate methodology for a given situation, with a brief summary on advanced multiscale modeling and measurement techniques.

## REFERENCES

---

1. H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids*, 2nd ed., Clarendon Press, Oxford, 1959.
2. M. N. Özışik, *Heat Conduction*, 2nd ed., Wiley, New York, 1993.
3. D. D. Joseph and L. Preziosi, "Heat waves," *Rev. Mod. Phys.*, **61**, 41–73, 1989; D. D. Joseph and L. Preziosi, "Addendum to the paper 'heat waves'," *Rev. Mod. Phys.*, **62**, 375–391, 1990.
4. M. N. Özışik and D. Y. Tzou, "On the wave theory in heat conduction," *J. Heat Transfer*, **116**, 526–535, 1994.
5. W. K. Yeung and T. T. Lam, "A numerical scheme for non-Fourier heat conduction, Part I: One-dimensional problem formulation and applications," *Numer. Heat Transfer B*, **33**, 215–233, 1998.
6. A. Haji-Sheikh, W. J. Minkowycz, and E. M. Sparrow, "Certain anomalies in the analysis of hyperbolic heat conduction," *J. Heat Transfer*, **124**, 307–319, 2002.
7. J. Gembarovic and J. Gembarovic, Jr., "Non-Fourier heat conduction modeling in a finite medium," *Int. J. Thermophys.*, **25**, 41261–41268, 2004.
8. M. B. Rubin, "Hyperbolic heat conduction and the second law," *Int. J. Eng. Sci.*, **30**, 1665–1676, 1992.
9. C. Bai and A. S. Lavine, "On hyperbolic heat conduction and the second law of thermodynamics," *J. Heat Transfer*, **117**, 256–263, 1995.
10. A. Barletta and E. Zanchini, "Hyperbolic heat conduction and local equilibrium: A second law analysis," *Int. J. Heat Mass Transfer*, **40**, 1007–1016, 1997.
11. D. Jou, J. Casas-Vazquez, and G. Lebon, *Extended Irreversible Thermodynamics*, 2nd ed., Springer, Berlin, 1996.
12. A. Majumdar, "Microscale heat conduction in dielectric thin films," *J. Heat Transfer*, **115**, 7–16, 1993.
13. S. Volz, J.-B. Saulnier, M. Lallemand, B. Perrin, P. Depondt, and M. Mareschal, "Transient Fourier-law deviation by molecular dynamics in solid argon," *Phys. Rev. B*, **54**, 340–347, 1996.
14. J. Xu and X. Wang, "Simulation of ballistic and non-Fourier thermal transport in ultra-fast laser heating," *Physica B*, **351**, 213–226, 2004.
15. D. Y. Tzou, *Macro- to Microscale Heat Transfer: The Lagging Behavior*, Taylor & Francis, Washington DC, 1997.
16. P. J. Antaki, "Solution for non-Fourier dual phase lag heat conduction in a semi-infinite slab with surface heat flux," *Int. J. Heat Mass Transfer*, **41**, 2253–2258, 1998.
17. D. W. Tang and N. Araki, "Wavy, wavelike, diffusive thermal responses of finite rigid slabs to high-speed heating of laser-pulses," *Int. J. Heat Mass Transfer*, **42**, 855–860, 1999.
18. D. Y. Tzou and K. S. Chiu, "Temperature-dependent thermal lagging in ultrafast laser heating," *Int. J. Heat Mass Transfer*, **44**, 1725–1734, 2001.

19. W. J. Minkowycz, A. Haji-Sheikh, and K. Vafai, "On departure from local thermal equilibrium in porous media due to a rapid changing heat source: The Sparrow number," *Int. J. Heat Mass Transfer*, **42**, 3373–3385, 1999.
20. W. Kaminski, "Hyperbolic heat conduction equation for materials with a nonhomogeneous inner structure," *J. Heat Transfer*, **112**, 555–560, 1990.
21. J. Callaway, "Model for lattice thermal conductivity at low temperatures," *Phys. Rev.*, **113**, 1046–1951, 1959.
22. R. A. Guyer and J. A. Krumhansl, "Solution of the linearized phonon Boltzmann equation," *Phys. Rev.*, **148**, 766–778, 1966; "Thermal conductivity, second sound, and phonon hydrodynamic phenomena in nonmetallic crystals," *Phys. Rev.*, **148**, 778–788, 1966.
23. J. Shiomi and S. Maruyama, "Non-Fourier heat conduction in a single-walled carbon nanotube: Classical molecular dynamics simulations," *Phys. Rev. B*, **73**, 205420, 2006.
24. D. H. Tsai and R. A. MacDonald, "Molecular-dynamics study of second sound in a solid excited by a strong heat pulse," *Phys. Rev. B*, **14**, 4714–4723, 1976.
25. X. Wang and X. Xu, "Thermoelastic wave induced by pulsed laser heating," *Appl. Phys. A*, **73**, 107–114, 2001; X. Wang, "Thermal and thermomechanical phenomena in picosecond laser copper interaction," *J. Heat Transfer*, **126**, 355–364, 2004.
26. S. I. Anisimov, B. L. Kapeliovich, and T. L. Perel'man, "Electron emission from metal surfaces exposed to ultrashort laser pulses," *Sov. Phys. JETP*, **39**, 375–377, 1974.
27. T. Q. Qiu and C. L. Tien, "Short-pulse laser heating on metals," *Int. J. Heat Mass Transfer*, **35**, 719–726, 1992; T. Q. Qiu and C. L. Tien, "Size effect on nonequilibrium laser heating of metal films," *J. Heat Transfer*, **115**, 842–847, 1993; T. Q. Qiu, T. Juhasz, C. Suarez, W. E. Bron, and C. L. Tien, "Femtosecond laser heating of multi-layer metals—II. Experiments," *Int. J. Heat Mass Transfer*, **37**, 2799–2808, 1994.
28. J. L. Hostetler, A. N. Smith, D. M. Czajkowsky, and P. M. Norris, "Measurement of the electron-phonon coupling factor dependence on film thickness and grain size in Au, Cr, and Al," *Appl. Opt.*, **38**, 3614–3620, 1999.
29. S. Link, C. Burda, Z. L. Wang, and M. A. El-Sayed, "Electron dynamics in gold and gold-silver alloy nanoparticles: The influence of a nonequilibrium electron distribution and the size dependence of the electron-phonon relaxation," *J. Chem. Phys.*, **111**, 1255–1264, 1999.
30. A. N. Smith and P. M. Norris, "Influence of intraband transition on the electron thermoreflectance response of metals," *Appl. Phys. Lett.*, **78**, 1240–1242, 2001; R. J. Stevens, A. N. Smith, and P. M. Norris, "Measurement of thermal boundary conductance of a series of metal-dielectric interfaces by the transient thermoreflectance techniques," *J. Heat Transfer*, **127**, 315–322, 2005.
31. D. G. Cahill, K. Goodson, and A. Majumdar, "Thermometry and thermal transport in micro/nanoscale solid-state devices and structures," *J. Heat Transfer*, **124**, 223–241, 2002; D. G. Cahill, W. K. Ford, K. Goodson, et al., "Nanoscale thermal transport," *J. Appl. Phys.*, **93**, 793–818, 2003.
32. D. M. Riffe, X. Y. Wang, M. C. Downer, et al., "Femtosecond thermionic emission from metals in the space-charge-limited regime," *J. Opt. Soc. Am. B*, **10**, 1424–1435, 1993.
33. A. N. Smith, J. L. Hostetler, and P. M. Norris, "Nonequilibrium heating in metal films: An analytical and numerical analysis," *Numer. Heat Transfer A*, **35**, 859–874, 1999.
34. D. Y. Zhou and K. S. Chiu, "Temperature-dependent thermal lagging in ultrafast laser heating," *Int. J. Heat Mass Transfer*, **44**, 1725–1734, 2001.
35. L. Jiang and H.-L. Tsai, "Energy transport and nanostructuring of dielectrics by femtosecond laser pulse trains," *J. Heat Transfer*, **128**, 926–933, 2006.
36. T. Klitsner, J. E. VanCleve, H. E. Fischer, and R. O. Pohl, "Phonon radiative heat transfer and surface scattering," *Phys. Rev. B*, **38**, 7576–7594, 1988.
37. R. B. Peterson, "Direct simulation of phonon-mediated heat transfer in a Debye crystal," *J. Heat Transfer*, **116**, 815–822, 1994.
38. E. T. Swartz and P. O. Pohl, "Thermal boundary resistance," *Rev. Mod. Phys.*, **61**, 605–668, 1989.
39. W. A. Little, "The transport of heat between dissimilar solids at low temperatures," *Can. J. Phys.*, **37**, 334–349, 1959.
40. G. Chen and C. L. Tien, "Thermal conductivity of quantum well structures," *J. Thermophys. Heat Transfer*, **7**, 311–318, 1993; G. Chen, "Size and interface effects on thermal conductivity of



- superlattices and periodic thin-film structures,” *J. Heat Transfer*, **119**, 220–229, 1997; G. Chen, “Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices,” *Phys. Rev. B*, **57**, 14958–14973, 1998.
41. G. Chen and T. Zeng, “Nonequilibrium phonon and electron transport in heterostructures and superlattices,” *Microscale Thermophys. Eng.*, **5**, 71–88, 2001; T. Zeng and G. Chen, “Phonon heat conduction in thin films: Impacts of thermal boundary resistance and internal heat generation,” *J. Heat Transfer*, **123**, 340–347, 2001.
  42. S. Sinha and K. E. Goodson, “Review: Multiscale thermal modeling in nanoelectronics,” *Int. J. Multiscale Comp. Eng.*, **3**, 107–133, 2005.
  43. R. A. Escobar, S. S. Ghai, M. S. Jhon, and C. H. Amon, “Multi-length and time scale thermal transport using the lattice Boltzmann method with application to electronics cooling,” *Int. J. Heat Mass Transfer*, **49**, 97–107, 2006.
  44. E. M. Sparrow and R. D. Cess, *Radiation Heat Transfer*, Augmented ed., McGraw-Hill, New York, 1978.
  45. M. F. Modest, *Radiative Heat Transfer*, McGraw-Hill, New York, 1993.
  46. M. M. Yovanovich, “Four decades of research on thermal contact, gap, and joint resistance in microelectronics,” *IEEE Trans. Compon. Packag. Technol.*, **28**, 182–206, 2005.
  47. P. E. Phelan, “Application of diffuse mismatch theory to the prediction of thermal boundary resistance in thin-film high-Tc superconductors,” *J. Heat Transfer*, **120**, 37–43, 1998; L. De Bellis, P. E. Phelan, and R. S. Prasher, “Variations of acoustic and diffuse mismatch models in predicting thermal-boundary resistance,” *J. Thermophys. Heat Transfer*, **14**, 144–150, 2000.
  48. H. J. Lee, Private communication, which provided a detailed derivation of the correct expression of the specularity  $p$ . A sequence of typos were found in Ziman’s book, Ref. [23] in Chap. 5, leading to the erroneous expression of  $p = \exp(-16\pi^3\sigma_{\text{rms}}^2/\lambda^2)$ .
  49. S. Mazumdar and A. Majumdar, “Monte Carlo study of phonon transport in solid thin films including dispersion and polarization,” *J. Heat Transfer*, **123**, 749–759, 2001.
  50. C.-J. Twu and J.-R. Ho, “Molecular-dynamics study of energy flow and the Kapitza conductance across an interface with imperfection formed by two dielectric thin films,” *Phys. Rev. B*, **67**, 205422, 2003.
  51. S. R. Phillpot, P. K. Schelling, and P. Keblinski, “Interfacial thermal conductivity: Insights from atomic level simulation,” *J. Mater. Sci.*, **40**, 3143–3148, 2005.
  52. Y. Chen, D. Li, J. R. Lukes, Z. Ni, and M. Chen, “Minimum superlattice thermal resistivity from molecular dynamics,” *Phys. Rev. B*, **72**, 174302, 2005.
  53. H. Zhong and J. R. Lukes, “Interfacial thermal resistance between carbon nanotubes: Molecular dynamics simulations and analytical thermal modeling,” *Phys. Rev. B*, **74**, 125403, 2006.
  54. J. D. Chung, A. J. H. McGaughey, and M. Kaviany, “Role of phonon dispersion in lattice thermal conductivity,” *J. Heat Transfer*, **126**, 376–380, 2004.
  55. S. V. J. Narumanchi, J. Y. Murthy, and C. H. Amon, “Simulation of unsteady small heat source effects in sub-micron heat conduction,” *J. Heat Transfer*, **125**, 896–903, 2003; S. V. J. Narumanchi, J. Y. Murthy, and C. H. Amon, “Submicron heat transport model in silicon accounting for phonon dispersion and polarization,” *J. Heat Transfer*, **126**, 946–955, 2004.
  56. J. Lai and A. Majumdar, “Concurrent thermal and electrical modeling of sub-micrometer silicon devices,” *J. Appl. Phys.*, **79**, 7353–7361, 1996.
  57. P. G. Sverdrup, Y. S. Ju, and K. E. Goodson, “Sub-continuum simulation of heat conduction in silicon-on-insulator transistors,” *J. Heat Transfer*, **123**, 130–137, 2001.
  58. S. Sinha, E. Pop, R. W. Dutton, and K. E. Goodson, “Non-equilibrium phonon distribution in sub-100 nm silicon transistors,” *J. Heat Transfer*, **128**, 638–647, 2006.
  59. M. R. Abel, T. L. Wright, W. P. King, and S. Graham, “Thermal metrology of silicon microstructures using Raman spectroscopy,” *IEEE Trans. Comp. Pack. Technol.*, accepted 2007.
  60. J. Lee, T. Beechem, T. L. Wright, B. A. Nelson, S. Graham, and W. P. King, “Electrical, thermal, and mechanical characterization of silicon microcantilever heaters,” *J. Microelectromech. Syst.*, **15**, 1644, 2007; J. Lee, T. L. Wright, M. R. Abel, et al., “Thermal conduction from microcantilever heaters in partial vacuum,” *J. Appl. Phys.*, **101**, 014906, 2007.
  61. K. Park, J. Lee, Z. M. Zhang, and W. P. King, “Frequency-dependent electrical and thermal response of heated atomic force microscope cantilevers,” *J. Microelectromech. Syst.*, accepted 2007.

62. K. Park, A. Marchenkov, Z. M. Zhang, and W.P. King, "Low temperature characterization of heated microcantilevers," *J. Appl. Phys.*, accepted 2007.

## PROBLEMS

- 7.1.** What is the characteristic length for heat conduction along a thin film? Why is local equilibrium a good assumption in this case, even though the film thickness is less than the mean free path of the heat carriers? Why does the thermal conductivity depend on the thickness of the film?
- 7.2.** Why do we say that Fourier's law is a fundamental physical law, like Newton's laws in mechanics, but the Cattaneo equation is not? Comment on the paradox of infinite speed of heat diffusion by considering the feasibility of exciting the surface temperature or depositing a heat flux to the surface instantaneously.
- 7.3.** Consider a 1-D semi-infinite medium initially at uniform temperature  $T_i$ . The surface temperature is suddenly changed to a constant temperature,  $T(0,t) = T_s$ . The analytical solution of the heat diffusion equation gives

$$\theta(x,t) = \frac{T(x,t) - T_i}{T_s - T_i} = \operatorname{erfc}\left(\frac{x}{2\sqrt{\alpha t}}\right)$$

For silicon at various temperatures, use the properties given in Example 5-6 to estimate how long it will take for a given location to gain a temperature rise that is  $\theta = 10^{-12}$ , or one part per trillion of the maximum temperature difference. Estimate the average thermal diffusion speed in terms of  $x$  and  $T_i$ . [Hint:  $\operatorname{erfc}(5.042) = 1.00 \times 10^{-12}$ .]

- 7.4.** Repeat Problem 7.3, using copper instead of silicon as the material, based on the properties given in Example 5-5. Discuss why the average thermal diffusion speed is different under different boundary conditions, i.e., constant heat flux and constant temperature. From an engineering point of view, do you think heat diffusion is a fast or slow process? Why?
- 7.5.** (a) Derive Eq. (7.4), the hyperbolic heat equation from the Cattaneo equation.  
(b) Derive Eq. (7.14), the lagging heat equation, based on the dual-phase-lag model.
- 7.6.** Take GaAs as an example. How would you compare the speed of sound with the average thermal diffusion speed at different temperatures and length scales? This problem requires some literature search on the properties.
- 7.7.** Assume the hyperbolic heat equation would work for transient heat transfer in glass (Pyrex), at near room temperature. Given  $\kappa = 1.4 \text{ W/(m} \cdot \text{K)}$ ,  $\rho = 2500 \text{ kg/m}^3$ ,  $c_p = 835 \text{ J/(kg} \cdot \text{K)}$ , and  $v_a = 5640 \text{ m/s}$ .  
(a) At what speed would the temperature wave propagate?  
(b) For an excimer laser with a pulse width  $t_p = 10 \text{ ns}$ , 0.1 ns after the pulse starts, could the hyperbolic equation be approximated by the parabolic equation?  
(c) Suppose we have an instrument available to probe the timescale below  $\tau_q$ , will the hyperbolic heat equation be able to describe the observation?
- 7.8.** Derive Eq. (7.13b) from Eq. (7.13a). Discuss the conditions for these equations to be reduced to Fourier's law or the Cattaneo equation.
- 7.9.** Show that Eq. (7.17) satisfies Eq. (7.16). Discuss the conditions for Eq. (7.17) to represent Fourier's law or the Cattaneo equation.
- 7.10.** Derive Eq. (7.18a), Eq. (7.18b), and Eq. (7.18c).
- 7.11.** Derive Eq. (7.27a) and Eq. (7.27b). Calculate  $\tau$ ,  $\tau_q$ , and  $\tau_T$  of copper, for  $T_e = 300, 1000,$  and  $5000 \text{ K}$ , assuming the lattice temperature  $T_s = 300 \text{ K}$ .
- 7.12.** Calculate the electron-phonon coupling constant  $G$  for aluminum, copper, gold, and silver near room temperature. Discuss the dependence of  $\kappa$  and  $G$  upon the electron and lattice temperatures  $T_e$  and  $T_s$ .
- 7.13.** At  $T_e = 1000, 3000,$  and  $6000 \text{ K}$ , estimate the energy transfer by thermionic emission from a copper surface, assuming that the electrons obey the equilibrium distribution function at  $T_e$ .
- 7.14.** Based on Example 7-3, evaluate the heat flux in a thin silicon film. How thin must it be in order for it to be considered as in the radiative thin limit? Calculate the medium temperature  $T$ . Plot the left-hand

side and the right-hand side of Eq. (7.43). Furthermore, assuming Eq. (7.43) to be true for each frequency, find a frequency-dependent temperature  $T(\omega)$  of the medium. At what frequency does  $T(\omega) = T$ ? Is there any physical significance of  $T(\omega)$ ?

**7.15.** Derive Eq. (7.53), using Eq. (7.38), Eq. (7.49), and Eq. (7.50).

**7.16.** In principle, one should be able to study nonequilibrium electrical and thermal conduction in the direction perpendicular to the plane, and use the BTE to determine the effective conductivities. This could be a team project, in which a few students work together to formulate the necessary equations. As an individual assignment, describe how to set up the boundary conditions, as well as the steps you plan to follow, without actually deriving the equations.

**7.17.** For a diamond type IIa film,  $v_l = 17,500$  m/s,  $v_t = 12,800$  m/s, and  $\kappa = 3300$  W/(m · K), near 300 K. Assume that the boundaries can be modeled as blackbodies for phonons. For boundary temperatures  $T_1 = 350$  K and  $T_2 = 250$  K, calculate and plot the heat flux  $q_x''$  and the effective thermal conductivity  $\kappa_{\text{eff}}$  across a film of thickness  $L$ , which varies from 0.05 to 50  $\mu\text{m}$ .

**7.18.** Calculate the TBR between high-temperature superconductor  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$  and MgO substrate, at an average temperature between 10 and 90 K, using both the AMM and the DMM without considering the electronic effect. The following parameters are given for  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ :  $v_l = 4780$  m/s,  $v_t = 3010$  m/s,  $\rho = 6338$  kg/m<sup>3</sup>, and  $\Theta_D = 450$  K; and for MgO:  $v_l = 9710$  m/s,  $v_t = 6050$  m/s,  $\rho = 3576$  kg/m<sup>3</sup>, and  $\Theta_D = 950$  K.

**7.19.** Evaluate the effective thermal conductivity near room temperature of a GaAs/AlAs superlattice, with a total thickness of 800 nm, using the DMM to compute the transmission coefficient. Assume the end surfaces are blackbodies to phonons; consider that (a) each layer is 4 nm thick and (b) each layer is 40 nm thick. The following parameters are given, considering phonon dispersion on thermal conductivity, for GaAs:  $C = 880$  kJ/(m<sup>3</sup> · K),  $v_g = 1024$  m/s, and  $\Lambda = 145$  nm; and for AlAs:  $C = 880$  kJ/(m<sup>3</sup> · K),  $v_g = 1246$  m/s, and  $\Lambda = 236$  nm. How is the result compared with a single layer of either GaAs or AlAs?

**7.20.** Evaluate the effective thermal conductivity near room temperature of a Si/Ge superlattice, with a total thickness of 1000 nm, using the DMM to compute the transmission coefficient. Assume the end surfaces are blackbodies to phonons; consider that (a) each layer is 5 nm thick and (b) each layer is 50 nm thick. The following parameters are given, considering phonon dispersion on thermal conductivity, for Si:  $C = 930$  kJ/(m<sup>3</sup> · K),  $v_g = 1804$  m/s, and  $\Lambda = 260$  nm; and for Ge:  $C = 870$  kJ/(m<sup>3</sup> · K),  $v_g = 1042$  m/s, and  $\Lambda = 199$  nm. How is the result compared with a single layer of either Si or Ge?

---

## CHAPTER 8

---

# FUNDAMENTALS OF THERMAL RADIATION

---

Radiation is one of the fundamental modes of heat transfer. However, the concepts of thermal radiation are much more complicated and, hence, very difficult to perceive. The main features of radiation that are distinct from conduction and convection are as follows: (a) Radiation can transfer energy with and without an intervening medium; (b) The radiant heat flux is not proportional to the temperature gradient; (c) Radiation emission is wavelength dependent, and the radiative properties of materials depend on the wavelength and the temperature; and (d) The radiant energy exchange and the radiative properties depend on the direction and orientation.<sup>1,2</sup>

The dual theory explains the nature of radiation as either electromagnetic waves or a collection of particles called photons. Although radiation can travel in vacuum, it originates from matter. All forms of matter emit radiation through complicated mechanisms (e.g., molecular vibration in gases, and electron and lattice vibrations in solids). In most solids and some liquids, radiation emitted from the interior is strongly absorbed by adjoining molecules. Therefore, radiation from or to these materials is often treated as *surface phenomena*, while radiation in gases and some semitransparent solids or liquids has to be treated as *volumetric phenomena*. Nevertheless, one must treat solids or liquids as a medium (i.e., volumetrically) to understand the mechanisms of reflection and emission, to predict the radiative properties of thin films and small particles, and to calculate radiation heat transfer between objects placed in close vicinity. *Thermal radiation* refers to a type of radiation where the emission is directly related to the temperature of the body (or surface).

There are numerous engineering applications where radiation heat transfer is important, such as furnaces, combustion, high-temperature materials processing and manufacturing, solar energy, space cooling and insulation, and cryogenic systems. Even at room temperature, radiative heat transfer may be of the same order of magnitude as convective heat transfer. The study of thermal radiation went along with the study of light phenomena and led to some major breakthroughs in modern physics. It is instructive to give a brief survey of major historical developments related to thermal radiation.

Quantitative understanding of the nature of light began in the seventeenth century with the discoveries of Snell's law of refraction, Fermat's least-time principle of light path, Huygens' principle of constructing the wavefront from secondary waves, and Newton's prism that helped him prove white light consists of many different types of rays. In the dawn of the nineteenth century, Sir Frederick Herschel (1738–1822), a German-born English astronomer, discovered infrared radiation.<sup>3</sup> His original objective was to find a suitable color for a glass filter, which could transmit most of light but the least amount of heat, for use in solar observations. By moving a thermometer along the spectrum of solar radiation that passed through a prism, Herschel accidentally found that the temperature of the thermometer would rise even though it was placed beyond the red end of the visible light. He published several papers in *Philosophical Transactions of the Royal Society of London* in

1800 and called the unknown radiation *invisible light* or *heat-making rays*. Young's double-slit experiment in 1801 demonstrated the interference phenomenon and the wave nature of light, followed by intensive studies on polarization and reflection phenomena, led by French physicist Augustin-Jean Fresnel (1788–1827) who contributed significantly to the establishment of the wave theory of light. In 1803, radiation beyond the violet end of the visible spectrum via chemical effects was also discovered. The ultraviolet, visible, and infrared spectra were thus associated with chemical, luminous, and heating effects, respectively. Yet, the common nature of the different types of radiation was not known until the late nineteenth century.

One of the obstacles of accurately measuring infrared radiation (or heat radiation, as it was called in those days) was the lack of sensitive detectors. In the earlier years, measurements were performed using thermometers with blackened bulbs. In 1829, Italian physicists Leopoldo Nobili (1784–1835) and Macedonio Melloni (1798–1854) invented the thermopile, which is made by connecting a number of thermocouples in series, that is much more sensitive and faster than the thermometer. Melloni used the device to study the infrared radiation from hot objects and the sun. Gustav Kirchhoff (1824–1887), a German physicist, contributed greatly to the fundamental understanding of spectroscopy and the thermal emission by heated objects. In 1862, he coined the term “black body” radiation and established Kirchhoff's law, which states that the emissivity of a surface equals its absorptivity at thermal equilibrium.

Many famous physicists and mathematicians have contributed to electromagnetism. The complete equations of electromagnetic waves were established in 1873 by Scottish physicist James Clerk Maxwell (1831–1879), and later confirmed experimentally by German physicist Heinrich Hertz (1857–1894), who discovered radio waves due to electrical vibrations. Before the existence of electrons had been proven, Dutch physicist Hendrik Lorentz (1853–1928) proposed that light waves were due to oscillations of an electric charge in the atom. He received the Nobel Prize in Physics in 1902, for his mathematical theory relating electron wave motion and light. The 1902 Nobel Prize was shared with his student Pieter Zeeman (1865–1943) for the experimental study about the effect of magnetic fields on atomic structures that has resulted in the splitting of spectral lines of the produced light. The electromagnetic wave theory has played a central role in radio, radar, television, microwave technology, telecommunication, thermal radiation, and physical optics. Albert Einstein arrived at the famous formula  $E = mc^2$  in 1905, after connecting the relativity principle with the Maxwell equations.

In 1881, Samuel Langley (1834–1906), the American astronomer, physicist, and aeronautics pioneer, invented a highly sensitive device called *bolometer* for detection of thermal radiation. The bolometer used two platinum strips, connected in a Wheatstone bridge circuit with a sensitive galvanometer, to read the imbalance of the bridge caused by the exposure of one of the strips to radiation. Langley was the first to make an accurate map of the solar spectrum up to a wavelength of 2.8  $\mu\text{m}$ . The Stefan-Boltzmann law of blackbody radiation is a result of the empirical relation obtained by Slovenian physicists Joseph Stefan (1835–1893) in 1879, based on observation of experiments, and the theoretical proof given by Austrian physicist Ludwig Boltzmann (1844–1906) in 1884, based on thermodynamic relations of a Carnot cycle with radiation as a working fluid using the concept of radiation pressure. In the late nineteenth century, German physicist Wilhelm Wien (1864–1928) derived the displacement law in 1893 by considering a piston moving within a mirrored empty cylinder filled with thermal radiation. Wien also derived a spectral distribution of blackbody radiation, called Wien's formula, which is applicable to the short-wavelength region of the blackbody spectrum but deviates toward long wavelengths. Wien received the Nobel Prize in 1911 “for his discoveries regarding the laws governing the radiation of heat.” In 1900, Lord Rayleigh (1842–1919), British physicist and Nobel Laureate in Physics in 1904, used the equipartition theorem to show that the blackbody emission should be directly proportional to temperature but inversely proportional to the fourth power of wavelength. Sir James Jeans (1877–1946), British physicist, astronomer, and mathematician,

derived a more complete expression in 1905. The Rayleigh-Jeans formula agreed with experiments at sufficiently high temperatures and long wavelengths, where Wien's formula failed, but disagreed with experiments at short wavelengths. It is noteworthy that Rayleigh made great contributions to light scattering and wave phenomena, such as the discovery of Rayleigh scattering by small objects that explains why the sky is blue and the sunset appears orange glow. Rayleigh also predicted the existence of *surface waves*, sometimes called *Rayleigh waves*, which propagate along the interface between two different media. The amplitude of the wave, however, reduced in each media as the distance from the interface increases.

In an effort to obtain a better agreement with measurements at long wavelengths, German physicist Max Planck (1858–1947) in 1900 used the maximum entropy principle, based on Boltzmann's entropy expression, to derive an equation, known as Planck's law, which agrees with experiments in the whole spectral region. Planck obtained his expression independently of Rayleigh's work published several months earlier, while the complete derivation of Rayleigh-Jeans formula was obtained several years later. In his book *The Theory of Heat Radiation*, Planck showed that his formula would reduce to Wien's formula at small  $\lambda T$  and Rayleigh-Jeans formula at very large  $\lambda T$ .<sup>4</sup> In his derivation, Planck used a bold assumption that is controversial to classical electrodynamics. His hypothesis was that energy is not infinitely divisible but must assume discrete values, which are proportional to the frequency. This concept would have been easily accepted for a system consisting of particles, like atoms or gas molecules, but not for oscillators that radiate electromagnetic energy. Planck's work opened the door to quantum mechanics. The idea of quantization of radiation was further developed by Einstein, who applied it to explain the photoelectric effect in 1905. Planck was awarded the Nobel Prize in Physics in 1918 for the discovery of energy quanta. In 1924, Indian mathematical physicist Satyendra Nath Bose (1894–1974) modified the Boltzmann statistics of ideal molecular gases, by treating photons as indistinguishable particles in order to derive Planck's distribution function. With the help of Einstein, Bose's work was published in *Zeitschrift für Physik* in 1924. Einstein further extended Bose's theory to atoms and predicted the existence of a phenomenon, known as Bose-Einstein condensate, as discussed in Chap. 3. It is clear that the path of quest for the truth in understanding thermal radiation has led to important discoveries in modern physics.

This chapter contains an introduction to the electromagnetic wave theory, blackbody radiation, plane wave reflection and refraction at the boundary between two semi-infinite media, and various models used to study the optical properties of different materials. The materials covered in the following sections are intended to provide a detailed background for more in-depth discussion on the applications to micro/nanosystems in subsequent chapters.

## 8.1 ELECTROMAGNETIC WAVES

---

### 8.1.1 Maxwell's Equations

The propagation of electromagnetic waves in any media is governed by a set of equations, first stated together by Maxwell. The macroscopic Maxwell equations can be written in the differential forms as follows:<sup>5-7</sup>

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (8.1)$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \quad (8.2)$$

$$\nabla \cdot \mathbf{D} = \rho_e \quad (8.3)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (8.4)$$

Based on the SI units,  $\mathbf{E}$  in V/m is the electric field,  $\mathbf{H}$  in A/m is the magnetic field,  $\mathbf{J}$  in A/m<sup>2</sup> is the electric current density (i.e., electric charge flux),  $\mathbf{D}$  in C/m<sup>2</sup> is the electric displacement,  $\mathbf{B}$  in Wb/m<sup>2</sup> is the magnetic flux density (also called magnetic induction), and  $\rho_e$  in C/m<sup>3</sup> is the charge density. Note that in magnetism, 1 tesla (T) = 1 Wb/m<sup>2</sup>, and 1 weber (Wb) = 1 V · s. The charge conservation or continuity equation,  $\nabla \cdot \mathbf{J} + \partial\rho_e/\partial t = 0$ , is implicitly included in the Maxwell equations, because it can be obtained by taking the divergence of Eq. (8.2) and then applying Eq. (8.3). The constitutive relations for a linear isotropic medium are

$$\mathbf{D} = \epsilon_m \mathbf{E} \quad (8.5)$$

$$\mathbf{B} = \mu_m \mathbf{H} \quad (8.6)$$

where  $\epsilon_m$  in F/m is the electric permittivity and  $\mu_m$  in N/A<sup>2</sup> is the magnetic permeability of the medium. Note that farad (F) is the SI unit of capacitance: 1 F = 1 C/V. The permittivity and permeability values of free space (vacuum) are  $\epsilon_0 = 8.854 \times 10^{-12}$  F/m and  $\mu_0 = 4\pi \times 10^{-7}$  N/A<sup>2</sup>, respectively. For anisotropic media,  $\epsilon_m$  and  $\mu_m$  are dyadic tensors. The microscopic form of Ohm's law gives

$$\mathbf{J} = \sigma \mathbf{E} \quad (8.7)$$

where  $\sigma$  in A/(V · m) is the electric conductivity.

A brief discussion on the physical interpretation of Maxwell's equations is given next. Equation (8.1) is an expression of Faraday's law of induction, which states that a time varying magnetic field produces an electric field in a coil. In other words, through any closed electric field line, there is a time varying magnetic field. Combining Eq. (8.1) with Green's theorem, Eq. (B.71), we see that the integral of the electric field around a closed loop is equal to the negative of the integral of the time derivative of the magnetic induction, over the area enclosed by the loop. Equation (8.2) is the general Ampere law, which includes Maxwell's displacement current ( $\partial\mathbf{D}/\partial t$ ). It states that through any closed magnetic field line, there is an electric current density  $\mathbf{J}$  or a displacement current or both. Conversely, circulating magnetic fields are produced by passing an electrical current through a conductor or changing electric fields or both. Equation (8.3) is Gauss's law, which implies that the electric field diverges from electric charges. Using Gauss's theorem, Eq. (B.70), it can be seen from Eq. (8.3) that the integral of the electric field over a closed surface is proportional to the electric charges enclosed by that surface. If there are no electric charges inside a closed surface, there is no net electric field penetrating the surface. Equation (8.4) is an analogy to Gauss's law for magnetic field. However, since there exist no isolated magnetic poles, called magnetic monopoles, the integration of magnetic field over any closed surface is zero.

The interpretations given in the preceding paragraph are straightforward since all variables and coefficients are considered as real quantities. However, Maxwell's equations are mostly useful when all quantities are expressed in complex variables. The material properties, such as  $\epsilon_m$  and  $\mu_m$ , are generally complex and frequency dependent. To facilitate the understanding, we will start with simple cases first and then generalize the theory for more realistic problems.

### 8.1.2 The Wave Equation

Sometimes called free charge density,  $\rho_e$  in Eq. (8.3) should be treated as excess charges or net charges per unit volume. Because the number of electrons equals the number of protons in the nuclei, in most media, we can assume  $\rho_e = 0$ . For a nonconductive material,  $\sigma = 0$ . We further assume that  $\epsilon_m$  and  $\mu_m$  are both real and independent of position, time, and the field strength. This is true for a nondissipative (lossless), homogeneous, and linear material. If  $\mu_m = \mu_0$ , the material is said to be nonmagnetic. Therefore, a nonconductive and

nonmagnetic material is a dielectric for which only  $\epsilon_m$  is needed to characterize its electromagnetic behavior. Materials with both  $\epsilon_m$  and  $\mu_m$  being real but  $\mu_m \neq \mu_0$  are sometimes called general dielectrics or dielectric-magnetic media. Substituting the constitutive relations into Maxwell's equations and then combining Eq. (8.1) and Eq. (8.2), we obtain

$$\nabla^2 \mathbf{E} = \mu_m \epsilon_m \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (8.8)$$

where the vector identity given in Eq. (B.64),  $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E}$ , has been employed. Equation (8.8) is the *wave equation*, which can also be written in terms of the magnetic field. The wave equation has infinite number of solutions (see Problem 8.1). The solution of Eq. (8.8) for a monochromatic plane wave can be written as

$$\mathbf{E} = \mathbf{E}_0 e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})} \quad (8.9)$$

where  $\mathbf{E}_0$  is the amplitude vector,  $\omega$  is the angular frequency,  $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$  is the position vector, and  $\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}}$  is the *wavevector*, which points toward the direction of propagation. In order for Eq. (8.9) to be a solution of Eq. (8.8), the magnitude of  $\mathbf{k}$  must be  $k = \omega\sqrt{\mu_m \epsilon_m}$ . The complex form of the electric field is used in Eq. (8.9) to facilitate mathematical manipulation. The actual electric field may be expressed as the real part of Eq. (8.9), viz.,

$$\text{Re}(\mathbf{E}) = \text{Re}(\mathbf{E}_0)\cos\phi + \text{Im}(\mathbf{E}_0)\sin\phi \quad (8.10)$$

where Re or Im stands for taking the real part or the imaginary part, and  $\phi = \omega t - \mathbf{k} \cdot \mathbf{r}$  is the phase. Equation (8.9) is a time-harmonic solution at a fixed frequency. Because any time-space-dependent function can be expressed as a Fourier series of many frequency components, we can integrate Eq. (8.9) over all frequencies to obtain the total electric field at any time and position. Therefore, understanding the nature of Eq. (8.9) is very important to the study of electromagnetic wave phenomena.

When Eq. (8.9) is substituted into Maxwell's equations, a time derivative  $\partial/\partial t$  can be replaced by a multiplication of  $-i\omega$  and the operator  $\nabla$  can be replaced by  $i\mathbf{k}$ . Hence, the first two Maxwell equations can be written as

$$\mathbf{k} \times \mathbf{E} = \omega\mu_m \mathbf{H} \quad (8.11a)$$

and

$$\mathbf{k} \times \mathbf{H} = -\omega\epsilon_m \mathbf{E} \quad (8.11b)$$

The two equations suggest that  $\mathbf{E}$ ,  $\mathbf{H}$ , and  $\mathbf{k}$  are orthogonal and form a right-handed triplet, when both  $\epsilon_m$  and  $\mu_m$  are positive. On the surface normal to the wavevector  $\mathbf{k}$ , the electric or magnetic field is a function of time only, because  $\mathbf{k} \cdot \mathbf{r} = \text{const}$ . This surface is called a *wavefront*. In the  $\mathbf{k}$  direction, the wavefront travels at the speed given by

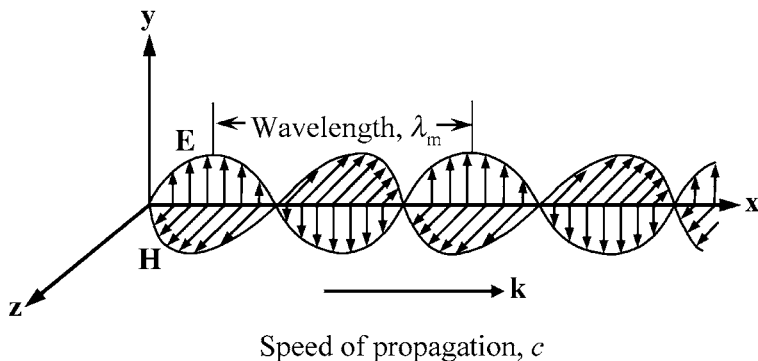
$$c = \frac{\omega}{k} = \frac{1}{\sqrt{\mu_m \epsilon_m}} \quad (8.12)$$

which is called *phase speed*, and it is the smallest speed at which the phase of the wave propagates.<sup>8</sup> The phase velocity is the phase speed times the unit wavevector.

Figure 8.1 illustrates a plane wave, propagating in the positive  $x$  direction, whose electric field is parallel to the  $y$  direction and magnetic field parallel to the  $z$  direction. In such cases,  $k = k_x$  and  $\mathbf{k} \cdot \mathbf{r} = kx$ . The wavefront is perpendicular to the  $x$  direction. It can be seen clearly that the wavevector is related to the wavelength  $\lambda_m$  in the medium by  $k = 2\pi/\lambda_m$ .

In free space, the speed of electromagnetic wave is given by  $c_0 = 1/\sqrt{\mu_0 \epsilon_0}$ . The speed of light in vacuum was instated as an exact number,  $c_0 = 299,792,458$  m/s, by the General Conference on Weights and Measures (abbreviated as CGPM for Conférence Générale des





**FIGURE 8.1** Illustration of a linearly polarized electromagnetic wave.

Poids et Mesures) in 1983. The SI base unit meter has since been defined as the distance that light travels in vacuum during a time interval of  $1/299,792,458$  s. The NIST reference on constant, units, and uncertainty can be found on the web page: <http://physics.nist.gov/cuu/index.html>, which contains detailed discussions about the fundamental physical constants and the base SI units. For most calculations, it suffices to use  $c_0 = 2.998 \times 10^8$  m/s. The refractive index of the medium is given as  $n = \sqrt{\mu_m \epsilon_m / \mu_0 \epsilon_0} = c_0 / c$ . Therefore,  $c = c_0 / n$  and  $\lambda_m = \lambda / n$ , where  $\lambda$  is the wavelength in vacuum. For nonmagnetic materials  $\mu_m / \mu_0 = 1$ ; thus,  $n = \sqrt{\epsilon_m / \epsilon_0}$ .

Notice that  $n$  of a medium is a function of frequency (or wavelength) and is in general temperature dependent. For polychromatic light, the phase speed usually depends on wavelength because  $n = n(\lambda)$  in a dispersive medium. In vacuum, the energy propagation velocity is the same as the phase velocity. For polychromatic waves in a dispersive medium, the group velocity  $\mathbf{v}_g$  determines the direction and speed of energy flow and is defined as

$$\mathbf{v}_g = \nabla_k \omega = \frac{d\omega}{dk} = \frac{\partial \omega}{\partial k_x} \hat{\mathbf{x}} + \frac{\partial \omega}{\partial k_y} \hat{\mathbf{y}} + \frac{\partial \omega}{\partial k_z} \hat{\mathbf{z}} \quad (8.13)$$

which is the gradient of  $\omega$  in the  $k$ -space. In a homogeneous and isotropic medium,  $v_g = c_0 / (n + \omega dn/d\omega)$  and the direction of the group velocity will be the same as that of the wavevector  $\mathbf{k}$ . In a nondispersive medium, where  $n$  is not a function of frequency, it is clear that  $v_g = c = c_0 / n$ . When light is refracted from a nondispersive medium to a dispersive medium, the group velocity can have a component parallel to the group fronts, and hence, the energy flow is not necessarily perpendicular to the group fronts.<sup>8</sup> Notice that the wave equation is also applicable to other types of waves such as acoustic waves, which are matter waves with a longitudinal and two transverse modes, as mentioned in Chap. 5.

### 8.1.3 Polarization

A simple transverse wave will oscillate perpendicular to the wavevector. Because electromagnetic waves have two field vectors that can change their directions during propagation, the polarization behavior may be complicated. It is important to understand the nature of polarization in order to fully characterize an electromagnetic wave. There are two equivalent ways to interpret a complex vector  $\mathbf{A}$ . The first method considers it as a vector whose components are complex, i.e.,

$$\mathbf{A} = A_x \hat{\mathbf{x}} + A_y \hat{\mathbf{y}} + A_z \hat{\mathbf{z}} \quad (8.14)$$

where  $A_x, A_y,$  and  $A_z$  are complex numbers:

$$A_x = A'_x + iA''_x, \quad A_y = A'_y + iA''_y, \quad \text{and} \quad A_z = A'_z + iA''_z \quad (8.14a)$$

The second method decomposes it into two real vectors such that

$$\mathbf{A} = \mathbf{A}' + i\mathbf{A}'' \quad (8.15)$$

where  $\mathbf{A}'$  and  $\mathbf{A}''$  are the real and imaginary parts of the complex vector, given by

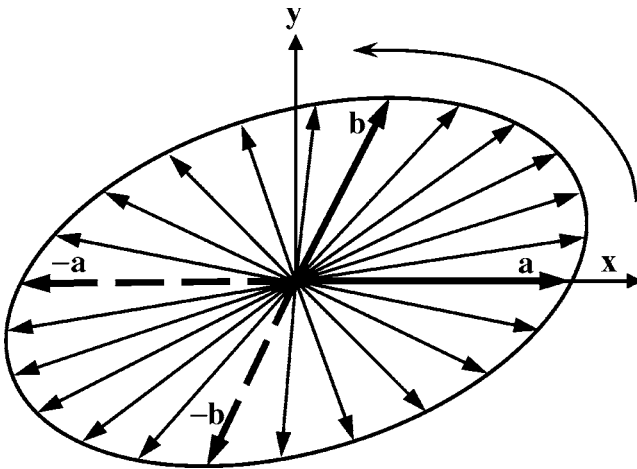
$$\mathbf{A}' = A'_x\hat{x} + A'_y\hat{y} + A'_z\hat{z} \quad \text{and} \quad \mathbf{A}'' = A''_x\hat{x} + A''_y\hat{y} + A''_z\hat{z} \quad (8.15a)$$

In either case, a complex vector has six real scalar terms.

For the time being, let us assume all the material properties to have real values and  $\mathbf{k}$  to be a real vector. Both  $\mathbf{E}$  and  $\mathbf{H}$  are complex, according to Eq. (8.9). To ensure that  $\mathbf{k} \cdot \mathbf{E} = 0$  at any time and location, both  $\text{Re}(\mathbf{E}_0)$  and  $\text{Im}(\mathbf{E}_0)$  must be perpendicular to  $\mathbf{k}$ . The same is true for the magnetic vector. Because  $\mathbf{H}$  can be obtained from Eq. (8.11a), the state of polarization can be based on how the electric field varies in time and along the  $\mathbf{k}$  direction in space. In order to study the time dependence of the electric field, rewrite Eq. (8.10) as

$$\text{Re}(\mathbf{E}) = \mathbf{a} \cos(\omega t) + \mathbf{b} \sin(\omega t) \quad (8.16)$$

where  $\mathbf{a} = \text{Re}(\mathbf{E}_0 e^{-i\mathbf{k}\cdot\mathbf{r}})$  and  $\mathbf{b} = \text{Im}(\mathbf{E}_0 e^{-i\mathbf{k}\cdot\mathbf{r}})$  are both real vectors and perpendicular to  $\mathbf{k}$ . In general, the electric field will vary with time in an ellipse, called the *vibration ellipse*, as shown in Fig. 8.2. If  $\mathbf{a}$  and  $\mathbf{b}$  are parallel or, equivalently,  $\text{Re}(\mathbf{E}_0)$  and  $\text{Im}(\mathbf{E}_0)$  are parallel to



**FIGURE 8.2** Illustration of polarization by the vibration ellipse, for a plane wave propagating in the positive  $z$  direction (out of the paper). The electric field vector is plotted at an increment of  $\omega\Delta t = \pi/12$ .

each other, then the electric field will not change its directions. The wave is said to be *linearly polarized*, and either  $\mathbf{a}$  or  $\mathbf{b}$  specifies the direction of polarization. An example of a linearly polarized wave is the wave shown in Fig. 8.1. When  $\mathbf{a} \perp \mathbf{b}$  and  $|\mathbf{a}| = |\mathbf{b}|$ , the vibration ellipse is a circle and the wave is said to be *circularly polarized*. In general, a monochromatic wave

described by Eq. (8.10) is *elliptically polarized*. For circularly or elliptically polarized light, if  $\mathbf{a} \times \mathbf{b}$  is in the same direction as  $\mathbf{k}$ , the vibration ellipse will rotate counterclockwise (left-handed), as viewed toward the light source; and if  $\mathbf{a} \times \mathbf{b}$  is opposite to the direction of propagation, the vibration ellipse will rotate clockwise (right-handed).<sup>6,7</sup> Similarly, one can consider the polarization of the electric field at a fixed time, and observe the vibration ellipse along the direction of propagation as an exercise (see Problem 8.2).

Because of the random nature of thermal radiation, the Fourier component does not vary with time exactly following  $e^{-i\omega t}$  but with some fluctuations in the amplitude. The polarization may become completely random, which is said to be *unpolarized*, *randomly polarized*, or *completely uncorrelated*. In any case, the electric field can be decomposed into the two orthogonal directions on the vibration ellipse. This is particularly useful for calculating energy transfer. A complete description of polarization is based on Stokes parameters, which are important in the study of light scattering and will be discussed in Chap. 9.

### 8.1.4 Energy Flux and Density

The energy conservation for electromagnetic field can be obtained from Maxwell's equations, according to English physicist John Poynting (1852–1914). To derive Poynting's theorem, one can dot multiply Eq. (8.1) and Eq. (8.2) by  $-\mathbf{H}$  and  $\mathbf{E}$ , respectively, and then add up each side. Using the vector identity in Eq. (B.63), we have  $\nabla \cdot (\mathbf{E} \times \mathbf{H}) = (\nabla \times \mathbf{E}) \cdot \mathbf{H} - (\nabla \times \mathbf{H}) \cdot \mathbf{E}$ . After simplifications, we obtain

$$-\nabla \cdot (\mathbf{E} \times \mathbf{H}) = \frac{\partial}{\partial t} \left( \frac{1}{2} \epsilon_m \mathbf{E} \cdot \mathbf{E} + \frac{1}{2} \mu_m \mathbf{H} \cdot \mathbf{H} \right) + \mathbf{E} \cdot \mathbf{J} \quad (8.17)$$

The left-hand term represents the energy flow into a differential control volume, the first term on the right is the rate of change of the stored energy (associated with the electric and magnetic fields), and the last term is the dissipated electromagnetic work or Joule heating. The *Poynting vector* is defined as

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (8.18a)$$

The Poynting vector is essentially the energy flux, which gives both the direction and the rate of energy flow per unit projected surface area. Equation (8.17) and Eq. (8.18a) can be easily extended to the complex field notation. Although it is easy to write the Poynting vector (which is always real) as  $\mathbf{S} = \text{Re}(\mathbf{E}) \times \text{Re}(\mathbf{H})$ , it is not very helpful because one would have to evaluate the real parts of  $\mathbf{E}$  and  $\mathbf{H}$  individually. Besides, the frequency of oscillation is usually too high to be measured. For harmonic fields, the time-averaged Poynting vector can be expressed as

$$\langle \mathbf{S} \rangle = \frac{1}{2} \text{Re}(\mathbf{E} \times \mathbf{H}^*) \quad (8.18b)$$

where \* signifies the complex conjugate. Similarly, the time-averaged energy density for time-harmonic fields can be expressed as<sup>5</sup>

$$\langle u \rangle = \frac{1}{4} \epsilon_m \mathbf{E} \cdot \mathbf{E}^* + \frac{1}{4} \mu_m \mathbf{H} \cdot \mathbf{H}^* \quad (8.19)$$

For an absorbing or dissipative medium, a more complete description of the energy density can be found in Cui and Kong (*Phys. Rev. B*, **70**, 205106, 2004).

**Example 8-1.** Prove that Eq. (8.18b) is the time-averaged Poynting vector for time-harmonic fields.

**Solution.** Let  $\mathbf{E} = \mathbf{E}(\mathbf{r})e^{-i\omega t}$  and  $\mathbf{H} = \mathbf{H}(\mathbf{r})e^{-i\omega t}$ , where  $\mathbf{E}(\mathbf{r})$  and  $\mathbf{H}(\mathbf{r})$  are complex vectors. Integrating the Poynting vector over a period  $T$ , we have

$$\begin{aligned} \langle \mathbf{S} \rangle &= \frac{1}{T} \int_T \text{Re}(\mathbf{E}) \times \text{Re}(\mathbf{H}) dt \\ &= \frac{1}{4T} \int_T \left[ \mathbf{E}(\mathbf{r})e^{-i\omega t} + \mathbf{E}^*(\mathbf{r})e^{i\omega t} \right] \times \left[ \mathbf{H}(\mathbf{r})e^{-i\omega t} + \mathbf{H}^*(\mathbf{r})e^{i\omega t} \right] dt \\ &= \frac{1}{4} (\mathbf{E} \times \mathbf{H}^* + \mathbf{E}^* \times \mathbf{H}) = \frac{1}{2} \text{Re}(\mathbf{E} \times \mathbf{H}^*) \end{aligned}$$

### 8.1.5 Dielectric Function

The conductivity is large at low frequencies for metals, due to free electrons. Even for good conductors, however, the electrons are not completely free but will be scattered by defects and phonons. At high frequencies, the current density  $\mathbf{J}$  and the electric field  $\mathbf{E}$  are not in phase anymore, suggesting that the conductivity should be a complex number. For insulators such as crystalline or amorphous dielectrics, electromagnetic waves can interact with bound electrons or lattice vibrations to transfer energy to the medium. At optical frequencies, the distinction between a conductor and an insulator becomes ambiguous unless the optical response over a large frequency region is considered. For example, a dielectric material can be highly reflective at a certain frequency region in the mid-infrared. On the other hand, a good conductor will be highly reflective in a much broader wavelength region from the near-infrared to the microwave. Let us first take the conductivity and the permittivity to be real, for a nonmagnetic material. The wave equation for  $\sigma \neq 0$  and  $\mu_m = \mu_0$  has the following form:

$$\nabla^2 \mathbf{E} = \mu_0 \sigma \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \epsilon_m \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (8.20)$$

Suppose Eq. (8.9) is a solution of this equation. We can substitute  $\partial \mathbf{E} / \partial t = -i\omega \mathbf{E}$ ,  $\partial^2 \mathbf{E} / \partial t^2 = -\omega^2 \mathbf{E}$ , and  $\nabla^2 \mathbf{E} = -k^2 \mathbf{E}$  into Eq. (8.20) to obtain

$$k^2 = i\omega \mu_0 \sigma + \omega^2 \mu_0 \epsilon_m \quad (8.21)$$

Therefore, the wavevector becomes complex:  $\mathbf{k} = \mathbf{k}' + i\mathbf{k}''$ , where  $\mathbf{k}' = k'_x \hat{\mathbf{x}} + k'_y \hat{\mathbf{y}} + k'_z \hat{\mathbf{z}}$  and  $\mathbf{k}'' = k''_x \hat{\mathbf{x}} + k''_y \hat{\mathbf{y}} + k''_z \hat{\mathbf{z}}$  are real vectors. Note that Eq. (8.21) tells us the value of  $k^2 = \mathbf{k} \cdot \mathbf{k} = k_x^2 + k_y^2 + k_z^2$ , where each wavevector component may be complex, but does not specify the individual components. The *complex dielectric function* is defined as

$$\epsilon = \epsilon' + i\epsilon'' = \frac{\epsilon_m}{\epsilon_0} + i \frac{\sigma}{\omega \epsilon_0} \quad (8.22)$$

For a nonmagnetic material, the *complex refractive index*  $\tilde{n} = n + i\kappa$  is related to the complex dielectric function by  $\epsilon = (n + i\kappa)^2$ . The imaginary part  $\kappa$  of the complex refractive index is called the extinction coefficient. By definition, we have

$$\epsilon' = n^2 - \kappa^2 \quad \text{and} \quad \epsilon'' = 2n\kappa \quad (8.23)$$

The refractive index  $n$  and the extinction coefficient  $\kappa$  are also called *optical constants*,<sup>9</sup> although none of them are constant over a large wavelength region for real materials. The

dielectric function is also called relative permittivity, with respect to the permittivity of vacuum  $\epsilon_0$ . One can consider the  $\sigma/\omega$  term in Eq. (8.22) as the imaginary part of the permittivity. Some texts used  $\epsilon = \epsilon' - i\epsilon''$  for the dielectric function and  $\tilde{n} = n - i\kappa$  for the complex refractive index. In doing so, Eq. (8.9) must be revised to  $\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k} \cdot \mathbf{r})}$ . In either convention,  $\epsilon''$  and  $\sigma$  must be nonnegative for a passive medium. Equation (8.21) can be rewritten as

$$k = \tilde{n}\omega/c_0 \quad (8.24)$$

For simplicity, we will remove the tilde and simply use  $n$  for the complex refractive index, where it can be clearly understood from the context.

By substituting  $i\mathbf{k}$  for  $\nabla$  and  $-i\omega$  for  $\partial/\partial t$ , we can rewrite Maxwell's curl equations as

$$\mathbf{k} \times \mathbf{E} = \omega\mu_0\mathbf{H} \quad (8.25)$$

and

$$\mathbf{k} \times \mathbf{H} = -\omega\epsilon_0\epsilon\mathbf{E} \quad (8.26)$$

Similar to the definition of the complex dielectric function, one may choose to define a complex conductivity that satisfies Ohm's law at high frequencies,  $\mathbf{J} = \tilde{\sigma}\mathbf{E}$ , where

$$\tilde{\sigma} = \sigma' + i\sigma'' = \sigma - i\omega\epsilon_m \quad (8.27)$$

because we have assumed that  $\sigma$  is the real part of  $\tilde{\sigma}$ . Therefore,

$$\sigma'' = -\omega\epsilon_0\epsilon' \quad \text{and} \quad \epsilon'' = \sigma'/\omega\epsilon_0 \quad (8.28)$$

Equation (8.26) can be recast in terms of the complex conductivity as

$$\mathbf{k} \times \mathbf{H} = -i\tilde{\sigma}\mathbf{E} \quad (8.29)$$

In the subsequent discussion, we will omit the tilde above  $\sigma$ , when the context is sufficiently clear. The complex conductivity and the complex dielectric function are related to each other. For a linear, isotropic, and homogeneous nonmagnetic material, only two frequency-dependent functions are needed to fully characterize the electromagnetic response. The function pairs often found in the literature are  $(n, \kappa)$ ,  $(\sigma', \epsilon')$ ,  $(\epsilon', \epsilon'')$ , and  $(\sigma', \sigma'')$ . The principle of causality, which states that the effect cannot precede the cause, or no output before an input, imposes additional restrictions on the frequency dependence of the optical properties so that the real and imaginary parts are not completely independent, but related to each other. In general, the relative permeability, which is complex and frequency dependent, can be expressed as

$$\mu = \mu' + i\mu'' = \mu_m/\mu_0 \quad (8.30)$$

The complex refractive index for magnetic materials should be defined as follows:

$$n = \sqrt{\epsilon\mu} \quad (8.31)$$

The amplitude of the complex wavevector is  $k = n\omega/c_0$ , the same as Eq. (8.24). One can verify that Eq. (8.9) is a solution of the wave equation. The relative permittivity  $\epsilon$  and permeability  $\mu$  will be used to formulate the general equations later in this chapter. In most sections of this chapter, we deal with nonmagnetic materials, such as metals, dielectrics, and semiconductors. However, we will devote the discussion of the optical properties of magnetic materials in Sec. 8.4.6, because of the emerging interest in *metamaterials*, which are synthesized materials with magnetic response at the microwave and higher frequencies (see Problem 8.6, for example).

### 8.1.6 Propagating and Evanescent Waves

In an absorbing nonmagnetic medium, the electric and magnetic fields will attenuate exponentially. As an example, consider a wave that propagates in the positive  $x$  direction, with its electric field polarized in the  $y$  direction. Then,

$$\mathbf{E} = \hat{\mathbf{y}}E_0e^{-i(\omega t - k'x)}e^{-k''x} \quad (8.32)$$

where  $k' = \omega n/c_0$  and  $k'' = \omega\kappa/c_0$  are the real and imaginary parts of the wavevector, respectively; that is,  $\mathbf{k} = (k' + ik'')\hat{\mathbf{x}}$ . Equation (8.32) suggests that the amplitude of the electric field will decay exponentially according to  $e^{-(2\pi\kappa/\lambda)x}$ . The magnetic field can be obtained from Eq. (8.25) as

$$\mathbf{H} = \hat{\mathbf{z}}\frac{n + i\kappa}{\mu_0c_0}E_0e^{-i(\omega t - k'x)}e^{-k''x} \quad (8.33)$$

By substituting Eq. (8.32) and Eq. (8.33) into Eq. (8.18b), we obtain the time-averaged energy flux in the  $x$  direction as

$$\langle S \rangle = \frac{n}{2\mu_0c_0}E_0^2e^{-2k''x} = \frac{n}{2\mu_0c_0}E_0^2e^{-a_\lambda x} \quad (8.34)$$

where  $a_\lambda = 4\pi\kappa/\lambda$  is called the *absorption coefficient*. The inverse of  $a_\lambda$  is called the *radiation penetration depth* (or photon mean free path) given by

$$\delta_\lambda = \frac{1}{a_\lambda} = \frac{\lambda}{4\pi\kappa} \quad (8.35)$$

It is the distance through which the radiation power is attenuated by a factor of  $1/e$  ( $\approx 37\%$ ). (See Problem 8.5 for some typical values of the penetration depth in various materials at different wavelengths.)

When  $\mathbf{k}$  is complex, the plane normal to  $\mathbf{k}'$  is the constant-phase plane and the plane normal to  $\mathbf{k}''$  is the constant-amplitude plane because

$$\mathbf{E} = \mathbf{E}_0e^{-i(\omega t - \mathbf{k}' \cdot \mathbf{r})}e^{-\mathbf{k}'' \cdot \mathbf{r}} \quad (8.36)$$

When  $\mathbf{k}' \times \mathbf{k}'' = 0$ , the wave is said to be *homogeneous*; otherwise, the constant-phase planes will not be parallel to the constant-amplitude planes, and the wave is said to be *inhomogeneous*. An example of a homogeneous wave is given in Eq. (8.32). Next, we will discuss an example of an inhomogeneous wave. Consider a wave, defined in the  $z \geq 0$  half plane of vacuum, with a wavevector  $\mathbf{k} = 2\omega/c_0\hat{\mathbf{x}} + i\sqrt{3}\omega/c_0\hat{\mathbf{z}}$ . The electric field is linearly polarized in the  $y$  direction; thus,  $\mathbf{E} = \hat{\mathbf{y}}E_0e^{-i(\omega t - \mathbf{k} \cdot \mathbf{r})}$ . It can be shown that  $\mathbf{k} \cdot \mathbf{k} = k^2 = \omega^2/c_0^2$ ; hence,  $\mathbf{k}$  is indeed a valid wavevector of vacuum. The electric field can be written as

$$\mathbf{E} = \hat{\mathbf{y}}E_0e^{-i(\omega t - k_x x)}e^{-\gamma z} \quad (8.37)$$

Here,  $k_x = 2k = 4\pi/\lambda$ , and  $\gamma = 2\sqrt{3}\pi/\lambda$ . Clearly, the wave has a constant phase for any constant- $x$  plane and a constant amplitude for any constant- $z$  plane. Furthermore, the amplitude decays exponentially toward positive  $z$  direction and becomes negligible, when  $z > \lambda$ , as shown schematically in Fig. 8.3. Such a wave is called an *evanescent wave*, which exists in waveguides and is important for near-field optics and nanoscale radiation heat transfer. It can be shown that the time-averaged Poynting vector is parallel to the  $x$  direction so that no energy is transported toward the  $z$  direction (see Problem 8.7).

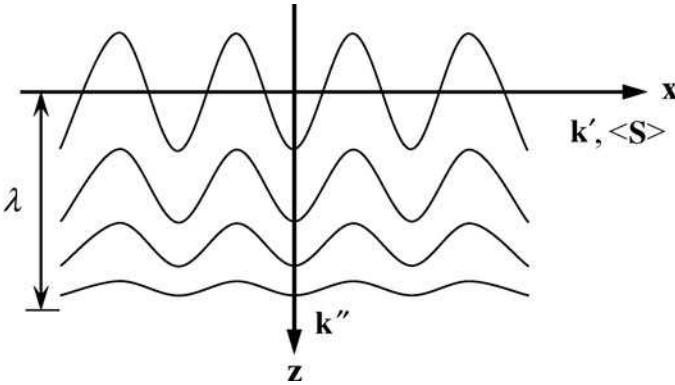


FIGURE 8.3 Schematic of an evanescent wave.

## 8.2 BLACKBODY RADIATION: THE PHOTON GAS

### 8.2.1 Planck's Law

Consider an enclosure of volume  $V$ , whose walls are at a uniform temperature  $T$ , as shown in Fig. 8.4a. The enclosure may contain a medium (such as a molecular gas), which may be

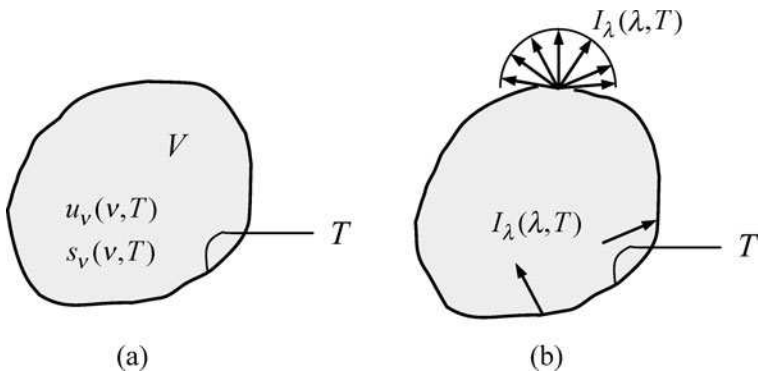


FIGURE 8.4 An isothermal enclosure (blackbody cavity): (a) without an opening and (b) with a small opening on the wall that has little effect on the equilibrium distribution.

evacuated (vacuum). Inside the enclosure, there exist electromagnetic fields, which may be viewed either as many transverse waves at different frequencies or as a large number of quanta with different energies. The particle theory treats radiation as a collection of photons. The energy and the momentum of each photon are related to the frequency and the speed of light, by  $\varepsilon = h\nu$  and  $p = h\nu/c$ , respectively. We are interested in finding the equilibrium distribution of photons with respect to photon energy or frequency or momentum.

Photons obey Bose-Einstein statistics, without requiring the total number be conserved because the number of photons depends on temperature; therefore,

$$\frac{dN}{V} = \frac{dg}{V} \frac{1}{e^{h\nu/k_b T} - 1} \tag{8.38}$$

The quantum states in the phase space, consisting of a volume  $V$  and a spherical shell in the momentum space (from  $p$  to  $p + dp$ ), are given by  $dg = 2V(4\pi p^2 dp)/h^3$ , where the factor 2 accounts for the two polarization states of electromagnetic waves. Thus, we can write

$$dg = \frac{8\pi V \nu^2 d\nu}{c^3} \tag{8.39}$$

Notice that  $c$  is the speed of light in the medium and the density of states is  $D(\nu) = dg/V$ . The number of photons per unit volume per unit frequency interval is therefore

$$f(\nu) = \frac{1}{V} \frac{dN}{d\nu} = f_{BE}(\nu)D(\nu) = \frac{8\pi\nu^2}{c^3(e^{h\nu/k_b T} - 1)} \tag{8.40}$$

In essence, the number of photons is the number of quantum states that are occupied. Therefore, the photon concentration (number density) is often referred to as the *occupation number* in quantum statistics. Since each photon has an energy  $h\nu$ , the spectral energy density (energy per unit volume per unit frequency interval) can be written as

$$u_\nu = \frac{8\pi h\nu^3}{c^3(e^{h\nu/k_b T} - 1)} \tag{8.41}$$

For an area element inside the enclosure, the radiant energy flux is related to the energy density and the speed of light by

$$q''_{rad,\nu} = \frac{u_\nu c}{4} \tag{8.42}$$

If a blackbody is placed inside the enclosure, it will absorb all incoming radiant energy that reaches its surface; at thermal equilibrium, it must emit the same amount of energy. After substituting Eq. (8.41) into Eq. (8.42), we obtain the *spectral emissive power* of a blackbody as

$$e_{b,\nu}(\nu,T) = \frac{2\pi h\nu^3}{c^2(e^{h\nu/k_b T} - 1)} \tag{8.43}$$

Substituting  $\nu = c/\lambda$ ,  $d\nu = -cd\lambda/\lambda^2$ , and  $e_{b,\nu}d\nu = -e_{b,\lambda}d\lambda$  into the above expression, we obtain the blackbody distribution function in terms of wavelength:

$$e_{b,\lambda}(\lambda,T) = \frac{2\pi hc^2}{\lambda^5(e^{hc/\lambda k_b T} - 1)} = \frac{C_1}{\lambda^5(e^{C_2/\lambda T} - 1)} \tag{8.44}$$

which is called *Planck's law* or *Planck's distribution* (of blackbody radiation). In Eq. (8.44),  $C_1$  and  $C_2$ , called the first and second radiation constants, are used for convenience. It should be noted that the blackbody intensity is  $I_{b,\lambda}(\lambda,T) = e_{b,\lambda}(\lambda,T)/\pi$ , as in Eq. (2.48), and isotropic inside the whole cavity. Furthermore, when there is a small opening, the emitted radiation is diffuse and obeys the blackbody distribution, as shown in Fig. 8.4b. The requirement is that the opening should be sufficiently small compared with the size of the enclosure, but large enough compared to the wavelengths of interest. The concept of blackbody cavity was made clear by Wien in his 1911 Nobel lecture, as seen from the excerpt below ([http://nobelprize.org/nobel\\_prizes/physics/laureates/1911/wien-lecture.html](http://nobelprize.org/nobel_prizes/physics/laureates/1911/wien-lecture.html)):



... there must exist, in a cavity surrounded by bodies of equal temperature, a radiation energy that is independent of the nature of the bodies. If in the walls surrounding this cavity a small aperture is made through which radiation issues, we obtain a radiation which is independent of the nature of the emitting body, and is wholly determined by the temperature. The same radiation would also be emitted by a body which does not reflect any rays and which is therefore designated as completely black, and this radiation is called the radiation of a black body or black-body radiation.

Equation (8.43) or Eq. (8.44) can be integrated over the whole spectrum to obtain the Stefan-Boltzmann law,  $e_b = \sigma_{SB} T^4$ , in vacuum. In Fig. 8.5,  $e_{b,\lambda}/\sigma_{SB} T^5$  has been plotted as a

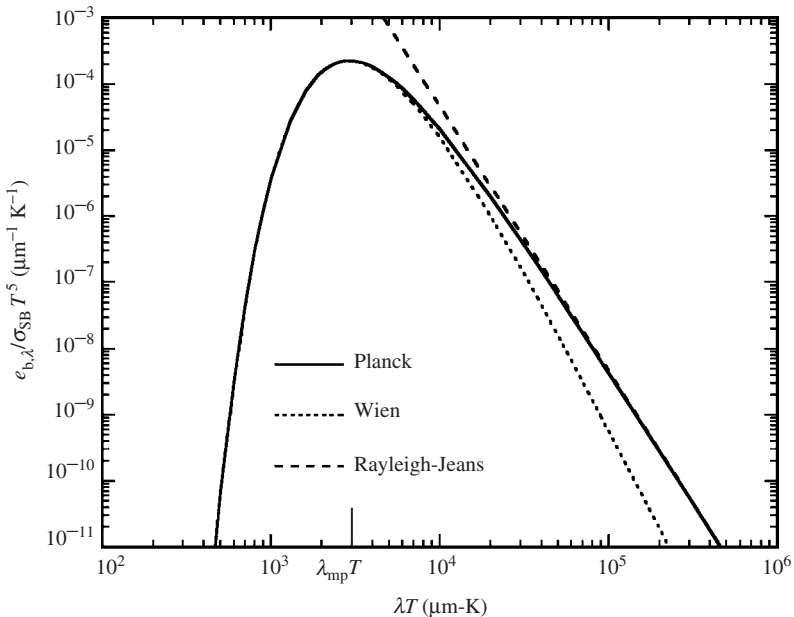


FIGURE 8.5 Planck's law for blackbody emissive power.

function of  $\lambda T$  so that the area under Planck's distribution (solid curve) is  $\int_0^\infty [e_{b,\lambda}(\lambda, T)/\sigma_{SB} T^5] d(\lambda T) = (1/\sigma_{SB} T^4) \int_0^\infty e_{b,\lambda}(\lambda, T) d\lambda = 1$ . The Planck distribution has a peak and approaches zero at extremely short or long wavelengths. If  $C_2/\lambda T \gg 1$ , the right-hand side of Eq. (8.44) can be approximated by  $C_1 \lambda^{-5} e^{-C_2/\lambda T}$ . This is called Wien's formula, which gives approximately correct result, even beyond the maximum emissive power, as can be seen from Fig. 8.5. At very long wavelengths, Wien's formula underpredicts the emissive power and asymptotically approaches to  $C_1 \lambda^{-5}$ , suggesting that the emissive power is independent of temperature. Note that the right-hand side of Eq. (8.44) approaches  $C_1 T/(C_2 \lambda^4)$  if  $C_2/\lambda T \ll 1$ , since  $e^x - 1 \approx x$  for  $x \ll 1$ . This is called the Rayleigh-Jeans formula, which is applicable at very long wavelengths, as shown in Fig. 8.5. The significance of the Rayleigh-Jeans formula is that it correctly predicts the temperature dependence of the blackbody spectrum, at very long wavelengths, where Wien's formula fails. The failure of

the Rayleigh-Jeans formula at short wavelengths is called the *ultraviolet catastrophe*. The significance of Planck’s formula is more than a correct unified mathematical formulation. It was derived based on the hypothesis of energy quanta that do not exist in classical Newtonian mechanics and Maxwell’s electrodynamics. It should be noted that the preceding derivation is based on statistical thermodynamics, presented in Chap. 3, rather than Planck’s original semi-classical oscillator model.

**Example 8-2.** Find the wavelength  $\lambda_{mp}$  at which Planck’s distribution reaches a maximum. What is the ratio of the energy emitted at  $\lambda < \lambda_{mp}$  to that at  $\lambda > \lambda_{mp}$ ?

**Solution.** By setting the derivative of Eq. (8.44) equal to zero, i.e.,  $de_{b,\lambda}/d\lambda = 0$ , we have  $x + 5e^{-x} - 5 = 0$ , where  $x = hc/(k_B\lambda T)$ . This equation can be solved numerically to yield

$$\lambda_{mp} [\mu\text{m}] \approx \frac{2898 \mu\text{m} \cdot \text{K}}{T[\text{K}]} \tag{8.45}$$

This is Wein’s displacement law. The location of  $\lambda_{mp}T$  is also marked on Fig. 8.5. To find out the ratio of the energy emitted at  $\lambda < \lambda_{mp}$  to that at  $\lambda > \lambda_{mp}$ , we can numerically evaluate  $\int_0^{\lambda_{mp}} e_{b,\lambda}(\lambda,T)d\lambda / \int_{\lambda_{mp}}^{\infty} e_{b,\lambda}(\lambda,T)d\lambda$ . The numerical result is approximately 1:3 and independent of temperature. For a medium of refractive index  $n$ , the speed of light  $c$  should be replaced by  $c_0/n$  in Eq. (8.43). In the previous discussion, we have assumed a nondispersive medium with  $n \equiv 1$ , which is true for vacuum only. Corrections are rarely needed if the medium is a gas, but would be necessary for radiation inside solids or liquids. Furthermore, in a dispersive medium, the group velocity needs to be considered in deriving the density of states  $D(\nu)$  in Eq. (8.40) and the energy flux in Eq. (8.42); see Prasher (*Appl. Phys. Lett.*, **86**, 071914, 2005).

**Example 8-3.** Assuming the sun to be a blackbody at 5800 K, calculate the emitted power at the following wavelength intervals:  $\lambda < 0.3 \mu\text{m}$ ,  $0.3 \mu\text{m} < \lambda < 0.4 \mu\text{m}$ ,  $0.4 \mu\text{m} < \lambda < 0.7 \mu\text{m}$ ,  $0.7 \mu\text{m} < \lambda < 3 \mu\text{m}$ , and  $\lambda > 3 \mu\text{m}$ . Neglect the absorption by the atmosphere. What is the radiant power arriving at the earth’s surface from the sun?

**Solution.** The total emissive power is  $\sigma_{SB}T_{sun}^4 = 5.67 \times 10^{-8} \times 5800^4 \approx 64 \text{ MW/m}^2$ . We can obtain the emitted power in each spectral region by integrating Eq. (8.44), as listed in the following table. Note that  $F_{\lambda_1 \rightarrow \lambda_2}$  represents the fraction of radiation falling between  $\lambda_1$  and  $\lambda_2$ .

$\lambda(\mu\text{m})$	< 0.3	0.3–0.4	0.4–0.7	0.7–3	> 3	Total
$\lambda_2 T (\mu\text{m} \cdot \text{K})$	1740	2320	4060	17400	$\infty$	–
$F_{0 \rightarrow \lambda_2}$	0.03	0.12	0.49	0.98	1	–
$F_{\lambda_1 \rightarrow \lambda_2}$	0.03	0.09	0.37	0.49	0.02	1
$\Delta E_b(\text{MW/m}^2)$	1.9	5.8	23.7	31.4	1.3	64.1

The total power emitted by the sun equals the emissive power multiplied by the surface area of the sun. The fraction of the power that reaches the earth equals the solid angle of the earth divided by  $4\pi$ . Note that the radius of the sun  $r_{sun} = 6.955 \times 10^8 \text{ m}$ , the radius of the earth  $r_{earth} = 6.378 \times 10^6 \text{ m}$ , and the earth-sun distance  $R_{earth-sun} = 1.496 \times 10^{11} \text{ m}$ . Therefore, the total power that will reach the earth’s surface, if the absorption by the atmosphere is neglected, is

$$\dot{Q} = 4\pi r_{sun}^2 \cdot \sigma_{SB}T_{sun}^4 \cdot \frac{\pi r_{earth}^2}{4\pi R_{earth-sun}^2} \approx 1.8 \times 10^{17} \text{ W}$$

The average irradiation on the earth is:  $G = \dot{Q}/\pi r_{earth}^2 \approx 1377 \text{ W/m}^2$ . This value is very close to the total solar irradiance (TSI), measured outside the earth’s atmosphere.

Because of the broad spectral region of electromagnetic waves, alternative units are often used, such as wavelength  $\lambda$  (in vacuum), wavenumber  $\bar{\nu} = 1/\lambda$ , frequency  $\nu = c_0/\lambda$ , angular frequency  $\omega = 2\pi\nu$ , and photon energy  $E = h\nu$ . Generally speaking, optical radiation covers the spectral region of ultraviolet (UV), visible (VIS), near-infrared (NIR),

**TABLE 8.1** Spectral Regions Expressed in Different Units

	UV from-to	VIS up to	NIR up to	MIR up to	FIR up to	MW up to
Wavelength, $\lambda$ ( $\mu\text{m}$ ) <sup>*</sup>	0.01–0.38	0.76	2.5	25	$10^3$	$10^5$
Wavenumber, $\bar{\nu}$ ( $\text{cm}^{-1}$ )	$10^6$ –( $2.6 \times 10^4$ )	$1.3 \times 10^4$	$4 \times 10^3$	400	10	0.1
Frequency, $\nu$ (THz)	( $3 \times 10^4$ )–790	395	120	12	0.3	$3 \times 10^{-3}$
Angular frequency, $\omega$ (rad/s)	( $2 \times 10^5$ )–( $5 \times 10^3$ )	$2.5 \times 10^3$	750	75	1.9	0.02
Photon energy, $E$ (eV) <sup>†</sup>	124–3.3	1.63	0.5	0.05	$1.2 \times 10^{-3}$	$1.2 \times 10^{-5}$

<sup>\*</sup>The wavelength will be reduced in a medium whose refractive index  $n$  is not unity.

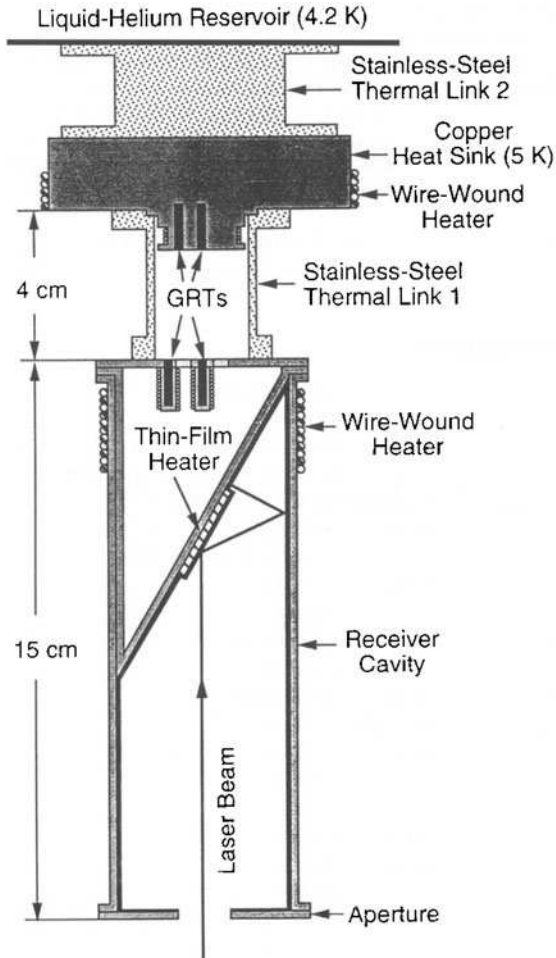
<sup>†</sup>The conversion from the vacuum wavelength  $\lambda$  in  $\mu\text{m}$  to the photon energy  $E$  in eV is  $E = 1.240/\lambda$ .

mid-infrared (MIR), and far-infrared (FIR). Table 8.1 outlines the subdivisions of the spectral region in different units from ultraviolet (UV) to microwave (MW).

## 8.2.2 Radiation Thermometry

The developments of the absolute temperature scale and radiation thermometry are among the most important applications of blackbody radiation. The Stefan-Boltzmann law  $e_b = \sigma_{\text{SB}} T^4$  defines an absolute thermodynamic temperature, which is consistent with the one defined by the ideal-gas law and the Carnot cycle. While radiation thermometry can serve as a primary standard, most practical radiation thermometers are not absolute instruments because of other considerations such as fast response, easy operation, and low cost. High-temperature furnaces are commonly used as calibration standards. The cavity is a hollow cylinder, made of graphite for example, with a conical ending and a small aperture. The most accurate calibration source is the fixed-point heat pipe blackbody, for which a pure metal is melted outside the graphite cylinder to maintain a constant temperature in a two-phase state. The freezing temperatures are then used to define the temperature scales (1234.93 K for Ag, 1337.33 for Au, and 1357.77 K for Cu).

To measure the absolute temperature of a thermally radiative body, two blackbody cavities at different temperatures would be needed: one serves as the emitter (blackbody source) and the other as the receiver (radiometer). Quinn and Martin used a blackbody source and a cryogenic radiometer to directly determine the thermodynamic temperatures and measure the Stefan-Boltzmann constant.<sup>10</sup> The experimentally obtained Stefan-Boltzmann constant was  $(5.66967 \pm 0.00076) \times 10^{-8} \text{ W}/(\text{m}^2 \cdot \text{K}^4)$ . The difference is 0.13% of the theoretical value  $(5.67040 \pm 0.00004) \times 10^{-8} \text{ W}/(\text{m}^2 \cdot \text{K}^4)$ , based on Planck's constant, Boltzmann's constant, and the speed of light. Since the early 1990s, the National Institute of Standards and Technology (NIST) has developed a high-accuracy cryogenic radiometer (HACR) facility to serve as the primary standard for optical radiation measurements. A schematic of the original HACR receiver is shown in Fig. 8.6. The receiver is mounted at the bottom of a liquid-helium cryostat in an evacuated chamber, and the optical access is through a Brewster window below the cavity. The HACR facility has gone through some major upgrades in recent years. The receiver cavity is made of copper with a high thermal conductivity and low specific heat at cryogenic temperatures. The inner wall of the cavity is coated with a specular black paint to absorb the incident radiation with an effective absorptance greater than 99.998%. The electrical-substitution technique links the radiant power to the electric power to achieve an overall uncertainty within 0.02% for optical power measurements. Detailed descriptions can be found from Pearson and Zhang and references therein.<sup>11</sup> The cosmic microwave background radiation, measured with cryogenic bolometers, can be



**FIGURE 8.6** Schematic of the receiver cavity of an absolute cryogenic radiometer, where GRT stands for germanium resistance thermometer, from Pearson and Zhang.<sup>11</sup>

fitted to the blackbody distribution at 2.7 K, which is the temperature of the universe at the present time. The discovery of cosmic radiation background in 1964 and the subsequent measurements and theoretical studies have been recognized by the Nobel Prizes in Physics to Arno Penzias and Robert Wilson in 1978 and to John Mather and George Smoot in 2006.

Most radiation thermometers are based on spectral measurements rather than on the measurement of the total irradiance from the target. When a radiation thermometer is used to measure the temperature of a real surface, the unknown emissivity of the surface and the influence of the surrounding radiation are the major issues that affect the measurement. Various methods have been developed to deal with these problems, including the creation of a blackbody cavity on the surface, the two-color method, and the use of a

controlled reference source.<sup>12</sup> The development of optical fibers has allowed radiometric temperature measurements for surface locations that are otherwise inaccessible by imaging radiometers.

The *measurement equation* of a spectral radiation thermometer can be approximated as follows:

$$V_d = C_1 I_{\text{ex},\lambda}(\lambda) \quad (8.46)$$

where  $V_d$  is the detector output signal and  $C_1$  is an instrument constant that is independent of the target material and temperature. The term  $I_{\text{ex},\lambda}(\lambda)$  is called the *exitent spectral radiance*, which includes the radiation emitted by the target and the surroundings, as well as that reflected by the target. The *radiance temperature*  $T_\lambda$  (also called the *brightness temperature*) is defined according to

$$I_{\text{b},\lambda}(\lambda, T_\lambda) = I_{\text{ex},\lambda}(\lambda) \quad (8.47)$$

where  $I_{\text{b},\lambda}(\lambda, T_\lambda)$  is the blackbody intensity at the wavelength  $\lambda$  and temperature  $T_\lambda$ . If the surrounding emission and absorption can be neglected, the exitent spectral radiance is due only to the emission; therefore,

$$I_{\text{ex},\lambda}(\lambda) = I_{\text{e},\lambda}(\lambda, T) = \varepsilon'_\lambda I_{\text{b},\lambda}(\lambda, T) \quad (8.48)$$

where  $\varepsilon'_\lambda$  is the directional-spectral emissivity, and  $I_{\text{e},\lambda}(\lambda, T)$  is the intensity emitted by the target. By combining Eq. (8.47) and Eq. (8.48) and applying Wien's formula, the surface temperature is related to the radiance temperature by

$$\frac{1}{T} = \frac{1}{T_\lambda} + \frac{\lambda}{C_2} \ln \varepsilon'_\lambda \quad (8.49)$$

The uncertainty in the measured temperature due to an uncertainty in the emissivity is

$$\frac{\delta T}{T} = -\frac{\lambda T}{C_2} \frac{\delta \varepsilon'_\lambda}{\varepsilon'_\lambda} \quad (8.50)$$

The effect of the emissivity uncertainty on the temperature accuracy decreases as  $\lambda$  decreases. However, the wavelength at which  $I_{\text{b},\lambda}(\lambda, T)$  is a maximum is given by Wien's displacement law. In practice, the choice of the operating wavelength should also be based on the material's properties and the surrounding radiation, and requires a detailed analysis of different effects. If the surrounding radiation is not negligible,  $I_{\text{ex},\lambda}(\lambda)$  is the sum of the emitted and reflected spectral radiances, and may be affected by participating medium emission and absorption.

**Example 8-4.** Rapid thermal processing is a semiconductor single-wafer manufacturing technique. Lightpipe radiation thermometer, at  $\lambda = 0.95 \mu\text{m}$ , is used to measure the wafer temperature. The emissivity of a plain silicon wafer is approximately 0.7 at this wavelength. Neglect the reflected radiation from the wafer. If the wafer is at a temperature of 1200 K, what is the radiance temperature? If the temperature needs to be determined within an uncertainty of 1 K, how much tolerance on the emissivity error is acceptable?

**Solution.** From Eq. (8.49),  $T_\lambda \approx 1167 \text{ K}$ , which differs from the actual temperature by approximately 33 K. One can also solve Eq. (8.47) and Eq. (8.48), using Planck's law, and the result is essentially the same. Based on Eq. (8.50), to obtain a temperature within an uncertainty of 1 K, the emissivity must be determined within an uncertainty of  $\delta \varepsilon'_\lambda = 0.0074$ . Zhou et al. (*Int. J. Heat Mass Transfer*, 45, 1945, 2002) developed a model to predict the effective emissivity of silicon wafers in rapid thermal processing furnaces and showed that the temperature measurement uncertainty can be significantly reduced by using a reflective cavity.

### 8.2.3 Entropy and Radiation Pressure

Like other particles, the photon gas also has the property of entropy and can be related to other properties in equilibrium states. Express the energy density in an enclosure of volume  $V$ , at thermodynamic equilibrium, with a temperature  $T$  as  $u = U/V = 4\sigma_{\text{SB}}T^4/c$ . It can be seen that the specific heat at constant volume is  $c_v = (\partial u/\partial T)_V = 16\sigma_{\text{SB}}T^3/c$ . The entropy can therefore be obtained as

$$S = \int_0^T V c_v \frac{dT}{T} = \frac{16}{3c} V \sigma_{\text{SB}} T^3 \quad (8.51a)$$

or 
$$s = \frac{16}{3c} \sigma_{\text{SB}} T^3 \quad (8.51b)$$

Note that  $T = (\partial U/\partial S)_V$  is satisfied. The Helmholtz free energy  $A = U - TS = -\frac{4}{3}V\sigma_{\text{SB}}T^4/c$ . Thus, the radiation pressure is

$$P = -\left(\frac{\partial A}{\partial V}\right)_T = \frac{4}{3c}\sigma_{\text{SB}}T^4 \quad (8.52)$$

The force by the radiation pressure, albeit small, has some important applications in trapping and manipulating atomic and molecular particles. This technique is called optical traps or optical tweezers; see Lang and Block (*Am. J. Phys.*, **71**, 201, 2003) for a bibliographical review.

If each photon mode (frequency) is individually considered, the *spectral entropy density* for unpolarized radiation can be expressed as follows:

$$s_\nu(\nu, T) = \frac{8\pi k_B \nu^2}{c^3} \left[ \frac{x}{e^x - 1} + \ln\left(\frac{e^x}{e^x - 1}\right) \right] \quad (8.53)$$

where  $x = h\nu/k_B T^4$ . Note that  $1/T = (\partial s_\nu/\partial u_\nu)_\nu = (k_B/h\nu) \ln(1 + 8\pi h\nu^3/u_\nu c^3)$ , which is consistent with Eq. (8.41). Similar to the energy flux (emissive power) and intensity, the *radiation entropy flux* can be obtained by multiplying a factor  $c/4$  to Eq. (8.51b) and Eq. (8.53), and the *radiation entropy intensity* can be obtained by dividing the flux by  $\pi$ , because of the isotropic nature of blackbody radiation. Clearly, electromagnetic radiation carries both energy and entropy.

**Example 8-5.** Consider the radiation heat transfer between two parallel plates at  $T_1$  and  $T_2$ , respectively. Assume each plate has an area of  $A$  and both plates are blackbodies. The separation distance is much smaller than  $\sqrt{A}$  but much greater than the wavelength of thermal radiation.

(a) How much entropy is generated at each plate? Evaluate the ratio of entropy generation assuming that  $T_1 = 2T_2$ .

(b) If a thermophotovoltaic receiver is mounted on the lower-temperature side to convert thermal radiative energy to electricity (work), what is its maximum achievable efficiency?

**Solution.** (a) The net energy flow from plate 1 to 2 is  $\dot{Q}_{12} = A\sigma_{\text{SB}}(T_1^4 - T_2^4)$ . The entropy of plate 1 will decrease at the rate of  $dS_1/dt = -\dot{Q}_{12}/T_1$ , and the entropy of plate 2 will increase at the rate of  $dS_2/dt = \dot{Q}_{12}/T_2$ . On the other hand, the net entropy flow from plate 1 to 2 can be calculated as  $\dot{S}_{12} = \frac{4}{3}A\sigma_{\text{SB}}(T_1^3 - T_2^3)$ . Therefore,  $\dot{S}_{\text{gen},1} = -\dot{Q}_{12}/T_1 + \dot{S}_{12} = A\sigma_{\text{SB}}(\frac{1}{3}T_1^3 - \frac{4}{3}T_2^3 + T_2^4/T_1)$ ,  $\dot{S}_{\text{gen},2} = A\sigma_{\text{SB}}(\frac{1}{3}T_2^3 - \frac{4}{3}T_1^3 + T_1^4/T_2)$ , and the combined total entropy generation is equal to  $\dot{Q}_{12}(1/T_2 - 1/T_1)$ , as expected. It can be shown that the entropy generation at each plate is always greater than zero if  $T_1 \neq T_2$ , or equal to zero if  $T_1 = T_2$ . When  $T_1 = 2T_2$ , the entropy generation by plate 1 is about one-quarter and that by plate 2 is about three-quarters of the total entropy generated.

(b) The available energy or exergy of thermal radiation is defined as the maximum work that can be produced by a system with respect to a large reservoir. In the present example, we may assume that the reservoir is at the same temperature as  $T_2$ . Suppose an amount of heat is taken from the high-temperature plate; we would like to find out the maximum work that can possibly be produced. Let us consider a reversible heat engine at  $T_2$ . The radiative energy leaving surface 1 can still be described by  $\dot{Q}_1 = A\sigma_{\text{SB}}(T_1^4 - T_2^4)$ , and the entropy leaving surface 1 is  $\dot{S}_1 = \frac{4}{3}A\sigma_{\text{SB}}(T_1^3 - T_2^3)$ . Therefore, the entropy generation in plate 1 cannot be eliminated. In other words, it is impossible to achieve the Carnot efficiency of  $\eta_{\text{Carnot}} = 1 - T_2/T_1$ . The maximum work can be obtained when the irreversibility at the lower-temperature plate is negligible and the heat engine is also reversible. It can easily be shown that the maximum work  $W_{\text{max}} = \dot{Q}_1 - T_2\dot{S}_1$ , and the optimal efficiency is given by

$$\eta_{\text{opt}} = \frac{\dot{W}_{\text{max}}}{\dot{Q}_1} = 1 - \frac{4(1 + y + y^2)}{3(1 + y)(1 + y^2)} \quad (8.54)$$

where  $y = T_1/T_2 \geq 0$ . When  $y = 2$ , we obtain an optimal efficiency  $\eta_{\text{opt}} = 37.8\%$ , which is less than the Carnot efficiency of 50%, because of the unrecoverable irreversibility at plate 1. A comprehensive discussion can be found from the review of Landsberg and Tonge.<sup>13</sup>

The next question is whether temperature can be defined for laser radiation. The answer is *yes*, and the temperature for high-intensity lasers can be very high. An intuitive guess is to define the temperature, based on the intensity  $I_\nu$  of the laser or the monochromatic radiation, by setting  $I_\nu = I_{\text{b},\nu}(\nu, T_\nu)$ . The definitions of entropy and thermodynamic temperature for optical radiation are very important for analyzing optical energy conversion systems, such as solar cells, thermophotovoltaic generators, luminescence devices, and laser cooling apparatus.<sup>13,14</sup> Assume that the monochromatic radiation is from a thermodynamic equilibrium state, such as a resonance cavity that allows only a single mode to exist. The spectral entropy intensity of unpolarized radiation can be written as follows:<sup>13</sup>

$$L_\nu = \frac{2k_{\text{B}}\nu^2}{c^2} \left[ \left( 1 + \frac{c^2 I_\nu}{2h\nu^3} \right) \ln \left( 1 + \frac{c^2 I_\nu}{2h\nu^3} \right) - \frac{c^2 I_\nu}{2h\nu^3} \ln \left( \frac{c^2 I_\nu}{2h\nu^3} \right) \right] \quad (8.55)$$

Thermodynamically, the *monochromatic radiation temperature* can be defined by

$$\frac{1}{T_\nu(\nu)} = \left( \frac{\partial L_\nu}{\partial I_\nu} \right)_\nu = \frac{k_{\text{B}}}{h\nu} \ln \left( 1 + \frac{2h\nu^3}{c^2 I_\nu} \right) \quad (8.56)$$

This is indeed Planck's distribution of intensity at the same temperature. The expressions can be modified for polarized radiation. When the energy intensity is very high, Eq. (8.56) approaches  $T_\nu(\nu) = c^2 I_\nu / (2k_{\text{B}}\nu^2)$ , which is in the Rayleigh-Jeans limit. The radiation temperature will be proportional to the intensity of the monochromatic radiation and can exceed  $10^{10}$  K, with a 1-mW He-Ne laser at 632.8-nm wavelength.<sup>15</sup> Therefore, for lasers with a moderate intensity,  $T_\nu$  tends to be so high that the entropy is nearly zero; hence, the interaction of a laser beam with a material can be considered as work interaction. If a collimated beam is randomly scattered by a rough surface, the scattered radiation will have a much lower intensity because of the increase in the solid angle. The process is accompanied with an entropy increase and is thus irreversible. It is not possible to increase the intensity of the scattered light, back to their original intensity, without leaving any net effect on the environment of the photon system. On the other hand, if a nearly collimated light is split into two beams with a beamsplitter, the transmitted and reflected beams can interfere with each other to reconstruct the original beam. This process is reversible because the two beams are *correlated*. The correlated beams have lower entropy than those with the same intensity at thermodynamic equilibrium. The concept of temperature is applicable only if the maximum-entropy state has been reached.<sup>15</sup> While the definition of the monochromatic radiation temperature is similar to that of the radiance temperature, the two concepts are quite different. In the definition of radiance temperature, the quality (entropy) does not

enter into play. On the other hand, the definition of the monochromatic temperature for incoherent radiation is for a state that is equilibrium in a certain wavelength and angular ranges.

Consider a gray-diffuse body, for which the emissive power is proportional to the blackbody emissive power, at any frequency and angle of emission. The monochromatic temperature calculated from Eq. (8.56), however, is frequency dependent. This is because the emitted radiation, as a whole, cannot be considered as a blackbody at any temperature. Thermal radiation of this type has been called *dilute blackbody radiation*.<sup>13</sup> This simple example shows that photons at any given frequency can be considered as in a thermodynamic equilibrium but not necessarily in equilibrium with photons at other frequencies. When radiation has two linear polarizations with a different intensity, the monochromatic temperature will be different, even for the two polarizations. In general, it is a function of frequency, direction, and polarization. The requirement is that each subsystem be in a thermodynamic equilibrium, even though it is not in equilibrium with other subsystems at the same spatial location. Photons at different frequencies, with different polarization states, or propagating toward different directions, can coexist in their own equilibrium state without any interaction with each other. The concept may be called partial equilibrium, as in the case when the two parts of a cylinder were separated by a moveable adiabatic wall. The mechanical equilibrium would be established to maintain the same pressure on each side, but the temperatures may be different from each other because thermal equilibrium is reached only inside each portion but not between them. Another example is in ultrafast laser heating of metals, as discussed in Chap. 7, where the electron and phonon systems can be treated as in separate equilibrium states but not in equilibrium with each other.

The concept of entropy intensity has recently been applied by Caldas and Semiao to study the entropy generation in an absorbing, emitting, and scattering medium, based on the equation of radiative transfer (ERT) introduced in Sec. 2.4.3.<sup>16</sup> The key is that the change in entropy in an elemental path length equals the change in intensity divided by the radiation temperature. The entropy change at steady state can be obtained from Eq. (2.53) in Chap. 2 as follows:

$$\frac{dL_\lambda}{d\xi} = \frac{a_\lambda I_{b,\lambda}}{T_\lambda(I_\lambda)} - \frac{(a_\lambda + \sigma_\lambda)I_\lambda}{T_\lambda(I_\lambda)} + \frac{\sigma_\lambda}{4\pi} \int_{4\pi} \frac{I_\lambda(\Omega')}{T_\lambda(I_\lambda)} \Phi(\Omega', \Omega) d\Omega' \quad (8.57)$$

Like  $I_\lambda$ , the entropy intensity  $L_\lambda$  is a function of wavelength, location, and direction. Note that  $I_{b,\lambda} = I_{b,\lambda}(\lambda, T_g)$ , where  $T_g$  is the local temperature. For an anisotropic radiation field,  $T_\lambda(I_\lambda)$  would be different for different directions. For nonblackbody radiation,  $T_\lambda(I_\lambda)$  will be a function of wavelength. The term  $I_\lambda/T_\lambda(I_\lambda)$ , however, is not the same as  $L_\lambda$ . Integration of Eq. (8.57) over the solid angle of  $4\pi$  at all wavelengths in a volume element yields the entropy that is transferred out of the control volume. Furthermore, the entropy change in the control volume is equal to the total energy absorbed divided by  $T_g$ . The energy rate received per unit volume can be expressed as

$$\dot{q} = \int_0^\infty \int_{4\pi} a_\lambda (I_\lambda - I_{b,\lambda}) d\Omega d\lambda \quad (8.58)$$

Because the entropy change is the sum of the net entropy transferred into the system and the entropy generation by irreversibility, we can express the volumetric entropy generation rate as

$$\begin{aligned} \dot{s}_{\text{gen}} = & \int_0^\infty \int_{4\pi} a_\lambda I_{b,\lambda} \left[ \frac{1}{T_\lambda(I_\lambda)} - \frac{1}{T_g} \right] d\Omega d\lambda - \int_0^\infty \int_{4\pi} \left[ \frac{a_\lambda + \sigma_\lambda}{T_\lambda(I_\lambda)} - \frac{a_\lambda}{T_g} \right] I_\lambda d\Omega d\lambda \\ & + \int_0^\infty \int_{4\pi} \left[ \frac{\sigma_\lambda}{4\pi} \left[ \int_{4\pi} \frac{I_\lambda(\Omega')}{T_\lambda(I_\lambda)} \Phi(\Omega', \Omega) d\Omega' \right] \right] d\Omega d\lambda \end{aligned} \quad (8.59)$$



For an isotropic field,  $I_\lambda$  is independent of the direction, and scattering does not contribute to the entropy generation. In this case, the entropy generation becomes

$$\dot{s}_{\text{gen}} = \int_0^\infty \int_{4\pi} a_\lambda (I_{b,\lambda} - I_\lambda) \left[ \frac{1}{T_\lambda(I_\lambda)} - \frac{1}{T_g} \right] d\Omega d\lambda \quad (8.60)$$

The entropy generation is always greater than zero, because the intensity is an increasing function of temperature, unless the medium is at thermal equilibrium. When a surface is involved in radiative heat transfer, the entropy generation rate per unit area can be expressed as

$$s''_{\text{gen}} = \int_0^\infty \int_0^{2\pi} \int_0^{\pi/2} \left[ \frac{I_{\text{in},\lambda} - I_{\text{out},\lambda}}{T_w} - (L_{\text{in},\lambda} - L_{\text{out},\lambda}) \right] \cos\theta \sin\theta d\theta d\phi d\lambda \quad (8.61)$$

where  $T_w$  is the wall temperature, subscripts “in” and “out” signify the energy or entropy intensity to and from the surface, respectively. If the surface is not a blackbody, the outgoing intensity includes both the emitted and reflected intensities. An alternative approach is to integrate the intensity over the whole sphere with a solid angle of  $4\pi$ . In Eq. (8.61), the entropy intensity is related to the energy intensity by Eq. (8.55), which is recast in terms of wavelength as follows:

$$L_\lambda(\lambda, I_\lambda) = \frac{2k_B c}{\lambda^4} \left[ \left( 1 + \frac{\lambda^5 I_\lambda}{2hc^2} \right) \ln \left( 1 + \frac{\lambda^5 I_\lambda}{2hc^2} \right) - \frac{\lambda^5 I_\lambda}{2hc^2} \ln \left( \frac{\lambda^5 I_\lambda}{2hc^2} \right) \right] \quad (8.62)$$

The use of Eq. (8.62) may be disputed when multiple reflections occur. The intensity of the emitted radiation is less than that of the blackbody and is reduced by each reflection. The question still remains as whether the blackbody intensity should be used to calculate the entropy or the actual intensity after each reflection or the combined intensity at any given location. An example is a system of two large parallel plates, separated by vacuum. One of the plates is at a temperature  $T_1$  and is diffuse-gray with an emissivity of 0.5. The other plate is insulated and is a perfect reflector (i.e., zero emissivity). It is clear that a thermal equilibrium will be established in the cavity after a long time. Again, the separation distance is much larger than the thermal radiation wavelengths. The radiation leaving surface 1 includes the emitted rays, as well as the first-order and higher-order reflected rays. An attempt to define the entropy of the emitted ray and each reflected ray will result in a total entropy intensity greater than the entropy intensity calculated based on the blackbody intensity  $I_{b,\lambda}(\lambda, T_1)$ . Therefore, to apply the previous analysis in a consistent way and to obtain meaningful results, we must make the following hypotheses:

- The intensity at any given location is additive regardless of where it originates from, as long as it falls within the same solid angle and wavelength intervals. While this sounds obvious, it is untrue when interference effects become important. The resulting intensity is called the combined intensity.
- The monochromatic radiation temperature  $T_\lambda$ , defined in Eq. (8.56), is a function of the combined intensity and is in general dependent on the direction and wavelength. The effect of polarization is neglected to simplify the problem. Equation (8.56) must not be applied to each of the reflected or scattered rays. The physical significance is that all the photons, with the same wavevector and frequency, can be considered as a subsystem that is at thermodynamic equilibrium with the temperature  $T_\lambda[I_\lambda(\lambda, \theta, \phi)]$ .
- The entropy intensity is defined based on the combined intensity, according to Eq. (8.62). While entropy must be additive, the entropy of all individual rays must be calculated

based on the monochromatic temperature of the combined intensity. Because the number of photons, intensity, and entropy are additive, the fraction of the entropy of each ray is the same as the ratio of the intensity of that ray to the combined intensity.

With the theories presented in this section, one should be able to perform a second law thermodynamic analysis for a given system, involving radiative transfer of energy. Zhang and Basu investigated entropy flow and generation considering incoherent multiple reflections.<sup>17</sup> There exist different approximations in analyzing the entropy of radiation. For example, the method of dilute blackbody radiation uses a dilution factor and defines an effective temperature for each wavelength.<sup>13</sup> When the process is very complicated, it appears that such an effective temperature cannot be easily defined and this definition cannot be applied to multiple reflections. Entropy generation is usually accompanied by the generation of heat, such as heating by friction, electrical resistor, chemical reaction, or absorption of solar radiation. On the other hand, it appears that entropy generation can occur in radiation without the generation of heat, such as by scattering. The definition of inelastic scattering is based on the conservation of energy (wavelength) and momentum, which does not impose any constraints on the reversibility. Further research is much needed in order to better understand the nature of entropy of radiation and determine the ultimate efficiency of photovoltaic cells and other radiative processes, including laser cooling and trapping. Another area of possible application of radiation entropy is in nanoscale heat conduction using the EPRT, as discussed in Chap. 7. The entropy concept may be extended to the phonon system by defining radiation entropy and entropy intensity of phonons.

### 8.2.4 Limitations of Planck's Law

The concept that *a blackbody surface absorbs all radiant energy that is incident upon it* is purely from the geometric-optics point of view, in which light travels in a straight line and cannot interact with an object that does not intercept the light ray. Another example of the geometric-optics viewpoint is that the transmittance of an iris (open aperture) should be 1, i.e., *all the radiation incident on the opening will go through*. However, for an aperture whose diameter is comparable to the wavelength of the incident radiation, diffraction may become important and, as a result, the transmittance can be less or even greater than 1. Due to the diffraction effect, a particle that is sufficiently small compared to the wavelength will interact with the radiation field, according to the scattering and absorption cross sections, which can be greater than the projected surface area. In some cases, it is possible for the object to absorb more energy than the product of the radiant flux and the projection area. The absorptance can be greater than 1 and thus exceeds the limit set by a blackbody. When such an object is placed in an isothermal enclosure, the emitted energy will be greater than that from a blackbody having the same dimensions. This anomaly has been discussed in detail by Bohren and Huffman.<sup>7</sup>

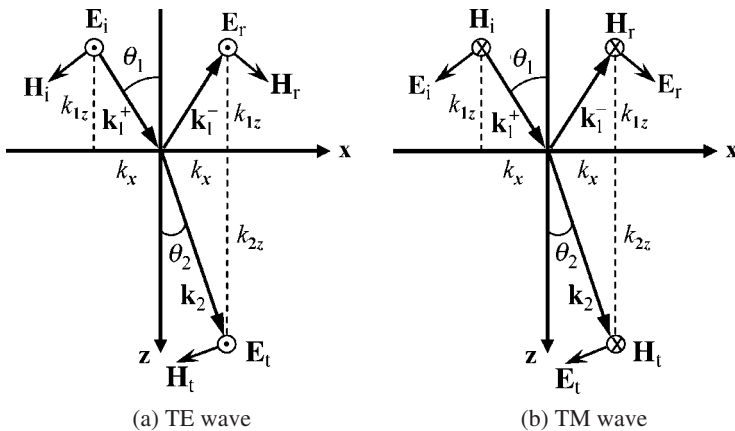
The energy density near the surface within a distance less than the wavelength can be much greater than that given by Eq. (8.41) and increases as the distance is further reduced. When two objects are placed at a distance much smaller than the characteristic wavelength of thermal radiation, i.e., in the near field, photon tunneling can occur and cause significant enhancement of the energy transfer. In recent years, there have been numerous studies of light transmission through small apertures, radiation heat transfer at nanometer distances, and light emission from nanostructures.<sup>18</sup> This is still an open field with many new developments as well as controversies. We will study these phenomena and the underlying physics in the following two chapters. The entropy concept and the second law limitation have not been applied to the study of near-field energy transfer.

### 8.3 RADIATIVE PROPERTIES OF SEMI-INFINITE MEDIA

#### 8.3.1 Reflection and Refraction of a Plane Wave

Consider radiation incident from one medium to another at the interface or the boundary. The boundary that separates the media is assumed to be a smooth plane and extends to infinity. Each medium is homogeneous and isotropic; so, there is no scattering within the medium. Therefore, the electric response can be characterized by the relative permittivity or dielectric function  $\epsilon$ , and the magnetic response can be characterized by the relative permeability  $\mu$ . For nonmagnetic materials, the refractive index is related to the dielectric function by  $n = \sqrt{\epsilon}$ . Keep in mind that these quantities are, in general, complex and frequency dependent. The real and imaginary parts of the refractive index are often called the optical constants. In this section, we present the general formulation for both magnetic and nonmagnetic materials. For certain crystalline and amorphous solids, like quartz and glass, the refractive index is real in a wide spectral region and is the only parameter needed to fully characterize the optical response of the material. In such a case, the expression can be largely simplified and the results can be easily comprehended. The reduced results will also be presented because of their importance to numerous engineering problems.

The incident radiation is a monochromatic plane wave with an angular frequency  $\omega$ . As shown in Fig. 8.7, the wavevector of the incident wave is  $\mathbf{k}_1^+ = (k_{1x}, 0, k_{1z})$ , and the surface



**FIGURE 8.7** Illustration of reflection and transmission at an interface: (a) TE wave or *s* polarization. (b) TM wave or *p* polarization.

normal defines the *plane of incidence*, which is the *x-z* plane. The wavevectors of the reflected and transmitted waves must lie in the same plane. The angle of incidence  $\theta_1$  is the angle between the incident wavevector and the *z* direction, i.e.,  $\sin \theta_1 = k_{1x}/k_1$  and  $\cos \theta_1 = k_{1z}/k_1$ , where  $k_1^2 = k_{1x}^2 + k_{1z}^2 = \mu_1 \epsilon_1 \omega^2 / c_0^2$ . It is common to study the reflection and the refraction for linearly polarized waves, with either the electric or magnetic field being parallel to the *y*-axis, because other polarizations can be decomposed into the two polarization components.

When the electric field is in the *y* direction, as shown in Fig. 8.7a, the wave is called a transverse-electric (TE) wave or is said to be perpendicularly (*s*) polarized. The incident electric field can be expressed as follows by omitting the time-harmonic term of  $e^{-i\omega t}$  hereafter:

$$\mathbf{E}_i = \hat{\mathbf{y}} E_i e^{ik_{1z}z + ik_{1x}x} \tag{8.63}$$

The boundary conditions state that the tangential components of both  $\mathbf{E}$  and  $\mathbf{H}$  must be continuous at the interface. This implies that the  $x$  component of the wavevector must be the same for the incident, reflected, and transmitted waves, i.e.,  $k_{1x} = k_{2x} = k_x$ . Because the angle of reflection must be the same as the angle of incidence (specular reflection), we have  $\mathbf{k}_1^- = (k_x, 0, -k_{1z})$ . For the transmitted or refracted wave, we have  $\mathbf{k}_2 = (k_x, 0, k_{2z})$  and

$$\sin\theta_2 = \frac{k_x}{k_2} = \frac{n_1 \sin\theta_1}{n_2} \quad (8.64)$$

which is called Snell's law. It can be easily visualized by observing the bended image of a chopstick in a bowl of water. Note that  $k_{2z}^2 = k_2^2 - k_x^2 = \mu_2 \epsilon_2 \omega^2 / c_0^2 - k_x^2 = k_2^2 \cos^2\theta_2$ . Generally speaking, the wavevector components and the refractive indices may be complex. Complex angles can be defined so that Eq. (8.64) is always valid. Near the interface, the nonzero components of the electric and magnetic fields are

$$E_y = \begin{cases} (E_i e^{ik_{1z}z} + E_r e^{-ik_{1z}z}) e^{ik_x x} & \text{for } z < 0 \\ E_t e^{ik_{2z}z} e^{ik_x x} & \text{for } z > 0 \end{cases} \quad (8.65)$$

$$H_x = \begin{cases} -\frac{k_{1z}}{\omega\mu_0\mu_1} (E_i e^{ik_{1z}z} - E_r e^{-ik_{1z}z}) e^{ik_x x} & \text{for } z < 0 \\ -\frac{k_{2z}}{\omega\mu_0\mu_2} E_t e^{ik_{2z}z} e^{ik_x x} & \text{for } z > 0 \end{cases} \quad (8.66)$$

$$H_z = \begin{cases} \frac{k_x}{\omega\mu_0\mu_1} (E_i e^{ik_{1z}z} + E_r e^{-ik_{1z}z}) e^{ik_x x} & \text{for } z < 0 \\ \frac{k_x}{\omega\mu_0\mu_2} E_t e^{ik_{2z}z} e^{ik_x x} & \text{for } z > 0 \end{cases} \quad (8.67)$$

and

where  $E_i$ ,  $E_r$ , and  $E_t$  are, respectively, the amplitudes of the incident, reflected, and transmitted electric fields at the interface. It is further assumed that  $k_x$  is real so that the amplitude of the field is independent of  $x$ . The Fresnel reflection and transmission coefficients for a TE wave are defined as  $r_{12,s} = E_r/E_i$  and  $t_{12,s} = E_t/E_i$ , respectively. Boundary conditions require that  $E_y$  and  $H_x$  be continuous at  $z = 0$ . From Eq. (8.65) and Eq. (8.66), we obtain  $1 + r_{12,s} = t_{12,s}$  and  $(k_{1z}/\mu_1)(1 - r_{12,s}) = (k_{2z}/\mu_2)t_{12,s}$ ; thus,

$$r_{12,s} = \frac{E_r}{E_i} = \frac{k_{1z}/\mu_1 - k_{2z}/\mu_2}{k_{1z}/\mu_1 + k_{2z}/\mu_2} \quad (8.68a)$$

and

$$t_{12,s} = \frac{E_t}{E_i} = \frac{2k_{1z}/\mu_1}{k_{1z}/\mu_1 + k_{2z}/\mu_2} \quad (8.68b)$$

which are generally applicable, as long as each medium is homogeneous and isotropic.<sup>6</sup> For nonmagnetic materials, the previous equations can be written as follows:

$$r_{12,s} = \frac{n_1 \cos\theta_1 - n_2 \cos\theta_2}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \quad (8.69a)$$

and

$$t_{12,s} = \frac{2n_1 \cos\theta_1}{n_1 \cos\theta_1 + n_2 \cos\theta_2} \quad (8.69b)$$

The directional-hemispherical spectral reflectivity, or simply reflectivity,  $\rho'_\lambda$  is given by the ratio of the reflected energy flux to the incident energy flux, and the directional-spectral absorptivity  $\alpha'_\lambda$  is the ratio of the transmitted energy flux to the incident energy flux, since all the photons transmitted through the interface will be absorbed inside the second medium. We use terms ending with “-ivity” only for a perfect interface and those with “-tance” for surfaces with roughness and coatings. The energy flux is related to the time-averaged Poynting vector, defined in Eq. (8.18b). From Eq. (8.65) to Eq. (8.67), the  $x$  and  $z$  components of the Poynting vector at the interface ( $z \rightarrow 0$ ) in medium 1 are

$$\langle S_{1x} \rangle = \frac{1}{2} \operatorname{Re} \left[ \frac{k_x^*}{\omega \mu_0 \mu_1^*} (E_i + E_r)(E_i^* + E_r^*) \right] \quad (8.70a)$$

and

$$\langle S_{1z} \rangle = \frac{1}{2} \operatorname{Re} \left[ \frac{k_{1z}^*}{\omega \mu_0 \mu_1^*} (E_i + E_r)(E_i^* - E_r^*) \right] \quad (8.70b)$$

It can be seen that, in general, the reflected wave and the incident wave are coupled and the energy flow cannot be separated by a reflected flux and an incident flux. Under the assumption that medium 1 is lossless (nonabsorbing or nondissipative) and  $k_x^2 < k_1^2$ , we can write

$$\langle S_{1z} \rangle = \langle S_{iz} \rangle - \langle S_{rz} \rangle \quad (8.71)$$

where

$$\langle S_{iz} \rangle = \frac{k_{1z}}{2\omega \mu_0 \mu_1} |E_i|^2 \quad \text{and} \quad \langle S_{rz} \rangle = \frac{k_{1z}}{2\omega \mu_0 \mu_1} |E_r|^2 \quad (8.72)$$

If medium 1 is lossy, there will be additional terms associated with  $E_i E_r^*$  and  $E_r^* E_i$ . In this case, the power flow normal to the interface cannot be separated as forward and backward terms because of the cross-coupling terms. Therefore, the lossless condition in medium 1 is required in order to properly define the energy reflectivity; see Salzberg (*Am. J. Phys.*, **16**, 444, 1948) and Zhang (*J. Heat Transfer*, **119**, 645, 1997). This is usually not a problem when radiation is incident from air or a dielectric prism onto a medium. The power reflectivity can be defined based on the  $z$  components of the reflected and incident Poynting vectors; therefore,

$$\rho'_{\lambda,s}(\theta_1) = |E_r|^2 / |E_i|^2 = |r_{12,s}|^2 \quad (8.73)$$

The Poynting vector at the interface in medium 2 can be written as

$$\langle \mathbf{S}_i \rangle = \frac{1}{2\omega \mu_0} \operatorname{Re} \left( \frac{k_x^* \hat{\mathbf{x}} + k_{2z}^* \hat{\mathbf{z}}}{\mu_2^*} \right) |E_i|^2 \quad (8.74)$$

which is not parallel to  $\operatorname{Re}(\mathbf{k}_2)$  unless  $\operatorname{Im}(\mu_2) = 0$ . Recall that the plane of constant phase is perpendicular to  $\operatorname{Re}(\mathbf{k}_2)$ . If medium 2 is dissipative,  $\operatorname{Im}(\mathbf{k}_2)$  is parallel to the  $z$ -axis and the amplitude will vary along the  $z$  direction. The wave becomes inhomogeneous in medium 2, except when  $k_x = 0$  (normal incidence). The definition of the transmitted energy flux at the interface is based on the projected Poynting vector in the  $z$  direction. Hence, the absorptivity is the ratio of the  $z$  components of the transmitted and incident Poynting vectors, viz.,

$$\alpha'_{\lambda,s}(\theta_1) = \frac{\operatorname{Re}(k_{2z}/\mu_2)}{\operatorname{Re}(k_{1z}/\mu_1)} |t_{12,s}|^2 \quad (8.75)$$

Note that  $\operatorname{Re}(k_{2z}/\mu_2) = \operatorname{Re}(k_{2z}^*/\mu_2^*)$ , and  $\operatorname{Re}(k_{1z}/\mu_1) = k_{1z}/\mu_1$  since medium 1 is lossless. It can be shown that  $\rho'_{\lambda,s} + \alpha'_{\lambda,s} = 1$ , as required by energy conservation:  $\langle S_{1z} \rangle = \langle S_{2z} \rangle$  at  $z = 0$ . For nonmagnetic and nearly nondissipative materials, we have

$$\alpha'_{\lambda,s}(\theta_1) = \frac{n_2 \cos \theta_2}{n_1 \cos \theta_1} |t_{12,s}|^2 \quad (8.76)$$

The reflection and transmission coefficients for the transverse-magnetic (TM) wave or parallel ( $p$ ) polarization are defined as the ratios of the magnetic fields:  $r_{12,p} = H_r/H_i$  and  $t_{12,p} = H_t/H_i$ , respectively. Hence,

$$r_{12,p} = \frac{H_r}{H_i} = \frac{k_{1z}/\epsilon_1 - k_{2z}/\epsilon_2}{k_{1z}/\epsilon_1 + k_{2z}/\epsilon_2} \tag{8.77a}$$

$$t_{12,p} = \frac{H_t}{H_i} = \frac{2k_{1z}/\epsilon_1}{k_{1z}/\epsilon_1 + k_{2z}/\epsilon_2} \tag{8.77b}$$

In the case of nonmagnetic materials, we obtain

$$r_{12,p} = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \tag{8.78a}$$

and

$$t_{12,p} = \frac{2n_2 \cos \theta_1}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \tag{8.78b}$$

At normal incidence, the reflection coefficients in Eq. (8.69a) and Eq. (8.78a) are related by

$$r_{12,s} = \frac{n_1 - n_2}{n_1 + n_2} = -r_{12,p} \tag{8.79}$$

When both  $n_1$  and  $n_2$  are real and  $n_1 < n_2$ , the electric field will experience a phase reversal (phase shift of  $\pi$ ) upon reflection but the magnetic field will not. On the other hand, if  $n_1 > n_2$ , it is the magnetic field that will experience a phase reversal. In fact, based on Maxwell's equations, the electric and magnetic quantities obey a duality, when  $\rho_e = 0$ , and can be interchanged with the following substitutions:  $\mathbf{E} \rightarrow \mathbf{H}$  and  $\mathbf{H} \rightarrow -\mathbf{E}$ . Note that  $\epsilon$  and  $\mu$ , as well as the polarization states  $s$  and  $p$ , should also be interchanged. The Poynting vector for a TM wave is  $\langle \mathbf{S} \rangle = \text{Re}(\mathbf{k}/\epsilon) |H_y|^2 / (2\omega\epsilon_0)$ , which is not parallel to  $\text{Re}(\mathbf{k})$  when  $\text{Im}(\epsilon_2) \neq 0$ . Upon refraction into an absorbing medium, the waves become inhomogeneous and the Poynting vectors for different polarizations may split into different directions; see Halevi and Mendoza-Hernandez (*J. Opt. Soc. Am.*, **71**, 1238, 1981). Nevertheless, the constant-amplitude plane is always perpendicular to the  $z$  direction because the amplitude cannot change along the  $x$ - $y$  plane. The reflectivity for  $p$  polarization is

$$\rho'_{\lambda,p}(\theta_1) = |r_{12,p}|^2 \tag{8.80}$$

Hence, the absorptivity becomes

$$\alpha'_{\lambda,p}(\theta_1) = \frac{\text{Re}(k_{2z}/\epsilon_2)}{\text{Re}(k_{1z}/\epsilon_1)} |t_{12,p}|^2 \tag{8.81}$$

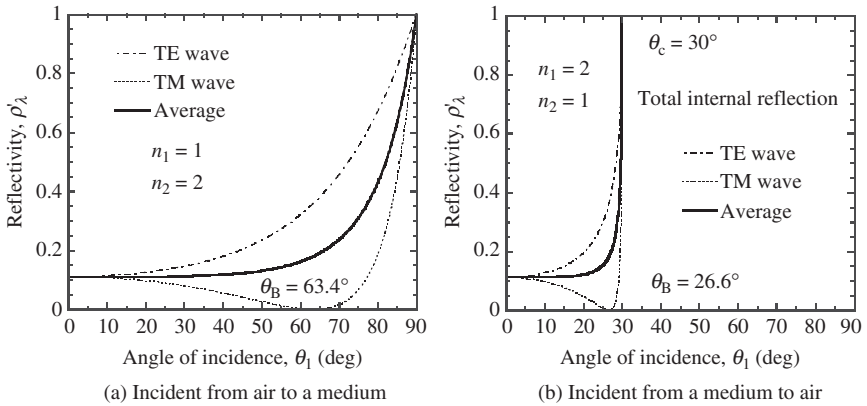
For nonmagnetic and nearly nonabsorbing materials, we have

$$\alpha'_{\lambda,p}(\theta_1) = \frac{n_1 \cos \theta_2}{n_2 \cos \theta_1} |t_{12,p}|^2 \tag{8.82}$$

If the incident wave is unpolarized or circularly polarized, the reflectivity can be obtained by averaging the values for  $p$ - and  $s$ -polarized waves, i.e.,

$$\rho'_\lambda = \frac{\rho'_{\lambda,p} + \rho'_{\lambda,s}}{2} \tag{8.83}$$

The reflectivity for radiation incident from air ( $n_1 \approx 1$ ) to a dielectric medium ( $n_2 = 2$ ) and that from the dielectric to air are shown in Fig. 8.8 for each polarization as well as for the



**FIGURE 8.8** Reflectivity versus the angle of incidence between air and a dielectric.

unpolarized incident radiation. When  $n_1 > n_2$ , the reflectance will reach 1 at  $\theta_1 = \theta_c = \sin^{-1}(n_2/n_1)$ . This angle is called the *critical angle*, and *total internal reflection* occurs at angles of incidence greater than the critical angle. This is the principle commonly used in optical fibers and waveguides, since light is trapped inside the high-index material and propagates along the medium. It can be seen that in total internal reflection,  $k_x > k_2$  while  $k_{z2}$  becomes purely imaginary. The amplitude of the wave is exponentially attenuating in the positive  $z$  direction. This is similar to Eq. (8.37) and is an evanescent wave, as shown in Fig. 8.3. The time-averaged Poynting vector is zero in the  $z$  direction. Hence, no energy is transmitted across the boundary.

For the TE wave, the reflectivity increases monotonically with the angle of incidence and reaches 1 at the grazing angle ( $90^\circ$ ) or at the critical angle when  $n_1 > n_2$ . The reflectivity for the TM wave, on the other hand, goes through a minimum that is equal to zero. The angle at which  $\rho'_{\lambda,p} = 0$  is called the Brewster angle, given by  $\theta_B = \tan^{-1}(n_2/n_1)$  for nonmagnetic materials. For  $p$  polarization, all the incident energy will be transmitted into medium 2, without reflection at the Brewster angle. This phenomenon has been used to build polarizers and transmission windows in absolute cryogenic radiometers. The physical mechanism of reflection can also be understood as the re-emission by the *induced electric dipoles* in the medium, based on the *Ewald-Oseen extinction theorem*. At the Brewster angle, the electric dipoles induced in the material align in the direction of the reflected wave, and the refracted wave is perpendicular to the reflected wave (i.e.,  $\theta_1 + \theta_2 = 90^\circ$ ). The reflective power goes to zero because an electric dipole cannot radiate along its own axis. The situation is changed when magnetic materials are involved, such as a negative index material. The fields radiated by both the induced electric dipoles and *magnetic dipoles* are responsible for the reflection. The Brewster angle can occur for either polarization when the radiated fields cancel each other. A detailed discussion can be found from the publication of Fu et al.<sup>19</sup> In an absorbing medium, there is a drop in reflectance for  $p$  polarization, but the minimum is not zero. Furthermore, there exists a *principal angle* at which the phase difference between the two reflection coefficients equals to  $90^\circ$  and the ratio of the reflectance for the TM and TE waves is minimized.<sup>6</sup>

The reflectivity for radiation incident from air ( $n_1 \approx 1$ ) or vacuum, at normal incidence, becomes

$$\rho'_{\lambda,n} = \frac{(n_2 - 1)^2 + \kappa_2^2}{(n_2 + 1)^2 + \kappa_2^2} \tag{8.84}$$

for any polarization. It can be seen that the normal reflectivity will be close to 1, when either  $n_2 \ll 1$  or  $n_2 \gg 1$ . The reflectivity is often large for most metals in the infrared because both  $n_2$  and  $\kappa_2$  are large, whereas the reflectivity of a conventional superconductor approaches to 1 when the frequency is lower than that of the superconducting energy gap, since  $n_2 \rightarrow 0$  in this case. On the other hand,  $\rho'_{\lambda,n} \rightarrow 0$  when  $n_2 \approx 1$  and  $\kappa_2 \ll 1$ . This can occur in a dielectric material at a wavelength in the infrared and for most metals in the x-ray region.

### 8.3.2 Emissivity

Real materials have finite thicknesses. The assumption of semi-infinity or opaqueness requires that the thickness is much greater than the radiation penetration depth. This is usually not a problem for a metal in the visible or infrared spectral regions. When this is not the case, we are dealing with a transparent or semitransparent material, like a glass window. The radiative properties of semitransparent layers and thin films will be studied in the next chapter. Laser beams or light from a spectrophotometer do not extend to infinity and are not perfectly collimated. Nevertheless, as long as the diameter of the beam spot is much greater than the wavelength and the beam divergence is not very large, the directional-spectral reflectivity and absorptivity, calculated from the previous section, are applicable to most situations. According to Kirchhoff's law, the directional-spectral emissivity is equal to the directional-spectral absorptivity of a material.<sup>1,2</sup> This can be shown by placing the object into an enclosure at the thermal equilibrium. When the material is not at thermal equilibrium with the surroundings, the emissivity is defined based solely on the spontaneous emission and is an intrinsic material property that does not depend on the surroundings. On the other hand, the absorptivity is defined based on the net absorbed energy by treating stimulated or induced emission as negative absorption. Under proper definitions, Kirchhoff's law is always valid in terms of the directional-spectral properties for any given polarization.<sup>1</sup> The only assumptions are (a) the material under consideration is at a uniform temperature, at least within several penetration depths near its surface and (b) the external field is not strong enough to alter the material's intrinsic properties, as in a nonlinear interaction. We can then compute the directional emissivity for an opaque surface or semi-infinite media, from the directional-hemispherical reflectivity for incidence from air or vacuum, using the following relation:

$$\epsilon'_\lambda = 1 - \rho'_\lambda \tag{8.85}$$

The emissivity is commonly calculated by averaging over the two polarizations. The preceding equation can be integrated to obtain the hemispherical emissivity

$$\epsilon_\lambda = \frac{1}{\pi} \int_0^{2\pi} \int_0^{\pi/2} \epsilon'_\lambda \cos\theta \sin\theta \, d\theta \, d\phi \tag{8.86}$$

It can be seen from Fig. 8.8a that, when averaged over the two polarizations, the reflectivity changes little until the Brewster angle and then increases to 1 when the incidence angle approaches 90°. The hemispherical emissivity for a nonmetallic surface is about 10% smaller than the normal emissivity. On the other hand, the hemispherical emissivity for



metallic surfaces is about 20% greater than the normal emissivity. Diffuse emission is a good first-order approximation, even though the surface is smooth. Thus, the hemispherical emissivity may be approximated by the normal emissivity. In most studies, the emissivity is calculated from the indirect method, based on the reflectivity and Kirchhoff's law, discussed earlier. Direct calculations can be accomplished by considering the emission from the material, and the internal absorption and transmission. Another method is based on the fluctuation-dissipation theorem, in which the emission arises from the thermally induced fluctuating currents inside the material. The fluctuational electrodynamics is essential to the study of near-field radiation and will be discussed in detail in Chap. 10. The total-hemispherical emissivity can be evaluated using Planck's distribution. Therefore,

$$\varepsilon_{\text{tot}} = \frac{\int_0^{\infty} \varepsilon_{\lambda}(\lambda) e_{\text{b},\lambda}(\lambda, T) d\lambda}{\int_0^{\infty} e_{\text{b},\lambda}(\lambda, T) d\lambda} = \frac{\int_0^{\infty} \varepsilon_{\lambda}(\lambda) e_{\text{b},\lambda}(\lambda, T) d\lambda}{\sigma_{\text{SB}} T^4} \quad (8.87)$$

The total emissivity depends on the surface temperature and the spectral dependence of the optical constants. Pure metals usually have a very low emissivity, and the emissivity increases due to surface oxidation. Spectrally selective materials that appear to be reflective to the visible light may exhibit a large total emissivity, greater than 0.9, at room temperature; examples are white paint and paper. An earlier compilation of the radiative properties of many engineering materials can be found in Touloukian and DeWitt.<sup>20</sup> The use of surface microstructure to modify the emission characteristics will be discussed in the next chapter.

### 8.3.3 Bidirectional Reflectance

Real surfaces contain irregularities or surface roughnesses that depend on the processing method. A surface appears to be smooth if the wavelength is much greater than the surface roughness height. A highly polished surface can have a roughness height on the order of nanometers. Some surfaces that appear "rough" to human eyes may appear to be quite "smooth" for far-infrared radiation. The reflection of radiation by rough surfaces is more complicated. For randomly rough surfaces, there often exist a peak around the direction of specular reflection, an off-specular lobe, and a diffuse component. When the surface contains periodic structures, such as patterned surfaces or micromachined surfaces, diffraction effects may become important and several peaks may appear.

The bidirectional reflectance distribution function (BRDF), which is a function of the angles of incidence and reflection, fully describes the reflection characteristics from a rough surface at a given wavelength. As illustrated in Fig. 8.9, the BRDF is defined as the reflected radiance (intensity) divided by the incident irradiance (flux) at the surface, i.e.,

$$f_r(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) = \frac{dI_r}{I_i \cos \theta_i d\Omega_i} \quad [\text{sr}^{-1}] \quad (8.88)$$

where  $(\theta_i, \phi_i)$  and  $(\theta_r, \phi_r)$  denote the directions of incident and reflected beams, respectively,  $I_i$  is the incident irradiance (radiant flux), and  $dI_r$  is the reflected radiance (intensity). In the experiment, the detector output signal is proportional to the solid angle  $d\Omega_i$ . The denominator of Eq. (8.88) gives the incident radiant power reaching the detector. Hence, the BRDF can be obtained from the following measurement equation:

$$f_r = \frac{1}{P_i} \frac{P_r}{\cos \theta_i d\Omega_i} \quad (8.89)$$

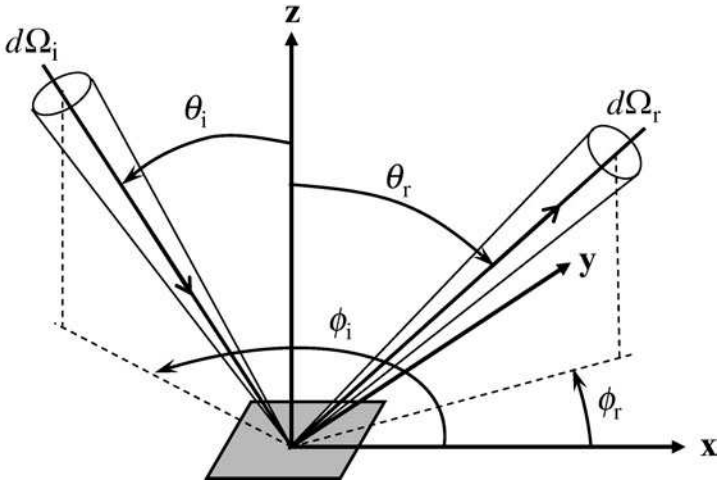


FIGURE 8.9 Geometry of the incident and reflected beams in defining the BRDF.

where  $P_i$  and  $P_r$  are the incident and reflected powers reaching the detector.<sup>21</sup> The directional-hemispherical reflectance can be obtained by integrating the BRDF over the hemisphere:<sup>1,2</sup>

$$\rho'_\lambda = \int_{2\pi} f_r \cos\theta_r d\Omega_r \tag{8.90}$$

An important principle of the BRDF is reciprocity, which states symmetry of the BRDF, with regard to reflection and incidence angles. In other words, the reflectance for energy incident from  $(\theta_i, \phi_i)$  and reflected to  $(\theta_r, \phi_r)$  is equal to that for energy incident from  $(\theta_r, \phi_r)$  and reflected to  $(\theta_i, \phi_i)$ . Therefore,

$$f_r(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) = f_r(\lambda, \theta_r, \phi_r, \theta_i, \phi_i) \tag{8.91}$$

For a *diffuse* or *Lambertian* surface, the BRDF is independent of  $(\theta_r, \phi_r)$  and is related to the directional-hemispherical reflectance as  $f_{r,\text{diff}} = \rho'_\lambda/\pi$ . On the other hand, the BRDF for an ideal specular, or mirrorlike, reflector can be represented as

$$f_{r,\text{spec}} = \frac{\rho'_\lambda}{\cos\theta_i} \delta_\theta(\theta_r - \theta_i) \delta_\phi(\phi_r - \phi_i - \pi) \tag{8.92}$$

where the Dirac delta function  $\delta(\xi)$  is zero everywhere, except at  $\xi = 0$ . Furthermore, the delta functions are normalized such that  $\int_{2\pi} \delta_\theta(\theta_r - \theta_i) \delta_\phi(\phi_r - \phi_i - \pi) d\Omega_r = 1$ . These examples clearly demonstrate that the BRDF is applicable to any kind of surfaces. In the next chapter, we will study the BRDF models based on geometric optics and physical optics, as well as rigorous solutions of the Maxwell equations. We will also discuss the effect of surface microstructures on the BRDF and how to characterize a rough surface.

## 8.4 DIELECTRIC FUNCTION MODELS

Unlike in dilute gases where the molecules are far apart, in solids, the closely packed atoms form band structures. Absorption in solids usually happens in a much broader frequency region or band. Free electrons in metals can interact with the incoming electromagnetic waves or photons, and cause a broadband absorption from the visible (or even ultraviolet) all the way to the microwave and longer wavelengths. For semiconductors especially with high impurity (doping) concentrations or at elevated temperatures, both the free electrons and holes contribute to the absorption process. The absorption of a photon makes the electron or the hole to transit to a higher-energy state within the same band. Therefore, free-carrier absorption is caused by *intra-band transitions*. In order to conserve momentum, the carriers must also collide with ionized impurities, phonons, other carriers, grain boundaries, interfaces, and so forth. The collisions act as a *damping force* on the motion of carriers. The Drude model describes the oscillatory movement of an electron, driven by a harmonic field, which is subjected to a damping force. The model is simple in form and predicts the dielectric function of some metals fairly well in a broad spectral region, especially in the mid- and far-infrared.

Absorption by lattice vibrations or bound electrons, which is important for insulators and lightly doped semiconductors, is due to the existence of electric dipoles formed by the lattice. A maximum absorption is achieved when the frequency equals the vibrational mode of the dipole, i.e., the resonance frequency, which is usually in the mid- to far-infrared region of the spectrum. The contribution of bound electrons is often modeled by the Lorentz model.

Interband transition is the *fundamental absorption process* in semiconductors. An electron can be excited from the valence band to the conduction band by absorbing a photon, whose energy is greater than the energy gap  $E_g$ . Because the absorption by electrons is usually weak in semiconductors, a strong absorption edge is formed near the bandgap. In this transition process, both the energy and the momentum must be conserved.

This section discusses the formulation for different contributions to the dielectric function. It should be noted that the real and imaginary parts of the dielectric function are interrelated according to the causality, which is discussed first. Because all naturally occurring and most of the synthesized materials are nonmagnetic at high frequencies, only nonmagnetic materials are considered so that  $\mu = 1$  and  $n = \sqrt{\epsilon}$  in the following, except in Sec. 8.4.6.

### 8.4.1 Kramers-Kronig Dispersion Relations

The real and imaginary parts of an analytic function are related by the Hilbert transform relations. Hendrik Kramers and Ralph Kronig were the first to show that the real and imaginary parts of the dielectric function are interrelated. These relations are called the Kramers-Kronig dispersion relations or K-K relations for simplicity. The K-K relations can be interpreted as the causality in the frequency domain and are very useful in obtaining optical constants from limited measurements. The principle of causality states that *the effect cannot precede the cause, or no output before input*. Some important relations are given here, and a detailed derivation and proofs can be found from Jackson,<sup>5</sup> Born and Wolf,<sup>6</sup> and Bohren and Huffman.<sup>7</sup>

The real part  $\epsilon'$  and the imaginary part  $\epsilon''$  of a dielectric function are related by

$$\epsilon'(\omega) - 1 = \frac{2}{\pi} \wp \int_0^{\infty} \frac{\zeta \epsilon''(\zeta)}{\zeta^2 - \omega^2} d\zeta \quad (8.93a)$$

and

$$\epsilon''(\omega) - \frac{\sigma_0}{\epsilon_0 \omega} = -\frac{2\omega}{\pi} \wp \int_0^{\infty} \frac{\epsilon'(\zeta) - 1}{\zeta^2 - \omega^2} d\zeta \quad (8.93b)$$

where  $\sigma_0$  is the dc conductivity,  $\wp$  denotes the principal value of the integral, and  $\zeta$  is a dummy frequency variable. These relations can be written in terms of  $n$  and  $\kappa$  as

$$n(\omega) - 1 = \frac{2}{\pi} \wp \int_0^{\infty} \frac{\zeta \kappa(\zeta)}{\zeta^2 - \omega^2} d\zeta \quad (8.94a)$$

$$\kappa(\omega) = -\frac{2\omega}{\pi} \wp \int_0^{\infty} \frac{n(\zeta) - 1}{\zeta^2 - \omega^2} d\zeta \quad (8.94b)$$

Equation (8.93) and Eq. (8.94) are the K-K relations, which relate the real part of a causal function to an integral of its imaginary part over all frequencies, and vice versa. A number of sum rules can be derived based on the K-K relations and are useful in obtaining or validating the dielectric function of a given material. The K-K relations can be applied to reflectance spectroscopy to facilitate the determination of optical constants from the measured reflectivity of a material in vacuum.<sup>9</sup> For radiation incident from vacuum on a material at normal incidence, the Fresnel reflection coefficient is

$$r(\omega) = |r(\omega)|e^{i\phi(\omega)} = \frac{1 - n(\omega) - i\kappa(\omega)}{1 + n(\omega) + i\kappa(\omega)} \quad (8.95)$$

where  $|r|$  is the amplitude and  $\phi$  the phase of the reflection coefficient. The directional-hemispherical spectral reflectivity, expressed in terms of  $\omega$ , is

$$\rho'_\omega(\omega) = rr^* = |r|^2 \quad (8.96)$$

The amplitude and the phase are related, and it can be shown that

$$\phi(\omega) = -\frac{\omega}{\pi} \wp \int_0^{\infty} \frac{\ln \rho'_\omega(\zeta)}{\zeta^2 - \omega^2} d\zeta \quad (8.97)$$

The refractive index and the extinction coefficient can be calculated, respectively, from

$$n(\omega) = \frac{1 - \rho'_\omega}{1 + \rho'_\omega + 2\cos\phi \sqrt{\rho'_\omega}} \quad (8.98)$$

and

$$\kappa(\omega) = \frac{2\sin\phi \sqrt{\rho'_\omega}}{1 + \rho'_\omega + 2\cos\phi \sqrt{\rho'_\omega}} \quad (8.99)$$

### 8.4.2 The Drude Model for Free Carriers

The Drude model describes frequency-dependent conductivity of metals and can be extended to free carriers in semiconductors. In the absence of an electromagnetic field, free electrons move randomly. When an electromagnetic field is applied, free electrons acquire a nonzero average velocity, giving rise to an electric current that oscillates at the same frequency as the electromagnetic field. The collisions with the stationary atoms result in a damping force on the free electrons, which is proportional to their velocity. The equation of motion for a single free electron is then

$$m_e \ddot{\mathbf{x}} = -m_e \gamma \dot{\mathbf{x}} - e\mathbf{E} \quad (8.100)$$

where  $e$  is the absolute charge of an electron,  $m_e$  is the electron mass, and  $\gamma$  denotes the strength of the damping due to collision, i.e., the *scattering rate* or the inverse of the relaxation time  $\tau$ . Assume the electron motion under a harmonic field  $\mathbf{E} = \mathbf{E}_0 e^{-i\omega t}$  is of the form  $\mathbf{x} = \mathbf{x}_0 e^{-i\omega t}$  so that  $\dot{\mathbf{x}} = -i\omega \mathbf{x}$ . We can rewrite Eq. (8.100) as

$$\dot{\mathbf{x}} = \frac{e/m_e}{i\omega - \gamma} \mathbf{E}$$

The electric current density is  $\mathbf{J} = -n_e e \dot{\mathbf{x}} = \tilde{\sigma}(\omega) \mathbf{E}$ ; therefore, the complex conductivity is

$$\tilde{\sigma}(\omega) = \frac{n_e e^2 / m_e}{\gamma - i\omega} = \frac{\sigma_0}{1 - i\omega/\gamma} \quad (8.101)$$

where  $\sigma_0 = n_e e^2 \tau / m_e$  is the dc conductivity, as discussed in Chap. 5. Equation (8.101) is called the Drude free-electron model, which describes the frequency-dependent complex conductivity of a free-electron system in terms of the dc conductivity and the scattering rate, in a rather simple form. The electrical conductivity approaches to the dc conductivity at very low frequencies (or very long wavelengths). The dielectric function is related to the conductivity by Eq. (8.28); thus,

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{\sigma_0 \gamma}{\varepsilon_0 (\omega^2 + i\gamma\omega)} \quad (8.102)$$

where  $\varepsilon_\infty$ , which is on the order of 1, is included to account for contributions, other than the contribution of the free electrons, that are significant at high frequencies. There exist several transitions at the ultraviolet and visible regions for metals, such as *interband transitions*. Note that when  $\omega \rightarrow \infty$ , the real part of the dielectric function of all materials should approach unity, as can be seen from Eq. (8.93a). In the low-frequency limit when  $\omega \ll \gamma$ ,  $\tilde{\sigma}(\omega \rightarrow 0) \approx \sigma_0$  and  $\varepsilon'' \gg \varepsilon'$ . Therefore,

$$n \approx \kappa \approx \sqrt{\frac{\sigma_0}{2\varepsilon_0\omega}} \quad (8.103)$$

This is the Hagen-Rubén equation and is applicable at very long wavelengths.<sup>1</sup> Both the refractive index and the extinction coefficient will increase with the square root of wavelength in vacuum. It is interesting to note that the radiation penetration depth  $\delta_\lambda = \lambda/(4\pi\kappa)$  will also increase with the square root of wavelength. As an example, consider gold at  $\lambda = 4 \mu\text{m}$  with  $\kappa = 25$ . The penetration depth is 13 nm at this wavelength. If the wavelength is increased to 4 cm, which is well into the microwave region, the penetration depth will increase to 1.3  $\mu\text{m}$ . Generally speaking, metals are highly reflecting in the infrared wavelength region.

The *plasma frequency* is defined according to  $\omega_p^2 = \sigma_0 \gamma / \varepsilon_0 = n_e e^2 / m_e \varepsilon_0$ . Using the plasma frequency, we can write Eq. (8.102) in a more compact form as follows:

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\gamma)} \quad (8.104)$$

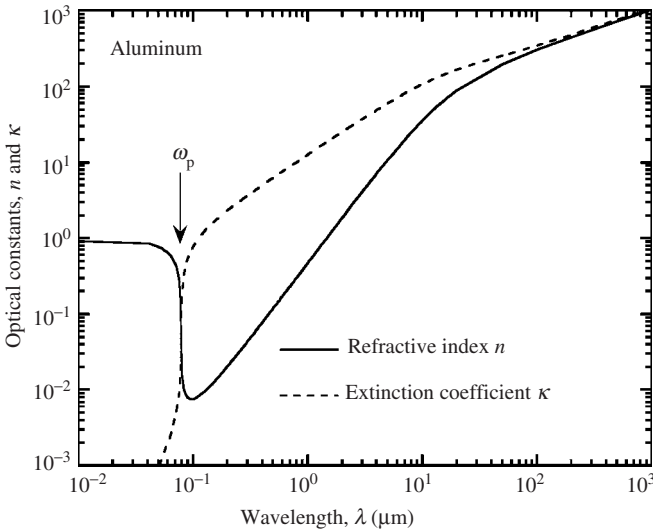
If  $\omega \gg \gamma$ , the dielectric function can be approximated as

$$\varepsilon(\omega) \approx \varepsilon_\infty - \frac{\omega_p^2}{\omega^2} \left( 1 - i \frac{\gamma}{\omega} \right), \quad \text{when } \omega \gg \gamma \quad (8.105)$$

The plasma frequency falls in the ultraviolet region for most metals. For example, the wavelength corresponding to the plasma frequency is approximately 80 nm for aluminum and 200 nm for tungsten. When  $\omega \gg \omega_p$ , as in the x-ray region,  $\epsilon(\omega) \rightarrow 1 + i\gamma\omega_p^2/\omega^3$ . Thus, metals behave more absorbing than reflecting. Take tungsten as an example. At  $\lambda = 1 \text{ nm}$ , the optical constants are  $n \approx 1$  and  $\kappa = 4 \times 10^{-4}$ . The penetration depth is calculated to be  $\delta_\lambda = 200 \text{ nm}$ . Because the refractive index is similar to that of air, the reflection is very low so that all incident radiation will be absorbed within 1- $\mu\text{m}$  skin depth. Some metals become rather transparent; for example, the radiation penetration depth in lithium is close to 100  $\mu\text{m}$  at  $\lambda = 1 \text{ nm}$ . The Center for X-Ray Optics at Lawrence Berkeley National Laboratory maintains a website on x-ray properties, which can be accessed at [http://www-cxro.lbl.gov/optical\\_constants](http://www-cxro.lbl.gov/optical_constants). If  $\omega < \omega_p$ , the real part of the dielectric function  $\epsilon'$  becomes negative, and the extinction coefficient is much greater than the refractive index, i.e.,  $\kappa \gg n$ . This corresponds to a high reflectivity, according to Eq. (8.84). A vanishing real part of the refractive index corresponds to a longitudinal collective oscillation of the electron gas, i.e., a *plasma oscillation*. Plasma oscillations originate from a long-range correlation of electrons caused by Coulomb forces.

**Example 8-6.** From Table 5.2, calculate the plasma frequency and the electron scattering rate for aluminum. Calculate the dielectric function, and compare the normal reflectivity with data.

**Solution.** For aluminum, at near room temperature,  $n_e = 18.1 \times 10^{28} \text{ m}^{-3}$  and  $\sigma_0 = 1/r_c = 3.75 \times 10^7 \text{ (m} \cdot \Omega)^{-1}$ . From Appendix A,  $e = 1.602 \times 10^{-19} \text{ C}$ ,  $m_e = 9.109 \times 10^{-31} \text{ kg}$ , and  $\epsilon_0 = 8.854 \times 10^{-12} \text{ C}^2/(\text{N} \cdot \text{m}^2)$ . Hence,  $\gamma = n_e e^2/m_e \sigma_0 = 1.4 \times 10^{14} \text{ rad/s}$ , or the scattering time  $\tau = 7.2 \times 10^{-14} \text{ s}$ , and  $\omega_p = 2.4 \times 10^{16} \text{ rad/s}$ , which corresponds to a wavelength of 79 nm. The exact parameters may differ slightly in different references, and sometimes, an effective mass is used which is slightly larger than the electron mass  $m_e$ . The predicted optical constants are plotted in Fig. 8.10,



**FIGURE 8.10** Optical constants of aluminum, calculated from the Drude model.

assuming  $\epsilon_\infty = 1$ . It can be seen that as the wavelength exceeds 100  $\mu\text{m}$ , the difference between  $n$  and  $\kappa$  diminishes. In the region  $0.1 \mu\text{m} < \lambda < 200 \mu\text{m}$ ,  $n < \kappa$  so that the real part of the dielectric

function  $\epsilon' = n^2 - \kappa^2$  becomes negative. A sharp transition occurs at the plasma frequency so that  $n \rightarrow 1$  and  $\kappa$  decreases rapidly toward higher frequencies.

The reflectivity calculated from Eq. (8.84) is compared with the measured data for an aluminum film, prepared by ultrahigh vacuum deposition, and measured in high vacuum to avoid oxidation.<sup>9</sup> The results agree very well at wavelengths greater than 2  $\mu\text{m}$  (see Fig. 8.11). For  $\lambda < 1 \mu\text{m}$ , the

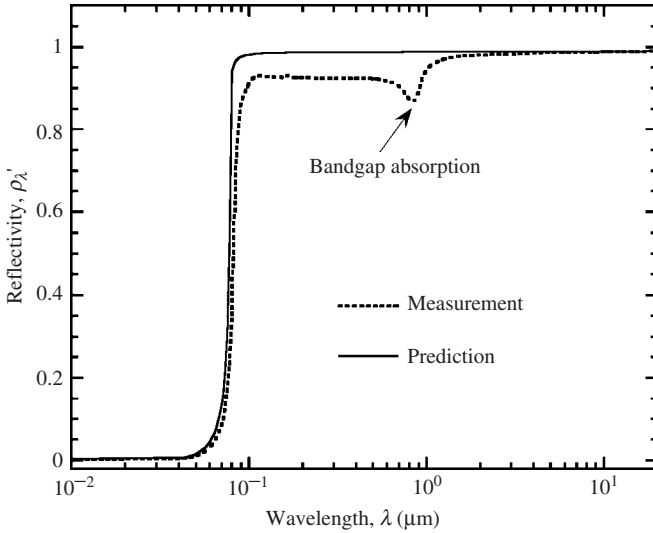


FIGURE 8.11 Normal spectral reflectivity of aluminum.

contribution from the bandgap transition and bound electrons causes a reduction in the reflectivity. Note that the Drude model did not include these effects and is applicable for long wavelengths only. The established optical constants of metals are based on the measured reflectivity in a broad spectral region and the K-K transformation discussed in Sec. 8.4.1. The results for a large number of samples are tabulated in *Handbook of the Optical Constants of Solids*, with pertinent references.<sup>9</sup>

In some studies, the Drude model is modified by considering the temperature and frequency dependence of the scattering rate and the effective mass. While the Drude model predicts well the radiative properties at room temperature or above, caution should be taken at extremely low temperatures. If the electron mean free path becomes comparable to the distance over which the electric field varies, i.e., the field penetration depth, nonlocal effects become important and the Drude theory breaks down. This can occur at cryogenic temperatures, and a more complex theory called the *anomalous skin effect* theory must then be applied.<sup>22</sup>

### 8.4.3 The Lorentz Oscillator Model for Lattice Absorption

Vibrations of lattice ions and bound electrons contribute to the dielectric function in a certain frequency region, often in the mid-infrared. The refractive index can be calculated using the Lorentz oscillator model, which assumes that a bound charge  $e$  is accelerated by the local electric field  $\mathbf{E}$ . In contrast to free electrons, a bound charge experiences a restoring force determined by a spring constant  $K_j$ . The oscillator is further assumed to have a

mass  $m_j$  and a damping coefficient  $\gamma_j$ , as shown in Fig. 8.12. The force balance yields the equation of motion for the oscillator:

$$m_j \ddot{\mathbf{x}} + m_j \gamma_j \dot{\mathbf{x}} + K_j \mathbf{x} = e \mathbf{E} \tag{8.106}$$

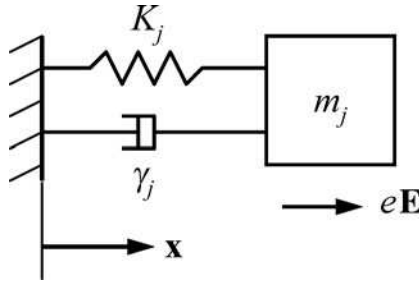


FIGURE 8.12 The classical oscillator model.

when  $\mathbf{E}$  is a harmonic field. The change  $e$  is conventionally taken as positive in the Lorentz model. There exists a solution, valid at timescales longer than the relaxation time, given by

$$\mathbf{x} = \frac{e/m_j}{\omega_j^2 - i\gamma_j\omega - \omega^2} \mathbf{E} \tag{8.107}$$

where  $\omega_j = (K_j/m_j)^{1/2}$  is the resonance frequency of the  $j$ th oscillator. The motion of the single oscillator causes a dipole moment  $e\mathbf{x}$ . If the number density of the  $j$ th oscillator is  $n_j$ , the polarization vector, or the dipole moment per unit volume, is  $\mathbf{P} = \sum_{j=1}^N n_j e\mathbf{x}$ , where  $N$  is the total number of active phonon modes (oscillators). The constitutive relation gives the polarization as  $\mathbf{P} = (\epsilon - 1)\epsilon_0 \mathbf{E}$ . It can be shown that

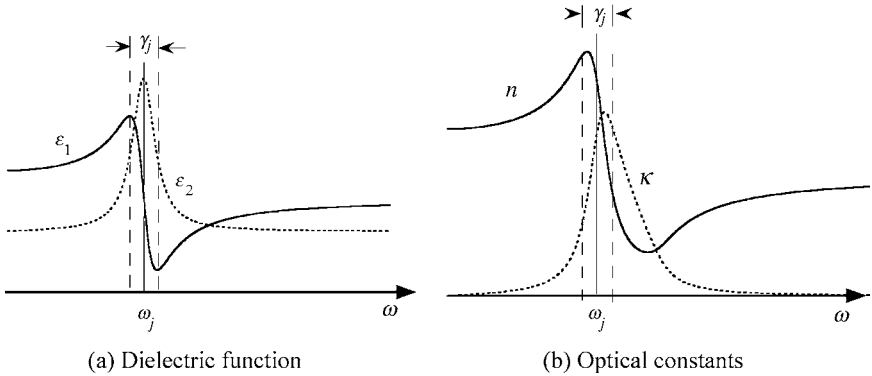
$$\epsilon(\omega) = \epsilon_\infty + \sum_{j=1}^N \frac{S_j \omega_j^2}{\omega_j^2 - i\gamma_j\omega - \omega^2} \tag{8.108}$$

where  $\epsilon_\infty$  is a high-frequency constant and  $S_j = \omega_{pj}^2/\omega_j^2 = n_j e^2/(\epsilon_0 m_j \omega_j^2)$  is called the *oscillator strength*.

At very low frequencies,  $\epsilon(0) = \epsilon_\infty + \sum_{j=1}^N S_j$ , which is called the dielectric constant.

The real and imaginary parts of the dielectric function and the refractive index for a simple oscillator are illustrated in Fig. 8.13, near the resonance frequency for  $\epsilon_\infty = 1$ . It can be seen from Eq. (8.108) and Fig. 8.13 that, for frequencies much lower or much higher than the resonance frequency, the extinction coefficient of the oscillator is negligible. Only within an interval of  $\gamma_j$  around the resonance frequency is the absorption appreciable. Within the absorption band, the real part of the refractive index decreases with frequency; this phenomenon is called *anomalous dispersion*. It follows that in an interval of width  $\gamma_j$  around the resonance frequency, the Lorentz oscillator is highly reflecting and absorbing, while for higher or lower frequencies, it acts as a transparent material. A more complicated treatment based on quantum mechanics yields a four-parameter model.<sup>23</sup> The previous





**FIGURE 8.13** The Lorentz oscillator model. (a) The dielectric function. (b) Optical constants.

classical oscillator model can be considered as the approximation when the relaxation time of the longitudinal and transverse optical phonons are the same. In some studies, frequency- and temperature-dependent scattering rate is also considered to model the infrared spectra.

In practice, it is much more difficult to predict the parameters of Lorentz oscillators from other data of the material than to predict the Drude parameters. In practice, the oscillator parameters are often treated as adjustable parameters that are determined by fitting Eq. (8.108) to the measured reflectivity data. The Lorentz model has been applied to a large number of dielectric materials by fitting the reflectance spectra.<sup>9</sup> Zhang et al. (*J. Opt. Soc. Am. B*, **11**, 2252, 1994; *Int. J. Thermophys.*, **19**, 905, 1998) obtained the Lorentz parameters for several perovskite crystals (LaAlO<sub>3</sub>, LaGaO<sub>3</sub>, and NdGaO<sub>3</sub>) and thin polyimide films.

**Example 8-7.** The Lorentz model for SiC at room temperature for ordinary ray is given as follows:

$$\epsilon(\omega) = \epsilon_\infty \left[ 1 + \frac{\omega_{\text{LO}}^2 - \omega_{\text{TO}}^2}{\omega_{\text{TO}}^2 - i\gamma\omega - \omega^2} \right] \quad (8.109)$$

where  $\omega_{\text{LO}} = 969 \text{ cm}^{-1}$  and  $\omega_{\text{TO}} = 793 \text{ cm}^{-1}$  are the frequencies corresponding to the longitudinal and transverse optical phonons, respectively,  $\gamma = 4.76 \text{ cm}^{-1}$ , and  $\epsilon_\infty = 6.7$ .<sup>24</sup> What is the refractive index at the high- and low-frequency limits? Calculate the normal reflectivity, and compare it with the experimental result.

**Solution.** Comparing Eq. (8.108) and Eq. (8.109), we see that the resonance frequency corresponds to the TO phonon frequency, and the oscillation strength is  $S_1 = \epsilon_\infty(\omega_{\text{LO}}^2/\omega_{\text{TO}}^2 - 1) = 3.3$ . The high-frequency limit of the refractive index is  $n \approx \sqrt{\epsilon_\infty} = 2.6$ , and the low-frequency limit is  $n = \sqrt{\epsilon_\infty + S_1} = 3.16$ . Note that transitions that occur in the visible and ultraviolet regions are not included so that the high-frequency limit is approximately  $1 \mu\text{m}$ . On the other hand, because there are no other transitions at long wavelengths, the dielectric constant is approximately the same for zero frequency. The normal reflectivity is calculated using Eq. (8.84) and compared with the data, as shown in Fig. 8.14. The agreement is excellent since the Lorentz parameters were fitted to the experimental data.<sup>24</sup> The phonon band causes a large  $\kappa$  value and hence a high reflectivity (very low emissivity) between  $\omega_{\text{LO}} = 969 \text{ cm}^{-1}$  and  $\omega_{\text{TO}} = 793 \text{ cm}^{-1}$ . This band is called *reststrahlen band* of the reststrahlen reflection. At  $\omega = 1000 \text{ cm}^{-1}$ , the reflectivity is nearly 0 so that the emissivity is almost 1. This happens at the edge of the reststrahlen band, when the refractive index increases passing 1 and the extinction coefficient decreases to a very small value. This wavelength is called *Christiansen wavelength*, and the associated phenomenon is called the *Christiansen effect*.<sup>7</sup>

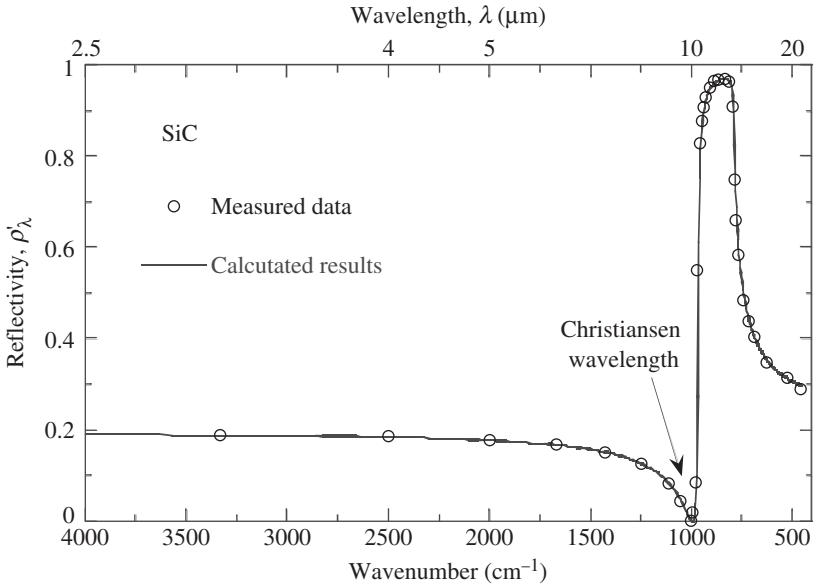


FIGURE 8.14 The calculated and measured normal reflectivity of SiC at room temperature.

### 8.4.4 Semiconductors

The absorption coefficient of lightly doped silicon is shown in Fig. 8.15 to illustrate the contribution of different mechanisms.<sup>17,25</sup> Let us look at the absorption of silicon in the visible

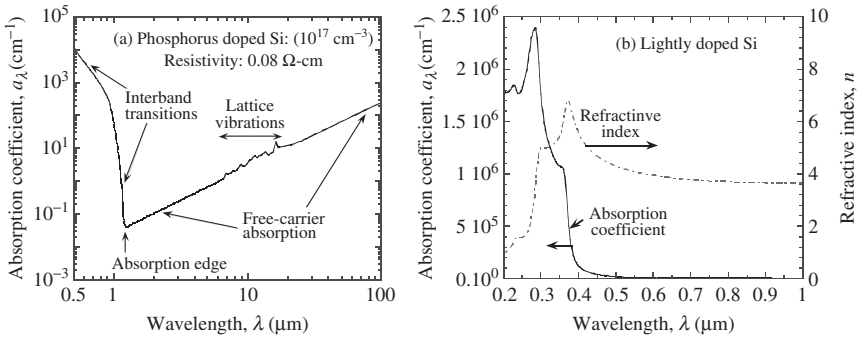
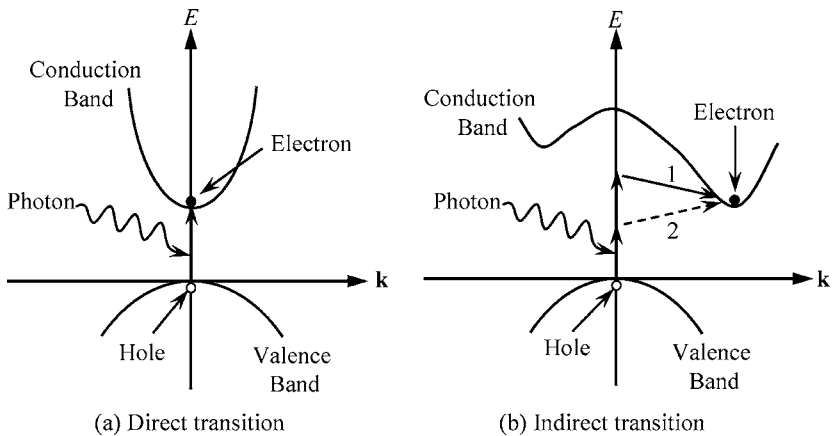


FIGURE 8.15 The absorption coefficient and the refractive index of Si at room temperature. (a) Absorption coefficient in the visible and the infrared. (b) Absorption coefficient and refractive index from the ultraviolet to the near infrared.

and the infrared first, as shown in Fig. 8.15a. At short wavelengths, photon energies are large enough to excite electrons from the valence band to the conduction band. This interband transition causes the absorption coefficient to rise quickly as the photon energy  $h\nu$  is increased above the indirect bandgap, which is approximately  $E_g = 1.1$  eV at room

temperature and decreases somewhat as temperature increases. As the wavelength further increases beyond the absorption edge ( $\lambda \approx 1.1 \mu\text{m}$ ), the absorption coefficient is affected by the existence of impurities and defects, absorption by free-carriers (i.e., intraband or intersubband transitions by electrons and holes), and absorption by lattice vibrations. While the lattice vibration affects certain regions of the spectrum, the free-carrier contribution increases at longer wavelengths. For intrinsic silicon at low temperatures, the free-carrier concentration is very low, and thus, silicon is transparent at wavelengths longer than the bandgap wavelength. Lattice absorption occurs in the mid-infrared and introduces some absorption for  $6 \mu\text{m} < \lambda < 25 \mu\text{m}$ . Free-carrier absorption is important for doped silicon at longer wavelengths. Note that even for intrinsic silicon at high temperatures, thermally excited free carriers dominate the absorption at longer wavelengths; a 0.5-mm-thick silicon wafer is essentially opaque above 1000 K. The free-carrier concentration for intrinsic silicon is about  $10^{10} \text{ cm}^{-3}$  at 300 K and nearly  $10^{18} \text{ cm}^{-3}$  at 1000 K. As shown in Fig. 8.15b, the absorption coefficient continues to increase toward shorter wavelengths due to the interband transition associated with the direct bandgap, which dominates the optical characteristics of silicon in the ultraviolet region. This transition also affects the refractive index of silicon at longer wavelengths. Beyond 500 nm, the refractive index of lightly doped Si decreases somewhat as the wavelength increases.

Modeling the *interband transitions* requires the quantum theory and is more complicated. In a direct-bandgap semiconductor, shown in Fig. 8.16a, the lowest point of the



**FIGURE 8.16** Interband transitions in semiconductors. (a) Direct transition without involving a phonon. (b) Indirect transition involving the emission or absorption of a phonon.

conduction band occurs at the same wavevector as the highest point of the valence. An electron can be excited from the top of the valence band to the bottom of the conduction band by absorbing a photon of energy that is at least equal to the bandgap energy. When the valence band and the conduction band are parabola-like, the absorption coefficient due to direct bandgap absorption can be expressed as

$$a_{\text{bg}} = A(\hbar\omega - E_g)^{1/2} \quad (8.110)$$

where  $A$  is a parameter that depends on the effective masses of the electrons and the holes, and the refractive index of the material.

When a transition requires a change in both energy and momentum, as in the case of an indirect bandgap semiconductor, shown in Fig. 8.16*b*, a phonon is either emitted (process 1) or absorbed (process 2) for momentum conservation because the photon itself cannot provide a change in momentum. This kind of transition is called *indirect interband transition*. With the involvement of phonons, the absorption coefficient is given as

$$a_a(\omega) = \frac{B(\hbar\omega - E_g + \hbar\omega_{\text{ph}})^2}{\exp(\hbar\omega_{\text{ph}}/k_B T) - 1}, \quad \hbar\omega > E_g - \hbar\omega_{\text{ph}} \quad (8.111)$$

and

$$a_e(\omega) = \frac{C(\hbar\omega - E_g - \hbar\omega_{\text{ph}})^2}{1 - \exp(-\hbar\omega_{\text{ph}}/k_B T)}, \quad \hbar\omega > E_g + \hbar\omega_{\text{ph}} \quad (8.112)$$

where  $a_a$  and  $a_e$  correspond to the absorption coefficients for transitions with phonon absorption and emission, respectively, and their values are nonzero only when the photon energy is greater than the bandgap energy subtracted (or added) by the phonon energy. In the preceding equations,  $B$  and  $C$  are temperature-dependent materials parameters. There may be several phonon modes that can cause indirect interband transitions, and their effects on the absorption coefficient can be superimposed;<sup>25</sup> also see Forouhi and Bloomer (*Phys. Rev. B*, **38**, 1865, 1988), Albrecht et al. (*Phys. Rev. Lett.*, **80**, 4510, 1998), Benedict et al. (*Phys. Rev. B*, **57**, R9385, 1998), and Rohlfing and Louie (*Phys. Rev. Lett.*, **81**, 2312, 1998).

The Drude model can be applied to model the free-carrier contribution for both intrinsic and doped silicon as given in the following:

$$\varepsilon(\omega) = \varepsilon_{\text{bi}} - \frac{N_e e^2 / \varepsilon_0 m_e^*}{\omega^2 + i\omega\gamma_e} - \frac{N_h e^2 / \varepsilon_0 m_h^*}{\omega^2 + i\omega\gamma_h} \quad (8.113)$$

where the first term on the right  $\varepsilon_{\text{bi}}$  accounts for contributions by transitions across the bandgap and lattice vibrations, the second term is the Drude term for transitions in the conduction band (free electrons), and the last term is the Drude term for transitions in the valence band (free holes).<sup>25,26</sup> Here,  $N_e$  and  $N_h$  are the concentrations,  $m_e^*$  and  $m_h^*$  the effective masses, and  $\gamma_e$  and  $\gamma_h$  the scattering rates of free electrons and holes, respectively. The effective masses are taken as  $m_e^* = 0.27m_e$  and  $m_h^* = 0.37m_e$ .

The value of  $\varepsilon_{\text{bi}}$  is determined using the refractive index and the extinction coefficient of intrinsic silicon. The refractive index of silicon changes from 3.6 at  $\lambda = 1 \mu\text{m}$  to 3.42 at  $\lambda > 10 \mu\text{m}$  at room temperature and increases slightly toward higher temperatures. Absorption by lattice vibrations occurs in silicon at wavelengths between 6 and 25  $\mu\text{m}$ . To account for the lattice absorption, the extinction coefficients are taken from the tabulated values in *Handbook of the Optical Constants of Solids*.<sup>9</sup> At elevated temperatures or for heavily doped silicon, the effect of absorption by lattice vibrations is negligible compared to the absorption by free carriers. The carrier concentrations and the scattering rates are functions of temperature and dopant concentrations. For bulk silicon, the scattering is caused by the collision of electrons or holes with the lattice (phonons) or ionized dopant sites (impurities or defects). The total scattering rates can be calculated by

$$\gamma_e = \gamma_{e\text{-ph}} + \gamma_{e\text{-d}} \quad \text{and} \quad \gamma_h = \gamma_{h\text{-ph}} + \gamma_{h\text{-d}} \quad (8.114)$$

Here again, the subscripts ph and d stand for phonon and defects, respectively. Generally speaking, the scattering rate increases with the defect concentration and temperature. The carrier concentrations depend on temperature and dopant concentrations. For intrinsic silicon, the concentration,  $N_{\text{th}}$ , of the thermally excited free electrons and holes are the same and can be found from the relation:

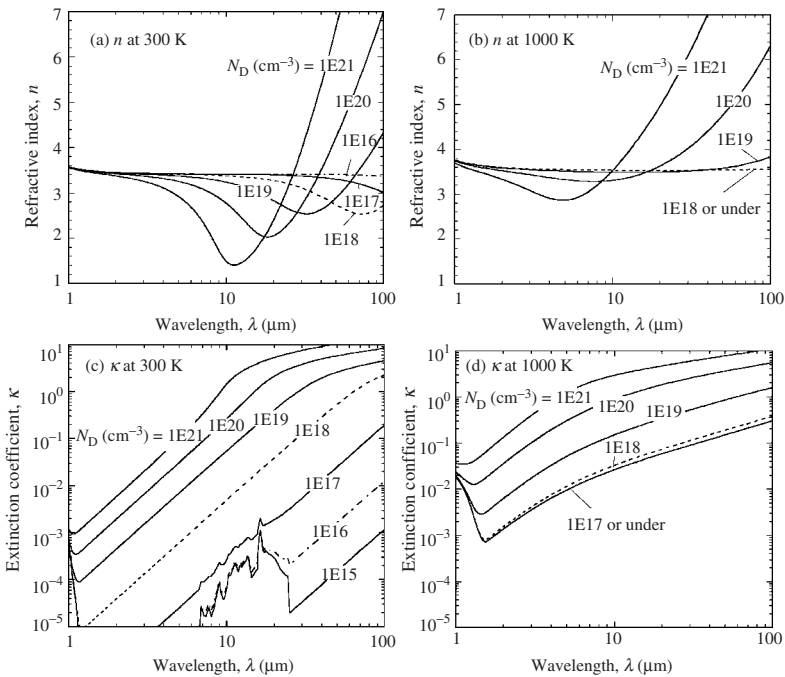
$$N_{\text{th}}^2 = N_c N_v \exp(-E_g/k_B T) \quad (8.115)$$

where  $N_C$  and  $N_V$  are the effective densities of states in the conduction band and the valence band, respectively, and for silicon,  $E_g = 1.17 - 0.000473T^2/(T + 636)$  eV. Note that  $N_C = 2.86 \times 10^{19} \text{ cm}^{-3}$  and  $N_V = 2.66 \times 10^{19} \text{ cm}^{-3}$  at 300 K, and both increase with temperature proportional to  $T^{3/2}$ . When the dopant concentrations are not very high, the free-carrier concentrations can be obtained from

$$N_e = \frac{1}{2} \left[ N_D - N_A + \sqrt{(N_D - N_A)^2 + 4N_{th}^2} \right] \tag{8.116}$$

and  $N_h = N_{th}^2/N_e$  when the majority impurities are  $n$ -type. When the majority impurities are  $p$ -type, the equations become  $N_h = \frac{1}{2} [N_A - N_D + \sqrt{(N_A - N_D)^2 + 4N_{th}^2}]$  and  $N_e = N_{th}^2/N_h$ . Equation (8.116) has been derived based on complete ionization, which may not hold for heavily doped semiconductors or at very low temperatures. Integration is needed to determine the concentration when complete ionization is not expected. The procedure has been implanted in the accompanied software *Rad-Pro* (radiative properties) for calculating the radiative properties of silicon at wavelengths longer than 0.5  $\mu\text{m}$ ; see [www.me.gatech.edu/~zzhang](http://www.me.gatech.edu/~zzhang) under Software Tool to download the free software *Rad-Pro*.

The calculated optical constants  $n$  and  $\kappa$  of silicon, for wavelengths in the range between 1 and 100  $\mu\text{m}$ , are shown in Fig. 8.17 at 300 K and 1000 K for  $n$ -type phosphorus donors. The



**FIGURE 8.17** Optical constants of  $n$ -type phosphorus-doped silicon, at 300 K and 1000 K, for different dopant concentrations.

refractive index changes little for lightly doped silicon, even at high temperatures. The refractive index for heavily doped silicon first decreases and then increases abruptly toward longer wavelengths. The carrier contribution to the extinction coefficient at 300 K is very

small for lightly doped silicon, and the lattice contribution can be clearly seen between 6 and 25  $\mu\text{m}$ . As the doping level exceeds  $10^{17} \text{ cm}^{-3}$ , these phonon features are screened out. This is also true for lightly doped silicon at 1000 K as the thermally excited carriers have a concentration of about  $10^{18} \text{ cm}^{-3}$ . At 1000 K,  $\kappa$  is essentially the same for  $N_D \leq 10^{17} \text{ cm}^{-3}$  and increases with higher dopant concentrations. At 300 K, however, except for the interband absorption ( $\lambda < 1.12 \mu\text{m}$ ) and in the lattice absorption ( $6 \mu\text{m} < \lambda < 25 \mu\text{m}$ ), the calculated  $\kappa$  decreases with reducing dopant concentration until  $N_D$  is less than  $10^{10} \text{ cm}^{-3}$ , when most carriers are from the thermal excitation rather than the doping. The significance is that the penetration depth, which is the inverse of the absorption coefficient, can be very large because of the small  $\kappa$  values. Generally speaking, for doping levels under  $10^{18} \text{ cm}^{-3}$ ,  $\kappa \ll n$  unless the wavelength is very long, and silicon behaves as a dielectric. For heavily doped silicon, on the other hand, the Drude model predicts that  $n \approx \kappa$  in the long-wavelength limit, just like in a metal. The accuracy of the simple Drude model is subject to a number of factors, such as the dependence of the effective mass on temperature, dopant concentration, and even frequency. The scattering rate may be frequency dependent as well. Furthermore, the ionization energy depends on the dopant concentration for heavily doped silicon. Nevertheless, this model has captured the essential features of the dielectric function of silicon, for wavelengths greater than 0.5  $\mu\text{m}$ , at temperatures from 300 to 1200 K, and with a doping level up to  $10^{19} \text{ cm}^{-3}$ .

### 8.4.5 Superconductors

A *superconductor* is a material that exhibits zero resistance and perfect diamagnetism when it is maintained at temperatures below the critical temperature  $T_c$ , under a bias current less than the critical current and an applied magnetic field less than the critical magnetic field. The discovery of high-temperature superconductors in the late 1980s has generated tremendous excitement in the public because the achievement of superconductivity above the boiling temperature of nitrogen (77 K at atmospheric pressure) offers many technological promises. More and more materials have been found to be superconducting at higher and higher temperatures. Extensive studies have been devoted to the infrared properties of superconducting films for applications as radiation detectors, optical modulators, and other optoelectronic devices.<sup>27</sup> High-temperature superconducting (HTS) materials are made of ceramic structures, such as  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ , where  $\delta$  is between 0 and 1. The Y-Ba-Cu-O compound behaves as an insulator when  $\delta > 0.6$  and as a conductor when  $\delta < 0.2$  at room temperature.

In the normal state ( $T > T_c$ ), the dielectric function  $\varepsilon(\omega)$  can be modeled as a sum of the free-electron contribution using the Drude model, an intraband absorption that is important for the mid-infrared region by using the Lorentz term, and a high-frequency constant:<sup>28</sup>

$$\varepsilon(\omega) = \varepsilon_\infty + \varepsilon_{\text{Mid-IR}} + \varepsilon_{\text{Drude}} \quad (8.117)$$

The expression of the Drude term is the same as Eq. (8.102) or Eq. (8.104). Although phonon contributions can be neglected compared to the large electronic contributions, a broadband mid-infrared electronic absorption often exists in the HTS materials, which is typically modeled with a Lorentz oscillator that has a large width, or a frequency-dependent scattering rate.

Many properties of superconductors can be explained in terms of a two-fluid model that postulates that a *fluid of normal electrons coexists with a superconducting electron fluid*. These two fluids coexist but do not interact. According to the BCS theory, interaction between a pair of free electrons and a phonon (or other thermally generated excitations) leads to the formation of an electron pair, called *Cooper pair*.<sup>29</sup> The Cooper pairs cannot be scattered by any sources as they move in the lattice structure. In the superconducting state, only a fraction of free electrons  $f_s$  are in the condensed phase (or superconducting state) and the remaining electrons are in the normal state. The value of  $f_s$  is temperature dependent

and goes to zero at  $T_c$ . The contribution of the superconducting electrons to the dielectric function is

$$\epsilon_{\text{Sup}} = -\frac{\omega_p^2}{\omega^2} + i\pi\delta(\omega)\frac{\omega_p^2}{\omega} \quad (8.118)$$

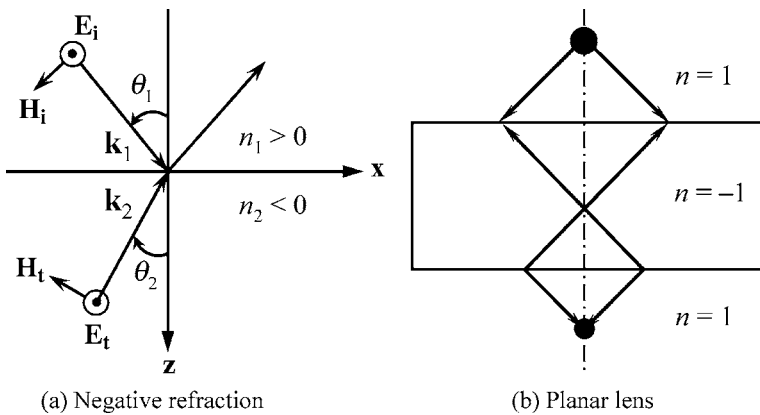
where  $\delta(\omega)$  is the Dirac delta function that equals zero at nonzero frequencies. The Drude term remains due to the presence of normal electrons with a number density of  $(1 - f_s)n_e$ . The dielectric function in the superconducting state can be modeled by

$$\epsilon(\omega) = \epsilon_\infty + \epsilon_{\text{Mid-IR}} + (1 - f_s)\epsilon_{\text{Drude}} + f_s\epsilon_{\text{Sup}} \quad (8.119)$$

The calculated results are usually fitted with the experimental measurements by adjusting the plasma frequency, the scattering rate, and the fraction of superconducting electrons. Excellent agreement has been observed between the predicted and experimental values of both the transmittance and the reflectance of superconducting films, at temperatures ranging from 300 down to 10 K.<sup>28</sup>

### 8.4.6 Metamaterials with a Magnetic Response

The concept of negative refractive index ( $n < 0$ ) was first postulated by Victor Veselago for a hypothetical material that has both negative permittivity and permeability in the same frequency region. In this case, the sign of  $n$  should be chosen as negative in  $n = \pm\sqrt{\epsilon\mu}$ . Many of the unique features associated with negative index materials (NIMs) were summarized in Veselago's original paper (*Sov. Phys. Usp.*, **10**, 509, 1968), such as negative phase velocity, reversed Doppler effect, and the prediction of a planar lens. As illustrated in Fig. 8.18a, if  $n$  is negative, the phase speed will be negative and light incident from a



**FIGURE 8.18** Unique features of a negative index material (NIM). (a) The refracted ray bends toward the same side as the incidence. (b) A slab of NIM can focus light like a lens does. Arrows indicate the wavevector directions. Note that the energy direction is the opposite of the wavevector direction in a NIM.

conventional positive index material (PIM) to a NIM will be refracted to the same side as the incidence. This is called *bending light in the wrong way*. Furthermore, if light can be bent differently, then a planar slab of a NIM can focus light as shown in Fig. 8.18b. The lack of simultaneous occurrence of negative  $\epsilon$  and  $\mu$  in natural materials hindered further

study on NIMs for some 30 years. On the basis of the theoretical work by John Pendry and coworkers in the late 1990s. Shelby et al. (*Science*, **292**, 77, 2001) first demonstrated that a metamaterial exhibits negative refraction at x-band microwave frequencies. In a NIM medium, the phase velocity of an electromagnetic wave is opposite to its energy flux. The electric field, the magnetic field, and the wavevector form a left-handed triplet. For this reason, NIMs are also called left-handed materials (LHMs). Because both  $\epsilon$  and  $\mu$  are simultaneously negative, NIMs are also called double negative (DNG) materials.

Pendry (*Phys. Rev. Lett.*, **85**, 3966, 2000) conceived that a NIM slab with  $\epsilon = \mu = -1$  would perform the dual function of correcting the phase of the propagating components and amplifying the evanescent components, which exist only in the near field of the object. The combined effects could make a perfect lens that eliminates the limitations on image resolution imposed by diffraction for conventional lenses. Despite the doubt cast by some researchers on the concept of “perfect lens” and even on negative refraction, both hypotheses of negative refraction and the ability to focus light by a slab of NIM have been verified by analytical, numerical, and experimental methods. Potential applications of NIMs range from nanolithography to novel Bragg reflectors, phase-compensated cavity resonators, waveguides, and enhanced photon tunneling for microscale energy conversion devices; see Zhang and Fu (*Appl. Phys. Lett.*, **80**, 1097, 2002). Ramakrishna gave an extensive bibliographic review on the theoretical and experimental investigations into NIMs and relevant materials.<sup>30</sup> There has been growing interest in the study of NIMs because of the promising new applications as well as the intriguing new physics. The search of new ways of constructing NIMs also calls for the development of new materials and processing techniques.

The ideal case, where  $\epsilon = \mu = -1$ , cannot exist at more than a single frequency because both  $\epsilon$  and  $\mu$  of a NIM must be inherently dependent on the frequency as required by the causality. In addition, real materials possess losses, and hence, both  $\epsilon$  and  $\mu$  are complex. The negative index can be realized by considering the complex plane, as illustrated in Fig. 8.19. Note that  $\epsilon = r_\epsilon e^{i\phi_\epsilon}$  and  $\mu = r_\mu e^{i\phi_\mu}$ . Then, we have

$$n = r_n e^{i\phi_n} = \sqrt{r_\epsilon r_\mu} e^{i(\phi_\epsilon + \phi_\mu)/2} \tag{8.120}$$

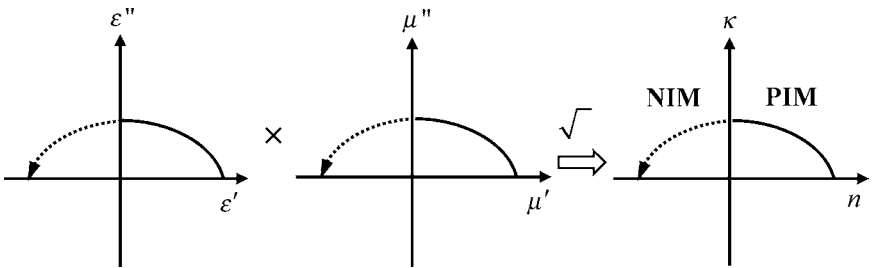


FIGURE 8.19 Illustration of a negative refractive index, using the complex planes.

Therefore, if both  $\epsilon'$  and  $\mu'$  are negative,  $n$  will be negative, but  $\kappa$  will always be positive. Note that a negative  $n$  can be obtained as long as  $\phi_n > \pi/2$ . Generally speaking, one would like to see all the phase angles to be close to  $\pi$  so that the loss is minimized. Note that the principal value of the phase is chosen from 0 to  $2\pi$  in the earlier discussion, rather than from  $-\pi$  to  $\pi$ . If the latter is chosen, one would obtain a negative  $\epsilon$  and a positive  $n$  for a NIM. Many metals and polar dielectrics have a negative  $\epsilon$  in the visible and the infrared. Furthermore, periodic structures of thin metal wires or strips can dilute the average concentration of electrons and shift the plasma frequency to the far-infrared or longer



wavelengths. Negative- $\mu$  materials rarely exist in nature, at the optical frequencies, but can be obtained using metamaterials consisting of split-ring resonator structures at microwave frequencies. These structures can be scaled down to achieve negative  $\mu$  toward higher frequencies. The combination of repeated unit cells of interlocking copper strips and split-ring resonators makes a metamaterial to exhibit a negative  $\epsilon$  and  $\mu$  simultaneously. Based on an effective-medium approach, the relative permittivity and permeability of a NIM can be expressed as functions of the angular frequency  $\omega$  as follows:

$$\epsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma_e\omega} \tag{8.121}$$

and

$$\mu(\omega) = 1 - \frac{F\omega^2}{\omega^2 - \omega_0^2 + i\gamma_m\omega} \tag{8.122}$$

where  $\omega_p$  is the effective plasma frequency,  $\omega_0$  is the effective resonance frequency,  $\gamma_e$  and  $\gamma_m$  are the damping terms, and  $F$  is the fractional area of the unit cell occupied by the split ring. From Eq. (8.121) and Eq. (8.122), both negative  $\epsilon$  and  $\mu$  can be realized in a frequency range between  $\omega_0$  and  $\omega_p$  for adequately small  $\gamma_e$  and  $\gamma_m$ . Here, the values of  $\omega_0$ ,  $\omega_p$ ,  $\gamma_e$ ,  $\gamma_m$  and  $F$  depend on the geometry of the unit cell that constructs the metamaterial. These structures can be scaled down to achieve negative index toward higher frequencies.

To illustrate the negative index behavior, Fig. 8.20 shows the calculated refractive index and the extinction coefficient of a hypothetical NIM using the following parameters:  $\omega_0 = 0.5\omega_p$ ,  $F = 0.785$ , and  $\gamma_e = \gamma_m = \gamma = 0.0025\omega_p$ .<sup>31</sup> Because of the scaling capability of the metamaterial, the frequency is normalized to  $\omega_p$ . It can be seen that in the frequency

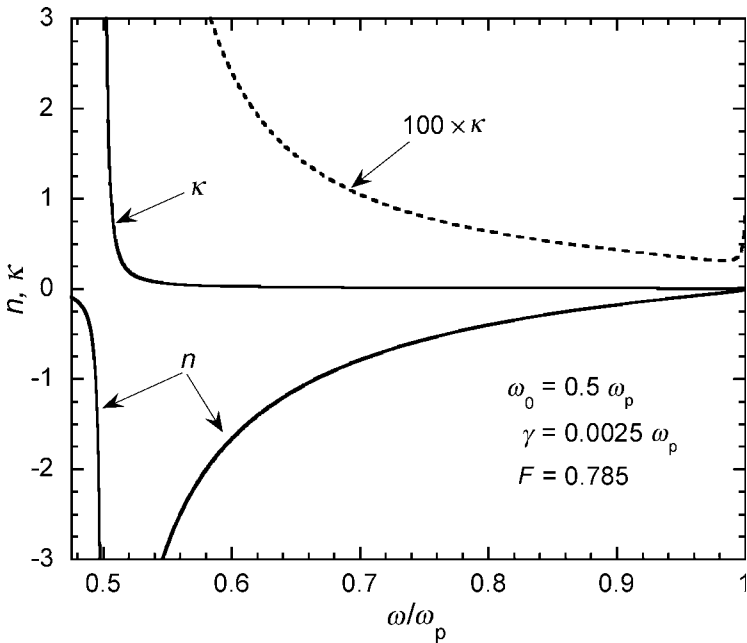


FIGURE 8.20 Calculated refractive index of a hypothetical negative index material (NIM).

range from  $\omega_0$  to  $\omega_p$ , where the real parts of  $\epsilon$  and  $\mu$  are negative,  $n$  is negative and  $\kappa$  (for small values of  $\gamma$ ) is small at frequencies not too close to  $\omega_0$ .

The study of NIMs is in its early stage, and new results are emerging every week. In addition to the use of lithographic techniques to scale down the split-ring dimensions, other approaches, such as nanowire pairs, self-assemblies, and photonic crystals, have also been proposed to realize NIMs in three dimensions and, more importantly, toward short wavelengths. There also exist theoretical challenges in terms of unambiguous determination of the electric and magnetic properties. How can the four parameters,  $\epsilon'$ ,  $\epsilon''$ ,  $\mu'$ , and  $\mu''$ , be determined from the reflectance measurement only? Can the K-K transform be applied to a magnetic material when there are two pairs of complex variables? If so, what is the minimum set of quantities that must be measured by experiments?

## 8.5 SUMMARY

---

In this chapter, we started from the Maxwell equations to derive the plane wave equation and defined the material's optical properties. The Planck law was derived using statistical mechanics, and the radiation entropy was then introduced. The reflection and refraction of waves at a smooth interface were discussed to relate radiative properties of surfaces with the electromagnetic wave theory. The last part of the chapter presented the dielectric functions for metals, dielectrics, semiconductors, and superconductors. At the very end, we also introduced the concept of NIM or DNG materials, as well as their unique features. In the subsequent two chapters, more extensive discussions will be given on thin films, gratings, rough surfaces, as well as evanescent waves, surface polaritons, and near-field energy transfer.

## REFERENCES

---

1. R. Siegel and J. R. Howell, *Thermal Radiation Heat Transfer*, 4th ed., Taylor & Francis, New York, 2002.
2. M. F. Modest, *Radiative Heat Transfer*, 2nd ed., Academic Press, San Diego, CA, 2003.
3. E. S. Barr, "Historical survey of the early development of the infrared spectral region," *Am. J. Phys.*, **28**, 42–54, 1960.
4. M. Planck, *The Theory of Heat Radiation*, Dover Publications, New York, 1959.
5. J. D. Jackson, *Classical Electrodynamics*, 3rd ed., Wiley, New York, 1998.
6. M. Born and E. Wolf, *Principles of Optics*, 7th ed., Cambridge University Press, Cambridge, UK, 1999.
7. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, New York, 1983.
8. Z. M. Zhang and K. Park, "Group front and group velocity in a dispersive medium upon refraction from a nondispersive medium," *J. Heat Transfer*, **126**, 244–249, 2004.
9. E. D. Palik (ed.), *Handbook of the Optical Constants of Solids*, Vols. I, II, and III, Academic Press, San Diego, CA, 1998.
10. T. J. Quinn and J. E. Martin, "A radiometric determination of the Stefan-Boltzmann constants and thermodynamic temperatures between  $-40^\circ\text{C}$  and  $+100^\circ\text{C}$ ," *Phil. Trans. Royal Soc. London A*, **316**, 85–189, 1985.
11. D. A. Pearson and Z. M. Zhang, "Thermal-electric modeling of absolute cryogenic radiometers," *Cryogenics*, **39**, 299–309, 1999.
12. Z. M. Zhang, "Surface temperature measurement using optical techniques," *Annu. Rev. Heat Transfer*, **11**, 351–411, 2000.

13. P. T. Landsberg and G. Tonge, "Thermodynamic energy conversion efficiencies," *J. Appl. Phys.*, **51**, R1–R20, 1980.
14. C. E. Mungan, "Radiation thermodynamics with application to lasing and fluorescent cooling," *Am. J. Phys.*, **73**, 315–322, 2005.
15. C. Essex, D. C. Kennedy, and R. S. Berry, "How hot is radiation?" *Am. J. Phys.*, **71**, 969–978, 2003.
16. M. Caldas and V. Semiao, "Entropy generation through radiative transfer in participating media: Analysis and numerical computation," *J. Quant. Spectrosc. Radiat. Transfer*, **96**, 423–437, 2005.
17. Z. M. Zhang and S. Basu, "Entropy flow and generation in radiative transfer between surfaces," *Int. J. Heat Mass Transfer*, **50**, 702–712, 2007.
18. Z. M. Zhang, C. J. Fu, and Q. Z. Zhu, "Optical and thermal radiative properties of semiconductors related to micro/nanotechnology," *Adv. Heat Transfer*, **37**, 179–296, 2003.
19. C. J. Fu, Z. M. Zhang, and P. N. First, "Brewster angle with a negative index material," *Appl. Opt.*, **44**, 3716–3724, 2005.
20. Y. S. Touloukian and D. P. DeWitt, *Thermal Radiative Properties*, Vols. 7, 8, and 9, in *Thermophysical Properties of Matter*, TPRC Data Series, Y. S. Touloukian and C. Y. Ho (eds.), IPI Plenum, New York, 1970–1972.
21. Y. J. Shen, Q. Z. Zhu, and Z. M. Zhang, "A scatterometer for measuring the bidirectional reflectance and transmittance of semiconductor wafers with rough surfaces," *Rev. Sci. Instrum.* **74**, 4885–4892, 2003.
22. W. M. Toscano and E. G. Cravalho, "Thermal radiation properties of the noble metals at cryogenic temperatures," *J. Heat Transfer*, **98**, 438–445, 1976.
23. F. Gervais and B. Piriou, "Temperature dependence of transverse- and longitudinal-optic modes in TiO<sub>2</sub> (rutile)," *Phys. Rev. B*, **10**, 1642–1654, 1974.
24. W. G. Spitzer, D. Kleinman, and D. Walsh, "Infrared properties of hexagonal silicon carbide," *Phys. Rev.*, **113**, 127–132, 1959.
25. P. J. Timans, "The thermal radiative properties of semiconductors," in *Advances in Rapid Thermal and Integrated Processing*, F. Roozeboom (ed.), pp. 35–101, Kluwer Academic Publishers, The Netherlands, 1996.
26. C. J. Fu and Z. M. Zhang, "Nanoscale radiation heat transfer for silicon at different doping levels," *Int. J. Heat Mass Transfer*, **49**, 1703–1718, 2006.
27. Z. M. Zhang and A. Frenkel, "Thermal and nonequilibrium responses of superconductors for radiation detectors," *J. Superconductivity*, **7**, 871–884, 1994.
28. A. R. Kumar, Z. M. Zhang, V. A., Boychev, D. B. Tanner, L. R. Vale, and D. A. Rudman, "Far-infrared transmittance and reflectance of YBa<sub>2</sub>Cu<sub>3</sub>O<sub>7-δ</sub> films on Si substrates," *J. Heat Transfer*, **121**, 844–851, 1999.
29. J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of superconductivity," *Phys. Rev.*, **108**, 1175–1204, 1957.
30. S. A. Ramakrishna, "Physics of negative refractive index materials," *Rep. Prog. Phys.*, **68**, 449–521, 2005.
31. C. J. Fu, Z. M. Zhang, and D. B. Tanner, "Energy transmission by photon tunneling in multilayer structures including negative index materials," *J. Heat Transfer*, **127**, 1046–1052, 2005.

## PROBLEMS

---

**8.1.** Write the wave equation in the 1-D scalar form as  $\partial^2\Psi/\partial x^2 = (1/c^2)\partial^2\Psi/\partial t^2$ , where  $c$  is a positive constant. Prove that any analytical function  $f$  can be its solution as long as  $\psi(x,t) = f(x \pm ct)$ . Plot  $\psi$  as a function of  $x$  for two fixed times  $t_1$  and  $t_2$ . Show that the sign determines the direction (either forward or backward) and  $c$  is the speed of propagation. Develop an animated computer program to visualize wave propagation.

**8.2.** Considering an electromagnetic wave propagating in the positive  $z$  direction, i.e.,  $\mathbf{k} = k\hat{\mathbf{z}}$ . Plot the vibration ellipse, and compare it with Fig. 8.2 for two cases: (1)  $\mathbf{a} = 3\hat{\mathbf{x}}$  and  $\mathbf{b} = \hat{\mathbf{x}} + 2\hat{\mathbf{y}}$  and (2)  $\mathbf{a} = 3\hat{\mathbf{x}}$  and  $\mathbf{b} = -2\hat{\mathbf{x}} + \hat{\mathbf{y}}$ . Consider the spatial dependence of the electric field at a given time,

say  $\omega t = 2\pi m$ , where  $m$  is an integer. Discuss how  $\mathbf{E}$  will change with  $kz$  for the following two cases: (3)  $\text{Re}(\mathbf{E}_0) = 3\hat{\mathbf{x}}$  and  $\text{Im}(\mathbf{E}_0) = 0$ , and (4)  $\text{Re}(\mathbf{E}_0) = 3\hat{\mathbf{x}}$  and  $\text{Im}(\mathbf{E}_0) = -3\hat{\mathbf{y}}$ . The polarization is said to be right handed if the end of the electric field vector forms a right-handed coil or screw in space at any given time. Otherwise, it is said to be left handed. Discuss the handedness for all the four cases.

**8.3.** Integrate Eq. (8.17) over a control volume to show that the energy transferred through the boundary into the control volume is equal to the sum of the storage energy change and energy dissipation. Write an integral equation using Green's theorem.

**8.4.** Derive the wave equation in Eq. (8.20) for a conductive medium; show Eq. (8.9) is a solution if  $k$  is complex, as given in Eq. (8.21). Many books use  $\mathbf{E} = \mathbf{E}_0 e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})}$  instead of Eq. (8.9) as the solution; how would you modify Eq. (8.21) and Eq. (8.22)? Show that the complex refractive index must be defined as  $\tilde{n} = n - i\kappa$ , where  $\kappa \geq 0$ .

**8.5.** Calculate the refractive index, absorption coefficient, and radiation penetration depth for the following materials, based on the dielectric function values at room temperature.

(a) Glass ( $\text{SiO}_2$ ):  $\epsilon = 2.1 + i0$  at  $1 \mu\text{m}$ ;  $\epsilon = 1.8 + i0.004$  at  $5 \mu\text{m}$ .

(b) Germanium:  $\epsilon = 21 + i0.14$  at  $1 \mu\text{m}$ ;  $\epsilon = 16 + i0.0003$  at  $20 \mu\text{m}$ .

(c) Gold:  $\epsilon = -10 + i1.0$  at  $0.65 \mu\text{m}$ ;  $\epsilon = -160 + i2.1$  at  $2 \mu\text{m}$ .

**8.6.** Consider a metamaterial with  $\mu = -1 + i0.01$  and  $\epsilon = -2 + i0.01$ ; determine the refractive index and the extinction coefficient. Calculate the radiation penetration depth. Do a quick Internet search on negative index materials, and briefly describe what you have learned.

**8.7.** Find the magnetic field  $\mathbf{H}$  for the wave given in Eq. (8.37). Show that the time-averaged Poynting vector is parallel to the  $x$ -axis. That is, the  $z$  component of  $\langle \mathbf{S} \rangle$  for such a wave vanishes. Briefly describe the features of an evanescent wave.

**8.8.** Write Planck's distribution in terms of wavenumber  $\bar{\nu} = c_0/\lambda$ , i.e., the emissive power in terms of the wavenumber:  $e_{b,\bar{\nu}}(\bar{\nu}, T)$ . What is the most probable wavenumber in  $\text{cm}^{-1}$ ? Compare your answer with the most probable wavelength obtained from Wien's displacement law in Eq. (8.45). Explain why the constants do not agree with each other. Cosmic background radiation can be treated as a blackbody radiation at  $2.7 \text{ K}$ ; what is the wavenumber corresponding to the maximum emissive power?

**8.9.** Based on the geometric parameters provided in Example 8-3 and neglecting the atmospheric effect, calculate the total intensity of the solar radiation arriving the earth's surface. Calculate the spectral intensity at  $628\text{-nm}$  wavelength. A child used a lens to focus the solar radiation to a small spot on a piece of paper and set fire this way. Does the beam focusing increase the intensity of the radiation? The lens diameter is  $5 \text{ cm}$ , and the distance between the lens and the paper is  $2.5 \text{ cm}$ . What are the focus size and the heat flux at the focus? Neglect the loss through the lens.

**8.10.** For a surface at  $T = 1800 \text{ K}$ , with an emissivity of  $0.6$ , what are the radiance temperatures at  $\lambda = 0.65 \mu\text{m}$  and  $1.5 \mu\text{m}$ ? If a conical hole is formed with a half cone angle of  $15^\circ$ , what is the effective emittance and the radiance temperature at  $\lambda = 0.65 \mu\text{m}$ ?

**8.11.** What is a radiometer? What is a calorimeter? What is a detector? What is a bolometer? If you are asked to buy a detector for infrared radiation measurement for the wavelength range between  $2$  and  $12 \mu\text{m}$ , discuss how you would select a detector and why.

**8.12.** Express Eq. (8.53) in terms of wavelength, i.e., as  $s_\lambda(\lambda, T)$ . Find an expression of the entropy intensity for blackbody radiation, i.e.,  $L_\lambda(\lambda, T)$ , and show that  $L_\lambda(\lambda, T) = (c/4\pi)s_\lambda(\lambda, T)$ .

**8.13.** Assume that all the blue light at  $\lambda$  in the range between  $420$  and  $490 \text{ nm}$  of the solar radiation is scattered by the atmosphere and uniformly distributed over a solid angle of  $4\pi \text{ sr}$ . What are the monochromatic temperatures of the scattered radiation at  $\lambda = 420 \text{ nm}$  and  $490 \text{ nm}$ ?

**8.14.** A diode-pumped solid state laser emits continuous-wave (cw) green light at a wavelength of  $532 \text{ nm}$  with a beam diameter of  $1.1 \text{ mm}$ . If the beam divergence is  $0.2 \mu\text{sr}$ , what would be the spot size at a distance of  $100 \text{ m}$  from the laser (without scattering)? If the output optical power is  $2 \text{ mW}$  and the spectral width is  $\delta\lambda = 0.1 \text{ nm}$  (assuming a square function), what is the average intensity of the laser beam? Find the monochromatic radiation temperature of the laser that is linearly polarized. Suppose the laser hits a rough surface and is scattered into the hemisphere isotropically. Find the radiation temperature of the scattered radiation and the entropy generation by scattering.

**8.15.** In Example 8-5, the two plates are blackbodies. Assume that the plates are diffuse-gray surfaces with emissivities  $\epsilon'_1$  and  $\epsilon'_2$ . Calculate the entropy generation rate in each plate per unit area. How will you determine the optimal efficiency for an energy conversion device installed at plate 2? For  $T_1 = 1500 \text{ K}$ ,  $T_2 = 300 \text{ K}$ , and  $\epsilon'_2 = 1$ , plot the optimal efficiency versus  $\epsilon'_1$ .

- 8.16.** The concept of dilute blackbody radiation can be used as an alternative method to calculate the entropy generation of a two-plate problem as in Problem 8.14. Assume that the multiply reflected rays are not in equilibrium with each other. Rather, each ray retains its original entropy, and can be treated as having an effective temperature of  $T_1$  or  $T_2$ , depending on which plate the ray is emitted from. How would you evaluate the entropy transfer from plate 1 to 2 and the entropy generation by each plate then?
- 8.17.** Calculate the entropy generation rate per unit volume for Example 2-7. Further, calculate the entropy generated at each surface, assuming that surface 2 is at 300 K.
- 8.18.** The conversion efficiency of thermophotovoltaic devices is wavelength dependent, and the optical constants are wavelength dependent as well. Perform a literature search to find some recent publications in this area. Use the entropy concept to determine the ultimate efficiency of a specific design. Based on your analysis, propose a few suggestions for further improvement of the particular design you have chosen.
- 8.19.** Derive the Fresnel reflection coefficient for a TM wave, following the derivation given in the text for a TE wave.
- 8.20.** Show that  $\rho'_{\lambda,s} + \alpha'_{\lambda,s} = 1$ , where  $\rho'_{\lambda,s}$  is given in Eq. (8.73) and  $\alpha'_{\lambda,s}$  is given in Eq. (8.75). Discuss why the  $z$  component of the time-averaged Poynting vector must be continuous at the boundary but not the  $x$  component.
- 8.21.** For nonmagnetic lossy media with  $\epsilon_1 = \epsilon'_1 + i\epsilon''_1$  and  $\epsilon_2 = \epsilon'_2 + i\epsilon''_2$ , expand Eq. (8.70b) and compare your results with Eq. (8.71).
- 8.22.** For plane wave incident from air to a nonmagnetic material with  $\epsilon = -2 + i0$  (negative real), show that the reflectivity is always 1 regardless of the angle of incidence and the polarization. What can you say about  $k_{z2}$  and  $\langle S_{z2} \rangle$ ? Is the wave in the medium a homogeneous wave or an evanescent wave?
- 8.23.** The refractive index of glass is approximately 1.5 in the visible region. What is the Brewster angle for glass when light is incident from air? Calculate the reflectance, and plot it against the incidence angle for  $p$  polarization,  $s$  polarization, and random polarization. Redo the calculation for incidence from glass to air, and plot the reflectance against the incidence angle. At what angle does total internal reflection begin, and what is this angle called?
- 8.24.** The principal angle is defined as the angle at which the ratio of the reflectance for TM and TE waves is minimized. For radiation incident from air to a medium with  $n = 2$  and  $\kappa = 1$ , determine the principal angle and show that the phase difference between the two reflection coefficients equals to  $\pi/2$  at this angle. [Hint: Use graphs to prove the existence of the principal angle.]
- 8.25.** Calculate and plot the emissivity (averaged over the two polarizations) versus the zenith angle for the materials and wavelengths given in Problems 8.5. Calculate and tabulate the normal and hemispherical emissivities for all cases.
- 8.26.** Calculate the optical constants and the radiation penetration depth for either gold or silver at room temperature, using the Drude model, and plot them as functions of wavelength. In addition, calculate the normal reflectivity and plot it against wavelength. Compare the results using the Hagen-Rubens equation. How will the scattering rate and the plasma frequency change if the temperature is raised to 600 K?
- 8.27.** Calculate the normal emissivity of MgO from 2000 to 200  $\text{cm}^{-1}$  (5 to 50  $\mu\text{m}$ ) using the Lorentz model with two oscillators having the following parameters:  $\epsilon_\infty = 3.01$ ;  $\omega_1 = 401 \text{ cm}^{-1}$ ,  $\gamma_1 = 7.62 \text{ cm}^{-1}$ , and  $S_1 = 6.6$ ;  $\omega_2 = 640 \text{ cm}^{-1}$ ,  $\gamma_2 = 102.4 \text{ cm}^{-1}$ , and  $S_2 = 0.045$ . Can you develop a program to calculate the hemispherical emissivity and plot it against the normal emissivity for a comparison?
- 8.28.** Using the accompanied software, *Rad-Pro*, to plot the absorption coefficient and the reflectivity of lightly doped silicon in the spectral region from 0.5 to 25  $\mu\text{m}$ , at two temperatures: 600 K and 900 K.
- 8.29.** Find the Brewster angles for light incident from air to a NIM with (a)  $\epsilon_2 = -2$  and  $\mu_2 = -2$ , (b)  $\epsilon_2 = -1$  and  $\mu_2 = -4$ , and (c)  $\epsilon_2 = -8$  and  $\mu_2 = -0.5$ .
- 8.30.** Suppose a NIM can be described by Eq. (8.121) and Eq. (8.122) with the following parameters:  $\omega_p = 4.0 \times 10^{14} \text{ rad/s}$  ( $\lambda_p = 4.71 \mu\text{m}$ ),  $\omega_0 = 2.0 \times 10^{14} \text{ rad/s}$  (i.e.,  $\lambda_0 = 9.42 \mu\text{m}$ ),  $\gamma = 0$ , and  $F = 0.785$ . Assume a wave is propagating in such a medium in the region of  $n < 0$  with a wavevector  $\mathbf{k} = k_x \hat{x}$ , where  $k_x = k = |n|\omega/c_0$ . Show that the group velocity is in the negative  $x$  direction. Also show that the Poynting vector is in the same direction as the group velocity.
- 8.31.** Suppose a NIM can be described by Eq. (8.121) and Eq. (8.122) with the following parameters:  $\omega_p = 4.0 \times 10^{14} \text{ rad/s}$  (i.e.,  $\lambda_p = 4.71 \mu\text{m}$ ),  $\omega_0 = 2.0 \times 10^{14} \text{ rad/s}$  (i.e.,  $\lambda_0 = 9.42 \mu\text{m}$ ), and  $F = 0.5$ . Calculate and plot the refractive index and the extinction coefficient in the spectral region from 2 to 15  $\mu\text{m}$ , for  $\gamma = 0, 10^{12}$ , and  $10^{13} \text{ rad/s}$ .

---

# CHAPTER 9

---

# RADIATIVE PROPERTIES OF NANOMATERIALS

---

Optical and thermal radiative properties are fundamental physical properties that describe the interaction between electromagnetic waves and matter from deep ultraviolet to far-infrared spectral regions. A large number of studies have been devoted to the measurement, analysis, modeling, and simulation of optical and radiative characteristics of materials in solid, liquid, gas, and plasma phases. The radiative properties of nanostructured materials are critical to the functionality and the performance of many devices, such as semiconductor lasers, radiation detectors, tunable optical filters, waveguides, solar cells, and selective emitters and absorbers. The use of microstructures not only modifies the optical properties for optoelectronic applications and processing control but also facilitates some important energy conversion devices, such as solar cells and thermophotovoltaic applications.

This chapter will start with the radiative properties of a single layer with or without considering the wave interference effect. The effect of partial coherence and surface scattering will be considered next. The approach will then be generalized to multilayered structures using the 1-D matrix formulation. Furthermore, periodic structures such as photonic crystals and gratings will be studied based on the Bloch wave equation. Subsequently, the effective medium formulations will be briefly discussed. Finally, the effect of surface roughness and microstructures on the radiative properties will be presented.

---

## **9.1 RADIATIVE PROPERTIES OF A SINGLE LAYER**

---

Crystalline films, from a few nanometers to several micrometers thick, have been deposited (by physical vapor deposition, chemical vapor deposition, sputtering, laser ablation, molecular beam epitaxy, rapid thermal processing, and other techniques) onto suitable substrates. These layered structures play important roles in contemporary technologies, such as integrated circuits, semiconductor lasers, quantum well detectors, superconductor/semiconductor hybrid devices, optical filters, and spectrally selective coatings for solar thermal applications. Radiative energy transport in thin films differs significantly from that at bulk solid surfaces and through thick windows because of multiple reflections and interference effects. The radiative properties of a lamina with smooth and parallel surfaces will be discussed first, with emphasis on different formulations for various applications. At the end of this section, the effect of surface scattering will be considered in the regime where the roughness is much smaller than the wavelength.

### 9.1.1 The Ray Tracing Method for a Thick Layer

A “thick” layer refers to the case where interference between multiply reflected waves can be neglected. In other words, the waves are *incoherent*. On the contrary, a “thin” film refers to the case where all multiply reflected waves are *coherent* and interfere with each other. The condition for being thick has often been commonly interpreted as that the layer thickness  $d$  is much greater than the wavelength. A more rigorous criterion is that the thickness is much greater than the coherence length, which can be much greater than the wavelength. The coherence length depends on the spectral width of the source and the spectral resolution of the spectrophotometer, such as a grating monochromator or a Fourier transform spectrometer. In addition, beam divergence, surface roughness, and nonparallelism of the surfaces further reduce the degree of coherence. Generally speaking, when the thickness is comparable to the wavelength, interference effects are important. However, this does not guarantee complete coherence because of the nature of the source and imperfect surfaces. Let us first consider the radiative properties of a layer or a slab, in the incoherent limit, because of its simplicity.

Either the ray tracing method or the net radiation method can be applied to find out the transmittance and the reflectance of a thick layer.<sup>1</sup> Consider a slab of thickness  $d$ , placed in air or vacuum, as shown in Fig. 9.1. The refractive index and the extinction coefficient of

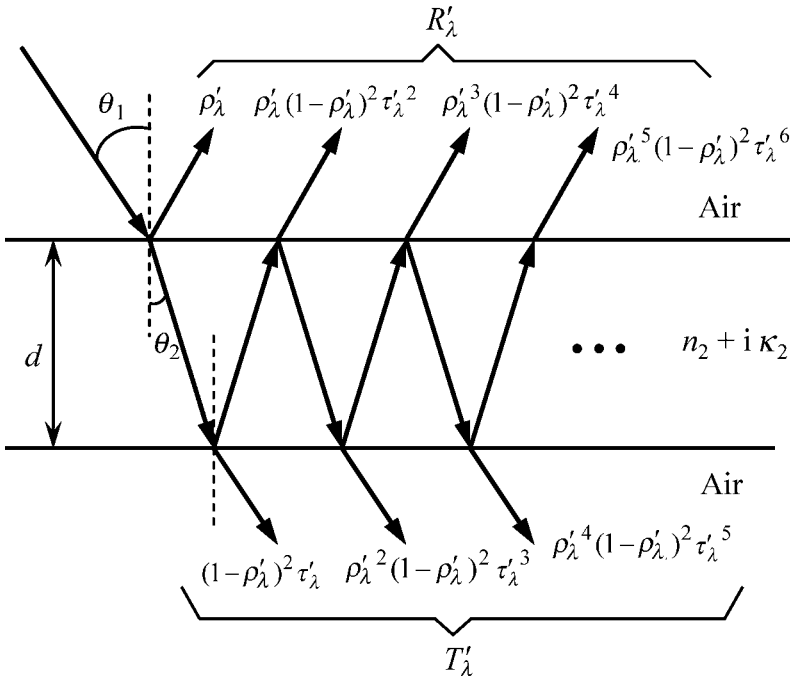


FIGURE 9.1 Transmittance and reflectance of a lamina as a result of multiple reflections.

the material are  $n_2$  and  $\kappa_2$ , respectively. As mentioned earlier, it is generally required that the thickness be much greater than the wavelength to avoid the interference effect. Because the intensity will attenuate exponentially inside an absorbing medium, the penetration depth  $\delta_\lambda = \lambda/4\pi\kappa$  should be greater than the layer thickness in order to have appreciable transmission. For this reason, the extinction coefficient is usually much smaller than the refractive

index, i.e.,  $\kappa \ll n$ . Therefore, we can limit our consideration to dielectric materials with a small loss, such as a glass window or a silicon wafer in the semitransparent region. For a given surface reflectivity  $\rho'_\lambda$  and an internal transmissivity  $\tau'_\lambda$ , ray tracing yields the directional-hemispherical spectral reflectance as

$$\begin{aligned} R'_\lambda &= \rho'_\lambda + \rho'_\lambda(1 - \rho'_\lambda)^2\tau'^2_\lambda + \rho'^3_\lambda(1 - \rho'_\lambda)^2\tau'^4_\lambda + \rho'^5_\lambda(1 - \rho'_\lambda)^2\tau'^6_\lambda + \dots \\ &= \rho'_\lambda + \frac{\rho'_\lambda(1 - \rho'_\lambda)^2\tau'^2_\lambda}{1 - \rho'^2_\lambda\tau'^2_\lambda} = \rho'_\lambda \left[ 1 + \frac{(1 - \rho'_\lambda)^2\tau'^2_\lambda}{1 - \rho'^2_\lambda\tau'^2_\lambda} \right] \end{aligned} \quad (9.1)$$

because the second term and beyond form a geometric series. Similarly, the directional-hemispherical spectral transmittance can be expressed as

$$T'_\lambda = \frac{(1 - \rho'_\lambda)^2\tau'_\lambda}{1 - \rho'^2_\lambda\tau'^2_\lambda} \quad (9.2)$$

Hence, the directional-spectral absorptance of the lamina at the given direction and wavelength is

$$A'_\lambda = 1 - T'_\lambda - R'_\lambda = \frac{(1 - \rho'_\lambda)(1 - \tau'_\lambda)}{1 - \rho'_\lambda\tau'_\lambda} \quad (9.3)$$

The reflectivity  $\rho'_\lambda$  can be calculated from Eq. (8.73) and Eq. (8.80), for each polarization, as a function of the angle of incidence  $\theta_1$  and the refractive index. For unpolarized incident radiation,  $R'_\lambda$ ,  $T'_\lambda$ , and  $A'_\lambda$  should be averaged over the two linear polarizations. The influence of  $\kappa_2$  on  $\rho'_\lambda$  is often negligibly small. On the other hand,  $\kappa_2$  affects the absorption through the internal transmissivity  $\tau'_\lambda$ , defined as

$$\tau'_\lambda = \exp\left(-\frac{4\pi\kappa_2 d}{\lambda \cos\theta_2}\right) \quad (9.4)$$

where  $\lambda$  is the wavelength in air or vacuum,  $\theta_2$  is the refraction angle inside the slab, and  $d/\cos\theta_2$  can be considered as the actual path length of the ray inside the layer. From Snell's law, we have  $\cos\theta_2 = \sqrt{1 - (1/n_2^2)\sin^2\theta_1}$ . Here again, the effect of  $\kappa_2$  is neglected. Figure 9.2 shows the transmittance at normal incidence for several semitransparent materials with a thickness  $d = 0.5$  mm, calculated using the tabulated optical constants from Palik.<sup>2</sup> It can be seen that SiO<sub>2</sub> glass is transparent in the visible region but opaque to infrared radiation beyond 5- $\mu\text{m}$  wavelength. On the other hand, silicon is opaque for visible light but has a transmittance of about 50% in the far-infrared region.

When there is no absorption, the reflectance and the transmittance are independent of the layer thickness  $d$ , and for normal incidence, the following simplified equation can be used:

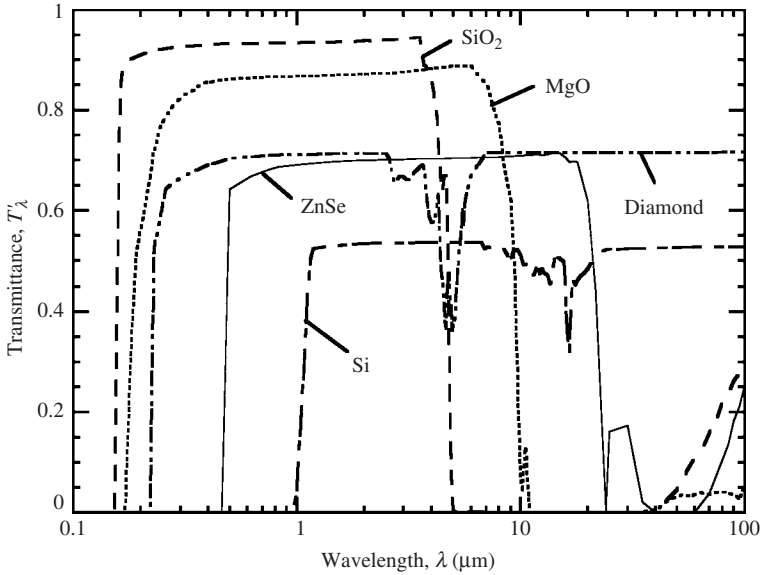
$$T'_\lambda = \frac{2n_2}{n_2^2 + 1} \quad (9.5)$$

For a fused silica (SiO<sub>2</sub>) window in the visible range, with a refractive index around 1.5, the transmittance is 0.923. For diamond with a refractive index of 2.4, the transmittance is 0.71. In some applications, antireflection coatings are often used to reduce reflectance and enhance transmittance, which will be discussed later for multilayer structures.

### 9.1.2 Thin Films

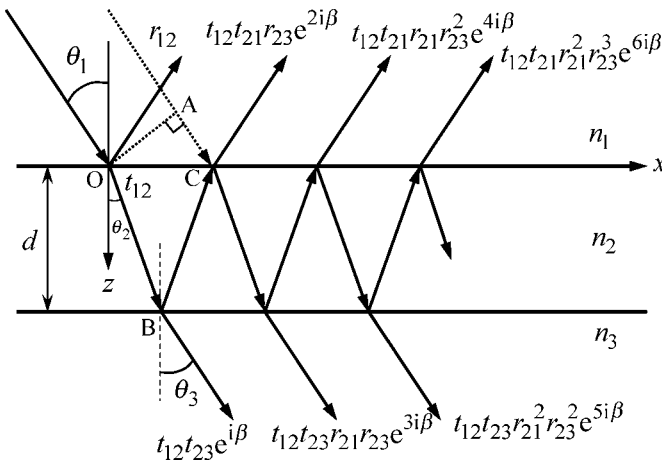
Thin-film coatings are of practical importance to the design of spectrally selective surfaces for solar energy utilization and space applications, optical filters, and antireflection coatings.





**FIGURE 9.2** Normal transmittance of several dielectric materials with 0.5-mm thickness at room temperature.

When the wavelength of radiation is comparable to the coherence length, which depends not only on the properties of the film but also on the characteristics of the source and the detector, wave interference becomes important. To consider the interference effect, the amplitude and the phase of the electric field (or the magnetic field) must be traced during multiple reflections. The method is usually referred to as thin-film optics, as illustrated in Fig. 9.3 for a thin film of thickness  $d$  between two semi-infinite media.<sup>3,4</sup> There are several practical configurations based on the structure shown in this figure. (a) The first is for a



**FIGURE 9.3** Illustration of interference between multiple reflections.

free-standing film in air. (b) The second is for radiation incident from air (medium 1) on a thin film (medium 2) coated onto a semi-infinite substrate (medium 3). (c) In the third configuration, media 1 and 3 are dielectrics but medium 2 is vacuum. This configuration is important for photon tunneling experiments to be discussed in Chap. 10. Let us first consider the lossless case, where the refractive indices are all real, and so are the angles of incidence and refraction. It will be seen later that the equations can easily be extended to absorbing media using complex variables. A plane wave with either  $p$  or  $s$  polarization is incident from medium 1. Note that  $t_{jk}$  and  $r_{jk}$ , where  $j, k = 1, 2$ , or  $3$ , are respectively the transmission and reflection coefficients between the media  $j$  and  $k$  for the given polarization. While the multiply reflected waves are illustrated with a spatial displacement, interference occurs at the same time and location between multiply reflected beams. For this reason, upon traversing the film, the wave acquires a phase shift  $\beta$  given by

$$\beta = \frac{2\pi d}{\lambda} \sqrt{n_2^2 - n_1^2 \sin^2 \theta_1} \quad (9.6)$$

Note again that  $\lambda$  is the wavelength in vacuum. This is to say that  $\beta = 2\pi(n_2/\lambda)d \cos \theta_2$ . The reason that  $\cos \theta_2$  is in the numerator, instead of the denominator, is because the phase for the same location  $x$  is considered when  $z$  is changed from  $0$  to  $d$ . The phase of the electric field is given by  $\mathbf{k} \cdot \mathbf{r}$ , and thus, the phase difference is  $\beta = k_2 d \cos \theta_2$ , where  $k_2 = 2\pi n_2/\lambda$ . Another way to understand the phase shift is to consider the plane of constant phase, as illustrated in Fig. 9.3 with the line OA. The first reflected wave is the wave from A to C that acquires a phase difference of  $(k_1 \sin \theta_1)(2d \tan \theta_2) = (k_2 \sin \theta_2)(2d \tan \theta_2)$  because  $k_x = k_j \sin \theta_j$  is the same in all media. The second reflected wave goes through the film twice (from O to B and then from B to C) and gains a phase difference of  $2k_2 d/\cos \theta_2$ . It can easily be shown that the phase shift between the first and the second reflected waves is  $2k_2 d(1/\cos \theta_2 - \sin^2 \theta_2/\cos \theta_2) = 2\beta$ . More detailed discussion can be found from Brewster.<sup>5</sup> After the superposition, the field reflection and transmission coefficients of the film can be expressed as

$$r = r_{12} + \frac{t_{12}t_{21}r_{23}e^{2i\beta}}{1 - r_{21}r_{23}e^{2i\beta}} \quad (9.7)$$

and

$$t = \frac{t_{12}t_{23}e^{i\beta}}{1 - r_{21}r_{23}e^{2i\beta}} \quad (9.8)$$

which are known as Airy's formulae.<sup>3,4</sup> It should be noted that these coefficients are defined based on the electric fields for  $s$  polarization and the magnetic fields for  $p$  polarization, respectively. The energy reflectance can be calculated by

$$R'_\lambda = rr^* = \left| r_{12} + \frac{t_{12}t_{21}r_{23}e^{2i\beta}}{1 - r_{21}r_{23}e^{2i\beta}} \right|^2 \quad (9.9)$$

For the incident radiation with random polarization, Eq. (9.9) should be averaged over the two linear polarizations by evaluating Fresnel's coefficients for each polarization separately. Furthermore, Eq. (9.6) through Eq. (9.9) are not limited to lossless situations as long as the absorption in medium 1 is negligible.<sup>6</sup> When  $n_2$  and  $n_3$  are complex, the phase shift given in Eq. (9.6) becomes complex. Note that the reflection and transmission coefficients in Eq. (9.7) and Eq. (9.8) are always complex. Waves inside an absorbing medium are inhomogeneous because the constant-phase planes are defined by the real part of the wavevector and the constant-amplitude planes are parallel to the interfaces. To determine the direction of energy flow, one needs to carefully evaluate the Poynting vector in medium 3. The expression of the energy transmittance is similar to those for the absorptivity in Eq. (8.75)

and Eq. (8.81). If medium 3 is lossless, we can write the transmittance in terms of the transmission coefficient as in the following:

$$T'_\lambda = \frac{n_3 \cos \theta_3}{n_1 \cos \theta_1} tt^*, \text{ for } s \text{ polarization} \quad (9.10a)$$

and

$$T'_\lambda = \frac{n_1 \cos \theta_3}{n_3 \cos \theta_1} tt^*, \text{ for } p \text{ polarization} \quad (9.10b)$$

For a free-standing film in air, since  $n_1 = n_3 = 1$ , the transmittance can be reduced to the following equation when the film is slightly absorbing (i.e.,  $\kappa_2 \ll n_2$ ):<sup>6</sup>

$$T'_\lambda = \frac{(1 - \rho'_\lambda)^2 \tau'_\lambda}{1 + \rho'^2_\lambda \tau'^2_\lambda - 2\rho'_\lambda \tau'_\lambda \cos(2\beta)} \quad (9.11)$$

In Eq. (9.11),  $\beta$  and  $\rho'_\lambda$  are calculated by neglecting  $\kappa_2$ , and  $\tau'_\lambda$  is from Eq. (9.4). The transmittance will oscillate even though the optical constants are unchanged. A change in wavelength, thickness, or refractive index can cause the transmittance to oscillate. The transmittance spectrum has peaks at  $\beta = m\pi$  and valleys at  $\beta = (m + \frac{1}{2})\pi$ , where  $m$  is a nonnegative integer. Figure 9.4 shows the calculated normal transmittance for  $d = 10 \mu\text{m}$

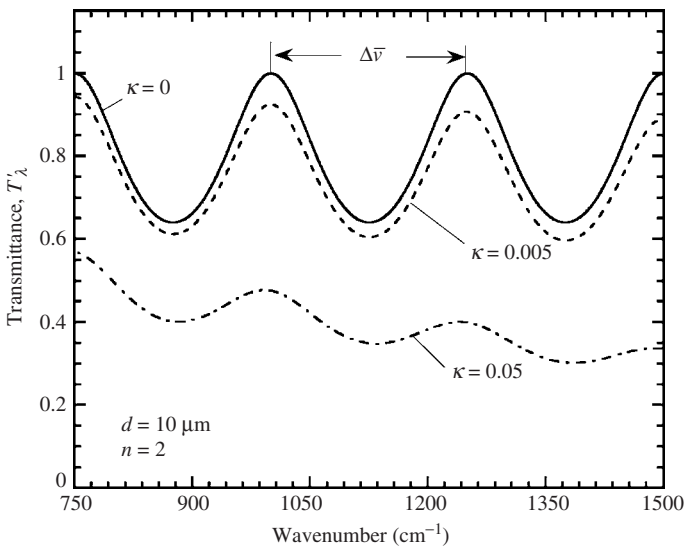


FIGURE 9.4 Calculated transmittance of a thin film of 10- $\mu\text{m}$  thickness with  $n = 2$

and  $n_2 = n + i\kappa$ , with  $n = 2$  and  $\kappa = 0, 0.005$ , and  $0.05$ . The subscript 2 is dropped for convenience. The results are plotted in terms of wavenumber between 750 and 1500  $\text{cm}^{-1}$ . The *free spectral range* is the frequency interval between two peaks. It is convenient to use the wavenumber instead of frequency. For normal incidence, the free spectral range in terms of wavenumber is given by

$$\Delta\bar{\nu} = \frac{1}{2nd} \quad (9.12a)$$

where  $d$  is in cm and  $\Delta\bar{\nu}$  is in  $\text{cm}^{-1}$ . When plotted in terms of wavelength, the free spectral range becomes

$$\Delta\lambda = \frac{\Delta\bar{\nu}}{\bar{\nu}^2} = \frac{\lambda^2}{2nd} \quad (9.12b)$$

which increases with wavelength for constant  $n$  and  $d$ . In the absence of absorption, the maximum transmittance is unity. The inclusion of a very small nonzero extinction coefficient  $\kappa$  can cause the transmittance to be reduced from the lossless situation, especially at shorter wavelengths. When  $\kappa = 0.05$ , the internal transmissivity  $\tau'_\lambda$  is a strong function of wavelength and the transmittance is significantly reduced. Furthermore, the *fringe contrast* is also reduced due to absorption. The fringe contrast  $\Phi$  is defined, based on the maximum transmittance  $T_{\max}$  and minimum transmittance  $T_{\min}$ , as

$$\Phi = \frac{T_{\max} - T_{\min}}{T_{\max} + T_{\min}} \quad (9.13)$$

For broadband or polychromatic radiation, the total transmittance is defined as the fraction of the energy transmitted. Suppose the spectral intensity is  $I_\lambda$ , then the total transmittance is

$$T'_{\text{tot}} = \int_0^\infty I_\lambda(\lambda) T'_\lambda(\lambda) d\lambda / \int_0^\infty I_\lambda(\lambda) d\lambda \quad (9.14)$$

In some practice, one needs to integrate the transmittance over a narrow band. An example is the radiation coming through a filter or a spectrometer with a finite resolution. The intensity is nearly constant within the small bandwidth; the transmittance can be averaged over a spectral width  $\Delta\lambda$  around  $\lambda$  for each wavelength, viz.,

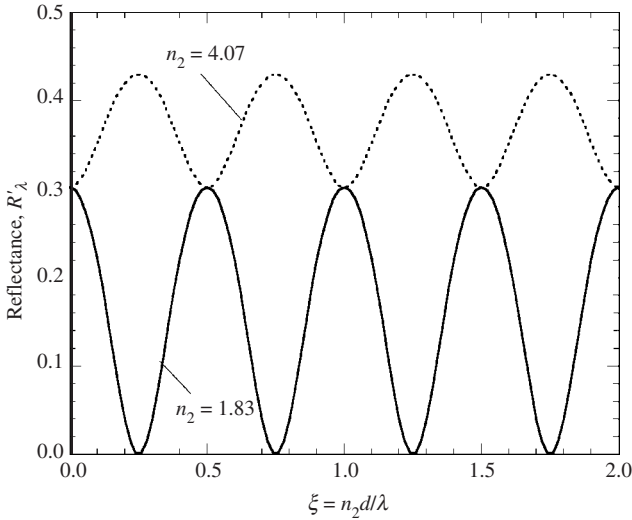
$$\bar{T}'_\lambda(\lambda) = \frac{1}{\Delta\lambda} \int_{\lambda - \Delta\lambda/2}^{\lambda + \Delta\lambda/2} T'_\lambda(\lambda) d\lambda \quad (9.15)$$

It can be shown that, integrating the coherence formula in Eq. (9.11) over a free spectral range  $\Delta\lambda = \Delta\bar{\nu}/\bar{\nu}^2$  gives the same result as the incoherence formula in Eq. (9.2). However, the fringe-averaged transmittance is not equal to the arithmetic average of the transmittance maximum and minimum. When  $d$  is much greater than the wavelength by a factor of, say, 1000, the free spectral range  $\Delta\lambda$  will become so small that most spectrophotometers do not have the sufficient resolution to discern the fringes. Furthermore, a slight variation in the film thickness or the wedge effect will cause the phases of multiple reflections to be canceled out. The measured transmittance will follow Eq. (9.2) without the high-frequency oscillation. That is why Eq. (9.2) has practical importance even though it can be obtained from Eq. (9.11) by spectral averaging. The spectral-averaging method is useful to obtain radiative properties in the partial coherence regime, to be discussed in Sec. 9.1.4.

It should be emphasized that for metallic films, when the extinction coefficient is not much smaller than the refractive index, Eq. (9.11) breaks down and the transmittance of a thin film must be calculated according to Eq. (9.10). Consider a 100-nm gold film with  $n_2 = 0.916 + i1.84$  at the wavelength  $\lambda = 0.5 \mu\text{m}$ . The penetration depth is  $\delta_\lambda = \lambda/(4\pi\kappa) = 21.6 \text{ nm}$ . At normal incidence,  $\rho'_\lambda = 0.481$  and  $\tau'_\lambda = 0.0098$ , and both Eq. (9.2) and Eq. (9.11) reduce to  $T'_\lambda \approx (1 - \rho'_\lambda)^2 \tau'_\lambda = 0.0026$ . This result, however, is incorrect because neither equation is applicable for large extinction coefficients. Using Eq. (9.10) and the complex Fresnel coefficients defined in Chap. 8, we have reevaluated the normal transmittance of the gold film to be  $T'_\lambda \approx 0.013$  in this case (see Zhang<sup>6</sup> for more discussion and Problem 9.5 as an exercise).

**Example 9-1.** Calculate the reflectance in terms of film thickness  $d$  for a dielectric film onto a silicon substrate with a refractive index of  $n_3 = 3.44$  near  $\lambda = 2.5 \mu\text{m}$ . Assume radiation is incident at normal incidence from air. Consider two cases:  $n_2 = 1.83$  (SiO) and  $n_2 = 4.07$  (Ge).

**Solution.** Equation (9.9) can be recast as  $R'_\lambda = \left| \frac{r_{12} + r_{23}e^{i2\beta_2}}{1 + r_{12}r_{23}e^{i2\beta_2}} \right|^2$ , where for normal incidence,  $r_{12} = (n_1 - n_2)/(n_1 + n_2)$  and  $r_{23} = (n_2 - n_3)/(n_2 + n_3)$ . While the Fresnel coefficients are for  $s$  polarization, a minus sign for both  $r_{12}$  and  $r_{23}$  for  $p$  polarization will not change the value of  $R'_\lambda$ . The results are plotted in terms of the dimensionless parameter  $\xi = n_2d/\lambda$  in Fig. 9.5. The reflectance



**FIGURE 9.5** Reflectance for a SiO or Ge film onto a Si substrate at  $\lambda = 2.5 \mu\text{m}$ .

oscillates with a period  $\Delta d = \lambda/(2n_2)$ . When  $n_2d = \lambda/2, 3\lambda/2, 5\lambda/2$ , and so on, the reflectance is reduced to that of silicon without coating, i.e.,  $R'_\lambda = (n_1 - n_3)^2/(n_1 + n_3)^2$ . When  $n_2 > n_3$  or  $n_2 < n_1$ , the reflectance is always greater than that without coating, and reaches a maximum at  $n_2d = \lambda/4, 3\lambda/4, 5\lambda/4$ , and so on. When  $n_1 < n_2 < n_3$ , the reflectance is always smaller than that without coating, and reaches a minimum at  $n_2d = \lambda/4, 3\lambda/4, 5\lambda/4$ , and so on. The values are determined by  $R'_\lambda = (n_1n_3 - n_2^2)^2/(n_1n_3 + n_2^2)^2$ . Note that the reflectance minimum becomes zero when  $n_2 = \sqrt{n_1n_3} = 1.885$ . Since the refractive index of SiO is close to this value, a nearly zero reflectance can be obtained. This is called the *antireflection effect* and has numerous applications in many optical systems including eye glasses. In addition, quarter-wave antireflection coatings can be used to improve the energy conversion efficiency for solar energy applications.

### 9.1.3 Partial Coherence

It should be noted that no source is perfectly coherent—even laser or atomic emission has a nonzero line width. Likewise, no source is completely incoherent—even the most chaotic blackbody radiation has a small *coherence length*. The coherence length is related to the distance that light travels within a *coherence time*. The concept of coherence is related to the situation where the wave nature will be preserved. When the time is longer than the coherence time or when waves travel a distance longer than the coherence length, fluctuations will manifest and thus undermine the interference effects.<sup>7</sup> Although complete incoherence and

coherence formulae can be applied to a variety of practical problems, there are situations that do not fall in either regime. An example is the measured transmittance spectra of a slab with a spectrometer, such as a grating spectrophotometer or the Fourier transform infrared (FTIR) spectrometer based on the Michelson interferometer with a beamsplitter, a fixed mirror, and a moving mirror. Due to the finite instrument resolution and imperfections of the sample surfaces (not perfectly parallel or smooth), the fringe contrast defined in Eq. (9.13) for transmittance is always less than that predicted by the coherence formula. A similar definition also applies to the reflectance spectrum.

Partial coherence theory was developed before 1960, and has gone through significant advancements after the first lasers in the 1960s, including the application to radiometry.<sup>7</sup> A brief introduction is given here with an emphasis on the radiative properties of thin films. The electric field can be expressed in either frequency domain as  $E(\nu)$  or time domain as  $E(t)$ , which are related by Fourier transforms. The *mutual coherence function* of any two waves is defined as

$$\langle E_j(t)E_k^*(t) \rangle = 4 \int_0^\infty G_{jk}(\nu) d\nu \quad (9.16)$$

where the angular bracket  $\langle \rangle$  symbolizes the time-averaging operation, i.e.,

$$\langle E_j(t)E_k^*(t) \rangle = \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \int_{-\tau}^{\tau} E_j(t)E_k^*(t) dt \quad (9.17)$$

and  $G_{jk}(\nu)$  is the mutual spectral density, given by

$$G_{jk} = \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \overline{E_j(\nu)E_k^*(\nu)} \quad (9.18a)$$

where the “long bar” denotes ensemble averaging. The spectral density of a wave is defined by

$$G(\nu) = \lim_{\tau \rightarrow \infty} \frac{1}{2\tau} \overline{E(\nu)E^*(\nu)} \quad (9.18b)$$

and the optical intensity, which is proportional to the radiant energy flux in a given medium, is

$$I = \langle E(t)E^*(t) \rangle = 4 \int_0^\infty G(\nu) d\nu \quad (9.19)$$

The complex degree of coherence is defined as

$$\gamma_{jk} = \frac{\langle E_j(t)E_k^*(t) \rangle}{\sqrt{\langle E_j(t)E_j^*(t) \rangle \langle E_k(t)E_k^*(t) \rangle}} \quad (9.20)$$

Note that  $|\gamma_{jk}| \leq 1$ . If there are only two waves, each with an optical intensity of  $I_1$  and  $I_2$ , the combined optical intensity of the two waves is given as follows:

$$I_c = I_1 + I_2 + \sqrt{I_1 I_2} (\gamma_{12} + \gamma_{12}^*) \quad (9.21)$$

Let us use Young’s double-slit experiment as an example, where light from a pinhole goes through two slits. Interference patterns will be projected on a screen. When the slits are of very small width and the source is nearly monochromatic, a sine wave pattern will be observed with alternate bright and dark fringes. This is because  $\gamma_{12} = \exp(i\delta)$ , where  $\delta$  is the phase difference between the two beams and varies with position on the screen. The outcome is completely coherent because  $|\gamma_{12}| = 1$ . On the other hand, when the source is polychromatic, the pattern will be the brightest at the center because constructive interference

occurs for all wavelengths only at the center. The interference fringes will fade away from the center and eventually disappear because of the lack of coherence. In this case,  $|\gamma_{12}|$  is position dependent. Partial coherence can also occur as the width of the slit is enlarged. If the slit width is comparable to or larger than the wavelengths, the screen will be evenly illuminated. This corresponds to a complete incoherence with  $|\gamma_{12}| = 0$  and  $I_c = I_1 + I_2$ .

Chen and Tien employed the partial coherence theory to calculate the radiative properties of a layer, by taking the forward propagating field in the film as composed of two components: the first transmitted wave and all the rest that are caused by multiple reflections.<sup>8</sup> Alternatively, the degree of coherence may be defined between any two multiply reflected waves, and the radiative properties in the partial coherence regime can be expressed in an infinite summation. Several factors affect the degree of coherence, such as the beam divergence, the thickness variation, or the finite spectral width of the instrument. The combined effect is that multiple reflections become less and less coherent, because the phase of the wave increases by  $2\beta$  each time it undergoes a round trip inside the film (see Fig. 9.3). Recently, Fu et al. obtained analytic formulae for the reflectance and the transmittance of a thin film using direct spectral integration.<sup>9</sup> The integral averaging of transmittance, calculated from wave optics over a finite frequency interval, yields the same result as the partial coherence formulation does. The spectral averaging of the transmittance can be evaluated by

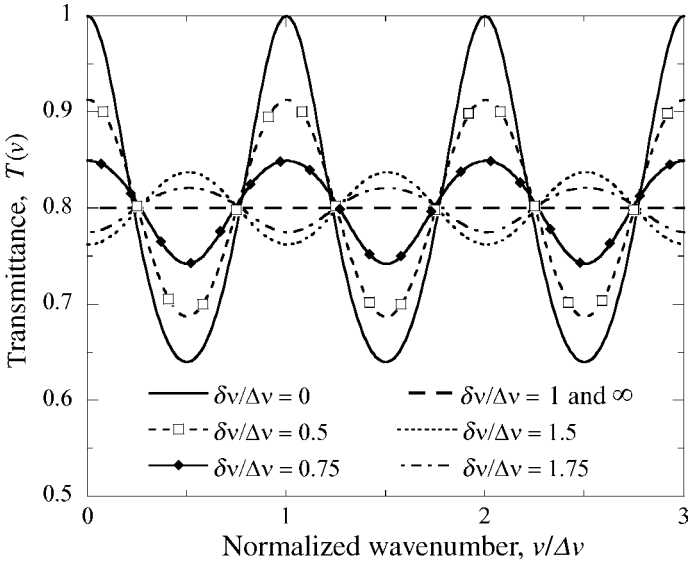
$$\bar{T}(\nu) = \frac{1}{\delta\nu} \int_{\nu - \delta\nu/2}^{\nu + \delta\nu/2} T(\nu') d\nu' \quad (9.22)$$

where  $\nu'$  is a dummy variable and  $\delta\nu$ , the frequency interval used for the averaging, is called the *coherence spectral width*.<sup>10</sup> The directional-hemispherical spectral transmittance is simply expressed as  $T(\nu)$  in Eq. (9.22) without any subscript or superscript for clarity. The frequency  $\nu$  is most conveniently expressed in  $\text{cm}^{-1}$  or in terms of wavenumber as done before. It should be emphasized that the spectrally averaged property is still a spectral property rather than a total property. It is inherently assumed that  $\delta\nu$  is a small bandwidth within which the source spectral intensity is independent of frequency. Furthermore,  $\delta\nu$  is related not only to the effective bandwidth, the resolution, and the sampling interval of the spectrometer but also to the conditions of the specimen. Figure 9.6 illustrates the effect of spectral averaging on the transmittance spectrum for a film with  $n = 2$  and  $k = 0$ , with various  $\delta\nu$  values, at normal incidence. Both the frequency  $\nu$  and the coherence spectral width  $\delta\nu$  are normalized by the free spectral range  $\Delta\nu$  so that the curves are independent of the film thickness and the frequency unit used. As  $\delta\nu/\Delta\nu$  increases from 0 (the coherent limit), the fringe contrast decreases until  $\delta\nu/\Delta\nu = 1$  when all the fringes disappear. When  $\delta\nu/\Delta\nu > 1$ , however, the fringes reappear but the peaks and the valleys invert from the original. The inversion is largest when  $\delta\nu/\Delta\nu = 1.5$ . When  $\delta\nu/\Delta\nu \gg 1$ , the fringe contrast becomes negligible, and the transmittance approximates the incoherent limit when geometric optics are applicable.

Although  $\delta\nu = 0$  and  $\delta\nu \rightarrow \infty$  correspond to the coherent and incoherent limits, respectively, the magnitude of  $\delta\nu$  is not directly related to the degree of coherence in the partial coherence regime. For example,  $\delta\nu/\Delta\nu = 1.5$  is more coherent than  $\delta\nu/\Delta\nu = 1$  (when all fringes disappear). The degrees of coherence are difficult to calculate even for smooth films and not applicable to films with rough surfaces. Lee et al. introduced a coherence function:

$$\phi = \frac{\bar{T}(\nu_{\max}) - \bar{T}(\nu_{\min})}{T_{\text{coh}}(\nu_{\max}) - T_{\text{coh}}(\nu_{\min})} \quad (9.23)$$

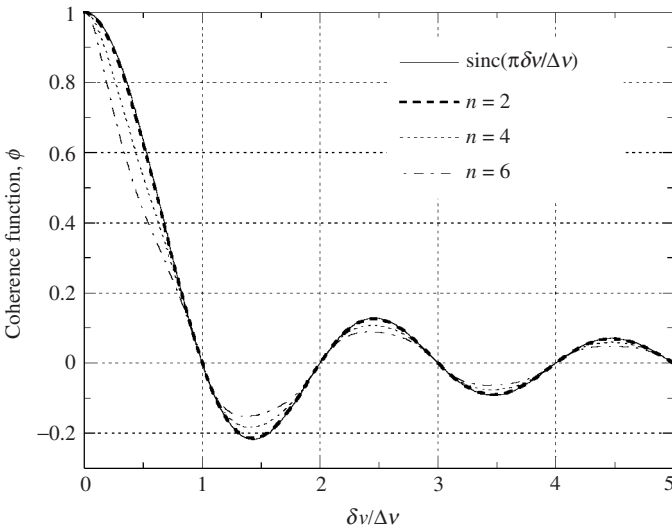
where  $T_{\text{coh}}$  is the transmittance calculated from the coherence formulation without scattering loss (i.e., thin-film optics),  $\bar{T}$  is the spectral averaging of transmittance calculated from Eq. (9.22) to include partial coherence, and  $\nu_{\max}$  and  $\nu_{\min}$  are the frequencies corresponding to transmittance maximum and minimum, respectively, in the coherent limit.<sup>10</sup> In essence,



**FIGURE 9.6** The effect of coherence spectral width on the spectrally averaged transmittance.

the denominator equals the difference between transmittance extrema in the coherent limit, and the numerator equals the difference in transmittance extrema, when partial coherence is considered.

The coherence function is plotted in Fig. 9.7 as a function of a dimensionless parameter  $\delta\nu/\Delta\nu$  for dielectric thin films. The film thickness is implicitly included in the parameters and

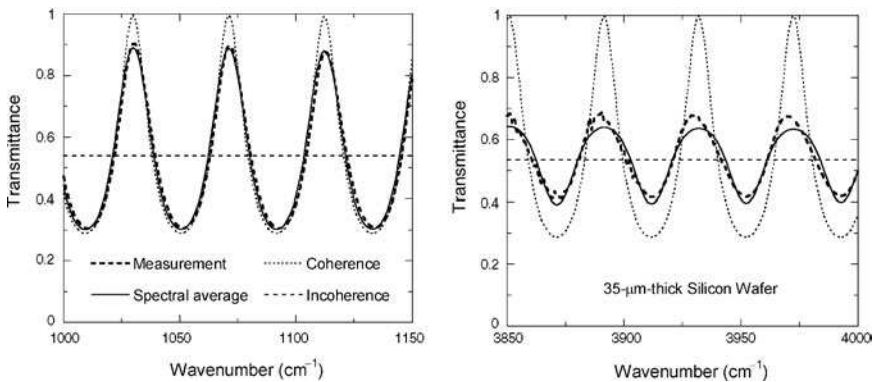


**FIGURE 9.7** Coherence function versus the ratio of the coherence spectral width to the free spectral range for different refractive indices.



does not affect the shape of the curves. The coherence function varies within  $(-1, 1)$ , and its magnitude quantifies the reduction in the fringe contrast from 1 in the coherent limit to 0 in the incoherent limit. The locations where  $\phi = 0$  correspond to  $\delta\nu = m\Delta\nu$  ( $m = 1, 2, 3 \dots$ ), when all fringes disappear in the transmittance spectra. When  $\phi < 0$ , the peaks and the valleys are inverted in the transmittance spectrum, resulting in fringe flipping. When  $n \leq 2$ , it can be seen from Fig. 9.7 that the coherence function is approximated by the sinc function:  $\text{sinc}(x) = \sin(x)/x$ . As refractive index increases, however, the coherence function becomes flatter and deviates from the sinc function. The coherence function serves the same role as the degree of coherence that helps determine which approach (i.e., wave optics, partial coherence formulation, or geometric optics) is most suitable for modeling the radiative properties for a particular case. In addition, Eq. (9.23) can also be applied to rough surfaces, as will be discussed in the next section.

Figure 9.8 shows the measured and predicted transmittance for a double-side polished silicon wafer in two narrow spectral regions as functions of the wavenumber. The trans-

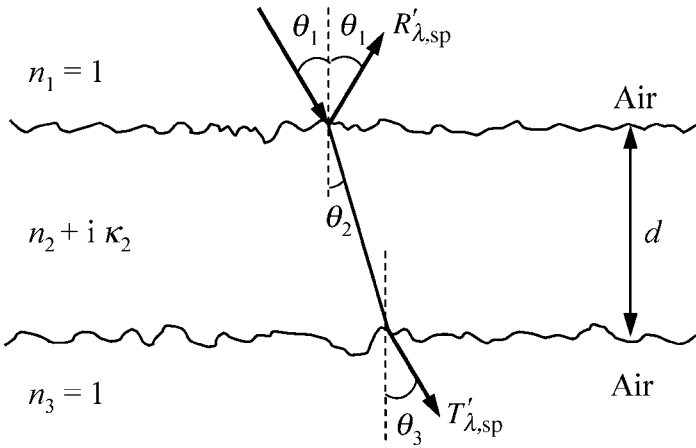


**FIGURE 9.8** Normal transmittance of a 35- $\mu\text{m}$ -thick Si wafer in two narrow spectral regions near the wavelengths of 10  $\mu\text{m}$  ( $1000 \text{ cm}^{-1}$ ) and 2.5  $\mu\text{m}$  ( $4000 \text{ cm}^{-1}$ ), respectively.<sup>10</sup>

mittance spectra in the coherent and incoherent limits are shown for comparison. Because the refractive index of silicon changes less than 1% ( $n = 3.432 \pm 0.011$ ), the free spectral range in wavenumber is  $\Delta\bar{\nu} \approx 41.3 \text{ cm}^{-1}$ , and the transmittance predicted by the incoherence formula is approximately 0.537. It can be seen that the transmittance is less coherent toward short wavelengths (increasing wavenumber). Therefore, a wavenumber-dependent coherence spectral width was used to fit the data obtained from the FTIR spectrometer.<sup>10</sup> The coherence spectral width  $\delta\bar{\nu}$  varies from  $10.4 \text{ cm}^{-1}$  at  $\bar{\nu} = 1000 \text{ cm}^{-1}$  to  $28.7 \text{ cm}^{-1}$  at  $\bar{\nu} = 4000 \text{ cm}^{-1}$ . The coherence function  $\phi$  calculated from Eq. (9.22) changes from 0.84 at  $\bar{\nu} = 1000 \text{ cm}^{-1}$  to 0.33 at  $\bar{\nu} = 4000 \text{ cm}^{-1}$ . The coherence spectral width is much greater than the instrument resolution of  $1 \text{ cm}^{-1}$ , suggesting that the surfaces of the wafer may be slightly nonparallel. The measured transmittance is also sensitive to the mechanical stress on the wafer.

### 9.1.4 Effect of Surface Scattering

In order to model the losses in the reflectance and transmittance due to scattering at the surfaces, shown in Fig. 9.9, the Fresnel coefficients can be modified by the scattering factors that depend on the rms roughness. Notice that the reflectance and transmittance obtained



**FIGURE 9.9** Geometry of a thin film with rough surfaces, in the model of the specular transmittance and reflectance, when  $\kappa_2 \ll n_2$  and  $\sigma_{\text{rms}} \ll \lambda$ .

this way are not directional-hemispherical properties. Because only the reflection and the transmission near the specular directions are considered, we will use *specular reflectance*  $R'_{\lambda,\text{sp}}$  and *specular transmittance*  $T'_{\lambda,\text{sp}}$ . The derivation of the scattering factor is based on the assumptions that the surface height follows the Gaussian distribution and the autocovariance function of surface roughness is also Gaussian. When both the rms roughness and the autocorrelation length are much less than the wavelength of the incident radiation, the *scalar scattering theory* may be applied to determine the reflection coefficients, considering scattering losses.<sup>11</sup> The modified Fresnel coefficients between the media  $j$  and  $k$  ( $j = 1, 2, \text{ or } 3; k = j \pm 1$ ) are given in the following:

$$r'_{jk} = r_{jk} S_{r,jk} \tag{9.24a}$$

and

$$t'_{jk} = t_{jk} S_{t,jk} \tag{9.24b}$$

where the prime refers to the modified Fresnel coefficients for a given polarization, and the scattering factors are defined as follows, based on real refractive indices only:

$$S_{r,jk} = \exp\left[-\frac{1}{2}\left(\frac{4\pi\sigma_{\text{rms}}n_j\cos\theta_j}{\lambda}\right)^2\right] \tag{9.25a}$$

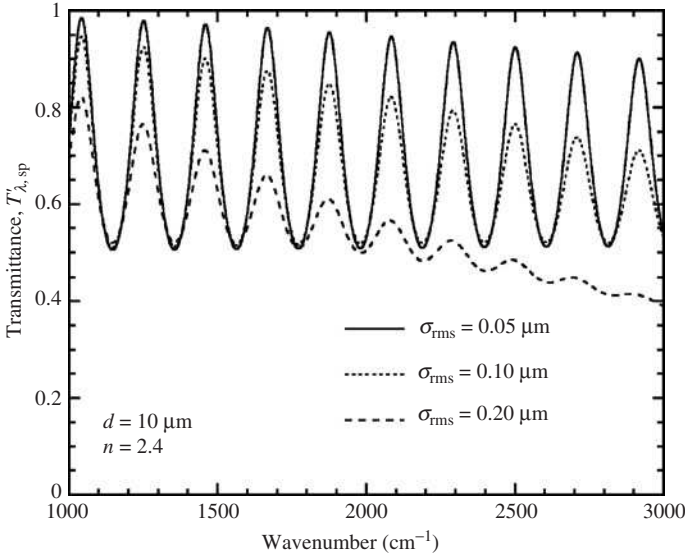
and

$$S_{t,jk} = \exp\left[-\frac{1}{2}\left(\frac{2\pi\sigma_{\text{rms}}(n_j\cos\theta_j - n_k\cos\theta_k)}{\lambda}\right)^2\right] \tag{9.25b}$$

where  $\sigma_{\text{rms}}$  is the rms roughness of the interface.<sup>11</sup> It should be noted that some relations of the Fresnel coefficients, such as  $r_{jk} = -r_{kj}$  and  $1 + r_{jk} = t_{jk}$ , do not hold after the modifications, because of scattering losses. The reflectance and the transmittance should be obtained from Eq. (9.9) and Eq. (9.10). Furthermore, the energy losses due to surface roughness increase toward shorter wavelengths, because of the  $\sigma_{\text{rms}}/\lambda$  term in the scattering factors; this yields a reduction in the fringe contrasts and a decrease in the overall transmittance. Even for a nonabsorbing film, the sum of the specular transmittance and reflectance is not equal to 1, because of scattering losses.

**Example 9-2.** Calculate the normal transmittance of a 10- $\mu\text{m}$  film with a refractive index  $n = 2.4$ , when there is no absorption, in the spectral range from 1000 and 3000  $\text{cm}^{-1}$ . Both surfaces are rough with a roughness  $\sigma_{\text{rms}}$  of 0.10  $\mu\text{m}$ . How does the  $\sigma_{\text{rms}}$  value affect the transmittance?

**Solution.** We can use Eq. (9.10) to calculate the transmittance but with the reflection and transmission coefficients modified by Eq. (9.24). The results are plotted in Fig. 9.10, for  $\sigma_{\text{rms}} = 0.05$ ,



**FIGURE 9.10** Transmittance of a dielectric thin film with surface roughness on both sides.

0.10, and 0.20  $\mu\text{m}$ , to examine the effect of roughness on the specular transmittance. It can be seen that surface roughness reduces both the peak transmittance and the fringe contrast. Furthermore, the reduction is more prominent toward shorter wavelengths.

An optically smooth surface has an rms roughness on the order of 10 nm. Some highly polished semiconductor wafers or thin films, grown by molecular beam epitaxy, can have an rms roughness less than 1 nm. On the other hand, chemical-vapor-deposited (CVD) diamond films and the backside of silicon wafers can have a roughness ranging from 100 nm to 1  $\mu\text{m}$ . The fringe contrast in the measured spectrum is often less than that predicted by wave optics after the modification of the Fresnel coefficient, due to the lack of parallelism between the two surfaces. In other words, when the effect of partial coherence is significant, the scalar scattering theory alone cannot accurately predict the transmittance of thin films. Lee et al. used the fringe-averaging method, along with the scalar scattering theory, to predict the specular transmittance for rough surfaces, and obtained excellent agreement with FTIR measurements for a CVD diamond film and several silicon wafers.<sup>10</sup> On the other hand, the scalar scattering theory cannot be applied when either the autocorrelation length or the rms roughness is comparable with the wavelength.

## 9.2 RADIATIVE PROPERTIES OF MULTILAYER STRUCTURES

Since many applications involve a thin film on a substrate or multilayer thin films, expressions of the radiative properties of multilayer structures are summarized in this section. The

matrix formulation for thin-film multilayer structures will be described, and its application to films on a thick substrate will also be discussed.

### 9.2.1 Thin Films with Two or Three Layers

Examples of two-layer thin films include a metallic coating on a thin dielectric substrate, especially in the long-wavelength region, where interference in the substrate cannot be ignored. The film can also be modeled as a sheet resistance for metallic films in the far-infrared and microwave regions. Nevertheless, thin-film optics is generally applicable to any spectral region and for different materials. The expressions of the reflectance and the transmittance of a thin film-substrate composite in vacuum are

$$R'_{\lambda,F} = \left| r_a + \frac{t_a t_b r_{S0} e^{i2\beta_s}}{1 - r_b r_{S0} e^{i2\beta_s}} \right|^2 \quad (9.26)$$

$$R'_{\lambda,S} = \left| r_{0S} + \frac{t_{0S} t_{S0} r_b e^{i2\beta_s}}{1 - r_b r_{S0} e^{i2\beta_s}} \right|^2 \quad (9.27)$$

and

$$T'_\lambda = \left| \frac{t_a t_{S0} e^{i\beta_s}}{1 - r_b r_{S0} e^{i2\beta_s}} \right|^2 \quad (9.28)$$

where the subscripts F and S indicate whether the incoming radiation is incident on the film or substrate, since the direction of incidence makes a difference for the reflectance,  $\beta_s$  is the complex phase shift inside the substrate;  $t_a$  and  $r_a$  are the transmission and reflection coefficients for incidence from vacuum to the film, when the substrate is assumed semi-infinite;  $t_b$  and  $r_b$  are the transmission and reflection coefficients for incidence from the substrate to the film; and subscripts S0 and OS refer to the Fresnel coefficients at the substrate-vacuum interface. The reflection and transmission coefficients  $r_a$ ,  $r_b$ ,  $t_a$ , and  $t_b$  are generally complex and should be calculated from Eq. (9.7) and Eq. (9.8) using the phase shift of the film. The absorbance also depends on which side the radiation is incident from. When there is another coating at the backside of the substrate, one can replace the Fresnel coefficient with the transmission and reflection coefficients of the film.

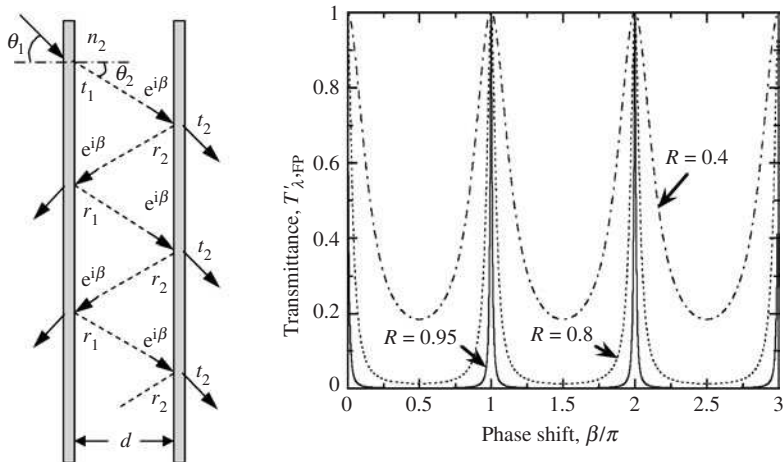
**Example 9-3.** A Fabry-Perot interferometer can be built with two mirrors made by coating highly reflecting materials (e.g., ultrathin metallic films) on both sides of a dielectric thin film, as illustrated on the left of Fig. 9.11. Derive a formula for the transmittance, and show that resonance in transmittance can be obtained within narrow spectral bands.

**Solution.** Before 1900, Charles Fabry and Alfred Perot constructed a device based on interference effect and published a series of papers on the possible applications in metrology and spectroscopy. This is the Fabry-Perot interferometer, also known as an optical cavity resonator or etalon. Like the Michelson interferometer, the Fabry-Perot interferometer is an important device used in spectroscopy, laser applications, and wavelength and frequency standards.<sup>11</sup> By considering the transmission and reflection coefficients  $t_1$ ,  $t_2$ ,  $r_1$ , and  $r_2$ , at each boundary of the dielectric film, the overall transmittance coefficient of the Fabry-Perot interferometer, shown in Fig. 9.11, can be expressed as follows:

$$t_{FP} = \frac{t_1 t_2 e^{i\beta}}{1 - r_1 r_2 e^{i2\beta}} \quad (9.29)$$

where  $\beta = 2\pi n_2 \bar{v} d_2 \cos \theta_2$  is the phase shift according to Eq. (9.6) Here,  $\bar{v} = 1/\lambda$  is the wavenumber in  $\text{cm}^{-1}$ . The energy transmittance can be written as follows:

$$T'_{\lambda,FP} = t_{FP} t_{FP}^* = \frac{T_1 T_2}{\left(1 - \sqrt{R_1 R_2}\right)^2 + 4\sqrt{R_1 R_2} \sin^2 \psi} \quad (9.30)$$



**FIGURE 9.11** Schematic of a Fabry-Perot interferometer (left) and the calculated transmittance for different  $R$  values (right).

where  $\psi = \beta + \arg(r_1)/2 + \arg(r_2)/2$  is a new phase angle,  $T_1 = t_1 t_1^*$  and  $T_2 = t_2 t_2^*$  are not exactly the transmittances through the coating, and  $R_1 = r_1 r_1^*$  and  $R_2 = r_2 r_2^*$  are indeed the reflectances for incidence from the dielectric to the left and right boundaries, respectively. When the loss can be neglected and the structure is symmetric, we have  $\psi = \beta$ ,  $R_1 = R_2 = R$ , and  $T_1 T_2 = (1 - R)^2$ ; thus, Eq. (9.30) can be simplified as

$$T'_{\lambda,FP} = \frac{(1 - R)^2}{(1 - R)^2 + 4R \sin^2 \beta} \quad (9.31)$$

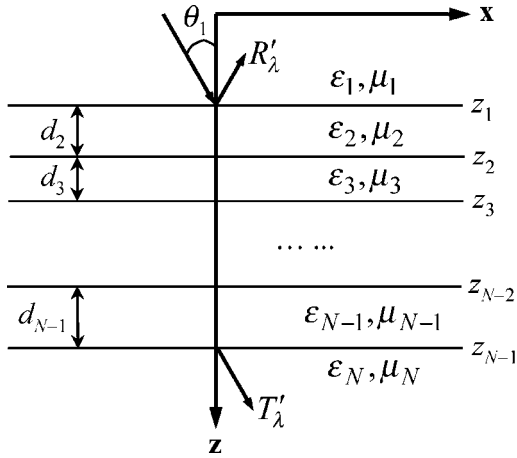
The results for different  $R$  values are shown on the right of Fig. 9.11. Clearly, a large  $R$  yields sharp transmission peaks at  $\beta = m\pi$ . Suppose the refractive index of the dielectric is kept constant and the change of the phase shift corresponds to the frequency variation, the free spectral range is the interval between two resonance peaks, given by  $\Delta\bar{\nu} = 1/(2n_2 d_2 \cos \theta_2)$ , similar to that of Eq. (9.12a). The full-width-at-half-maximum (FWHM),  $\delta\bar{\nu}$ , measures how sharp the peak is. The ratio  $Q = \Delta\bar{\nu}/\delta\bar{\nu}$  is called the *fineness* of the interferometer, which determines the resolving power. The fineness is known as the  $Q$ -factor of the resonator. For a lossless Fabry-Perot cavity, it can be shown that

$$Q = \frac{\Delta\bar{\nu}}{\delta\bar{\nu}} = \frac{\pi\sqrt{R}}{1 - R} \quad (9.32)$$

which is 313, when  $R = 0.99$ . Kumar et al. constructed a Fabry-Perot resonator, based on high-critical-temperature superconducting films on Si substrates, and demonstrated sharp transmission peaks in the far-infrared at cryogenic temperatures, when  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$  becomes superconducting.<sup>12</sup>

## 9.2.2 The Matrix Formulation

A multilayer structure containing  $N$  layers is shown in Fig. 9.12. In this section, the 1-D matrix formulation is present in such a way that magnetic materials can also be included. Each layer is assumed to be isotropic and homogeneous, and it can be fully described by a relative permittivity  $\epsilon_l$  and a relative permeability  $\mu_l$  ( $l = 1, 2, \dots, N$ ). For a monochromatic plane wave originated from layer 1, which is assumed to be lossless, the phase-matching condition requires that  $k_{lx} \equiv k_x = \omega n_1 \sin \theta_1/c$ . Consider a linearly polarized electromagnetic wave, whose plane of incidence is perpendicular to the  $y$ -axis. For  $s$  polarization or



**FIGURE 9.12** Schematic illustration of an  $N$ -layer structure, where the first and last layers are semi-infinite, and each layer is assumed to be homogeneous and isotropic.

TE wave, where the electric field is parallel to the  $y$ -axis, the electric field in the  $l$ th layer can be written as  $E_l(z)e^{i(k_x x - \omega t)}$ , where<sup>13</sup>

$$E_l(z) = A_l e^{ik_{lz}z} + B_l e^{-ik_{lz}z}$$

and 
$$E_l(z) = A_l e^{ik_{lz}(z-z_{l-1})} + B_l e^{-ik_{lz}(z-z_{l-1})}, \quad l = 2, 3, \dots, N \tag{9.33}$$

Here,  $A_l$  and  $B_l$  are the amplitudes of the forward and backward waves at the interface, respectively,  $z_l = z_{l-1} + d_l$  ( $l = 2, 3, \dots, N - 1$ ), and  $d_l$  is the layer thickness. The magnetic field can be obtained from the electric field using Maxwell’s equations. The expression of the wave component  $k_{lz}$  is calculated from  $k_x^2 + k_{lz}^2 = \epsilon_l \mu_l \omega^2 / c^2$ . The only condition imposed is that the imaginary part of  $k_{lz}$  must not be less than zero. This will ensure that the wave will decay toward positive  $z$ . After applying boundary conditions at the interface, we obtain the field amplitudes of adjacent layers relate as

$$\begin{pmatrix} A_l \\ B_l \end{pmatrix} = \mathbf{P}_l \mathbf{D}_l^{-1} \mathbf{D}_{l+1} \begin{pmatrix} A_{l+1} \\ B_{l+1} \end{pmatrix}, \quad l = 1, 2, \dots, N - 1 \tag{9.34}$$

In Eq. (9.33),  $\mathbf{P}_l$  is the propagation matrix given by

$$\mathbf{P}_l = \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad l = 1$$

and 
$$\mathbf{P}_l = \begin{pmatrix} e^{-ik_{lz}d_l} & 0 \\ 0 & e^{ik_{lz}d_l} \end{pmatrix}, \quad l = 2, 3, \dots, N - 1 \tag{9.35}$$

$\mathbf{D}_l$  is called the dynamical matrix, and  $\mathbf{D}_l^{-1}$  is its inverse. For  $s$  polarization,  $\mathbf{D}_l$  is given in terms of  $k_{lz}$  and  $\mu_l$  as follows:

$$\mathbf{D}_l = \begin{pmatrix} 1 & 1 \\ k_{lz}/\mu_l & -k_{lz}/\mu_l \end{pmatrix}, \quad l = 1, 2, \dots, N \tag{9.36}$$

By successively applying Eq. (9.35) to all layers, we have

$$\begin{pmatrix} A_1 \\ B_1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} A_N \\ B_N \end{pmatrix} \quad (9.37)$$

where

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \prod_{l=1}^{N-1} \mathbf{P}_l \mathbf{D}_l^{-1} \mathbf{D}_{l+1} \quad (9.38)$$

The electric field transmission and reflection coefficients are obtained by setting  $B_N = 0$ , because the last layer is semi-infinite, and thus there is no backward wave. Simple algebraic manipulations give the expressions of the coefficients as

$$t = A_N/A_1 = 1/M_{11} \quad (9.39)$$

and

$$r = B_1/A_1 = M_{21}/M_{11} \quad (9.40)$$

Furthermore, the energy reflectance and transmittance are given as follows:

$$R'_\lambda = rr^* = \left| \frac{M_{21}}{M_{11}} \right|^2 \quad (9.41)$$

$$T'_\lambda = \frac{\operatorname{Re}(k_{Nz}/\mu_N^*)}{\operatorname{Re}(k_{1z}/\mu_1^*)} tt^* = \frac{\operatorname{Re}(k_{Nz}/\mu_N)}{\operatorname{Re}(k_{1z}/\mu_1)} \left| \frac{1}{M_{11}} \right|^2 \quad (9.42)$$

For  $p$  polarization or TM wave, the magnetic field is parallel to the  $y$ -axis. Equation (9.33) can be written in terms of the magnetic field. The previous procedure can then be applied to derive the transmission and reflection coefficients based on the magnetic fields. Then, the dynamical matrix  $\mathbf{D}_l$  given in Eq. (9.36) must be replaced by

$$\mathbf{D}_l = \begin{pmatrix} 1 & 1 \\ k_{1z}/\varepsilon_l & -k_{1z}/\varepsilon_l \end{pmatrix}, \quad l = 1, 2, \dots, N \quad (9.43)$$

The expression for the reflectance is the same as Eq. (9.41), and that for transmittance for  $p$  polarization becomes

$$T'_\lambda = \frac{\operatorname{Re}(k_{Nz}/\varepsilon_N)}{\operatorname{Re}(k_{1z}/\varepsilon_1)} tt^* = \frac{\operatorname{Re}(k_{Nz}/\varepsilon_N)}{\operatorname{Re}(k_{1z}/\varepsilon_1)} \left| \frac{1}{M_{11}} \right|^2 \quad (9.44)$$

The assumption that the first medium is lossless is necessary because the reflectance is ill-defined if the first medium is lossy, because of the coupling between the reflected and incident waves.<sup>6</sup> However, Eq. (9.41), Eq. (9.42), and Eq. (9.44) are applicable even though the last medium is lossy. Comparing Eq. (9.42) with Eq. (9.44), and Eq. (9.36) with Eq. (9.43), we immediately notice the duality of the electric and magnetic fields, since the only difference is the interchange of  $\varepsilon$  and  $\mu$  in these equations. Further applications of the matrix formulation will be discussed in subsequent sections as well as in the next chapter.

### 9.2.3 Radiative Properties of Thin Films on a Thick Substrate

Radiative properties of thin coatings on a substrate are important for a large number of applications, such as a thermal oxide on a Si substrate, antireflection coatings on the lens

of glasses, interference filters, metallic coatings, and superconducting films. In these cases, the coating thicknesses are on the order of nanometers and they must be considered as thin films. On the other hand, the substrate is usually thick enough to be considered incoherent, while being semitransparent for energy transfer consideration. Furthermore, the substrate is either lossless or slightly absorbing ( $\kappa_s \ll n_s$ ), as discussed earlier. Figure 9.13 shows

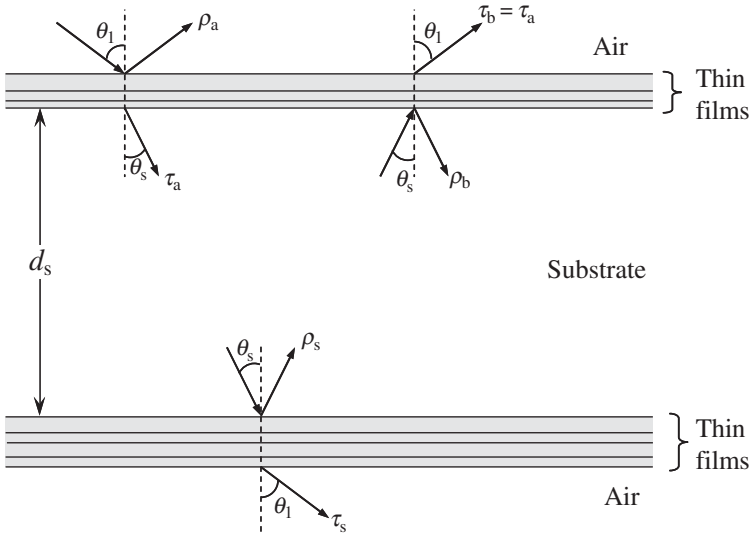


FIGURE 9.13 Radiative properties of multilayer thin films on an incoherent, thick substrate.

the geometry of an incoherent substrate of thickness  $d_s$  bounded by multilayer thin films on both sides. The refraction angle  $\theta_s$  in the substrate can be calculated from the incidence angle  $\theta_1$  by neglecting absorption of the substrate. In Fig. 9.13,  $\rho_a$  or  $\rho_b$  refers to the reflectance of the first multilayer structure for rays originated from air or the substrate, and  $\tau_a$  and  $\tau_b$  are the corresponding transmittance. Furthermore,  $\rho_s$  and  $\tau_s$  represent the reflectance and transmittance for rays originated from the substrate at the second multilayer structure. Since the coupling between the incident and reflected waves in the substrate is negligible, the transmittance is the same whether the ray is originated from air or the substrate, i.e.,  $\tau_b = \tau_a$ . In order to calculate these parameters, the matrix formulation discussed in the previous section can be separately applied to the multilayer structures for a given incident direction. The internal transmittance of the substrate is  $\tau = \exp(-4\pi\kappa_s d_s / \lambda \cos\theta_s)$ , where  $\lambda$  is the wavelength in vacuum. The reflectance and the transmittance of the multilayer structure can be calculated using the ray-tracing method and expressed as follows:

$$R'_\lambda = \rho_a + \frac{\rho_s \tau_a^2 \tau^2}{1 - \rho_s \rho_b \tau^2} \tag{9.45a}$$

$$T'_\lambda = \frac{\tau_a \tau_s \tau}{1 - \rho_s \rho_b \tau^2} \tag{9.45b}$$



Radiative properties of arbitrary numbers of thick and thin layers have been derived theoretically.<sup>14</sup> For each thin-film stack, the field reflection and transmission coefficients are obtained first, using the matrix formulation described previously. The power transmittance and reflectance at the interfaces of each thick layer can then be obtained. Using the net-radiation method, the energy transmittance and reflectance can be evaluated. Spectral averaging is another and perhaps more powerful technique of obtaining the transmittance and reflectance for systems involving thick and thin layers.

Lee and Zhang developed a reliable and easily accessible software tool, named *Rad-Pro* (for radiative properties), in an Excel-VBA environment.<sup>15</sup> It allows users to calculate the directional, spectral, and temperature dependence of the radiative properties for the multilayer structures of silicon, including doping effects, and related materials such as silicon dioxide, silicon nitride, and polysilicon. *Rad-Pro* contains various options such as coherence versus incoherence, spectral averaging, polarization status, and input of user-defined materials. This software is downloadable free of charge at the author's website: [www.me.gatech.edu/~zzhang](http://www.me.gatech.edu/~zzhang).

### 9.2.4 Local Energy Density and Absorption Distribution

The absorptance of the composite layers can be calculated by subtracting the reflectance and the transmittance from unity. The Poynting vector can be evaluated as a function of  $z$  to obtain the radiant energy flux  $S(z) = \frac{1}{2} \text{Re}(\mathbf{E} \times \mathbf{H}^*)$ . The fraction of energy absorbed between  $z_1$  and  $z_2$  is given by

$$\alpha_{z_1-z_2} = \frac{S_z(z_1) - S_z(z_2)}{S_{iz}} \quad (9.46)$$

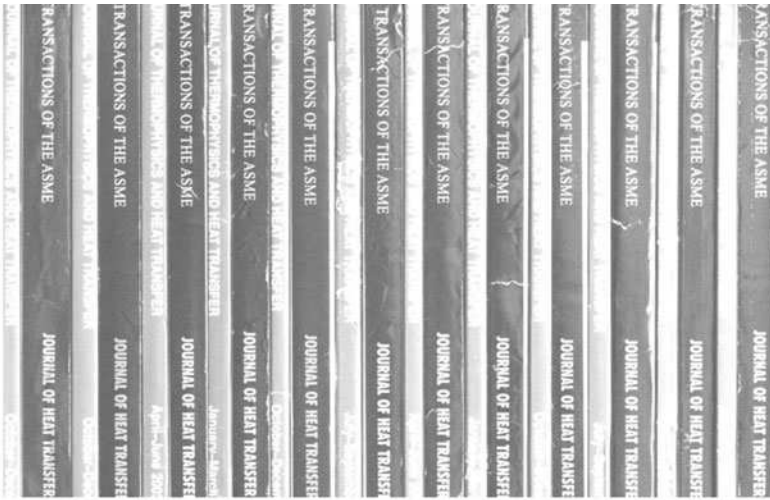
where  $S_{iz}$  is the incident radiant energy flux in the  $z$  direction. From Sec. 8.1.4, one can obtain the local energy density. The energy dissipated per unit volume is given by  $-\nabla \cdot \mathbf{S}$ .

## 9.3 PHOTONIC CRYSTALS

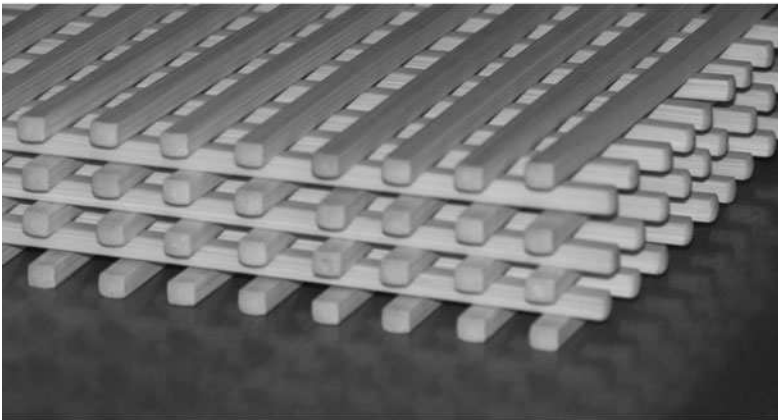
Recently, many studies have utilized the unique features of periodic microstructures (i.e., photonic crystals) to engineer the radiative properties for specific applications.<sup>16,17</sup> A *photonic crystal (PC)* is a periodic array of unit cells (i.e., photonic lattices in analog to those in real crystals), which replicate infinitely into one, two, or three dimensions. Figure 9.14a illustrates a 1-D PC using the arrangement of alternating *Journal of Heat Transfer* and *Journal of Thermophysics and Heat Transfer* issues in the author's bookshelf. To have a PC with a period of the order of infrared wavelengths, say 3  $\mu\text{m}$ , the thickness needs to be reduced by a factor of 6000. Figure 9.14b is a photo of a stack of chopsticks in three dimensions. Structures of 3-D tungsten PCs have been fabricated with a rod width of 1.2  $\mu\text{m}$  and rod-to-rod spacing of 4.2  $\mu\text{m}$ , for tuning the infrared thermal emission properties.<sup>18</sup>

From the analogy of the electron movement in crystals, electromagnetic wave propagation in a PC should also satisfy the Bloch condition, discussed in Chap. 6. Similarly, due to the periodicity, a PC exhibits band structures consisting of pass and stop bands when the frequency is plotted against the wavevector. In the pass band, for instance, waves can propagate inside a PC. Whereas in the stop band, no energy-carrier waves can exist inside a PC, and only oscillating but evanescently decaying fields possibly exist. The existence of stop bands enables a PC to be used in many optoelectronic devices such as band-pass filters and waveguides.<sup>4,5,13,19</sup> Most of the 1-D PCs are made with alternating layers of two lossless dielectrics, while metal-dielectric PCs have recently been developed and studied by several groups. In some cases, the dimension may be smaller than 100 nm for tuning the visible properties.

While 3-D PCs with complicated structures have been fabricated and used in a number of applications, the fundamental physics can be illustrated using 1-D PCs and can easily be



(a) 1-D structure made by alternating layers

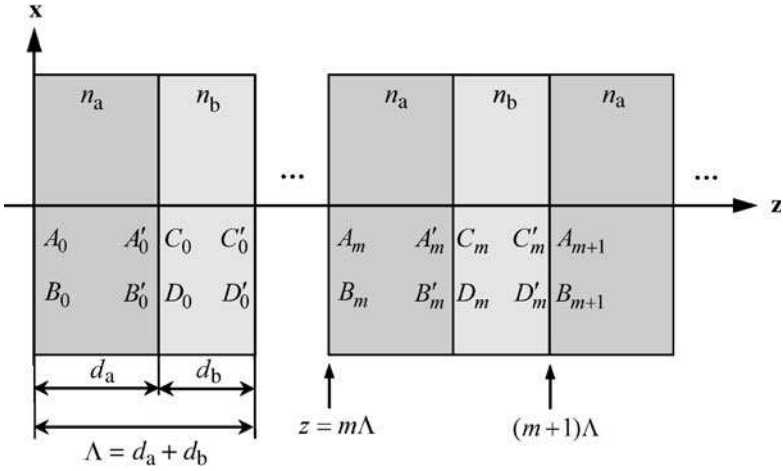


(b) 3-D structures made by stacking rod

**FIGURE 9.14** Illustration of 1-D and 3-D photonic crystal (PC) structures.

generalized for 2-D or 3-D structures. The 1-D PC, illustrated in Fig. 9.15, is a periodic multilayer structure, where  $\Lambda = d_a + d_b$  is the period of the PC or photonic lattice constant. The unit cell is composed of alternating dielectrics with different refractive indices  $n_a$  and  $n_b$ . It is assumed that all layers are infinitely extended in the  $x$ - $y$  plane, and the PC is in the positive- $z$  half space starting with  $m = 0$  at  $z = 0$ . From the analogy between wave propagation in a periodic media and the motion of electrons in crystalline materials, the electric field vector in the 1-D PC, for a monochromatic electromagnetic wave of angular frequency  $\omega$ , should satisfy the Bloch condition given by

$$\mathbf{E}(x, y, z, t) = \mathbf{u}(z) e^{iKz} e^{i(k_x x + k_y y - \omega t)} \tag{9.47}$$



**FIGURE 9.15** Amplitudes of the forward and backward waves in a semi-infinite 1-D PC, in the right half space. The unit cell of the 1-D PC is made of two dielectric layers: type a and type b, and has a period  $\Lambda = d_a + d_b$ .

where  $\mathbf{u}(z + \Lambda) = \mathbf{u}(z)$  is a periodic function of  $z$  with a period equal to the lattice constant of the photonic crystal,  $k_x$  and  $k_y$  are the parallel components of the wavevectors that must be the same in all layers as required by the phase-matching condition, and  $K$  represents the Bloch wavevector that is a scalar in the 1-D case. Here,  $K$  is a characteristic parameter of the PC that is the same for all layers. The wavevector components in the  $z$  direction are  $k_{az}$  and  $k_{bz}$  in media a and b, respectively, and are determined by the relations  $k_x^2 + k_y^2 + k_{az}^2 = k_a^2 = n_a^2 \omega^2 / c^2$  and  $k_x^2 + k_y^2 + k_{bz}^2 = k_b^2 = n_b^2 \omega^2 / c^2$ . From the Bloch condition, the electric field in the 1-D PC satisfies the following equation:

$$\mathbf{E}(x, y, z + \Lambda, t) = \mathbf{E}(x, y, z, t) e^{iK\Lambda} \tag{9.48}$$

The magnetic field is related to the electric field by Maxwell’s equations and must also follow the Bloch condition. Therefore, the fields inside the PC are not periodic functions.

Because of the axial symmetry, the coordinate can always be rotated around the  $z$ -axis to make  $k_y = 0$ . For  $s$  polarization, the electric field is parallel to the  $y$  direction and can be expressed as

$$E_y(x, z) = \begin{cases} [A_m e^{ik_{az}(z-m\Lambda)} + B_m e^{-ik_{az}(z-m\Lambda)}] e^{ik_x x}, & m\Lambda \leq z \leq (m\Lambda + d_a) \\ [C_m e^{ik_{bz}(z-m\Lambda-d_a)} + D_m e^{-ik_{bz}(z-m\Lambda-d_a)}] e^{ik_x x}, & (m\Lambda + d_a) \leq z \leq (m+1)\Lambda \end{cases} \tag{9.49}$$

where the time-dependent term  $\exp(-i\omega t)$  is omitted for simplicity,  $m$  is an integer,  $A_m$  and  $C_m$  are the amplitudes of forward waves, and  $B_m$  and  $D_m$  are the amplitudes of backward waves at the interfaces, as shown in Fig. 9.15.<sup>13</sup> The coefficients  $A'_m = e^{ik_{az} d_a} A_m$ ,  $B'_m = e^{-ik_{az} d_a} B_m$ ,  $C'_m = e^{ik_{bz} d_b} C_m$ , and  $D'_m = e^{-ik_{bz} d_b} D_m$  are the amplitudes at the other side of the boundary. Boundary conditions require that the tangential components of the electric and magnetic fields  $E_y$  and  $H_x$ , respectively, to be continuous at each interface. From the matrix formulation, the coefficients  $A_m$  and  $B_m$  at  $z = m\Lambda$  are related to those at  $z = (m+1)\Lambda$  with the propagation matrix  $\mathbf{P}$  and dynamical matrix  $\mathbf{D}$  as follows:

$$\begin{pmatrix} A_m \\ B_m \end{pmatrix} = (\mathbf{P}_a \mathbf{D}_a^{-1} \mathbf{D}_b) (\mathbf{P}_b \mathbf{D}_b^{-1} \mathbf{D}_a) \begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} \tag{9.50}$$

From Eq. (9.48), the ratio of the electric fields at two points separated by a period  $\Lambda$  along the  $z$  direction is equal to  $\exp(iK\Lambda)$ ; thus,

$$\begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} = e^{iK\Lambda} \begin{pmatrix} A_m \\ B_m \end{pmatrix} \quad (9.51)$$

The Bloch wavevector parameter  $K$  can be obtained by solving the eigenvalue equation:

$$\mathbf{M} \begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} = e^{-iK\Lambda} \begin{pmatrix} A_{m+1} \\ B_{m+1} \end{pmatrix} \quad (9.52)$$

where  $\mathbf{M} = (\mathbf{P}_a \mathbf{D}_a^{-1} \mathbf{D}_b)(\mathbf{P}_b \mathbf{D}_b^{-1} \mathbf{D}_a)$ . In general,  $K$  depends on the frequency  $\omega$  and the parallel wavevector component  $k_x$ , for a given geometry and refractive indices. Once  $K$  is determined, the electric field in the PC can be expressed in the Bloch wave form as

$$E_y(x, z) = u(z) e^{iKz} e^{ik_x x} \quad (9.53)$$

where  $u(z)$  is a periodic function of  $z$ . For  $m\Lambda \leq z \leq (m+1)\Lambda + d_a$ ,

$$u(z) = [A_0 e^{ik_x(z-m\Lambda)} + B_0 e^{-ik_x(z-m\Lambda)}] e^{-iK(z-m\Lambda)} \quad (9.54a)$$

and for  $(m+1)\Lambda \leq z \leq (m+2)\Lambda + d_a$ ,

$$u(z) = [C_0 e^{ik_x(z-m\Lambda-d_a)} + D_0 e^{-ik_x(z-m\Lambda-d_a)}] e^{-iK(z-m\Lambda)} \quad (9.54b)$$

Note that  $A_0$  and  $B_0$  are amplitudes of the first layer, i.e., at  $m = 0$ , and

$$\begin{pmatrix} C_0 \\ D_0 \end{pmatrix} = (\mathbf{P}_a \mathbf{D}_a^{-1} \mathbf{D}_b)^{-1} \begin{pmatrix} A_0 \\ B_0 \end{pmatrix} \quad (9.55)$$

The expressions for the magnetic field can be obtained from those of the electric field using Maxwell's equations. For  $p$  polarization, the magnetic field is parallel to the  $y$ -axis. The same procedure can be used to determine the magnetic field first and then the electric field. The amplitudes  $A_0$  and  $B_0$  depend on the boundary condition at  $z = 0$ , i.e., the interaction of the PC with the medium in the left half space.

For a given PC, the Bloch wavevector can be solved from the eigenvalue problem given in Eq. (9.52), for any real positive values of  $\omega$  and  $k_x$ . In general,  $K$  is complex. When  $K$  is purely real, i.e.,  $\text{Im}(K) = 0$ , the electric field oscillates in the direction of  $z$ , and the Bloch wave propagates into the positive  $z$  direction, which is called an *extended mode*. When  $\text{Im}(K) \neq 0$ , on the other hand, the amplitude of the Bloch wave decays exponentially along the positive  $z$  direction, and the wave is confined to the first few unit cells of the photonic crystal; this is called a *localized mode*.<sup>6,17</sup> For the localized mode, the field is localized in the vicinity of the defect or the edge. Notice that  $K = K(k_x, \omega)$ , and the regions with  $\text{Im}(K) = 0$  in the  $\omega$ - $k_x$  plane are called pass bands, and those with  $\text{Im}(K) \neq 0$  are called stop bands. Suppose light is incident from air (in the left half space) on the PC at  $z = 0$ . In the stop band, the PC will act like a perfect mirror, which is also called a Bragg reflector. A diagram in the  $\omega$ - $k_x$  domain, showing the different regions, allows one to study the band structures of a PC, as demonstrated in the following example.

**Example 9-4.** Consider the 1-D PC depicted in Fig. 9.15, with the following parameters:  $n_a = 2.4$ ,  $n_b = 1.5$ , and  $d_a = d_b$ . Construct the band structure for both polarizations, and calculate the normal reflectance.

**Solution.** The PC is semi-infinite, and the incidence is from air. The unit cell of the 1-D PC is defined by the thickness  $\Lambda = d_a + d_b$ . Following the previous discussion, we have calculated the band

structure of the 1-D PC for either polarization, and the results are shown in Fig. 9.16. Here, the parallel component of the wavevector is  $k_x = (\omega/c)\sin\theta$ , where  $\theta$  is the angle of incidence. The band

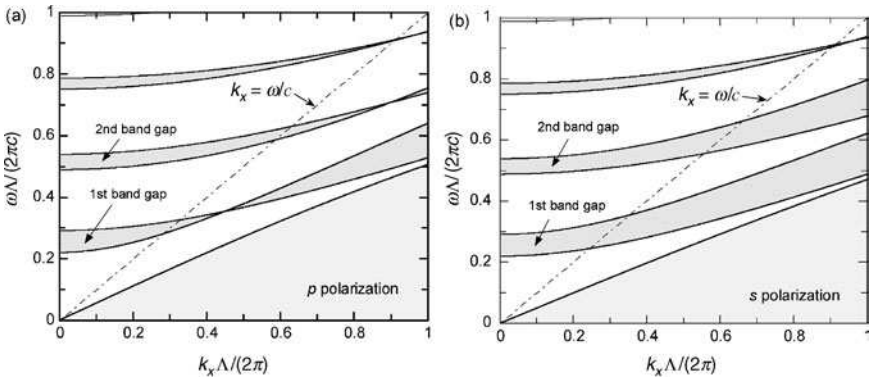


FIGURE 9.16 Band structures of a 1-D PC. (a) TM wave ( $p$  polarization). (b) TE wave ( $s$  polarization).

structure is expressed by the reduced frequency and wavevector; hence, it is independent of the period of the PC. For the calculation, it is assumed that the 1-D PC is a perfectly periodic structure infinitely extended into the  $z$  direction (i.e., no defect or edge exists in the PC). The shaded regions represent the stop bands, while unshaded regions are the pass bands. The light line in air, which corresponds to  $\theta = 90^\circ$ , is plotted as a dash-dot line based on  $\omega = k_x c$ . On the upper-left side of this line, propagating waves exist in air and  $\theta \leq 90^\circ$ . On the lower-right side of this line, evanescent waves exist in air since  $\theta$  becomes complex. Note that stop bands shrink to zero only for  $p$  polarization. The point where the top and bottom band edges merge together corresponds to the Brewster angle between the dielectric of types a and b of the PC. At the Brewster angle, the reflectivity at the interface between two dielectrics is zero; thus, waves or incident energy can propagate into the PC. For the 1-D PC considered here, because the Brewster angle is located on the lower-right side of the light line, the propagating waves in air will not be affected by the Brewster angle of the constituent dielectrics of the PC.

Figure 9.17 shows the reflectance of the 1-D PC structure with different numbers of periods ( $N = 30$  and  $300$ ), calculated using the 1-D matrix formulation. The wavelength is normalized to the period  $\Lambda$ . The reflectance approaches unity in the stop band (when  $N > 30$ ). In the pass band, interference causes oscillations in the reflectance. Since the free spectral range decreases as the total thickness of the PC increases, the oscillation frequency increases with the number of periods of the PC structure. A special type of 1-D PCs is the Bragg reflector, which is composed of alternating high- and low-index films, each at a thickness of one-quarter of the wavelength in the film, i.e.,  $d_a = \lambda/(4n_a)$  and  $d_b = \lambda/(4n_b)$ . Further discussion about surface waves and coherent emission characteristics of PC structures will be deferred to the next chapter.

## 9.4 PERIODIC GRATINGS

The diffraction grating is considered as one of the simplest and most important devices in optical metrology, and many studies have been performed on the effect of gratings on radiative property modification.<sup>20,21</sup> Nanoscale diffraction elements fabricated using nanolithography may enable many applications in biochemical sensing, surface diagnostics, and nanophotonics. Patterned semiconductor microelectronics has periodic structures on the surface with a period below 100 nm.<sup>22</sup> Understanding the radiative properties is essential for thermal processing and modeling in semiconductor manufacturing as the feature size continues to shrink.

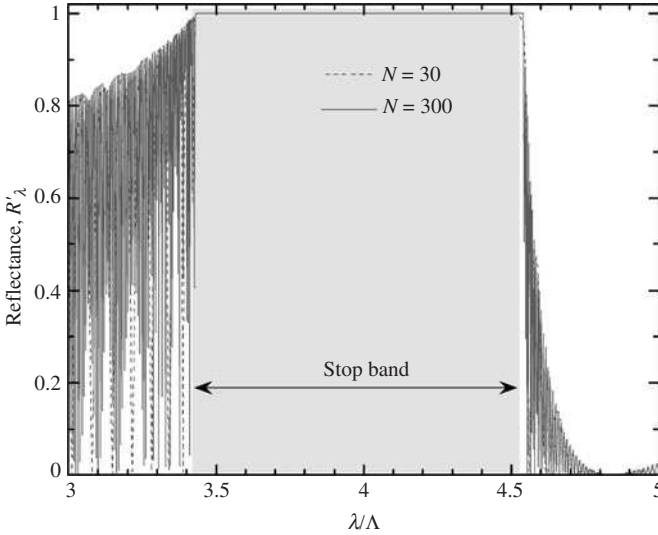


FIGURE 9.17 Reflectance of a 1-D PC, with different numbers of periods.

In the inhomogeneous region, where the permittivity  $\varepsilon$  and the permeability  $\mu$  are spatial functions, the monochromatic plane wave equations become more complicated. By assuming the solution is a time-harmonic plane wave, we can rewrite the Maxwell equations as follows:

$$\nabla \times \mathbf{E} = i\omega\mu\mu_0\mathbf{H} \quad (9.56)$$

$$\nabla \times \mathbf{H} = -i\omega\varepsilon\varepsilon_0\mathbf{E} \quad (9.57)$$

$$\varepsilon\nabla \cdot \mathbf{E} + \mathbf{E} \cdot \nabla\varepsilon = 0 \quad (9.58)$$

$$\mu\nabla \cdot \mathbf{H} + \mathbf{H} \cdot \nabla\mu = 0 \quad (9.59)$$

Since only isotropic media are considered here, both  $\varepsilon$  and  $\mu$  are scalars. By taking the curl of Eq. (9.56) and applying the vector identities in Appendix B.7 and Eq. (9.58) and Eq. (9.59), we obtain

$$\nabla^2\mathbf{E} + \nabla(\mathbf{E} \cdot \nabla\ln\varepsilon) + \nabla\ln\mu \times (\nabla \times \mathbf{E}) + k^2\mu\varepsilon\mathbf{E} = 0 \quad (9.60)$$

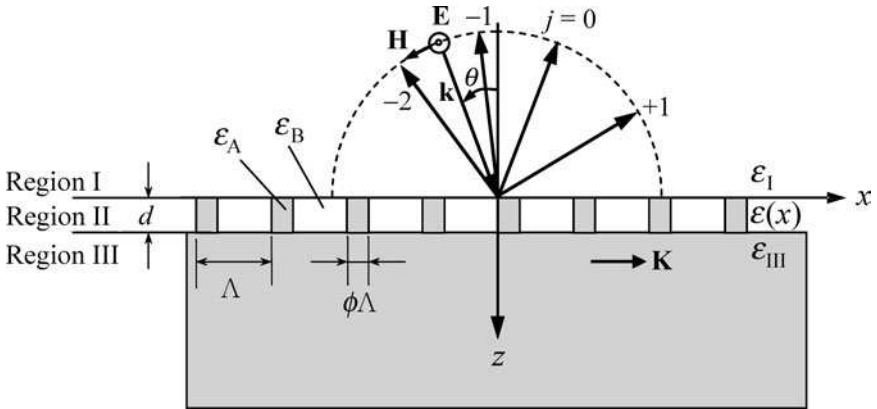
$$\nabla^2\mathbf{H} + \nabla(\mathbf{H} \cdot \nabla\ln\mu) + \nabla\ln\varepsilon \times (\nabla \times \mathbf{H}) + k^2\mu\varepsilon\mathbf{H} = 0 \quad (9.61)$$

where  $k = \omega/c$  is the wavevector in vacuum.<sup>13</sup>

These equations cannot be solved easily and numerical methods are often required. Among them are rigorous coupled-wave analysis (RCWA), finite-difference time-domain (FDTD), finite element method (FEM), boundary element method (BEM), as well as the volume integral method. Effective medium formulation is another approach that takes the average field by approximating the inhomogeneous medium with effective homogenous electric and magnetic properties. The concept of RCWA will be presented next because it is an effective tool for calculating the optical properties of the grating geometry with sufficient accuracy. We will then give a brief discussion of the effective medium formulation, which may be applied to both periodic and random media under restricted conditions.

### 9.4.1 Rigorous Coupled-Wave Analysis (RCWA)

We begin with a discussion restricted to  $s$  polarization and for incidence perpendicular to the gratings, as illustrated in Fig. 9.18. A plane wave is incident on a 1-D grating surface



**FIGURE 9.18** Schematic drawing for a TE wave incident on a grating layer, showing the reflected diffraction orders  $j = -2, -1, 0$ , and  $1$ .

from free space, region I with  $\epsilon_1 = n_1 = 1$  and  $\kappa_1 = 0$ . Region II is composed of binary materials A and B so that the dielectric function in region II is a periodic function of  $x$  with a period  $\Lambda$ , i.e., the grating period. The filling ratio  $\phi$  is the volume fraction of material A, and the lateral extension of the grating is assumed to be infinite. Region III is the substrate with a dielectric function  $\epsilon_{\text{III}}$ .

The wavevector  $\mathbf{k}$  defines the direction of incidence, and the angle between  $\mathbf{k}$  and the surface normal  $\hat{\mathbf{z}}$  is the angle of incidence  $\theta$ , also called the polar angle. The grating vector  $\mathbf{K}$  is defined in the positive  $x$  direction with a magnitude  $K = 2\pi/\Lambda$ . In the following discussion, it is assumed that the incident wavevector is on the  $x$ - $z$  plane, i.e., the  $y$  component of  $\mathbf{k}$  is zero. For  $s$  polarization, the electric field  $\mathbf{E}$  is parallel to the  $y$  direction and perpendicular to the grating vector  $\mathbf{K}$ . The magnitude of the incident electric field, after normalization, can be expressed as  $\exp(ik_x x + ik_z z - i\omega t)$ . For simplicity, the time-harmonic term  $\exp(-i\omega t)$  will be omitted hereafter. The magnitude of  $\mathbf{k}$  in regions I and III can be expressed as

$$k_1 = \frac{2\pi n_1}{\lambda} = \frac{2\pi}{\lambda} = k \quad \text{and} \quad k_{\text{III}} = \frac{2\pi n_{\text{III}}}{\lambda} = n_{\text{III}} k \quad (9.62)$$

where  $n_{\text{III}}$  is the refractive index of region III. There exists a phase difference of  $2\pi\Lambda \sin\theta/\lambda = k_x \Lambda$  between the incident wave at  $(x, z)$  and that at  $(x+\Lambda, z)$  due to a path difference of  $\Lambda \sin\theta$ . This condition must also be satisfied by each diffracted wave, i.e., the magnitude of the  $j$ th-order reflected wave can be written as  $r_j \exp(ik_{x,j} x - ik_{1z,j} z)$ , where  $r_j$  is the reflection coefficient, and  $k_{x,j}$  is determined from the Bloch-Floquet condition:<sup>22</sup>

$$k_{x,j} = \frac{2\pi}{\lambda} \sin\theta + \frac{2\pi}{\Lambda} j = k_x + Kj \quad (9.63a)$$

This equation can be expressed in terms of the angle of reflection given by

$$\sin\theta_j = \sin\theta + \frac{j\lambda}{\Lambda} \quad (9.63b)$$

where  $\theta_j = \sin^{-1}(k_{x,j}/k)$  is the  $j$ th-order diffraction angle for reflection and Eq. (9.63b) is the well-known *grating equation*. When  $k_{x,j} > k_1$ ,  $\sin \theta_j > 1$  and the  $j$ th-order reflected wave decays exponentially toward the negative  $z$  direction. This is an evanescent wave that exists only near the surface, within a distance on the order of the wavelength. Note that the  $z$  component of  $\mathbf{k}$  for the  $j$ th-order reflected wave is

$$k_{1z,j} = \begin{cases} (k_1^2 - k_{x,j}^2)^{1/2}, & k_1 > k_{x,j} \\ i(k_{x,j}^2 - k_1^2)^{1/2}, & k_{x,j} > k_1 \end{cases} \tag{9.64}$$

Because  $k_{x,j}$  must be the same in all media, similar criteria can be applied to the transmitted waves in region III to obtain  $k_{\text{III}z,j}$  by replacing I by III in the subscripts in Eq. (9.64).

The electric field in region I is a superposition of the incident and reflected waves; therefore,

$$E_1(x,z) = \exp(ik_x x + ik_z z) + \sum_j r_j \exp(ik_{x,j} x - ik_{1z,j} z) \tag{9.65}$$

The electric field in region III can be obtained by superimposing all transmitted waves as

$$E_{\text{III}}(x,z) = \sum_j t_j \exp[ik_{x,j} x + ik_{\text{III}z,j}(z - d)] \tag{9.66}$$

where  $t_j$  is the transmission coefficient for the  $j$ th-order transmitted wave.

The electric field in region II can be expressed as

$$E_{\text{II}}(x,z) = \sum_j \Psi_j(z) \exp(ik_{x,j} x) \tag{9.67}$$

where  $\Psi_j(z)$  is the amplitude of the  $j$ th space-harmonic component. Here, the order  $j$  is matched with the diffraction order in regions I and III. Due to the periodic structure, the dielectric function of region II can be expanded in the following Fourier series:

$$\varepsilon(x) = \sum_m \varepsilon_m \exp\left(i \frac{2m\pi}{\Lambda} x\right), \quad m = 0, \pm 1, \pm 2, \dots \tag{9.68}$$

where  $\varepsilon_m$  is the  $m$ th coefficient that can be calculated from

$$\varepsilon_0 = \phi \varepsilon_A + (1 - \phi) \varepsilon_B \text{ and } \varepsilon_m = \frac{(\varepsilon_A - \varepsilon_B) \sin(m\phi\pi)}{m\pi} \quad (m \neq 0) \tag{9.69}$$

for rectangular gratings depicted in Fig. 9.18. It should be noted that each  $\varepsilon_m$  is not a physical property of the material, and its imaginary part may be negative for a passive medium.

The coupled-wave formulation comes from the wave equation of the total electric field in region II. Due to the factors that  $\varepsilon$  is independent of  $y$  and  $\mathbf{E}$  is parallel to the  $y$ -axis, we have from Eq. (9.60) that

$$\nabla^2 E_{\text{II}}(x,z) + k^2 \varepsilon(x) E_{\text{II}}(x,z) = 0 \tag{9.70}$$

A differential equation can be obtained by substituting Eq. (9.67) and Eq. (9.68) into Eq. (9.70) as

$$\begin{aligned} & \sum_j \frac{d^2 \Psi_j}{dz^2} \exp(ik_{x,j} x) - \sum_j k_{x,j}^2 \Psi_j \exp(ik_{x,j} x) \\ & + k^2 \left[ \sum_m \varepsilon_m \exp\left(i \frac{2m\pi}{\Lambda} x\right) \right] \left[ \sum_p \Psi_p \exp(ik_{x,p} x) \right] = 0 \end{aligned} \tag{9.71}$$



Equation (9.71) can be rearranged in terms of  $\exp(ik_{x,j}x)$  for the  $j$ th order as follows:

$$\sum_j \left( \frac{d^2 \Psi_j}{dz^2} - k_{x,j}^2 \Psi_j + k^2 \sum_p \epsilon_{j-p} \Psi_p \right) \exp(ik_{x,j}x) = 0 \quad (9.72)$$

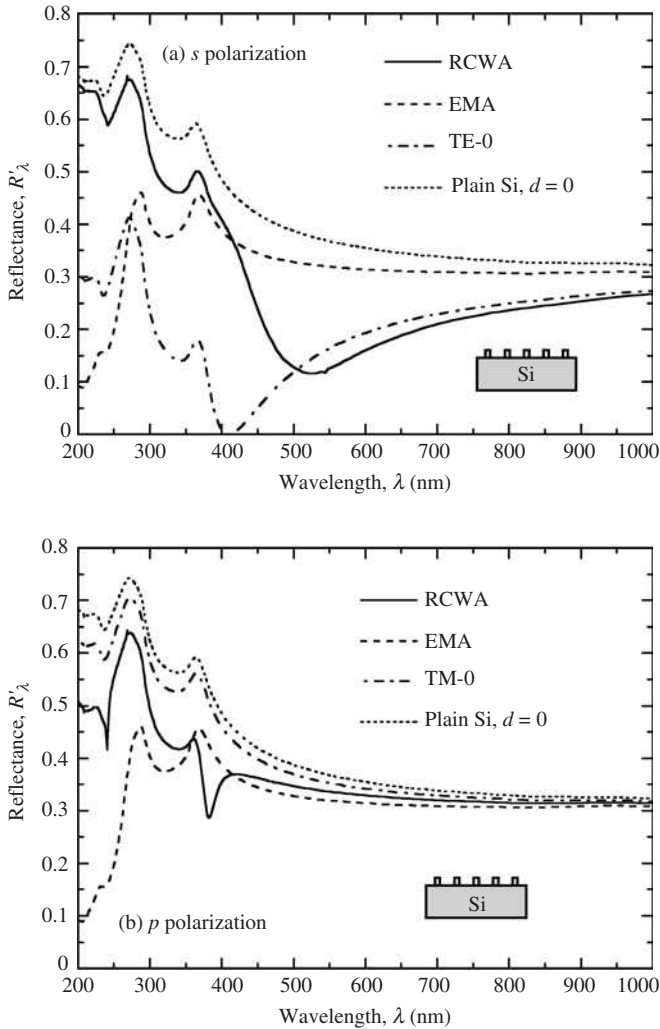
In order to satisfy this equation for any value of  $x$ , the coefficient of  $\exp(ik_{x,j}x)$  must be zero for all  $j$ 's. Hence, Eq. (9.72) is an infinite set of second-order coupled equations. Note that each space-harmonic term is coupled to other components through the harmonics of the grating. The numerical solution is obtained with sufficiently large number of diffraction orders. Suppose  $j = 0, \pm 1, \pm 2, \dots, \pm q$ , then there are  $N = 2q + 1$  diffraction orders so that  $p = 0, \pm 1, \pm 2, \dots, \pm q$  will also have  $N$  terms. Equation (9.72) can be represented by an  $N \times N$  matrix. The Fourier expansion of the dielectric function will have  $m = 0, \pm 1, \pm 2, \dots, \pm 2q$ , or  $4q + 1$ , terms. The magnetic field can be obtained from Eq. (9.67) and expressed in terms of  $\Psi_j$ . The  $N$  unknown functions  $\Psi_j$  ( $j = 0, \pm 1, \pm 2, \dots, \pm q$ ) can be expressed as summations of the eigenfunctions, which have  $2N$  unknown coefficients. Together with  $r_j$  and  $t_j$  ( $j = 0, \pm 1, \pm 2, \dots, \pm q$ ), there are  $4N$  unknowns. By matching the boundary conditions for the electric field and the tangential component of the magnetic field at the interface between regions I and II and that between regions II and III, the corresponding  $4N$  linear equations can be solved using the matrix method. An enhanced, numerically stable transmittance matrix approach was developed and applied to the implementation of RCWA for surface-relief and multilevel gratings, with detailed equations and solution procedures.<sup>22</sup> The derivation for a TM wave is more complicated because of the extra term in Eq. (9.61). Nevertheless, a corrected procedure has been proposed by Li.<sup>23</sup> Many researchers have considered the effect of azimuthal angle of incidence on the radiative properties of gratings, i.e., when the incident wavevector  $\mathbf{k}$  is not perpendicular to the grating grooves. RCWA has also been developed and applied for 2-D gratings as well as gratings of complicated geometries.

Once the reflection and transmission coefficients are obtained, it is possible to compute the fields inside and outside the grating structures, as well as to obtain the grating efficiency for each diffracted wave by calculating the time-averaged Poynting vector. The directional-hemispherical reflectance  $\rho'_\lambda$  is the summation of the reflectance of all orders. Furthermore, the directional absorptance can be calculated by  $\alpha'_\lambda = 1 - \rho'_\lambda$ , assuming region III is semi-infinite.

As an example, the reflectance at normal incidence of a silicon grating for both  $p$  and  $s$  polarizations is shown in Fig. 9.19. The grating region simulates polycrystalline silicon gates in the 65-nm devices used in high-performance complementary metal-oxide-semiconductor (CMOS) technology.<sup>24</sup> The grating period  $\Lambda = 240$  nm. The thickness of the grating (i.e., the height of the gates) is  $d = 50$  nm. The width of the gates is 30 nm, yielding a filling ratio of  $\phi = 1/6$ . The properties of the gates and the substrate are taken from Palik for single-crystal silicon at room temperature.<sup>2</sup> Comparison has been made to the reflectance of plain silicon, which is independent of the polarization, and that predicted by effective medium formulations (to be discussed later). When compared with plain silicon, the reflectance is significantly reduced by the thin grating layer, and the reduction depends strongly on the wavelength and the polarization. For a TE wave, the reduction is largest at  $\lambda = 520$  nm; whereas for a TM wave, the reduction is more significant at shorter wavelengths and grating anomalies occur at the wavelengths of 240 and 380 nm.

### 9.4.2 Effective Medium Formulations

When the grating period is much smaller than the wavelength, i.e.,  $\Lambda/\lambda < (n_{\text{III}} + \sin\theta)^{-1}$ , all the diffracted waves are evanescent waves, except the zeroth-order (specular direction) one. The reflection is similar to a smooth film with an effective uniform dielectric function.



**FIGURE 9.19** Calculated reflectance of silicon gratings. (a) TE wave (*s* polarization). (b) TM wave (*p* polarization).<sup>24</sup>

This approach is called the *method of homogenization*, and the underlying physics is based on the effective medium theory (EMT). Effective medium formulations have been used widely to describe the optical properties of inhomogeneous media. The EMT was first postulated by Garnett (*Phil. Trans. Royal Soc. London A*, **203**, 385, 1904) to obtain the effective dielectric function of metallic particles embedded in a dielectric medium. The general assumption is that the spacing separating the particles is sufficiently large or the filling ratio of the particles is small. Bruggeman (*Ann. der Physik*, **24**, 636, 1935) developed a different formulation by assuming that two materials are embedded in the effective medium and obtained an expression which has been successfully applied to study the effect of

porosity on refractive index and absorption coefficient of various materials. Bruggeman's expression is often called the effective medium approximation (EMA). The dielectric function of the effective medium  $\varepsilon_{\text{EMA}}$  is related to those of the two components by

$$\phi \frac{\varepsilon_{\text{EMA}} - \varepsilon_A}{\varepsilon_A + 2\varepsilon_{\text{EMA}}} + (1 - \phi) \frac{\varepsilon_{\text{EMA}} - \varepsilon_B}{\varepsilon_B + 2\varepsilon_{\text{EMA}}} = 0 \quad (9.73)$$

where  $\phi$  is the volume fraction (filling ratio) of material A. In 1956, Rytov (*Sov. Phys. JETP*, **2**, 466, 1965) first applied the EMT for a periodic structure by treating a stratified medium as a homogeneous uniaxial crystal and obtained the effective permittivity and permeability tensors. The zeroth order is considered to be applicable when  $\Lambda \ll \lambda$  and has been used for designing surfaces with antireflection and selective radiative properties. The expression has been extended to include higher-order terms for both 1-D and 2-D gratings. The effective medium formulation for gratings depends on the polarization. The zeroth-order expressions of the dielectric function for different polarizations are given below:

$$\varepsilon_{\text{TE},0} = \phi\varepsilon_A + (1 - \phi)\varepsilon_B, \text{ for TE waves} \quad (9.74)$$

$$\frac{1}{\varepsilon_{\text{TM},0}} = \frac{\phi}{\varepsilon_A} + \frac{1 - \phi}{\varepsilon_B}, \text{ for TM waves} \quad (9.75)$$

The results of the effective medium formulation are compared with those of the RCWA in Fig. 9.19, in which the reflectance predicted by the EMA is independent of the polarization. Both of the effective medium formulations cannot predict the radiative properties well at shorter wavelengths. The agreement between effective medium formulations and the RCWA is reasonable in the long-wavelength end, except that the EMA is worse for the TE wave. Chen et al. performed a detailed study on the effects of temperature, wavelength, polarization, and angle of incidence on the absorptance of nanoscale patterned wafers for the CMOS technology.<sup>24</sup> They also compared the configuration of combined polycrystalline silicon gates with SiO<sub>2</sub> trenches or a SiO<sub>2</sub> film. The results demonstrate nanostructures can have a significant impact on the radiative properties in unexpected ways. Hence, further research is much needed to fully understand the effect of complex nanostructures on radiative energy transfer and properties.

## 9.5 BIDIRECTIONAL REFLECTANCE DISTRIBUTION FUNCTION (BRDF)

The bidirectional reflectance, formally known as bidirectional reflectance distribution function (BRDF), is a fundamental radiative property, which describes the redistribution of energy reflected from a rough surface. Knowledge of BRDFs is essential for the analysis of radiative heat transfer between rough surfaces. Because the major heating source in rapid thermal processing is lamp radiation, knowledge of the radiative properties of materials is important for the thermal budget and temperature control during the process. A challenging problem is the accurate measurement of wafer temperature based on radiation thermometry, because it is nonintrusive and can achieve fast response. The accuracy of radiation thermometry can be affected by the emittance change and the background radiation, especially when the measured surface is rough, such as the backside of the silicon wafer. The surface roughness affects not only the emittance of the wafer but also the directional distribution of the reflected radiation by scattering. Therefore, a detailed understanding of the directional radiative properties of rough surfaces is essential to model the apparent emittance, considering the background radiation and multiple reflections.

Roughness is a measure of the topographic relief of a surface. It describes features of irregularities on the surface. Some common roughness parameters and functions include rms roughness  $\sigma_{\text{rms}}$ , *power spectral density* (PSD), autocorrelation length  $\tau_{\text{cor}}$ , and *slope distribution function* (SDF). A surface appears to be smooth if the wavelength is much greater than  $\sigma_{\text{rms}}$ . A highly polished surface can have an rms roughness on the order of nanometers. Some surfaces that look rough to human eyes may appear to be smooth for the far-infrared radiation. The reflection of radiation by rough surfaces is more complicated. For randomly rough surfaces, the scattered energy distribution or the BRDF often exhibits a peak around the direction of specular reflection, an off-specular lobe, and a diffuse component.

The BRDF of a surface can be predicted by solving the Maxwell equations if the surface roughness is fully characterized. The boundary integral method is commonly used to rigorously solve the Maxwell equations by matching the boundary conditions for the electric and magnetic fields. Since the rigorous electromagnetic wave solution generally requires a huge memory with a high-speed CPU, this approach is practically applicable to 1-D rough surfaces only, though in some cases, solutions for 2-D rough surfaces have been obtained. It is common to use approximation methods, such as the Rayleigh-Rice perturbation theory, the Kirchhoff approximation, and the geometric optics approximation. These approximations are appropriate only within certain ranges of roughness and wavelength.

The geometric illustration for the BRDF definition has been given in Chap. 8, Fig. 8.9. The Rayleigh-Rice perturbation theory can be used for relatively smooth surfaces, i.e., for surfaces with  $\sigma_{\text{rms}} \cos \theta_i / \lambda < 0.05$ , or small particles on surfaces. It is based on a statistical Fourier analysis of the surface, and predicts that the BRDF is directly proportional to the PSD and inversely proportional to the fourth power of the wavelength.<sup>25</sup> The Kirchhoff approximation is another physical-optics-based method that is often used to model the surface scattering with wave characteristics, like wave diffraction, by assuming that the radius of the surface curvature is smaller than the wavelength and there is no multiple scattering. The Kirchhoff approximation is applicable when the surface profile is slightly undulating (i.e., without sharp crests and deep valleys). The condition for this approximation to hold is that  $\sigma_{\text{rms}}$  must be relatively small compared with  $\lambda$  and  $\tau_{\text{cor}}$ . In the Kirchhoff approximation, the effects of shadowing and multiple scattering, which may be significant at large angles of incidence, are usually neglected. Most studies assumed that the roughness statistics is Gaussian.

The geometric optics approximation (GOA) neglects interference and diffraction effects and treats a rough surface as one with many small facets where an incident ray reflects specularly. Under these assumptions, the ray-tracing technique can be applied to predict the BRDF either with appropriate analytical expressions or with a Monte Carlo method. The shadowing and multiple scattering can be taken into account through a probability density function, called shadowing or masking function. Multiple scattering can be incorporated into the geometric optics formulation with the Monte Carlo method. The GOA is applicable to surfaces whose  $\sigma_{\text{rms}}$  and  $\tau_{\text{cor}}$  are greater than  $\lambda$ . There exists a good agreement between the simulation results employing the GOA and the rigorous electromagnetic wave solution. However, the simulation based on geometric optics requires much less computational resources and takes much less time than that based on the rigorous solution. In the following, the GOA-based analytical formulation and ray-tracing algorithms will be presented, and the results will be compared for anisotropic surfaces.

### 9.5.1 The Analytical Model

For the in-plane BRDF ( $\phi_r = \phi_i$  or  $\phi_r = \phi_i + 180^\circ$ ), referring to Fig. 8.9, Zhu and Zhang unified several analytical models considering first-order scattering only.<sup>26</sup> The expression of the BRDF is given in the following:

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{p(\zeta_x, \zeta_y) S(\theta_i) S(\theta_r)}{4 \cos \theta_i \cos \theta_r \cos^4 \alpha} \rho(n, \psi) \quad (9.76)$$

Here,  $p(\zeta_x, \zeta_y)$  is the 2-D SDF, and  $\zeta_x$  and  $\zeta_y$  are the slopes in  $x$  and  $y$  directions, given by

$$\zeta_x = \frac{\partial \zeta}{\partial x} = -\frac{\sin\theta_i \cos\phi_i + \sin\theta_r \cos\phi_r}{\cos\theta_i + \cos\theta_r} \quad (9.77a)$$

and

$$\zeta_y = \frac{\partial \zeta}{\partial y} = -\frac{\sin\theta_i \sin\phi_i + \sin\theta_r \sin\phi_r}{\cos\theta_i + \cos\theta_r} \quad (9.77b)$$

respectively. A shadowing function  $S$  is used in Eq. (9.76) to account for shadowing and re-striking, and is a function of the incidence or reflection zenith angles and the rms slope  $w$ , which equals  $\sqrt{2}\sigma_{\text{rms}}/\tau$  for Gaussian surfaces. Smith (*IEEE Trans. Ant. Prop.*, **15**, 668, 1967) derived a shadowing function based on Gaussian statistics. The Smith shadowing function is expressed as

$$S(\theta) = \frac{1 - 0.5\text{erfc}(\Gamma)}{1 - 0.5\text{erfc}(\Gamma) + \exp(-\Gamma^2)/(2\sqrt{\pi}\Gamma)}, 0 \leq \theta \leq 90^\circ \quad (9.78)$$

where  $\theta$  is the zenith angle of incidence (for shadowing) or reflection (for masking), and  $\Gamma = \tan(90^\circ - \theta)/(\sqrt{2}w)$ . The microfacet reflectance  $\rho(n, \psi)$ , where  $n$  is a complex refractive index and  $\psi$  is the local incidence angle, is calculated from Fresnel's reflection coefficients by averaging over the two polarizations. In the denominator of Eq. (9.76),  $\alpha$  is the inclination angle of the microfacet. While  $\alpha = (\theta_i + \theta_r)/2$  and  $\psi = |\theta_i - \theta_r|/2$  for  $\phi_r = \phi_i$ ,  $\alpha = |\theta_i - \theta_r|/2$  and  $\psi = (\theta_i + \theta_r)/2$  for  $\phi_r = \phi_i + 180^\circ$ . While the expression is simple, the GOA allows calculations for the in-plane BRDF with first-order scattering only.

## 9.5.2 The Monte Carlo Method

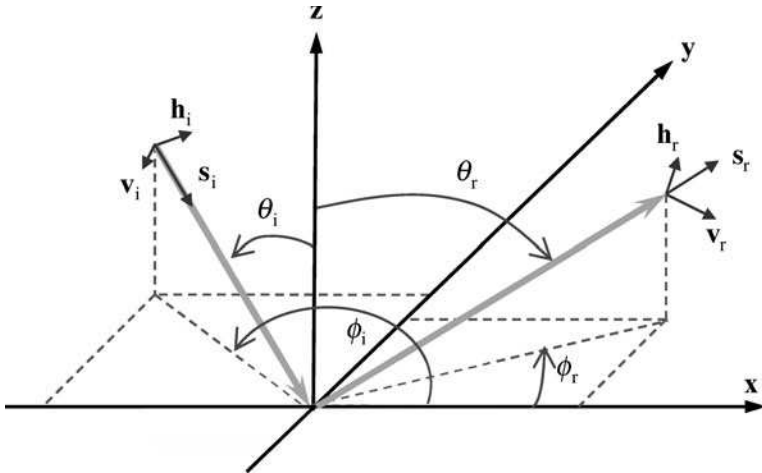
Lee et al. developed two ray-tracing techniques for modeling the BRDF in the Monte Carlo method, namely, the surface generation method (SGM) and the microfacet slope method (MSM).<sup>27</sup> The major difference lies in how to simulate the rough surfaces. The SGM is the most commonly used ray-tracing method, in which a surface realization (i.e., a numerically generated rough surface) is required prior to tracing the ray bundles. Therefore, the origin and direction of reflection is determined based on the physical location and orientation of the microfacet that the ray strikes. The BRDF is obtained from an ensemble average over a sufficiently large number of surface realizations. On the other hand, the MSM does not need to generate the entire surface *a priori*. In the MSM, ray tracing is performed by generating a normal vector of a microfacet for each ray bundle, based on the SDF and the direction of the incoming ray.<sup>28</sup> Because a surface profile does not exist in the MSM, the optical path of a propagating ray and whether the ray re-strikes the surface cannot be directly determined. Hence, the MSM relies on a shadowing function, which is the probability that a reflected ray re-strikes another surface facet, to model multiple scattering. Zhu et al. compared the two ray-tracing techniques with rigorous solutions of the electromagnetic wave equation, using the boundary integral method, for dielectric surfaces coated with a thin film.<sup>29</sup> Although the MSM is not applicable for very rough surfaces at oblique incidence, it takes less computational time and has the advantage for multiscale problems, such as light scattering from semitransparent materials, because the MSM algorithm is compatible in both micro- and macroscales.

The *spectral method* is commonly used for surface realization in the SGM by using the power spectrum. The power spectrum can be obtained from the roughness statistics. The autocorrelation function multiplied by  $\sigma_{\text{rms}}$  and the PSD are a Fourier transform pair. A rough surface, defined with the height distribution function and the autocorrelation function, are usually generated with the spectral method, regardless of whether the surface is Gaussian

or not. However, it is difficult to generate an anisotropic surface with this method. On the other hand, the surface topographic data from the AFM measurement are stored in a 2-D array of the height, which can be conveniently incorporated into the SGM algorithm without using the spectral method. The challenge is how to deal with the trade-off between the measurement area, spatial resolution, noise and artifacts in the AFM measurements, measurement time, and the number of measurements that will produce statistically meaningful results. The anisotropic SDF can be numerically evaluated as a 2-D histogram using topographic data for use in the MSM. A weight function must be included in generating microfacets because, statistically, the incident energy that is intercepted by a microfacet depends not only on the SDF but also on the projected area of the microfacet. The rejection method allows the generation of microfacets, following the weighted SDF, with uniform random numbers. The rejection method is suitable for any type of distribution function as long as a comparison function is appropriately selected. Meanwhile, the Smith shadowing function determines the probability of re-striking, in the MSM.

The polarization state may change upon reflection by a 2-D rough surface, because of the random orientation of the microfacets. When the microfacet reflectivity is calculated using Fresnel's reflection coefficients, the change of the polarization state should also be considered. In a 2-D rough surface, even though the incident radiation is purely *s* or *p* polarized, the radiation incident at the microfacet can have both polarization components in the local coordinates. Furthermore, depolarization may occur upon reflection so that the polarization of the scattered wave is different from that of the incident wave. The geometrical relations between wavevectors and polarization vectors delineate the contribution of each polarization to the reflectivity. As illustrated in Fig. 9.20, unit vectors in the direction of incidence and reflection, i.e.,  $\mathbf{s}_i$  and  $\mathbf{s}_r$ , respectively, are defined in the following:

$$\mathbf{s}_i = \begin{pmatrix} -\sin \theta_i \cos \phi_i \\ -\sin \theta_i \sin \phi_i \\ -\cos \theta_i \end{pmatrix} \quad \text{and} \quad \mathbf{s}_r = \begin{pmatrix} \sin \theta_r \cos \phi_r \\ \sin \theta_r \sin \phi_r \\ \cos \theta_r \end{pmatrix} \quad (9.79)$$



**FIGURE 9.20** Schematic of incident and scattered rays. Here, *x*, *y*, and *z* are the global coordinates, where the *x*-*y* plane is the mean plane of a rough surface.

The vectors  $\mathbf{s}_i$  and  $\hat{\mathbf{z}}$  define the plane of incidence in the global coordinates, and the vectors  $\mathbf{s}_r$  and  $\hat{\mathbf{z}}$  define the plane of reflection. A unit vector  $\mathbf{h}_i$  perpendicular and a unit vector  $\mathbf{v}_i$  parallel to the plane of incidence characterize the two polarizations of the incident wave. Here,  $\mathbf{h}_i$  indicates the electric field for  $s$  polarization while  $\mathbf{v}_i$  the electric field for  $p$  polarization. Similarly,  $\mathbf{h}_r$  and  $\mathbf{v}_r$  represent the two polarizations of the reflected wave. Hence,

$$\mathbf{h}_i = \frac{\hat{\mathbf{z}} \times \mathbf{s}_i}{|\hat{\mathbf{z}} \times \mathbf{s}_i|} = \begin{pmatrix} \sin\phi_i \\ -\cos\phi_i \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{h}_r = \frac{\hat{\mathbf{z}} \times \mathbf{s}_r}{|\hat{\mathbf{z}} \times \mathbf{s}_r|} = \begin{pmatrix} -\sin\phi_r \\ \cos\phi_r \\ 0 \end{pmatrix} \quad (9.80)$$

$$\mathbf{v}_i = \mathbf{h}_i \times \mathbf{s}_i = \begin{pmatrix} \cos\theta_i \cos\phi_i \\ \cos\theta_i \sin\phi_i \\ -\sin\theta_i \end{pmatrix} \quad \text{and} \quad \mathbf{v}_r = \mathbf{h}_r \times \mathbf{s}_r = \begin{pmatrix} \cos\theta_r \cos\phi_r \\ \cos\theta_r \sin\phi_r \\ -\sin\theta_r \end{pmatrix} \quad (9.81)$$

Calculation of the reflectivity involves two conversions of the polarization components. The  $s$ - and  $p$ -polarization components of the incident wave defined in the global coordinates are first converted to their counterparts in the local coordinates. The local polarization components are multiplied by Fresnel's reflection coefficients and then converted to the global components. Accordingly, the microfacet reflectivities for the co- and cross-polarizations can be expressed as follows:

$$\rho_{ss} = |(\mathbf{v}_r \cdot \mathbf{s}_i)(\mathbf{v}_i \cdot \mathbf{s}_r)r_s + (\mathbf{h}_r \cdot \mathbf{s}_i)(\mathbf{h}_i \cdot \mathbf{s}_r)r_p|^2 / |\mathbf{s}_i \times \mathbf{s}_r|^4 \quad (9.82a)$$

$$\rho_{sp} = |(\mathbf{h}_r \cdot \mathbf{s}_i)(\mathbf{v}_i \cdot \mathbf{s}_r)r_s - (\mathbf{v}_r \cdot \mathbf{s}_i)(\mathbf{h}_i \cdot \mathbf{s}_r)r_p|^2 / |\mathbf{s}_i \times \mathbf{s}_r|^4 \quad (9.82b)$$

$$\rho_{ps} = |(\mathbf{v}_r \cdot \mathbf{s}_i)(\mathbf{h}_i \cdot \mathbf{s}_r)r_s - (\mathbf{h}_r \cdot \mathbf{s}_i)(\mathbf{v}_i \cdot \mathbf{s}_r)r_p|^2 / |\mathbf{s}_i \times \mathbf{s}_r|^4 \quad (9.82c)$$

$$\rho_{pp} = |(\mathbf{h}_r \cdot \mathbf{s}_i)(\mathbf{h}_i \cdot \mathbf{s}_r)r_s + (\mathbf{v}_r \cdot \mathbf{s}_i)(\mathbf{v}_i \cdot \mathbf{s}_r)r_p|^2 / |\mathbf{s}_i \times \mathbf{s}_r|^4 \quad (9.82d)$$

where  $r$  denotes Fresnel's reflection coefficient. The subscripts  $s$  and  $p$  stand for each polarization. On the left-hand side, the double subscripts indicate the polarization for the incidence and the reflection, respectively.

In terms of the microfacet reflectivities, the reflected energies  $G_{r,s}$  and  $G_{r,p}$  are related to the incident energies  $G_{i,s}$  and  $G_{i,p}$  by

$$\begin{bmatrix} G_{r,s} \\ G_{r,p} \end{bmatrix} = \begin{bmatrix} \rho_{ss} & \rho_{ps} \\ \rho_{sp} & \rho_{pp} \end{bmatrix} \begin{bmatrix} G_{i,s} \\ G_{i,p} \end{bmatrix} \quad (9.83)$$

The reflectivity is defined as the ratio of the reflected energy  $G_r = G_{r,s} + G_{r,p}$  to the incident energy  $G_i = G_{i,s} + G_{i,p}$ ; thus, it depends on the polarization state of the incident wave. To facilitate the calculation, the incident energy of each ray bundle is set to unity such that  $(G_{i,s}, G_{i,p}) = (1, 0)$  for  $s$  polarization,  $(G_{i,s}, G_{i,p}) = (0, 1)$  for  $p$  polarization, and  $(G_{i,s}, G_{i,p}) = (0.5, 0.5)$  for random polarization (i.e., unpolarized incidence). For the first reflection,  $G_{r,s}$  and  $G_{r,p}$  are calculated from Eq. (9.83). For multiple reflections, the previously reflected energies are substituted for  $G_{i,s}$  and  $G_{i,p}$ , and the next reflected energy is updated according to Eq. (9.83). Each ray bundle is traced until it leaves the surface, and then, the information of its direction and energy for each polarization is stored in a database. Because the energy of the bundle is reduced after each reflection, there is no need to use random numbers to decide whether a ray bundle is reflected at the microfacet or not.

In a special case, when the planes of incidence and reflection are identical, the polarization state is maintained for either  $s$  or  $p$  polarization if only the first-order scattering has

been considered. This means that the vectors  $\mathbf{h}_i$  and  $\mathbf{h}_r$  are either parallel or antiparallel (refer to Fig. 9.20); consequently,  $\mathbf{h}_i \cdot \mathbf{s}_r = 0$  and  $\mathbf{h}_r \cdot \mathbf{s}_i = 0$ . It can be seen from Eq. (9.82) that  $\rho_{sp} = \rho_{ps} = 0$ ,  $\rho_{ss} = |r_s|^2$ , and  $\rho_{pp} = |r_p|^2$ . The corresponding BRDF is called the in-plane BRDF ( $\phi_r = \phi_i$  or  $\phi_r = \phi_i + 180^\circ$ ). Nevertheless, the cross-polarization term is nonzero for the in-plane BRDF when multiple scattering is significant. After a large number of ray bundles have been traced, the BRDF can be calculated in terms of the energy of the ray bundles:

$$f_r(\lambda, \theta_i, \phi_i, \theta_r, \phi_r) = \frac{1}{G_i(\theta_i, \phi_i)} \frac{\Delta G_r(\theta_r, \phi_r)}{\cos \theta_r \Delta \Omega_r} \quad (9.84)$$

where  $G_i(\theta_i, \phi_i)$  is the total energy of the incident ray bundles, and  $\Delta G_r(\theta_r, \phi_r)$  is the energy of the ray bundles leaving the surface within the solid angle  $\Delta \Omega_r$  in the direction  $(\theta_r, \phi_r)$ . The integration of the BRDF yields the directional-hemispherical reflectance. The directional emittance can be obtained according to the conservation of energy and Kirchhoff's law.

### 9.5.3 Surface Characterization

In most studies, surface roughness is assumed to satisfy Gaussian statistics in the derivation of the BRDF model and for the surface generation in the Monte Carlo simulation. Furthermore, the roughness statistics of 2-D rough surfaces is assumed to be isotropic in most publications so that the autocorrelation function is independent of the direction. However, the Gaussian distribution may miss important features of natural and man-made rough surfaces that are strongly anisotropic. Before the invention of the AFM, the surface profile was usually measured with a mechanical profiler that scans the surface line-by-line. Some mechanical stylus profilers can measure rough surfaces with a vertical resolution of a few nanometers. However, the lateral resolution is usually on the order of  $1 \mu\text{m}$  due to the large radius of the stylus probe. Because the radius of curvature of the probe tip is in the range from 5 to 50 nm, an AFM can provide detailed information on the topography of a small area on the microrough surfaces, with a vertical resolution of subnanometers and a lateral resolution around 10 nm. The result is stored in an array, containing the height information,  $z(m, n)$ , where  $m = 1, 2, \dots, M$  and  $n = 1, 2, \dots, N$  are the points along the  $x$  and  $y$  directions, respectively.

To evaluate the 2-D slope distribution  $p(\zeta_x, \zeta_y)$ , each surface element is determined by the four closest nodes in the data array. The four-node element can be considered as two triangular surfaces with a common side. The surface normals for the two triangles can be averaged to give the mean slope of the surface element such that

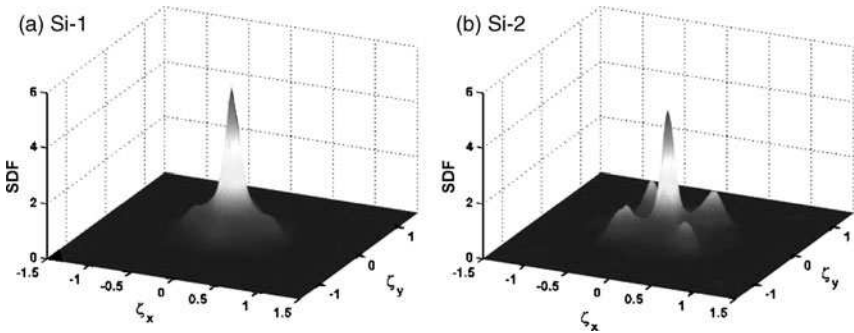
$$\zeta_x = \frac{z_{m+1,n} - z_{m,n}}{2l} + \frac{z_{m+1,n+1} - z_{m,n+1}}{2l} \quad (9.85a)$$

$$\zeta_y = \frac{z_{m,n+1} - z_{m,n}}{2l} + \frac{z_{m+1,n+1} - z_{m+1,n}}{2l} \quad (9.85b)$$

where  $l$  is the lateral distance between adjacent data points.<sup>26</sup> The SDF can be determined by evaluating the slopes of all measured surface elements. For a scan area of  $100 \times 100 \mu\text{m}^2$ , the lateral interval  $l \approx 0.2 \mu\text{m}$ , when the data are stored in a  $512 \times 512$  array.

The 2-D SDFs from the AFM measurement in the tapping mode, for two lightly doped  $<100>$  single-crystal silicon surfaces, are shown in Fig. 9.21.<sup>27</sup> In the contact mode, lateral or shear forces can distort surface features and reduce the spatial resolution. Thus, deep valleys may not be correctly measured. The AFM scanning performed in the tapping mode with sharper silicon tips allows measuring precipitous slopes. The two SDFs are non-Gaussian and anisotropic, although the anisotropy of Si-1 is not as striking as that of Si-2. The SDF of Si-1





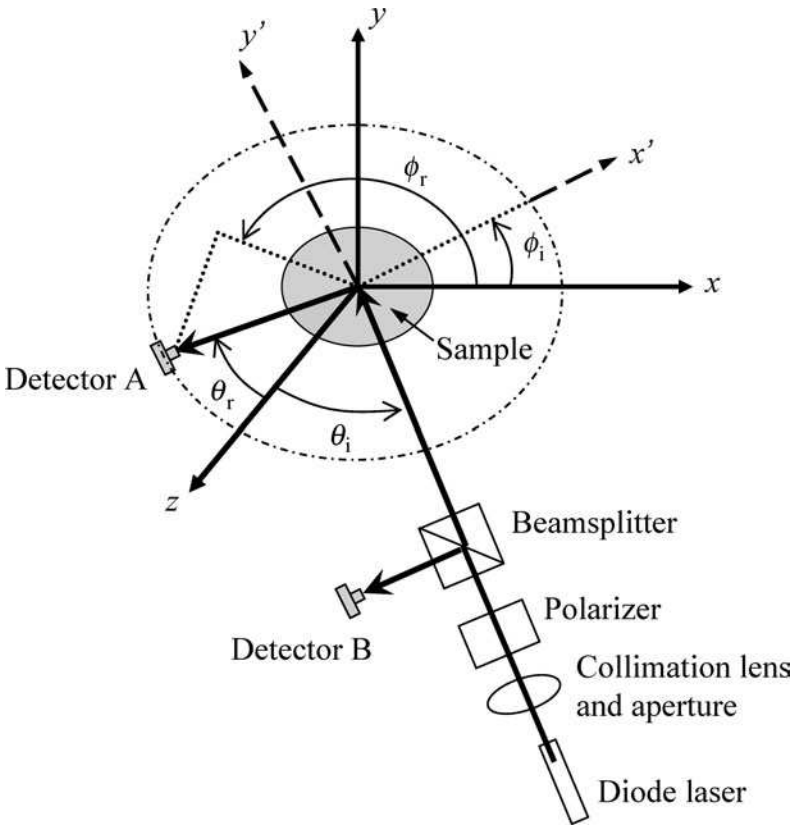
**FIGURE 9.21** 2-D slope distribution obtained from AFM topographic measurements for two samples: (a) Si-1; (b) Si-2.

contains only one dominant peak at the center, indicating that a large number of microfacets are only slightly tilted. The SDF of Si-2 also has a dominant peak at the center, though smaller than that of Si-1. Four side peaks can also be seen that are nearly symmetric. These side peaks are associated with the formation of  $\{311\}$  planes, during the chemical etching in the (100) crystalline wafer.<sup>26,27</sup> The angle between the (100) plane and any of the four (311) planes is  $\cos^{-1}(3/\sqrt{11}) = 25.2^\circ$ , which is close to the location of the observed side peaks.

#### 9.5.4 BRDF Measurements

The BRDF of silicon wafers was measured with a laser scatterometer, named as three-axis automated scatterometer (TAAS), shown schematically in Fig. 9.22.<sup>30</sup> The sample is vertically mounted. Three rotary stages, automatically controlled by a computer, are used to change incidence and reflection directions. One rotates the sample around the  $y$ -axis to change the incidence angle  $\theta_i$ , another rotates detector A in the  $x$ - $z$  (horizontal) plane to change the reflection angle  $\theta_r$ , and the third rotates the arm of detector A out of the  $x$ - $z$  plane to change the azimuthal angle  $\phi_r$  for out-of-plane measurements. Manual rotation of the sample on a sample holder around the  $z$ -axis adjusts the azimuthal angle  $\phi_i$ . The incident laser beam is parallel to the optical table ( $x$ - $z$  plane). A diode laser system serves as an optical source, and a lock-in amplifier, connected with a diode laser controller, modulates the output optical power at 400 Hz. The wavelength can be selected by replacing the fiber-coupled diode laser, and a number of diode lasers in the visible and the near-infrared are available. The diode laser is mounted on a thermoelectrically controlled stage to provide power stability within a standard deviation of 0.2%. An optical fiber is used to provide flexibility for optical access and alignment. The light from the output end of the fiber is in the horizontal plane.

As shown in Fig. 9.22, the beam first passes through a collimator with a pair of lenses and a small aperture. A linear polarizer mounted on a dial allows the selection of polarization for light incident on the sample. The beamsplitter then divides the laser beam into two passes: one goes to the sample and the other to a stationary reference detector B. The light scattered by the sample is measured by detector A. The beam spot size on the sample is a few millimeters in diameter, and the measurement can be considered as a spatial average over the beam diameter. Si and Ge photodiode detectors measure the radiant power in the wavelength range from 350 to 1100 nm and from 800 to 1800 nm, respectively. The power collected at each detector is sent to a trans-impedance preamplifier that has nine decades of amplification range. The preamplifier has a linear frequency response from dc (zero frequency) up to a certain maximum frequency that is much greater than 400 Hz. The lock-in



**FIGURE 9.22** Schematic of the three-axis automated scatterometer (TAAS) for BRDF measurements.

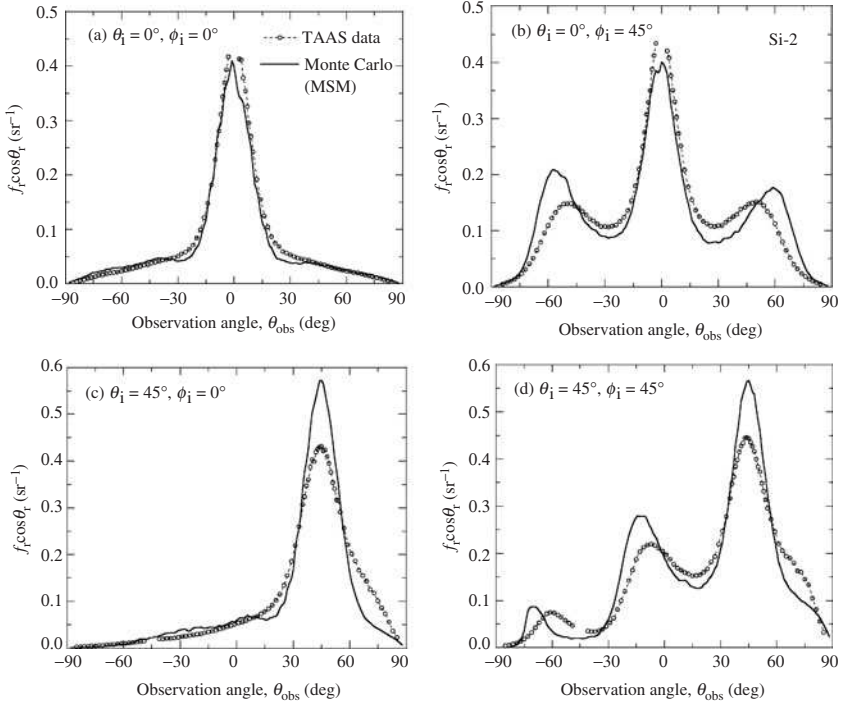
amplifier picks up only the phase-locked signals at 400 Hz, thereby eliminating the effect of background radiation or stray light without using a chopper. The measurement equation for the BRDF is given by

$$f_r(\theta_i, \phi_i, \theta_r, \phi_r) = C_1 \frac{V_A}{V_B \cos \theta_r \Delta \Omega_r} \quad (9.86)$$

where  $V_A$  and  $V_B$  are the outputs of detectors A and B, respectively, and  $\Delta \Omega_r$  is the reflection solid angle, which is  $1.84 \times 10^{-4}$  sr, as determined by the area of a precision-machined aperture in front of the detector and the distance between this aperture and the beam spot on the sample. An instrument constant  $C_1$  compensates the beamsplitter ratio and the difference in the responsivities of the two detectors. The BRDF within  $\pm 2.5^\circ$  of the retro-reflection direction ( $\theta_r = \theta_i$  and  $\phi_r = \phi_i$ ) cannot be measured since the movable detector blocks the incident beam. A PC performs the data acquisition and automatic rotary-stage control in a LabView environment. In the measurements,  $V_A$  and  $V_B$  are averaged over many measurements at a given position to reduce the random error. The relative uncertainty of the TAAS is estimated to be 5% for  $f_r > 0.1$  through intercomparison with a reference standard instrument at NIST.<sup>30</sup>

### 9.5.5 Comparison of Modeling with Measurements

Figure 9.23 compares the predicted BRDFs based on the slope distribution with the BRDFs measured using TAAS at  $\lambda = 635$  nm, for Si-2, which is strongly anisotropic.<sup>27</sup> For clarity,

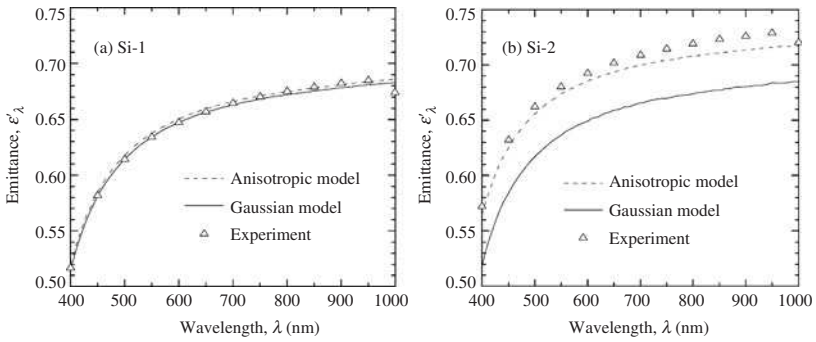


**FIGURE 9.23** Comparison of Monte Carlo model based on the MSM and the measured in-plane BRDF for Si-2. The observation angle  $\theta_{\text{obs}}$  is the same as the reflection polar angle when  $\phi_r = \phi_i + \pi$  and negative reflection polar angle when  $\phi_r = \phi_i$ .<sup>27</sup>

only the prediction using the MSM is presented. The predictions with the SGM and the analytical model yield a similar agreement with experiments.<sup>26,27</sup> As can be seen from Fig. 9.23a, the prediction and the measurement agree well, except near  $\theta_{\text{obs}} = 0^\circ$ , where the measurements can not be taken within  $\pm 2.5^\circ$  and the simulation has a large fluctuation. The simulation captures the general features and trends of the measured BRDF, while some discrepancies exist near the side peaks. For  $\theta_i = 0^\circ$  and  $\phi_i = 45^\circ$ , as shown in Fig. 9.23b, the BRDF contains two large side peaks associated with the side peaks in the SDF for Si-2 at  $|\zeta_x| \approx |\zeta_y| \approx 0.38$  in Fig. 9.21b. The Monte Carlo simulations also predict the side peaks located approximately at  $\theta_r = 57^\circ$ , which deviates somewhat from the measured value of  $50^\circ$ . Based on Snell's law, the inclination angle of microfacets is half of  $\theta_r$ , at  $\theta_i = 0^\circ$ . Therefore, the measured side peaks in the BRDF correspond to an inclination angle  $25^\circ$ , which is very close to the angle of  $25.2^\circ$  between any of the four  $\{311\}$  planes and the  $(100)$  plane. On the other hand, the predicted side peaks correspond to an inclination angle of  $28.5^\circ$ , which is almost the same as that calculated from the slope at  $|\zeta_x| = |\zeta_y| = 0.38$ . Consequently, the side peak position obtained from the BRDF measurement is more reliable than that predicted

by the Monte Carlo methods using the topographic data from the AFM measurement. Due to the artifacts in the AFM measurements, the BRDF values are underpredicted when  $15^\circ < \theta_r < 50^\circ$  and overpredicted when  $50^\circ < \theta_r < 80^\circ$ . When,  $\theta_i = 45^\circ$  the Monte Carlo method overpredicts the specular peak, presumably due to the limitation of geometric optics. The disagreement between the predicted and measured BRDFs, for  $60^\circ < \theta_{\text{obs}} < 85^\circ$ , may be due to the combined result of the artifacts in the AFM measurement, the limitation of the GOA, and multiple scattering. For  $\theta_i = 45^\circ$  and  $\phi_i = 45^\circ$ , a small side peak appears at  $\theta_{\text{obs}} = -60^\circ$  in the measured curve and at  $\theta_{\text{obs}} = -71^\circ$  in the predicted curve. This is believed to be due to microfacets with  $\{111\}$  orientation that have an inclination angle of  $54.7^\circ$ . The small side peak should occur around  $\theta_{\text{obs}} = -64.4^\circ$  based on simple geometric arguments.

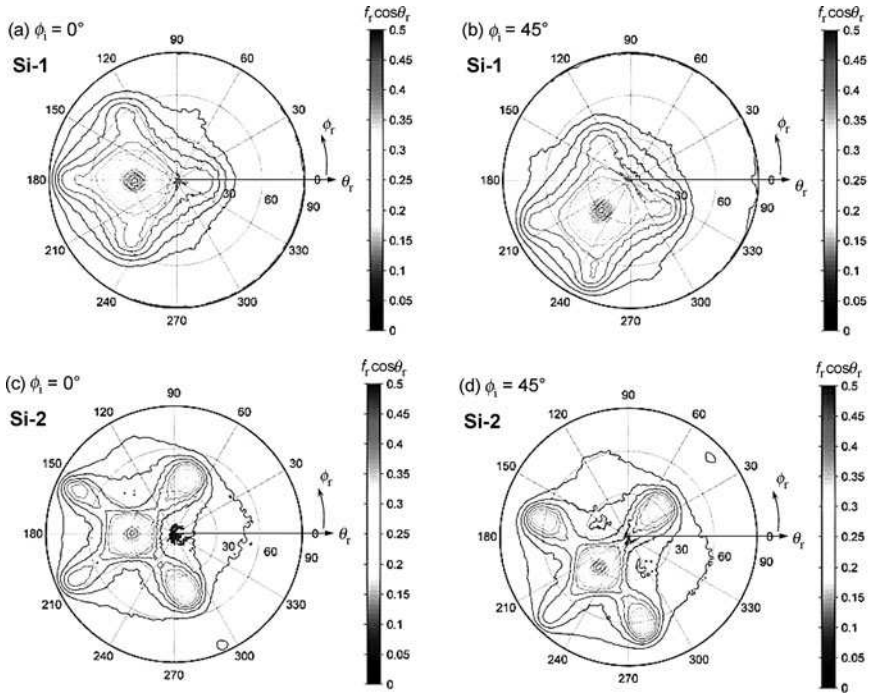
Figure 9.24 shows the directional-spectral emittance measured using an integrating sphere coupled with a monochromator.<sup>31</sup> The directional emittance was calculated from the



**FIGURE 9.24** Comparison of the predicted and measured emittance of Si-1 and Si-2, in a polar angle approximately equal to  $7^\circ$ .<sup>31</sup>

measured directional-hemispherical reflectance at an incidence angle of approximately  $7^\circ$ . The emittance values calculated from the models based on Gaussian distribution and anisotropic slope distribution are compared with those obtained from experiments. For Si-1, which is nearly isotropic, the difference between the models is small and the agreement with the experiment is excellent. The combined uncertainty in the measurement is estimated to be 0.01, except at  $\lambda = 1000$  nm, where the silicon wafer becomes slightly transparent. For Si-2, however, the Gaussian model underpredicts the emittance and there is a large enhancement of the emittance due to anisotropy. The Monte Carlo model, based on the MSM, significantly improves the prediction. Given the fact that the AFM surface topographic measurements may not perfectly match the actual surface slope distribution, an uncertainty of 0.01 has been estimated for the Monte Carlo model. It can be seen that the prediction agrees with the measurement better at short wavelengths, where geometric optics is more suitable.

The out-of-plane BRDFs of Si-1 and Si-2, calculated with the MSM at  $\theta_i = 30^\circ$ , are presented in Fig. 9.25 as contour plots in a polar coordinates system.<sup>27</sup> In these plots, the radial and azimuthal coordinates respectively correspond to  $\theta_r$  and  $\phi_r$ , and the  $z$ -axis represents  $f_r \cos \theta_r$ . The BRDFs depend little on  $\phi_r$  around the specular direction, but the dependence becomes large as the angular separation from the specular peak increases. The region where the BRDF is independent of  $\phi_r$  is broader for Si-1 than for Si-2. The predicted BRDFs for Si-2 display a strong specular reflection peak, together with the four large side peaks associated with  $\{311\}$  planes. In addition, a small side peak associated with a  $\{111\}$  plane appears at large  $\theta_r$ , as illustrated in Fig. 9.25c at  $\phi_r = 294^\circ$  and another in Fig. 9.25d at  $\phi_r = 45^\circ$ . The



**FIGURE 9.25** BRDF predicted by the MSM at  $\theta_i = 30^\circ$  for random polarization.<sup>27</sup> (a) Si-1 at  $\phi_i = 0^\circ$ . (b) Si-1 at  $\phi_i = 45^\circ$ . (c) Si-2 at  $\phi_i = 0^\circ$ . (d) Si-2 at  $\phi_i = 45^\circ$ . In the polar contour plots, the radial coordinate corresponds to  $\theta_r$ , and the azimuthal coordinate corresponds to  $\phi_r$ .

actual magnitudes of the small side peaks may be smaller than those predicted by the MSM, and their positions may shift toward smaller  $\theta_r$ . Nevertheless, Fig. 9.25 indicates that the Monte Carlo method is an effective technique for studying the BRDFs for anisotropic surfaces.

## 9.6 SUMMARY

This chapter provided a detailed treatment of the radiative properties of stratified media based on the electromagnetic wave theory, considering partial coherence, and extended to the discussion of periodic structures, i.e., photonic crystals. A discussion of the coupled-wave analysis was also present for periodic gratings. Moreover, a survey was given to some recent research on the bidirectional reflectance of anisotropic surfaces.

An important area that was not covered is light scattering by small particles and colloids for which there have been tremendous interests and extensive studies. The heat transfer community is very familiar with light scattering and radiative transfer in scattering and absorbing media. Recently, more and more studies on light scattering have employed rigorous treatments of the electromagnetic wave scattering in random media, considering polarization and surface plasmon resonance. Further discussions on evanescent waves, surface waves including surface plasmon and phonon polaritons, and near-field energy transfer by electromagnetic waves will be given in the next chapter.

**REFERENCES**

1. R. Siegel and J. R. Howell, *Thermal Radiation Heat Transfer*, 4th ed., Taylor & Francis, New York, 2002.
2. E. D. Palik (ed.), *Handbook of the Optical Constants of Solids*, Vols. I, II, and III, Academic Press, San Diego, CA, 1998.
3. O. S. Heavens, *Optical Properties of Thin Solid Films*, Dover Publications, New York, 1965.
4. Z. Knittl, *Optics of Thin Films*, Wiley, New York, 1976.
5. M. Q. Brewster, *Thermal Radiative Transfer and Properties*, Wiley, New York, 1992.
6. Z. M. Zhang, "Reexamination of the transmittance formulae of a lamina," *J. Heat Transfer*, **119**, 645–647, 1997; Z. M. Zhang, "Optical properties of a slightly absorbing film for oblique incidence," *Appl. Opt.*, **38**, 205–207, 1999.
7. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge University Press, Cambridge, UK, 1995.
8. G. Chen and C. L. Tien, "Partial coherence theory of thin film radiative properties," *J. Heat Transfer*, **114**, 636–643, 1992.
9. K. Fu, P.-f., Hsu, and Z. M. Zhang, "Unified analytical formulation of thin-film radiative properties including partial coherence," *Appl. Opt.*, **45**, 653–661, 2006.
10. B. J. Lee, V. P. Khuu, and Z. M. Zhang, "Partially coherent spectral radiative properties of dielectric thin films with rough surfaces," *J. Thermophys. Heat Transfer*, **19**, 360–366, 2005.
11. J. M. Vaughan, *The Fabry-Perot Interferometer: History, Theory, Practice and Applications*, Adam Hilger, Bristol, PA, 1989.
12. A. R. Kumar, V. A. Boychev, Z. M. Zhang, and D. B. Tanner, "Fabry-Perot resonators built with  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$  films on Si substrates," *J. Heat Transfer*, **122**, 785–791, 2000.
13. P. Yeh, *Optical Waves in Layered Media*, Wiley, New York, 1988; P. Yeh, A. Yariv, and C. S. Hong, "Electromagnetic propagation in periodic stratified media. I. General theory," *J. Opt. Soc. Am.*, **67**, 423–438, 1977.
14. C. L. Mitsas and D. I. Siapkas, "Generalized matrix method for analysis of coherence and incoherent reflectance and transmittance of multilayer structures with rough surfaces, interfaces, and finite substrates," *Appl. Opt.*, **34**, 1678–1683, 1995.
15. B. J. Lee and Z. M. Zhang, "Rad-Pro: effective software for modeling radiative properties in rapid thermal processing," in *Proc. 13th IEEE Annu. Int. Conf. Adv. Thermal Processing of Semiconductors (RTP'2005)*, pp. 275–281, Santa Barbara, CA, October 4–7, 2005.
16. J. D. Joannopoulos, R. D. Meade, and J. N. Winn, *Photonic Crystals*, Princeton University Press, Princeton, NJ, 1995.
17. K. Sakoda, *Optical Properties of Photonic Crystals*, Springer-Verlag, Berlin, 2001.
18. J. G. Fleming, S. Y. Lin, I. El-Kady, R. Biswas, and K. M. Ho, "All-metallic three-dimensional photonic crystals with a large infrared bandgap," *Nature*, **417**, 52–55, 2002; C. H. Seager, M. B. Sinclair, and J. G. Fleming, "Accurate measurements of thermal radiation from a tungsten photonic lattice," *Appl. Phys. Lett.*, **86**, 244105, 2005.
19. H. A. Macleod, *Thin Film Optical Filters*, 3rd ed., Institute of Physics, Bristol, UK, 2001.
20. D. Maystre (ed.), *Selected Papers on Diffraction Gratings*, SPIE Milestone Series 83, The International Society for Optical Engineering, Bellingham, WA, 1993.
21. R. Petit (ed.), *Electromagnetic Theory of Gratings*, Springer, Berlin, 1980.
22. M. G. Moharam, E. B. Grann, D. A. Pommet, and T. K. Gaylord, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," *J. Opt. Soc. Am. A*, **12**, 1068–1076, 1995; M. G. Moharam, D. A. Pommet, E. B. Grann, and T. K. Gaylord, "Stable implementation of the rigorous coupled-wave analysis for surface-relief gratings: Enhanced transmittance matrix approach," *J. Opt. Soc. Am. A*, **12**, 1077–1086, 1995.
23. L. F. Li, "Use of Fourier series in the analysis of discontinuous periodic structures," *J. Opt. Soc. Am. A*, **13**, 1870–1876, 1996.
24. Y. B. Chen, Z. M. Zhang, and P. J. Timans, "Radiative properties of patterned wafers with nanoscale linewidth," *J. Heat Transfer*, **129**, 79–90, 2007.

25. P. Beckmann and A. Spizzichino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Artech House, Norwood, MA, 1987.
26. Q. Z. Zhu and Z. M. Zhang, "Anisotropic slope distribution and bidirectional reflectance of a rough silicon surface," *J. Heat Transfer*, **126**, 985–993, 2004; Q. Z. Zhu and Z. M. Zhang, "Correlation of angle-resolved light scattering with the microfacet orientation of rough silicon surfaces," *Opt. Eng.*, **44**, 073601, 2005.
27. H. J. Lee, Y. B. Chen, and Z. M. Zhang, "Directional radiative properties of anisotropic rough silicon and gold surfaces," *Int. J. Heat Mass Transfer*, **49**, 4482–4495, 2006.
28. Y. H. Zhou and Z. M. Zhang, "Radiative properties of semitransparent silicon wafers with rough surfaces," *J. Heat Transfer*, **125**, 462–470, 2003; H. J. Lee, B. J. Lee, and Z. M. Zhang, "Modeling the radiative properties of semitransparent wafers with rough surfaces and thin-film coatings," *J. Quant. Spectros. Radiat. Transfer*, **93**, 185–194, 2005.
29. Q. Z. Zhu, H. J. Lee, and Z. M. Zhang, "Validity of hybrid models for the bidirectional reflectance of coated rough surfaces," *J. Thermophys. Heat Transfer*, **19**, 548–557, 2005.
30. Y. J. Shen, Q. Z. Zhu, and Z. M. Zhang, "A scatterometer for measuring the bidirectional reflectance and transmittance of semiconductor wafers with rough surfaces," *Rev. Sci. Instrum.*, **74**, 4885–4892, 2003.
31. H. J. Lee, A. C. Bryson, and Z. M. Zhang, "Measurement and modeling of the emittance of silicon wafers with anisotropic roughness," *Proc. 16th Symp. Thermophys. Properties*, Boulder, CO, July 30–August 4, 2006.

## PROBLEMS

---

**9.1.** A greenhouse looks like a small glass house used to grow plants in the winter. Based on the transmittance curve of fused silica ( $\text{SiO}_2$ ), shown in Fig. 9.2, explain why glass walls can keep the plants warm in the winter. Discuss the greenhouse effect in the atmosphere. What gases are responsible for the greenhouse effect?

**9.2.** Calculate the transmittance  $T$ , the reflectance  $R$ , and the absorptance  $A$  of a thick (without considering interference) silicon wafer (0.5 mm thick) at normal incidence. Plot  $T$ ,  $R$ , and  $A$  versus wavelength, in the range from 2.5 to 25  $\mu\text{m}$ . The refractive index and the extinction coefficient of the doped silicon are given in the following table:

Optical Constants of a Doped Silicon Wafer

Wavelength $\lambda$ ( $\mu\text{m}$ )	Refractive index $n$	Extinction coefficient $\kappa$
2.5	3.44	0
5.0	3.43	$1.0 \times 10^{-7}$
7.5	3.42	$8.4 \times 10^{-5}$
10.0	3.42	$2.1 \times 10^{-4}$
12.5	3.42	$4.0 \times 10^{-4}$
15.0	3.42	$5.0 \times 10^{-4}$
17.5	3.42	$9.0 \times 10^{-4}$
20.0	3.42	$1.0 \times 10^{-3}$
22.5	3.42	$1.1 \times 10^{-3}$
25.0	3.42	$1.3 \times 10^{-3}$

**9.3.** Calculate and plot the transmittance and reflectance for the same silicon wafer described in Problem 9.2 at  $\lambda = 5 \mu\text{m}$  as functions of the polar angle  $\theta$ . Consider the individual polarizations and their average. Compare your results with those by Zhang et al. (*Infrared Phys. Technol.*, **37**, 539, 1996).

**9.4.** Using data from the table in Problem 9.2, calculate and plot the normal transmittance of a 100- $\mu\text{m}$ -thick silicon wafer, near 10- $\mu\text{m}$  wavelength, considering interference.

- (a) Plot the transmittance in terms of wavelength ( $\mu\text{m}$ ) with an interval between the data spacing of 0.05 and 0.005  $\mu\text{m}$ , respectively, on one graph.
- (b) Plot the transmittance in terms of wavenumber ( $\text{cm}^{-1}$ ) with an interval between the data spacing of 5 and 0.5  $\text{cm}^{-1}$ , respectively, on one graph.
- (c) What is the fringe-averaged transmittance at 10- $\mu\text{m}$  wavelength?
- (d) What is the free spectral range in wavenumber and in wavelength? How will  $\Delta\bar{\nu}$  and  $\Delta\lambda$  change if the wavelength  $\lambda$  is changed to 20  $\mu\text{m}$ ?

**9.5.** For gold, the refractive index at  $\lambda = 0.5 \mu\text{m}$  is  $n = 0.916 + i1.84$ , and at  $\lambda = 2.0 \mu\text{m}$  is  $n = 0.85 + i12.6$ . Calculate the transmittance of a free-standing gold film at these wavelengths for  $d = 10, 20, 50,$  and  $100 \text{ nm}$ , using both Eq. (9.10) and Eq. (9.11). Which equation gives the correct results, and why?

**9.6.** For the three-layer structure shown in Fig. 9.3, calculate the normal reflectance for  $n_1 = 1.45$  (glass),  $n_2 = 1$  (air gap), and  $n_3 = 2$  (substrate) without any absorption at  $\lambda = 1 \mu\text{m}$ . Plot the reflectance as a function of the air-gap width  $d$ . Obtain the analytical formulae of the reflectance maximum and minimum.

**9.7.** Assume that glass has a refractive index of 1.46 without any absorption in the visible spectrum ( $0.4 \mu\text{m} < \lambda < 0.7 \mu\text{m}$ ). Design an antireflection coating (for normal incidence) that will minimize the reflectance from a semi-infinite glass. You need to determine the coating thickness and the refractive index (assuming it is independent of wavelength). Plot the normal reflectance of the coated glass surface in the spectral range from 0.4 to 0.7  $\mu\text{m}$ . What material would you recommend for use with the desired property?

**9.8.** To evaluate the effect of antireflection coating for oblique incidence, assume the antireflection coating has a refractive index of 1.21 and a thickness of 114 nm. What will be the reflectance, at 45° and 60°, for each polarization?

**9.9.** While the extinction coefficient is often related to absorption or loss, it should be noted that when  $\kappa \gg n$ , it is the real part of the refractive index that is related to the loss. This is because the dielectric function can be expressed as  $\epsilon = \epsilon' + i\epsilon'' = (n^2 - \kappa^2) + i2n\kappa$ , where  $\epsilon''$  is related to the dissipation. For a semi-infinite medium, a purely negative dielectric function means perfect reflection. The effect of  $n$  on the absorption by a thin film can be studied by considering a thin film of thickness  $d$  with a complex refractive index  $n_2 = n + i\kappa$ . For a wavelength of  $\lambda = 0.5 \mu\text{m}$  and at normal incidence, let  $d = 30 \text{ nm}$  and  $\kappa = 3.0$ . Plot the transmittance, the reflectance, and the emittance (which is the same as the absorbance), against the refractive index  $n$  ranging from 0.01 to 2. Discuss the effect of  $n$  on the absorption.

**9.10.** Use the dielectric function of SiC given in Example 8-7 to calculate the normal emittance for a SiC film at wavelengths from 9 to 15  $\mu\text{m}$ , for different film thicknesses:  $d = 1, 10, 100,$  and  $1000 \mu\text{m}$ . Assume the multiply reflected waves to be perfectly coherent.

**9.11.** Calculate the emittance as a function of the emission angle for a doped silicon wafer of 200- $\mu\text{m}$  thickness, at  $\lambda = 20 \mu\text{m}$  with  $n_2 = 3.42 + i0.001$ . Consider  $p$  and  $s$  polarizations separately, and then, take an average. Assume the multiply reflected waves to be perfectly coherent.

**9.12.** This problem concerns the transmission and reflection of infrared radiation of a YBCO ( $\text{YBa}_2\text{Cu}_3\text{O}_7$ ) film on a thin MgO substrate of 325- $\mu\text{m}$  thickness, at 300 K and normal incidence. For the YBCO film, use the properties for sample A from Kumar et al. (*J. Heat Transfer*, **121**, 844, 1999). For MgO, use the Lorentz model in Problem 8.26.

- (a) Plot the radiation penetration depth of the YBCO film,  $\delta_f(\lambda)$ , and that of MgO,  $\delta_s(\lambda)$ , for  $1 \mu\text{m} < \lambda < 1000 \mu\text{m}$ .
- (b) Neglecting the interference effect in the MgO substrate, calculate and plot the transmittance  $T$ , the film-side reflectance  $R_f$ , and the back-side reflectance  $R_s$ , for  $1 \mu\text{m} < \lambda < 1000 \mu\text{m}$ , with different film thicknesses: 0, 30, 48, 70, and 400 nm. Plot  $T$ ,  $R_f$ , and  $R_s$  in terms of both wavelength ( $\mu\text{m}$ ) and wavenumber ( $\text{cm}^{-1}$ ).
- (c) Repeat the previous calculation, considering the interference effects in the MgO substrate, for  $200 \mu\text{m} < \lambda < 1000 \mu\text{m}$  (50 to 10  $\text{cm}^{-1}$ ). Plot in terms of the wavenumber only. What happens with the interference fringes when the film thickness is 48 nm?

**9.13.** Calculate the normal transmittance of a 10- $\mu\text{m}$  film with a refractive index  $n = 2.4$  without any absorption in the spectral range from 1000 to 3000  $\text{cm}^{-1}$ . One surface of the film is polished, and the other surface has a roughness  $\sigma_{\text{rms}}$  of 0.10  $\mu\text{m}$ . How does the  $\sigma_{\text{rms}}$  value affect the transmittance? Compare your result with that shown in Fig. 9.10.



**9.14.** Reproduce Example 9-2 and Fig. 9.10. Suppose the coherence spectral width  $\delta\nu = 1.5\Delta\nu$ , where  $\Delta\nu$  is the free spectral range. Determine the fringe-averaged transmittance. Explain why the peaks and the valleys flip after fringe averaging.

**9.15.** Calculate and plot the transmittance of a Fabry-Perot resonance cavity, assuming the medium to be lossless with  $n_2 = 2$ ,  $d_2 = 100 \mu\text{m}$ , and  $R = 0.9$ , for normal incidence in the wavenumber region from  $950$  to  $1050 \text{ cm}^{-1}$ . What are the free spectral range, the FWHM of the peak, and the  $Q$ -factor of the resonator? Does the theoretically predicted FWHM match with the plot?

**9.16.** Group project: A reflectance Fabry-Perot cavity can be constructed by coating a  $\text{SiO}_2$  film onto a silver substrate first and then a thin silver film onto the  $\text{SiO}_2$  film. Derive a formula for the reflectance. Based on Kirchhoff's law, one can calculate the emissivity of the structure. Show that the emissivity exhibits sharp peaks close to unity at specific wavelengths for normal incidence. When the wavelength is fixed, calculate the emissivity versus the polar angle for each polarization. Plot and show that there exist angular lobes in the emissivity of such structures. Hint: Choose the thicknesses of the silver film (on the order of  $100 \text{ nm}$ ) and the  $\text{SiO}_2$  film (on the order of  $3000 \text{ nm}$ ), and the wavelength (around  $1 \mu\text{m}$ ). Use the optical constants from Palik.<sup>2</sup>

**9.17.** Group project: Develop a Matlab code for the multilayer radiative properties based on the matrix formulation described in the text for both TE and TM waves. Compare your results with those calculated by using Rad-Pro, downloadable from [www.me.gatech.edu/~zhang](http://www.me.gatech.edu/~zhang).

**9.18.** Group project: Evaluate and plot the band structures of a Bragg reflector made of quarter-wave high- and low-index materials GaAs,  $n = 3.49$ , and AlAs,  $n = 2.95$ , around the wavelength of  $1064 \text{ nm}$ . Optional: Plot the normal reflectance near  $1064\text{-nm}$  wavelength with 7, 17, and 27 periods, assuming that the substrate is GaAs.

**9.19.** Derive Eq. (9.60) and Eq. (9.61).

**9.20.** Based on Eq. (9.64), show that when the evanescent wave exists, it will decay toward negative  $z$ . Change the subscript from I to III, and show that when the evanescent wave exists, it will decay toward positive  $z$ .

**9.21.** Derive Eq. (9.71) and Eq. (9.72).

**9.22.** Use different effective medium formulations to compute the effective dielectric function for silicon with a filling ratio  $\phi = 1/16$  in air at  $\lambda = 300 \text{ nm}$  ( $n = 5.0$  and  $\kappa = 4.2$ ),  $\lambda = 400 \text{ nm}$  ( $n = 5.6$  and  $\kappa = 0.39$ ),  $\lambda = 500 \text{ nm}$  ( $n = 4.3$  and  $\kappa = 0.073$ ), and  $\lambda = 800 \text{ nm}$  ( $n = 3.7$  and  $\kappa = 0.0066$ ).

**9.23.** Consider a grating region consisting of Si, with a filling ratio of  $1/6$ , on a semi-infinite Si substrate. The height of the grating is  $50 \text{ nm}$ . Calculate the reflectance for normal incidence, using different effective medium formulations at the corresponding wavelengths given in Problem 9.22. Compare your results with those in Fig. 9.19.

**9.24.** Plot the shadowing function for a Gaussian distribution as a function of the polar angle  $\theta$  for the rms slopes  $w = 0.05, 0.1, 0.2, \text{ and } 0.3$ .

**9.25.** Calculate the BRDFs at  $\lambda = 0.5$  and  $2 \mu\text{m}$  based on the analytical model for a gold surface (opaque) with a Gaussian roughness statistics. The SDF is given by

$$p(\zeta_x, \zeta_y) = \frac{1}{2\pi w} \exp\left(-\frac{\zeta_x^2 + \zeta_y^2}{2w^2}\right)$$

Use the optical constants from Problem 9.5 and the rms slope  $w = 0.1$  and  $0.3$ .

**9.26.** Comment on the limitations of different analytical models for the BRDF, such as the Rayleigh-Rice perturbation theory, the Kirchhoff approximation, and the geometric optics approximation.

---

# CHAPTER 10

---

## NEAR-FIELD ENERGY TRANSFER

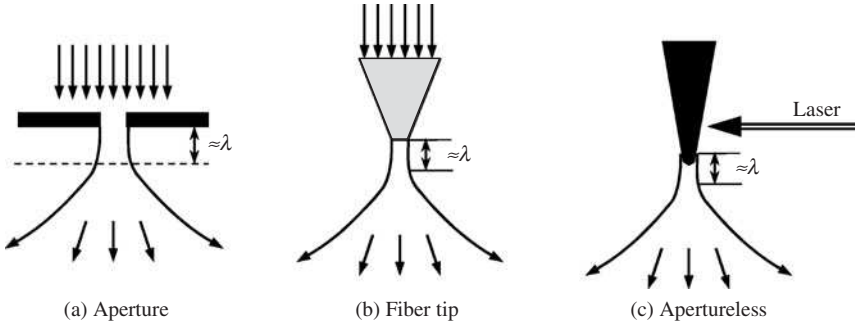
---

Near-field optics has played a significant role in nanoscience and nanobiotechnology in the past 20 years and continues to be an active research area, especially when dealing with field localization and resonances in micro/nanostructures, with applications in biochemical sensing and nanolithography. The preceding two chapters have laid the foundation of electromagnetic waves in bulk materials and nanostructures. The present chapter offers a more detailed treatment of the energy transfer by electromagnetic waves in the near field, as well as the coupling between near-field phenomena and far-field characteristics. The applications include nanomanufacturing, energy conversion systems, and nanoelectronics thermal management.

Ernst Abbe in 1873 and Lord Rayleigh in 1879 studied the required angular separation between two objects for their images to be resolved. The resolution of a conventional microscope is diffraction limited such that the smallest resolvable distance is approximately  $0.5\lambda/n$ , where  $\lambda$  is the wavelength in vacuum and  $n$  is the refractive index of the medium. Even with an immersion oil ( $n \approx 1.5$ ), the imaging sharpness is rather limited to the order of wavelength. The concept of near-field imaging was first described by Syge (*Phil. Mag.*, **6**, 356, 1928). This work elaborated the concept of using subwavelength aperture as small as 10 nm in diameter to introduce light to a specimen (e.g., a stained biological section), placed within 10-nm distance, which could move in its plane with a step size less than 10 nm. By measuring the transmitted light with a photoelectric cell and a microscope, an ultramicroscopic image could be constructed. In a subsequent paper (*Phil. Mag.*, **13**, 297, 1932), Syge described the idea of using piezoelectricity in microscopy. Syge's works, however, were largely unnoticed and the idea of near-field imaging was rediscovered many years later. Ash and Nicholls published a paper (*Nature*, **237**, 510, 1972) entitled "Super-resolution aperture scanning microscope." This work experimentally demonstrated near-field imaging with a resolution of  $\lambda/60$  using 10-GHz microwave radiation ( $\lambda = 3$  cm). In the 1980s, two groups have successfully developed near-field microscopes in the visible region.<sup>1,2</sup> The IBM group in Zurich formed the aperture through a quartz tip coated with a metallic film on its sides,<sup>1</sup> whereas the Cornell group used silicon microfabrication to form the aperture.<sup>2</sup> The fabrication process was later improved by using metal-coated tapered optical fibers. In the early 1990s, Betzig at Bell Labs and collaborators demonstrated single molecule detection and data storage capability of 45 gigabits per square inch.<sup>3</sup> Nowadays, near-field scanning optical microscope (NSOM), also known as scanning near-field optical microscope (SNOM), has become a powerful tool in the study of fundamental space- and time-dependent processes, thermal metrology, and optical manufacturing with a spatial resolution of less than 50 nm. NSOM is usually combined with the atomic force microscope (AFM) for highly controllable movement and position sensing. An alternative approach is to use a metallic AFM tip to couple the far-field radiation with the near-field electromagnetic waves in a subwavelength region underneath the tip. This is

the so called *apertureless NSOM*, which does not require an optical fiber or an aperture. Apertureless tips allow high-intensity laser energy to be focused to nanoscale dimensions for laser-assisted nanothermal manufacturing.<sup>4,5</sup>

Figure 10.1 illustrates three typical NSOM designs. The first is an aperture-based setup, where a very small opening is formed on an opaque plate and collimated light is incident



**FIGURE 10.1** Schematic illustration of different NSOM setups. (a) Aperture on an opaque plate. (b) Aperture at the end of a coated optical fiber. (c) Apertureless metallic tip. The opening or the tip is much smaller than the wavelength  $\lambda$ . The electric field is highly collimated in the near field within a distance of  $\lambda$  and diverges as the distance increases.

from the above. The second is based on a tapered optical fiber whose tip serves as an aperture. The third uses an apertureless metallic sharp tip, which reflects (scatters) the incident laser light. All of the three designs have one thing in common. The light is confined to a narrow region whose width may be much less than a wavelength. Furthermore, the electromagnetic field within one wavelength distance is very intense and highly collimated. In the near-field region, evanescent waves dominate. Because the amplitude of an evanescent wave decays exponentially away from the aperture or tip, the far-field, or the radiation field diverges and becomes very weak. Understanding the nature of evanescent waves and the localized fields is essential for the NSOM and other near-field optical devices.

Evanescent waves are also essential in energy transfer between adjacent objects, through photon tunneling, and in surface plasmon polaritons or surface phonon polaritons. Polaritons are elementary excitons in solids due to charge oscillations near the interface and can interact strongly with electromagnetic waves. In this chapter, we will first use total internal reflection to introduce evanescent waves, and then discuss polaritons or electromagnetic surface waves. The application to construct coherent thermal emission sources and radiation heat transfer at nanometer distances will be presented afterward.

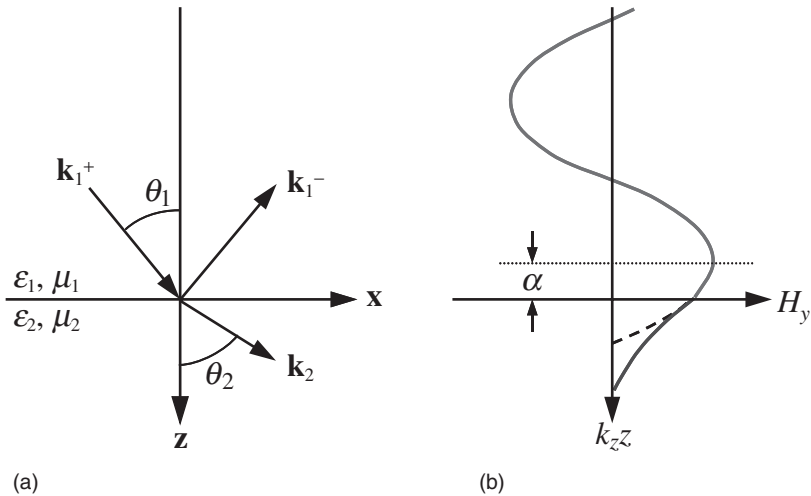
## 10.1 TOTAL INTERNAL REFLECTION, GUIDED WAVES, AND PHOTON TUNNELING

Total internal reflection occurs when light comes from an optically denser material to another material at incidence angles greater than the critical angle determined by Snell's law. As discussed in Chap. 8, the amplitude of the reflection coefficient becomes unity at incidence angles greater than the critical angle. Although no energy is transferred from medium 1 to medium 2, there exists an electromagnetic field in the second medium near the surface. This electromagnetic field can store as well as exchange energy with medium

1 at any instant of time. The time-averaged energy flux must be zero across the interface. Total internal reflection has important applications in optical fibers and waveguides. When medium 2 is not infinitely extended but a very thin layer sandwiched between the first medium and the third medium (which may be made of the same material as that of medium 1), photons can tunnel through the second medium into the third, even though the angle of incidence is greater than the critical angle. This phenomenon is called photon tunneling, radiation tunneling, or frustrated total internal reflection, and has been studied for over 300 years since Newton's time. Detailed descriptions of the original experiments and analyses by Isaac Newton can be found from his classical book, *Opticks* (reprinted by Dover Publications in 1952). The enhanced energy transfer by photon tunneling may have applications in thermophotovoltaic energy conversion devices as well as nanothermal manufacturing using heated AFM cantilever tips.

### 10.1.1 The Goos-Hänchen Shift

Evanescent waves can be illustrated by using the total internal reflection arrangement. Consider a plane wave of angular frequency  $\omega$  incident from a semi-infinite medium 1 to medium 2, as shown in Fig. 10.2a. The wavevector  $\mathbf{k}_1^+ = k_x \hat{\mathbf{x}} + k_{1z} \hat{\mathbf{z}}$ ,  $\mathbf{k}_1^- = k_x \hat{\mathbf{x}} - k_{1z} \hat{\mathbf{z}}$ , and



**FIGURE 10.2** Illustration of total internal reflection. (a) Schematic of the incident, reflected, and transmitted waves at the interface between two semi-infinite media. (b) The magnetic field distribution for a TM wave when total internal reflection occurs.

$\mathbf{k}_2 = k_x \hat{\mathbf{x}} + k_{2z} \hat{\mathbf{z}}$ , since the parallel wavevector component  $k_x$  must be the same as required by the phase-matching boundary condition. The magnitudes of the wavevectors are

$$k_1^2 = k_x^2 + k_{1z}^2 = \epsilon_1 \mu_1 \omega^2 / c^2 \quad (10.1a)$$

and

$$k_2^2 = k_x^2 + k_{2z}^2 = \epsilon_2 \mu_2 \omega^2 / c^2 \quad (10.1b)$$

where  $\epsilon$  and  $\mu$  are the relative (ratio to those of vacuum) permittivity and permeability, respectively, and  $c$  is the speed of light in vacuum (throughout this chapter). Let us assume

that the incident wave is  $p$  polarized or a TM wave, so that the only nonzero component of the magnetic field is in the  $y$  direction. The magnetic field of the incident wave may be expressed as  $\mathbf{H}_i = (0, H_y, 0)$ , where  $H_y(x, y, z, t) = H_i e^{ik_{1z}z + ik_1x - i\omega t}$ . For simplicity, let us omit  $\exp(-i\omega t)$  from now on. Recall that the Fresnel coefficients for a TM wave are defined as the ratios of the reflected or transmitted magnetic field to the incident magnetic field. For example, the Fresnel reflection coefficient is

$$r_p = \frac{H_r}{H_i} = \frac{k_{1z}/\epsilon_1 - k_{2z}/\epsilon_2}{k_{1z}/\epsilon_1 + k_{2z}/\epsilon_2} \quad (10.2)$$

The field in medium 1 is composed of the incident and reflected fields, and that in medium 2 is the transmitted field. Therefore,

$$\frac{H_y}{H_i} = \begin{cases} (e^{ik_{1z}z} + r_p e^{-ik_{1z}z}) e^{ik_x x} & \text{for } z \leq 0 \\ (1 + r_p) e^{ik_{2z}z} e^{ik_x x} & \text{for } z > 0 \end{cases} \quad (10.3)$$

The electric fields can be obtained by applying the Maxwell equations. Similar to Sec. 8.3.1, we can write the electric and magnetic fields in both media as follows:

$$\frac{E_x}{H_i} = \begin{cases} \frac{k_{1z}}{\omega \epsilon_1 \epsilon_0} (e^{ik_{1z}z} - r_p e^{-ik_{1z}z}) e^{ik_x x} & \text{for } z \leq 0 \\ \frac{k_{2z}}{\omega \epsilon_2 \epsilon_0} (1 + r_p) e^{ik_{2z}z} e^{ik_x x} & \text{for } z > 0 \end{cases} \quad (10.4)$$

and

$$\frac{E_z}{H_i} = \begin{cases} -\frac{k_x}{\omega \epsilon_1 \epsilon_0} (e^{ik_{1z}z} + r_p e^{-ik_{1z}z}) e^{ik_x x} & \text{for } z \leq 0 \\ -\frac{k_x}{\omega \epsilon_2 \epsilon_0} (1 + r_p) e^{ik_{2z}z} e^{ik_x x} & \text{for } z > 0 \end{cases} \quad (10.5)$$

Assume that  $\epsilon$ 's and  $\mu$ 's are real and furthermore,  $\epsilon_1 \mu_1 > \epsilon_2 \mu_2 > 0$ . From Eq. (10.1b), we have  $k_{2z}^2 = \epsilon_2 \mu_2 \omega^2 / c^2 - k_x^2$ . When  $\sqrt{\epsilon_2 \mu_2} < k_x c / \omega < \sqrt{\epsilon_1 \mu_1}$ , the incidence angle  $\theta_1$  is defined but the refraction angle is not, because  $k_{2z}$  becomes imaginary. One can write  $k_{2z} = i\eta_2$ , where  $\eta_2 = \sqrt{k_x^2 - \epsilon_2 \mu_2 \omega^2 / c^2}$  is a real positive number. In this case,  $|r_p| = 1$  and

$$r_p = e^{i\delta} = e^{-i2\alpha} \quad (10.6)$$

where  $\tan \alpha = (\eta_2 / \epsilon_2) / (k_{1z} / \epsilon_1)$ . Following Haus,<sup>6</sup> the magnetic field at  $x = 0$  in medium 1 can be written as

$$H_y = 2H_i e^{-i\alpha} \cos(k_{1z}z + \alpha), \quad z \leq 0 \quad (10.7a)$$

Similarly, in medium 2,  $H_y$  becomes

$$H_y = 2H_i e^{-i\alpha} \cos(\alpha) e^{-\eta_2 z}, \quad z > 0 \quad (10.7b)$$

The magnetic field at  $x = 0$  is plotted in Fig. 10.2b with respect to  $k_z z$ , at the instance of time when the phase of  $H_i e^{-i\alpha - \omega t}$  becomes zero. From this figure, one can see that the field decays exponentially in medium 2. As a result of total internal reflection, there is a phase shift in medium 1 so that the maximum field is shifted from the interface to  $k_z z = -\alpha$ . The phase angle of the reflection coefficient  $\delta = -2\alpha$  is called the *Goos-Hänchen phase shift*, which depends on the incidence angle  $\theta_1$  or  $k_x$ . The difference in  $\delta$  for TE and TM waves in a dielectric prism was used to construct a polarizer called *Fresnel's rhomb*, which can change a linearly polarized wave to a circularly polarized wave, or vice versa.<sup>7</sup>

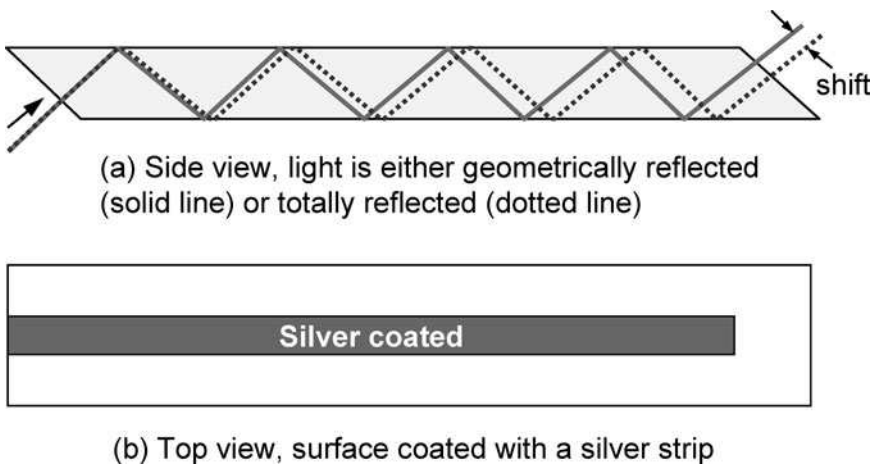
**Example 10-1.** Calculate the time-averaged Poynting vector near the interface in the case of total internal reflection.

**Solution.** Based on Example 8-1, it can be seen that the Poynting vector  $\mathbf{S} = \text{Re}(\mathbf{E}) \times \text{Re}(\mathbf{H})$  is in general a function of time. The time-dependent terms that oscillate with  $2\omega$ , however, become zero after integration. The time-averaged Poynting vector is  $\langle \mathbf{S} \rangle = \frac{1}{2} \text{Re}(\mathbf{E} \times \mathbf{H}^*)$ . For  $z > 0$ ,  $\langle S_z \rangle = \frac{1}{2} \text{Re}(E_x H_y^*) = 0$  because  $k_{2z}$  is purely imaginary. It can also be shown that  $\langle S_z \rangle = 0$  for  $z \leq 0$  (see Problem 10.2). Furthermore,

$$\langle S_x \rangle = -\frac{1}{2} \text{Re}(E_z H_y^*) = \begin{cases} \frac{k_x}{\omega \epsilon_1 \epsilon_0} |H_1|^2 [1 + \cos(2k_{1z}z + 2\alpha)], & z \leq 0 \\ \frac{k_x}{\omega \epsilon_2 \epsilon_0} |H_1|^2 [1 + \cos(2\alpha)] e^{-2\eta z}, & z > 0 \end{cases} \quad (10.8)$$

Note that  $\langle S_x \rangle$  does not have to be continuous at the interface. Depending on whether  $\epsilon$  is positive or negative, the sign of  $\langle S_x \rangle$  may be the same as or opposite to  $k_x$ . It should also be noted that  $\langle S_x \rangle$  is a sinusoidal function of  $z$  in medium 1 and decays exponentially in medium 2 as  $z$  approaches infinity.

Newton conjectured that, for the total internal reflection of light by the boundary, the beam of light would penetrate some distance into the optically rarer medium and then reenter the optically denser medium. In addition, he suspected that the path of the beam would be a parabola with its vertex in the rarer medium and, consequently, the actual reflected beam would be shifted laterally with respect to the geometric optics prediction. From the Poynting vector formulation given in Eq. (10.8), the energy must penetrate into the second medium to maintain the energy flow parallel to the interface and reenter the first medium so that no net energy is transferred across the interface. The actual beams have a finite extension so that the reflected beam in the far field can be separated from the incident beam since the Poynting vector is parallel to the wavevector. The effect of the parallel energy flow indeed causes the reflected beam to shift forward from that expected by the geometric optics analysis. Goos and Hänchen were the first to observe the lateral beam shift through a cleverly devised experiment [*Ann. Physik*, **6**(1), 333, 1947; **6**(5), 251, 1949]. A schematic of this experiment is shown in Fig. 10.3, in



**FIGURE 10.3** Illustration of the Goos-Hänchen experiment.

which a glass plate was used so that the incident light was multiply reflected by the top and bottom surfaces. In the middle of one or both of the surfaces, a silver strip was deposited. This way, the beam reflected by the silver film (solid line) would essentially follow geometric optics and that by total internal reflection would experience a lateral shift. Although the lateral shift is on the order of the wavelength, a large number of reflections (over 100 times) allowed the shift to be observed by a photographic plate. Lotsch published a series of papers on the comprehensive study of the Goos-Hänchen effect.<sup>8</sup> Puri and Birman provided an elegant review of earlier works, including several methods for analyzing the Goos-Hänchen effect.<sup>9</sup> A quantitative study of the Goos-Hänchen effect is presented next.

One way to model the lateral shift is to use a beam of finite width rather than an unbounded plane wave. Another method that is mathematically simpler considers the phase change of an incoming wave packet, which is composed of two plane waves with a slightly different  $k_x$ . Upon total internal reflection, the phase shift  $\delta = -2\alpha$  for a given polarization is a function of  $k_x$ . The difference in the phase shift will cause the reflected beam to exhibit a lateral shift along the interface ( $x$  direction) given as

$$D = -\frac{d\delta}{dk_x} = \frac{\varepsilon_1}{\varepsilon_2} \frac{2k_x}{\eta_2 k_{1z}} \frac{k_{1z}^2 + \eta_2^2}{k_{1z}^2 + (\eta_2 \varepsilon_1 / \varepsilon_2)^2} \quad (10.9)$$

where we have used  $\alpha = \tan^{-1}(\eta_2 \varepsilon_1 / k_{1z} \varepsilon_2)$ . In formulating Eq. (10.9),  $k_x$  is always taken as positive. Equation (10.8) clearly suggests that  $\langle S_x \rangle$  and  $k_x$  have the same sign when the permittivity is positive and different sign when the permittivity is negative.<sup>10</sup> When  $\varepsilon_1$  and  $\varepsilon_2$  have different signs, the lateral shift  $D$  will be negative, which implies that the lateral shift is opposite to  $\langle S \rangle_x$  of the incident beam. For a TE wave, one can simply replace  $\varepsilon$ 's by  $\mu$ 's in Eq. (10.9). For two dielectrics, we have  $\mu_1 = \mu_2 = 1$ ,  $\varepsilon_1 = n_1^2$ , and  $\varepsilon_2 = n_2^2$ , where  $n_1$  and  $n_2$  are the refractive indices of medium 1 and 2, respectively. Consequently, Eq. (10.9) reduces to the following:

$$D_s = \frac{2 \tan \theta_1}{\eta_2} \quad \text{for a TE wave} \quad (10.10a)$$

and

$$D_p = \frac{2 \tan \theta_1}{\eta_2 (n_1^2 \sin^2 \theta_1 / n_2^2 - \cos^2 \theta_1)} \quad \text{for a TM wave} \quad (10.10b)$$

At grazing incidence,  $k_{1z} \rightarrow 0$ , however, the shift in the direction parallel to the beam is  $D \cos \theta_1 = (2/\eta_2)(\varepsilon_2/\varepsilon_1) \sin \theta_1$ , which approaches a finite value and does not diverge. At the critical angle,  $\theta_1 = \theta_c = \sin^{-1}(n_2/n_1)$ ,  $\eta_2 = \delta = 0$ , and  $D$  approaches infinity. This difficulty can be removed by using the Gaussian beam incidence.<sup>11</sup> Quantum mechanics has also been applied to predict the lateral beam shift.<sup>8</sup> The Goos-Hänchen effect also has its analogy in acoustics and is of contemporary interest in dealing with negative index materials, waveguides, and photon tunneling.<sup>10,12,13</sup>

### 10.1.2 Waveguides and Optical Fibers

Optical fibers and waveguides are essential for optical communication and optoelectronics. There are numerous other applications such as noncontact radiation thermometry, near-field microscopy, and decoration lightings. According to a report in 2000, the total length of optical fiber wires that had been installed worldwide exceeded  $3.0 \times 10^{11}$  m, which equals the distance of a round trip from the earth to the sun. Optical fibers usually operate based on the principle of total internal reflection, as shown in Fig. 10.4. The fiber core is usually surrounded by a cladding material with a lower refractive index.

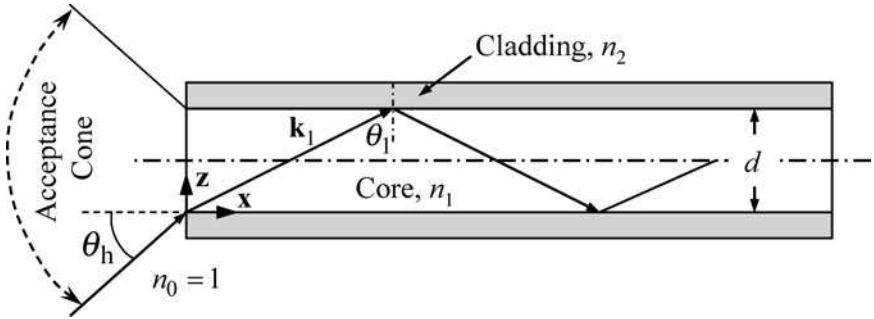


FIGURE 10.4 Schematic of a planar dielectric waveguide.

The numerical aperture  $NA$  is defined according to the half angle  $\theta_h$  of the acceptance cone, within which total internal reflection occurs. It can be seen from Fig. 10.4 that

$$NA = \sin\theta_h = n_1 \cos\theta_c = \sqrt{n_1^2 - n_2^2} \quad (10.11)$$

For example, if  $n_1 = 1.53$  and  $n_2 = 1.46$ , the critical angle  $\theta_c = 72.6^\circ$ , the maximum cone angle  $\theta_h = 27^\circ$ , and  $NA = 0.46$ . There are different types of waveguides, such as graded-index waveguides and metallic waveguides, in addition to the simple dielectric type. The cross section may be circular, annular, rectangular, or elliptical. In some cases, the diameter of the fiber is much greater than the wavelength and the electromagnetic waves inside the fiber are incoherent. These devices are sometimes called lightpipes, which are used for relatively short distances. Optical fibers in communication technology use very thin wires and transmit light with well-defined *modes*. In the following, the configuration of a 1-D dielectric slab between two media will be discussed to illustrate the basics of an optical waveguide. More detailed treatments can be found from the texts of Haus<sup>6</sup> and Kong.<sup>7</sup> The present author was fortunate to learn optoelectronics and the electromagnetic wave theory through graduate courses taught by these professors.

Consider the planar structure shown in Fig. 10.4 that is infinitely extended in the  $y$  direction. When the variation of  $d$  along the  $x$  direction is negligibly small compared to the wavelength, the electromagnetic waves inside the waveguide are coherent. A standing wave pattern must be formed in the  $z$  direction. This requires the phase change in the  $z$  direction, for the round trip including two reflections at the boundary, to be a multiple of  $2\pi$ , i.e.,

$$2k_{1z}d + 2\delta = 2m\pi, m = 0, 1, 2, \dots \quad (10.12)$$

where  $k_{1z} = (\omega/c)n_1 \cos\theta_1$ , and the phase shift upon total internal reflection is

$$\delta = -2\alpha = -2\tan^{-1}\left(g \frac{\sqrt{\sin^2\theta_1 - \sin^2\theta_c}}{\cos\theta_1}\right) \quad (10.13)$$

where  $g = 1$  for TE waves and  $g = n_2^2/n_1^2$  for TM waves.

The solutions of Eq. (10.12) give discrete values of  $\theta_1$  or  $k_x = (\omega/c)n_1 \sin\theta_1$ , at which waves can propagate through the fiber for a prescribed frequency. These are called *guided modes* of the optical fiber, and Eq. (10.12) may be regarded as the *mode equation*. The orders of mode are identified as  $TE_0, TE_1, \dots, TE_m$  or  $TM_0, TM_1, \dots, TM_m$  for a 1-D waveguide. For a 2-D waveguide, the subscripts consist of two indices “ $ml$ ” for each mode. As

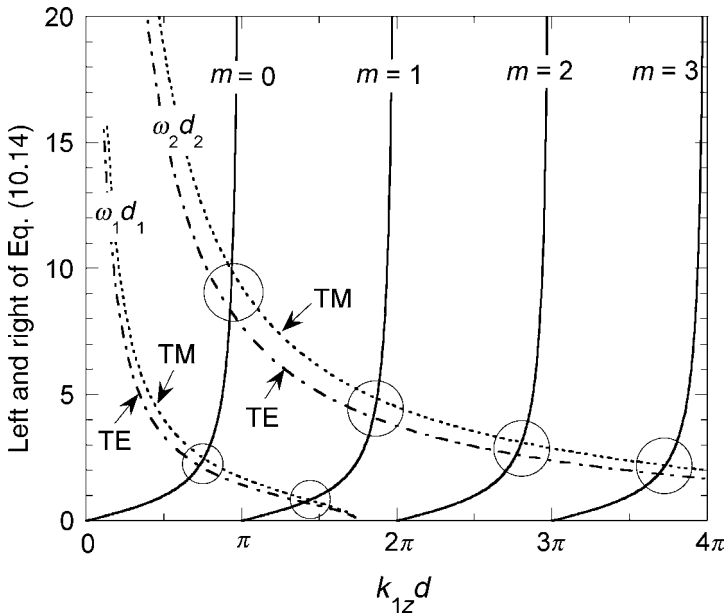


$\theta_1$  decreases from  $\pi/2$  to  $\theta_c$ ,  $k_{1z}$  increases and higher-order modes can be excited. One might wonder why  $\theta_1 = \pi/2$  or  $k_{x1} = k_1$  is *not* a guided mode. In this case, energy would go through the core, cladding, and air in a straight line. Any bending in the waveguide would result in some loss of energy transfer. On the other hand, the guided modes are much less affected by the bending. This is why an optical fiber can transfer signals to a very long distance while being flexible.

To illustrate the solution in terms of  $k_{1z}d$ , let us rearrange Eq. (10.12) as follows:

$$\tan\left(\frac{k_{1z}d}{2} - \frac{m\pi}{2}\right) = \tan\alpha = g \frac{\eta_2}{k_{1z}} = g\sqrt{\frac{(k_1d)^2 - (k_2d)^2}{(k_{1z}d)^2} - 1} \quad (10.14)$$

The left and right sides of Eq. (10.14) can be plotted in the same graph against  $k_{1z}d$ , as shown in Fig. 10.5, for two values of  $\omega d$ , assuming  $\omega_2d_2 > \omega_1d_1$ . The dash-dotted curves



**FIGURE 10.5** Solutions of the mode equation, when  $\omega_2d_2 > \omega_1d_1$ . The circles indicate the intersections between the curves described by the left and right sides of Eq. (10.14).

are for TE waves, and the dotted curves are for TM waves. The intersections within the circles identify the guided modes. It is noted that fewer modes are permitted with a smaller  $\omega d$  or  $d/\lambda$ . In the graph with  $\omega_1d_1$ , the possible modes are  $TE_0$ ,  $TE_1$ ,  $TM_0$ , and  $TM_1$  only. A fiber that supports only a single mode for a given frequency is called a *single-mode fiber*; otherwise, it is called a *multimode fiber*.

**Example 10-2.** Determine the range of  $d/\lambda$  so that only the  $TE_0$  and  $TM_0$  waves are guided in the planar waveguide with  $n_1 = 1.55$  and  $n_2 = 1.42$ . Moreover, if  $d/\lambda = 1000$ , how many TE and TM modes may be guided?

**Solution.** Because  $d/\lambda$  must be small enough so that the right-hand side of Eq. (10.14) becomes zero at  $k_{1z}d = \pi$ , we have  $(k_1d)^2 - (k_2d)^2 = \pi^2$ , or  $4\pi^2(n_1^2 - n_2^2)(d/\lambda)^2 = \pi^2$ . Finally, we find  $d/\lambda < 0.5(n_1^2 - n_2^2)^{-1/2} = 1.3$ . Moreover, from Fig. 10.5, we can estimate the highest-order mode

$M$  using  $k_{1z}d = M\pi$  and  $\cos \theta_1 = \cos \theta_c$  when  $d \gg \lambda$ . Hence,  $2\pi(d/\lambda) \cos \theta_c = M\pi$ , or  $M = 2(d/\lambda) \cos \theta_c = 801.8$ . There will be 802 TE modes and 802 TM modes including the zeroth-order modes.

Next, we will study the fields in a planar waveguide. Let us take a TE wave and write in the more general terms  $\epsilon_1, \mu_1, \epsilon_2$ , and  $\mu_2$ . The electric field is nonzero only in the  $y$  direction, and the  $y$ -component of the electric field is given by

$$E_y = \begin{cases} Ce^{\eta_2 z} e^{ik_x x}, & z < 0 \\ (Ae^{ik_{1z} z} + Be^{-ik_{1z} z}) e^{ik_x x}, & 0 \leq z \leq d \\ De^{-\eta_2(z-d)} e^{ik_x x}, & z > d \end{cases} \quad (10.15)$$

where the time-harmonic term  $\exp(-i\omega t)$  is again omitted for simplicity. The magnetic fields can be obtained as  $H_x = -(i\omega\mu_1\mu_0)^{-1}(\partial E_y/\partial z)$  and  $H_z = (i\omega\mu_1\mu_0)^{-1}(\partial E_y/\partial x)$ . There are four boundary conditions for the tangential components to be continuous at  $z = 0$  and  $z = d$ . We end up with a set of homogeneous linear equations of the coefficients  $A, B, C$ , and  $D$ . The solution exists only when the determinant of the characteristic  $4 \times 4$  matrix becomes zero and can be expressed in a combined equation as follows:

$$\tan(k_{1z}d) \left( \frac{k_{1z}^2}{\epsilon_1^2} - \frac{\eta_2^2}{\epsilon_2^2} \right) = 2 \left( \frac{k_{1z}\eta_2}{\epsilon_1\epsilon_2} \right) \quad (10.16)$$

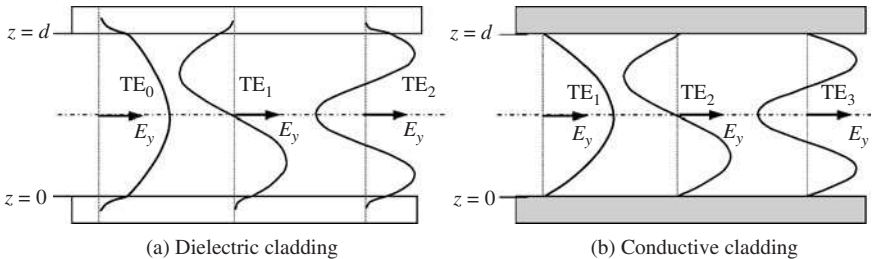
This is an equivalent expression of the mode equation. An easier way to solve Eq. (10.15) is by considering the condition of total internal reflection at the boundaries, i.e.,

$$A = Be^{i\delta} \text{ and } B = Ae^{i(2k_{1z}d + \delta)} \quad (10.17)$$

The combination gives  $e^{i(2k_{1z}d + 2\delta)} = 1$ , which is nothing but Eq. (10.12). After substituting  $A = Be^{-i2\alpha}$  into Eq. (10.15), boundary conditions require that

$$E_y = \begin{cases} 2e^{-i\alpha} B \cos\left(\frac{k_{1z}d}{2} - \frac{m\pi}{2}\right) e^{\eta_2 z} e^{ik_x x}, & z < 0 \\ 2e^{-i\alpha} B \cos\left(k_{1z}z - \frac{k_{1z}d}{2} - \frac{m\pi}{2}\right) (k_{1z}z - \alpha) e^{ik_x x}, & 0 \leq z \leq d \\ 2e^{-i\alpha} B \cos\left(\frac{k_{1z}d}{2} + \frac{m\pi}{2}\right) e^{-\eta_2(z-d)} e^{ik_x x}, & z > d \end{cases} \quad (10.18)$$

Figure 10.6a shows the electric field distribution for TE<sub>0</sub>, TE<sub>1</sub>, and TE<sub>2</sub>. The decaying fields inside the cladding are clearly demonstrated. For a cladding with the conductivity



**FIGURE 10.6** Electric field distribution  $E_y(z)$  in planar waveguides. For the conducting cladding,  $\sigma \rightarrow \infty$  and the lowest-order TE mode is the first order.

$\sigma \rightarrow \infty$ , the waves will be perfectly reflected at the interface without any phase shift and the electric field must vanish in the cladding. Only the odd  $m$ 's are guided modes. The first guided mode is  $TE_1$ , and the guided mode  $TE_q$  corresponds to  $q = (m + 1)/2$ , with  $m = 1, 3, 5, \dots$ . The electric fields for the conducting waveguide modes  $TE_1$ ,  $TE_2$ , and  $TE_3$  are shown in Fig. 10.6b for comparison with those for the first three modes in the dielectric waveguide. The difference lies in that no fields can penetrate into the conducting waveguide, whereas the fields can penetrate into the dielectric cladding.

**Example 10-3.** Determine the energy flux, phase velocity, and group velocity of the electromagnetic waves in a planar dielectric waveguide.

**Solution.** Obviously, there is no net energy flow in the  $z$  direction, and  $\langle S \rangle_z = \frac{1}{2} \text{Re}(E_y H_z^* - E_z H_y^*)$ . The second term on the right becomes zero for a TE wave; thus,  $\langle S \rangle_x = (k_x/2\omega\mu_0)E_y E_y^*$ . Integration of  $\langle S \rangle_x$  from  $z = -\infty$  to  $+\infty$  gives the power transmitted per unit length in the  $y$  direction. Note that a small portion of energy is transmitted through the cladding. The phase velocity along the  $x$  direction is  $v_p = \omega/k_x = c/(n_1 \sin\theta_1)$ . The group velocity for a given mode is given by  $v_g = (dk_x/d\omega)^{-1}$ , which requires the solution of Eq. (10.16) accounting for the frequency-dependent refractive index.

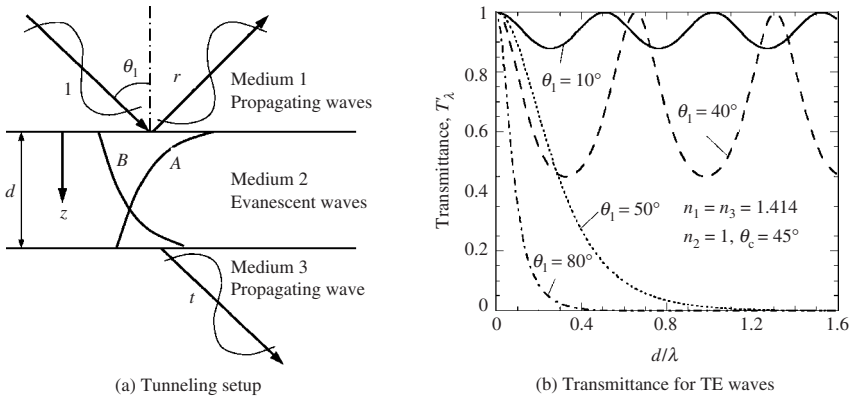
In Chap. 9, we introduced the concept of Fabry-Perot resonant cavities. Two- and three-dimensional optical cavities and microwave cavities support resonance modes, which are standing waves within the cavity. These devices are important for photonics and optoelectronics. Microcavities have also been used to modify the surface radiative properties. The quality factor, or the  $Q$ -factor, of a resonator is defined as the ratio of energy storage to the energy dissipation. High  $Q$ -factors can be achieved with the microfabricated microcavities for quantum electrodynamics (QED), enhancement and suppression of spontaneous emission, and biological and chemical sensing.<sup>14</sup> A special microcavity is made of spheres or disks, where the resonance is built up around a circumference in the form of a polygon. Total internal reflection traps the light inside the microsphere or the disk. At a particular wavelength, when resonance occurs, light undergoes multiple reflections, and a strong electric field which is confined near the perimeter can be built. This is the so-called *whispering gallery mode (WGM)*, named after the whispering gallery at St. Paul's Cathedral in London. A whispering gallery is a circular gallery under a dome where whisperers can be heard from the opposite side of the building. Optical fibers or waveguides are commonly used to couple the photon energy to or from the microcavities via evanescent waves. Ultrahigh  $Q$ -factors can be achieved with WGMs. The energy coupling mechanisms have recently been studied by Guo and Quan using a finite-element method.<sup>15</sup>

A recent development in fiber optics is the use of photonic crystals (PCs) to confine the light into a fiber, whose cladding region is made of PCs, rather than a solid low-index material. The fiber core may be either solid or hollow, and the PCs in the cladding region may contain air-filled holes in silica. For this reason, these fibers are called *photonic crystal fibers (PCFs)*, and some are called *holey fibers*.<sup>16</sup> In the stop band, waves cannot propagate inside the PC and thus effectively confine the propagating wave to the core region, where the modes can be guided, without using total internal reflection. One of the advantages of PCFs over conventional optical fibers is the spectral broadening that enables high-intensity pulses to be transmitted with less distortion or loss of the spectral information, which have important applications such as optical coherence spectroscopy and tomography. Another advantage is that the use of large guiding areas can provide low-loss high-power delivery for imaging, lithography, and astronomy. Other potential applications range from birefringence and nonlinear optics to atomic particle guidance.<sup>16</sup>

### 10.1.3 Photon Tunneling by Coupled Evanescent Waves

In the preceding sections, we clearly demonstrated that an evanescent wave exists inside the optically rarer medium, which can be air or vacuum, and decays exponentially away

from the surface. Furthermore, the evanescent wave or field does not carry energy in the direction normal to the interface. On the other hand, if another optically denser medium is brought to close proximity of the first medium, as shown in Fig. 10.7, energy can be



**FIGURE 10.7** Illustration of photon tunneling. (a) Schematic drawing of the three layers and fields. (b) Calculated transmittance for a TE wave, assuming  $n_1 = n_3 = 1.414$  and  $n_2 = 1$ . Note the distinct differences between the interference effect and the photon tunneling phenomenon, where the transmittance decreases with increasing  $d$  and becomes negligibly small for  $d > \lambda$ .

transmitted from the first to the third medium, even though the angle of incidence is greater than the critical angle. This phenomenon, known as *frustrated total internal reflection*, *photon tunneling*, or *radiation tunneling*, is very important for energy transfer between two bodies when the distance of separation is shorter than the dominant wavelength of the emitting source. Frustrated total internal reflection has been known since Newton's time and was theoretically investigated by Hall (*Phys. Rev. Ser. I*, **15**, 73, 1902). Cryogenic insulation is a practical example when photon tunneling may be significant.<sup>17</sup> Advances in micro/nanotechnologies have made it possible for the energy transfer by photon tunneling to be appreciable and even dominant at room temperature or above. This may have applications ranging from microscale thermophotovoltaic devices to nanothermal processing and nanoelectronics thermal management.<sup>18–20</sup>

While photon tunneling is analogous to electron tunneling, through a potential barrier, which may be explained by quantum mechanics, it can be understood by the coupling of two oppositely decaying evanescent waves.<sup>21</sup> Because of the second interface, a backward-decaying evanescent wave is formed inside layer 2, the optical rarer medium. The Poynting vector of the coupled evanescent fields has a nonzero normal component, suggesting that the energy transmission between the media is possible as long as the gap width is smaller than the wavelength. Beyond this wavelength, the field strength of the forward-decaying evanescent wave is too low when it reaches the second interface and the reflected evanescent field is negligible. The matrix formulation discussed in Chap. 9 can be used to calculate the transmittance and the reflectance through the gap (i.e., medium 2) as if there were propagating waves. To illustrate this, consider all three layers are dielectric. Taking the TM wave incidence as an example, let us write the magnetic field inside medium 2 as follows:

$$H_y(x, z) = (Ae^{ik_2z} + Be^{-ik_2z})e^{ik_x x}, \quad 0 \leq z \leq d \quad (10.19)$$

where  $A$  and  $B$  are determined by the incident field and boundary conditions. When two waves are combined, the Poynting vector of the field  $\langle \mathbf{S} \rangle = \frac{1}{2} \text{Re}[(\mathbf{E}_1 + \mathbf{E}_2) \times (\mathbf{H}_1^* + \mathbf{H}_2^*)]$  has four terms. Two of them can be associated with the power flux of each individual wave,

while the other two represent the interaction between the waves. After simplification, the normal component of the Poynting vector can be expressed as

$$\langle S_z \rangle = \frac{k_{2z}}{2\omega\epsilon_2\epsilon_0} (|A|^2 - |B|^2), \quad \text{when } k_{2z}^2 = k_2^2 - k_x^2 > 0 \quad (10.20a)$$

and 
$$\langle S_z \rangle = -\frac{\eta_2}{\omega\epsilon_2\epsilon_0} \text{Im}(AB^*), \quad \text{when } \eta_2^2 = -k_{2z}^2 = k_x^2 - k_2^2 > 0 \quad (10.20b)$$

Because there is no loss or absorption,  $\langle S_z \rangle$  is independent of  $z$  in medium 2, and the ratio of  $\langle S_z \rangle$  in medium 2 to that of the incidence in medium 1 is the transmittance. When propagating waves exist in medium 2 or the angle of incidence is smaller than the critical angle, interference will occur and the energy flux in the  $z$  direction can be represented by the forward- and backward-propagating waves, see Eq. (10.20a). The transmittance oscillates as the thickness of medium 2 is increased. When evanescent waves exist in medium 2 at incidence angles greater than the critical angle, the transmittance is a decaying function of the thickness of medium 2, as shown in Fig. 10.7b. While the individual evanescent wave does not carry energy, the coupling results in energy transfer, as suggested by Eq. (10.20b). Equation (9.8) through Eq. (9.10), derived in the previous chapter, can be used to calculate the transmittance and the reflectance. These equations are applicable to arbitrary electric and magnetic properties as long as the medium is isotropic and homogeneous within each layer. The phase shift  $\beta$  in these equations is purely imaginary when medium 2 is a dielectric.

**Example 10.4.** Assuming that the incident field has an amplitude of 1, determine  $A$  and  $B$  in Eq. (10.19) for  $\theta_1 > \theta_c = \sin^{-1}(n_2/n_1)$ , when all three media are dielectric with  $n_3 = n_1 > n_2$ . Find an expression of the tunneling transmittance using real variables only.

**Solution.** The tangential fields can be written as follows for the three-layer structure shown in Fig. 10.7a. Note that  $\eta_2 = \sqrt{k_x^2 - k_2^2} = (2\pi n_1/\lambda) \sqrt{\sin^2\theta_1 - \sin^2\theta_c}$ .

$$H_y = \begin{cases} (e^{ik_1z} + re^{-ik_1z})e^{ik_x x}, & z \leq 0 \\ (Ae^{-\eta_2 z} + Be^{\eta_2 z})e^{ik_x x}, & 0 < z \leq d \\ te^{ik_1z}e^{ik_x x}, & z > d \end{cases} \quad (10.21)$$

$$E_x = \begin{cases} \frac{k_{1z}}{\omega n_1^2 \epsilon_0} (e^{ik_1z} - re^{-ik_1z})e^{ik_x x}, & z \leq 0 \\ \frac{i\eta_2}{\omega n_2^2 \epsilon_0} (Ae^{-\eta_2 z} - Be^{\eta_2 z})e^{ik_x x}, & 0 < z \leq d \\ \frac{k_{1z}}{\omega n_1^2 \epsilon_0} te^{ik_1z}e^{ik_x x}, & z > d \end{cases} \quad (10.22)$$

The continuity of tangential fields at the two interfaces allow us to determine  $t$ ,  $r$ ,  $A$  and  $B$ . Note that because the incident field has an amplitude of 1, the preceding equations do not yield a set of homogeneous linear equations as in the case of guided waves. If we use Eq. (10.6) for  $r_p = e^{i\delta}$ , where  $\delta = -2\alpha$  and  $\cot(\alpha) = (k_{1z}/n_1^2)/(\eta_2/n_2^2)$  for a TM wave, we can rewrite Eq. (9.7) and Eq. (9.8) to obtain the reflection and transmission coefficients as follows:

$$r = \frac{e^{i\delta}(1 - e^{-2\eta_2 d})}{1 - e^{2i\delta}e^{-2\eta_2 d}} \quad (10.23)$$

$$t = \frac{(1 - e^{2i\delta})e^{-\eta_2 d}}{1 - e^{2i\delta}e^{-2\eta_2 d}} \quad (10.24)$$

where we have used the relationship of Fresnel's coefficients and set the phase shift in Eq. (9.6) to  $\beta = i\eta_2 d$ . After matching the boundary conditions at  $z = d$ , we have

$$A = 0.5t[1 - i\cot(\alpha)] \quad \text{and} \quad B = 0.5t[1 + i\cot(\alpha)]e^{-\eta_2 d} \quad (10.25)$$

It can be shown that the normal component of the Poynting vector is the same in media 2 and 3 (see Problem 10.8). The tunneling transmittance becomes

$$T'_\lambda = tt^* = \frac{2[1 - \cos(2\delta)]e^{-2\eta_2 d}}{1 + e^{-4\eta_2 d} - 2\cos(2\delta)e^{-2\eta_2 d}} \quad (10.26a)$$

$$\text{or} \quad T'_\lambda = \frac{\sin^2(\delta)}{\sin^2(\delta) + \sinh^2(\eta_2 d)} \quad (10.26b)$$

Clearly, the tunneling transmittance does not oscillate as  $d$  increases; rather, it decreases monotonically from 1 to 0 as  $d$  is increased from 0 to infinity. Equation (10.23) through Eq. (10.25) can be applied to TE waves by taking  $\cot(\alpha) = k_{1z}/\eta_2$ , which changes the Fresnel reflection coefficient  $r_p$  to  $r_s$  because only the dielectric media are considered here. Equation (10.26) is convenient for calculating the tunneling transmittance between dielectrics.

### 10.1.4 Thermal Energy Transfer between Closely Spaced Dielectrics

Energy exchange between closely spaced dielectric plates can be calculated by integrating Planck's function over all wavelengths as well as over the whole hemisphere using the directional-spectral transmittance. Let us use an example to illustrate the procedure and the effect of photon tunneling and interferences on the near-field thermal radiation.

**Example 10-5.** Calculate the hemispherical transmittance between two dielectrics of  $n_1 = n_3 = 3$ , separated by a vacuum gap  $d$  ( $n_2 = 1$ ). Use the results to calculate the radiative energy transfer between the two media, assuming  $T_1 = 1000$  K and  $T_3 = 300$  K.

**Analysis.** In the far field, we can use the following formula discussed in Chap. 2 (see Example 2-6) to calculate the net radiative heat flux:

$$q''_{13,d \rightarrow \infty} = \frac{\sigma_{\text{SB}} T_1^4 - \sigma_{\text{SB}} T_3^4}{1/\epsilon_1 + 1/\epsilon_3 - 1} \quad (10.27)$$

The hemispherical emissivity of each surface can be evaluated using Eq. (8.86), which can be rewritten as follows, considering that the emissivity is independent of the azimuthal angle  $\phi$ :

$$\epsilon_{\lambda,h} = 2 \int_0^{\pi/2} \epsilon'_\lambda(\theta) \cos \theta \sin \theta d\theta \quad (10.28)$$

One could average the directional-spectral emissivity over the two polarizations. However, the preferable way is to calculate the hemispherical emissivity for each polarization and use it to calculate the net heat flux by taking half of Eq. (10.27). The heat fluxes calculated for the two polarizations can then be added to obtain the total heat flux. The results give the far-field limit, which is always smaller than  $q''_{13, \text{BB}} = \sigma_{\text{SB}}(T_1^4 - T_3^4)$ , which is the net radiative heat flux between two blackbodies. This will not be the case in the near field when interference and tunneling effects are important.

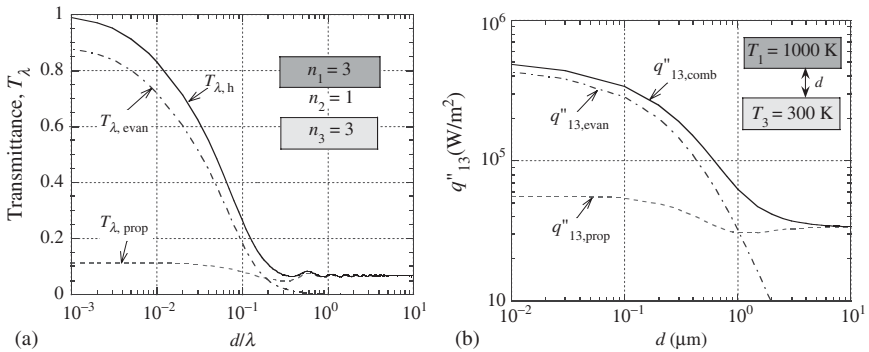
**Solution.** The hemispherical transmittance can be evaluated in the similar way by integration over the hemisphere. Note that only a small cone of radiation, originated from medium 1, will result in propagating waves in medium 2. This half cone angle is the critical angle, which is  $\theta_c = \sin^{-1}(n_2/n_1) \approx 19.5^\circ$ . Thus, we can divide the hemispherical transmittance in two parts to separately evaluate the transmittance. Keeping in mind that the transmittance is defined as the ratio of the transmitted energy to the incident energy, we can sum the two parts to obtain the hemispherical transmittance

$$T_{\lambda, h} = T_{\lambda, \text{prop}} + T_{\lambda, \text{evan}} \tag{10.29}$$

where 
$$T_{\lambda, \text{prop}} = 2 \int_0^{\theta_c} T'_{\lambda} \cos\theta \sin\theta d\theta \tag{10.29a}$$

and 
$$T_{\lambda, \text{evan}} = 2 \int_{\theta_c}^{\pi/2} T'_{\lambda} \cos\theta \sin\theta d\theta \tag{10.29b}$$

If  $n_1 \neq n_3$ ,  $\theta_c$  will depend on whether the incidence is from medium 1 or 3, and the resulting hemispherical transmittance will be the same. We can obtain the average transmittance for the two polarizations, as shown in Fig. 10.8a. The propagating wave contribution shows some oscillations but reaches a constant value when  $d/\lambda \rightarrow 0$  where all waves will be constructively added. At  $d/\lambda \gg 1$ , the constructive and destructive interferences cancel out so that  $T_{\lambda, \text{prop}}$  become a constant again. The



**FIGURE 10.8** Radiation heat transfer between dielectric surfaces in close proximity. (a) Contributions to hemispherical transmittance by interference and tunneling, where the transmittance is the average of both polarizations. (b) Net heat flux as a function of the distance of separation.

contribution of evanescent waves becomes important when  $d/\lambda < 1$  and starts to dominate over that of the propagating waves when  $d/\lambda \ll 1$ . When  $d/\lambda \rightarrow 0$ , the evanescent wave or tunneling contributes to nearly 90% of the transmittance when  $n_1 = 3$ . This explains why photon tunneling is very important for the near-field energy transfer.

Planck’s blackbody distribution function, given by Eq. (8.44), can be rewritten for each polarization in media 1 and 3, respectively, as

$$e_{b,\lambda}(\lambda, T_1) = \frac{n_1^2 C_1}{2\lambda^5 (e^{C_2/\lambda T_1} - 1)} \tag{10.30a}$$

and 
$$e_{b,\lambda}(\lambda, T_3) = \frac{n_3^2 C_1}{2\lambda^5 (e^{C_2/\lambda T_3} - 1)} \tag{10.30b}$$

where  $\lambda$  in  $\mu\text{m}$  is the wavelength in vacuum, and  $C_1 = 3.742 \times 10^8 \text{ W} \cdot \mu\text{m}^4/\text{m}^2$  and  $C_2 = 1.439 \times 10^4 \mu\text{m} \cdot \text{K}$  are the first and second radiation constants in vacuum. The emissive power in a nondispersive dielectric is increased by a factor of the square of the refractive index, as a result of the increased photon density of states. The factor 2 in the denominator is included because only single polarization has been considered. The net radiation heat flux from medium 1 to 3 is

$$q''_{1 \rightarrow 3} = \int_0^{\infty} e_{b,\lambda}(\lambda, T_1) T_{\lambda,h}(\lambda) d\lambda \quad (10.31a)$$

and that from medium 3 to 1 is

$$q''_{3 \rightarrow 1} = \int_0^{\infty} e_{b,\lambda}(\lambda, T_3) T_{\lambda,h}(\lambda) d\lambda \quad (10.31b)$$

where  $T_{\lambda,h}$  is obtained from Eq. (10.29). Hence, the net radiation heat transfer becomes

$$q''_{13} = q''_{1 \rightarrow 3} - q''_{3 \rightarrow 1} \quad (10.32)$$

One can also separately substitute the hemispherical transmittance of propagating and evanescent waves to Eq. (10.31). Equation (10.32) should be individually applied to TE and TM waves, and then summed together to get the net heat flux. The integration limits can be set such that the lower limit  $\lambda_L = 0.1\lambda_{mp}$  and the upper limit  $\lambda_H = 10\lambda_{mp}$ , where  $\lambda_{mp}$  is the wavelength corresponding to the maximum blackbody emissive power at the temperature of the higher-temperature medium as expressed in Eq. (8.45). The calculated results of the near-field radiative transfer are shown in Fig. 10.8b as a function of the separation distance  $d$ . Several important observations can be made. (a) When  $d \ll \lambda_{mp}$ , the propagating waves result in  $q''_{13,prop} = \sigma_{SB}(T_1^4 - \sigma T_3^4)$  and the evanescent waves result in  $q''_{13,evan} = (n_1^2 - 1)\sigma_{SB}(T_1^4 - T_3^4)$ . The combined net radiation heat transfer is  $q''_{13,comb} = n_1^2\sigma_{SB}(T_1^4 - T_3^4)$ . (b) As the distance increases, the evanescent wave contribution goes down monotonically and becomes negligible when  $d = \lambda_{mp}$ , which is about 3  $\mu\text{m}$ . (c) Due to interference effects, the energy transfer by propagating waves decreases slightly as  $d$  increases and then reaches the far-field limit, Eq. (10.27), when  $d \gg \lambda_{mp}$ .

If the media were conductive, the previous calculations are not appropriate because of the large imaginary part of the refractive index or the dielectric function. In fact, the near-field radiation heat transfer can be greatly enhanced with the presence of surface waves or if the media are semiconductors.<sup>18–20</sup> The treatment requires the knowledge of fluctuational electrodynamics, which will be discussed in Sec. 10.5 at length.

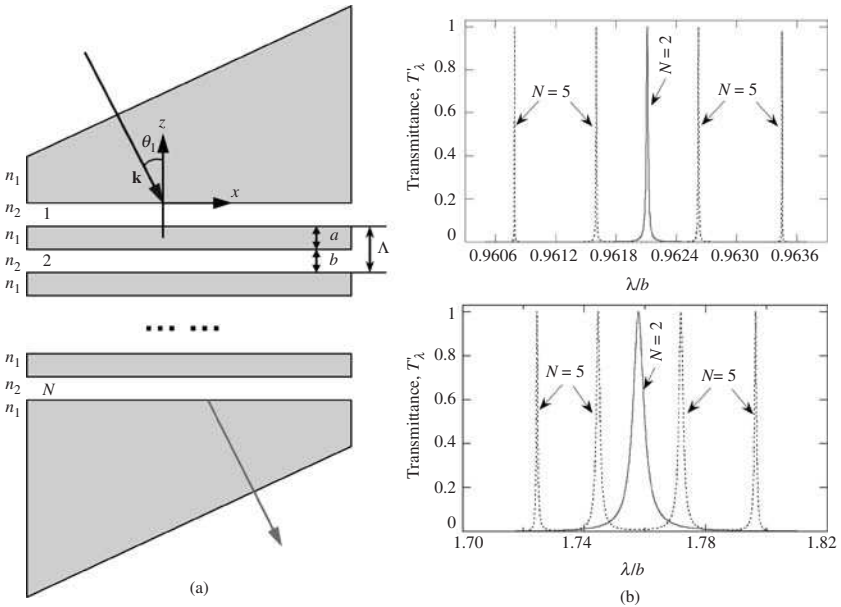
### 10.1.5 Resonance Tunneling through Periodic Dielectric Layers

There exists a photonic analogue of *resonance tunneling* of electrons in double-barrier quantum well structures. The geometry to illustrate resonance photon tunneling is depicted in Fig. 10.9a, with periodic layers of thicknesses  $a$  and  $b$ , like the photonic crystal (PC) structure discussed in Sec. 9.3, with a period  $\Lambda = a + b$ . For tunneling to occur, the double-prism structure can be used so that light is incident from medium 1 with a refractive index  $n_1$ . The barrier of thickness  $b$  is made of another dielectric with a refractive index  $n_2$  that is lower than  $n_1$ . There are  $N$  periods or unit cells in total between the end media. Light is incident at an incidence angle  $\theta_1 > \theta_c = \sin^{-1}(n_2/n_1)$ . Yeh performed a detailed analysis of this phenomenon and derived the equation of transmittance,<sup>22</sup> which can be expressed as

$$T'_\lambda = \frac{1}{1 + \frac{\sinh^2(\eta b) \sin^2(NK\Lambda)}{\sin^2(\delta) \sin^2(K\Lambda)}} \quad (10.33)$$

where  $K$  is the Bloch wavevector of the PC,  $\delta$  is the phase angle upon total internal reflection, and  $\eta$  is the imaginary part of the normal component of the wavevector in the lower-index dielectric, as defined in Example 10-4. It can be seen that Eq. (10.33) reduces to Eq. (10.26a) and Eq. (10.26b) for  $N = 1$ , where the transmittance is 1 at  $b = 0$ , and decreases monotonically with increasing  $b$ .





**FIGURE 10.9** Resonance tunneling. (a) Alternative high-index ( $n_1$ ) and low-index ( $n_2$ ) multiple dielectric layers for resonance tunneling. (b) Calculated transmittance spectra for  $N = 2$  and  $N = 5$ , at two wavelength regions. Calculation conditions are  $n_1 = 3$ ,  $n_2 = 2$ ,  $a = b/2$ , and  $\theta_1 = 45^\circ$ .

The following equation can be used to calculate  $K\Lambda$ :

$$\cos(K\Lambda) = \cos(k_{1z}a)\cosh(\eta b) + \cot(\delta)\sin(k_{1z}a)\sinh(\eta b) \tag{10.34}$$

where  $k_{1z}$  is the normal component of the wavevector in medium 1. While  $\cos(K\Lambda)$  is real,  $K\Lambda$  is in general complex. However, there exist regions or pass bands where  $|\cos(K\Lambda)| \leq 1$  so that  $K\Lambda$  is real. The transmittance does not oscillate in the pass bands, unlike what was shown in Fig. 9.17, where propagating waves exist in both types of dielectrics. Here, evanescent waves exist in the lower-index dielectric layers. However, the transmittance expressed in Eq. (10.33) becomes unity when the following equation holds:

$$\frac{\sin(NK\Lambda)}{\sin(K\Lambda)} = 0 \tag{10.35}$$

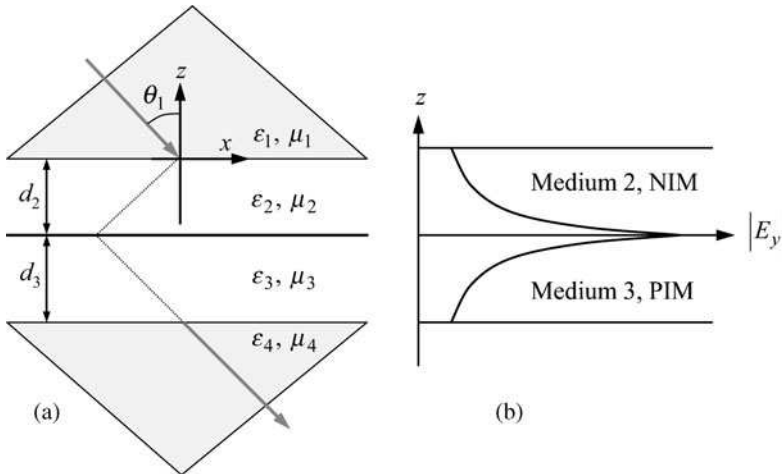
The denominator of this equation simply excludes the zeros in  $\sin(NK\Lambda)$  for  $K\Lambda = m\pi$ ,  $m = 0, \pm 1, \pm 2, \dots$ . It turned out that in each pass band, there exist  $(N - 1)$  solutions, with different combinations of  $\omega$ ,  $k_x$ , and the thicknesses  $a$  and  $b$ . As an example, Fig. 10.9b illustrated the transmittance as a function of  $\lambda/b$  when  $n_1 = 3$ ,  $n_2 = 2$ ,  $\theta_1 = 45^\circ$  and  $a/b = 0.5$ . Because of the narrow transmittance peaks, the plot is broken into two panels, each corresponding to a pass band. For  $N = 2$ , there is only one peak in each pass band, while for  $N = 5$ , there are four peaks. Yeh showed that the resonance frequencies correspond to the guided modes in the multilayer-waveguide equations.<sup>22</sup> Hence, the fields are highly localized near the higher-index layer. Total internal reflection causes very high

reflection on the surfaces of the higher-index layer and produces resonances similar to those in a Fabry-Perot cavity resonator. It should be noted that extremely sharp transmittance peaks can be obtained when  $\lambda$  is close to the gap thickness  $b$  (see the upper panel).

Further investigation on resonance tunneling is needed for the application in narrow band-pass filters. Due to the guided modes and the localized field, the magnitude of the evanescent wave may be amplified in the forward direction in some region (see Problem 10.11). Similar to the lateral shift by total internal reflection, due to the parallel energy flow in the high-index layer (waveguide), there must be a lateral shift of the transmitted light for finite beams. Little has been reported in the literature about the beam shift and the field distribution in dielectric multilayer structures, when resonance tunneling occurs.

### 10.1.6 Photon Tunneling with Negative Index Materials

Negative index materials (NIMs), for which the permittivity and the permeability become negative simultaneously in a given frequency region, can also be used to enhance photon tunneling.<sup>23</sup> The basics of NIMs has already been presented in Sec. 8.4.6. The structure is illustrated in Fig. 10.10a with a pair of layers in between two prisms. One of the layers has



**FIGURE 10.10** Photon tunneling with a layer of NIM. (a) The tunneling arrangement. (b) The field distribution in the middle layers for a TE wave.

a negative refractive index. Assume that one of the layers is vacuum and another has  $\epsilon = \mu = -1$ , so its refractive index is exactly  $-1$ . The transmittance becomes unity when the thickness of the NIM layer and that of the vacuum are the same, regardless of the angle of incidence and polarization. Let us use the full notation of  $\epsilon$  and  $\mu$  without using the refractive index. The transmission coefficient can be expressed as follows:<sup>23</sup>

$$t = \frac{8}{\xi_1 e^{-i\phi_1} + \xi_2 e^{i\phi_1} + \xi_3 e^{-i\phi_2} + \xi_4 e^{i\phi_2}} \quad (10.36)$$

Here, the phase angles  $\phi_1$  and  $\phi_2$  can be expressed as

$$\phi_1 = k_{2z} d_2 + k_{3z} d_3 \quad \text{and} \quad \phi_2 = k_{2z} d_2 - k_{3z} d_3 \quad (10.37)$$

where  $d_2$  and  $d_3$  are the thicknesses of layers 2 and 3, and  $k_{2z}$  and  $k_{3z}$  are the normal component of the wavevector in media 2 and 3, respectively. Note that when tunneling occurs,  $k_{2z}$  and  $k_{3z}$  become purely imaginary for the lossless case, as will be discussed later. For a TE wave, the coefficients in Eq. (10.36) are

$$\xi_1 = \left(1 + \frac{k_{2z}\mu_1}{k_{1z}\mu_2}\right) \left(1 + \frac{k_{3z}\mu_2}{k_{2z}\mu_3}\right) \left(1 + \frac{k_{4z}\mu_3}{k_{3z}\mu_4}\right) \quad (10.38a)$$

$$\xi_2 = \left(1 - \frac{k_{2z}\mu_1}{k_{1z}\mu_2}\right) \left(1 + \frac{k_{3z}\mu_2}{k_{2z}\mu_3}\right) \left(1 - \frac{k_{4z}\mu_3}{k_{3z}\mu_4}\right) \quad (10.38b)$$

$$\xi_3 = \left(1 + \frac{k_{2z}\mu_1}{k_{1z}\mu_2}\right) \left(1 - \frac{k_{3z}\mu_2}{k_{2z}\mu_3}\right) \left(1 - \frac{k_{4z}\mu_3}{k_{3z}\mu_4}\right) \quad (10.38c)$$

and

$$\xi_4 = \left(1 - \frac{k_{2z}\mu_1}{k_{1z}\mu_2}\right) \left(1 - \frac{k_{3z}\mu_2}{k_{2z}\mu_3}\right) \left(1 + \frac{k_{4z}\mu_3}{k_{3z}\mu_4}\right) \quad (10.38d)$$

For a TM wave, the transmission coefficient is defined based on the magnetic fields and the coefficients can be easily obtained by substituting  $\varepsilon$ 's for  $\mu$ 's in Eq. (10.38). The sign selection of  $k_{lz}$  was mentioned in Sec. 9.2.2 in the discussion of the matrix formulation. Basically, when there exist propagating waves in medium  $l$ ,  $k_{lz} = (2\pi n_l/\lambda) \sqrt{1 - (n_1/n_l)^2 \sin^2 \theta_1}$ , whose sign becomes negative in a NIM. On the other hand, if the waves become evanescent in medium  $l$ , we use  $k_{lz} = i(2\pi/\lambda) \sqrt{n_l^2 \sin^2 \theta_1 - n_1^2} = i\eta_l$ . Here,  $\eta_l$  is always positive in a lossless medium, even in a NIM. Assume that the prisms are made of the same materials so that properties of medium 1 and medium 4 are identical. Furthermore, layer 2 is made of a NIM with index-matching conditions, i.e.,  $\varepsilon_2 = -\varepsilon_3$  and  $\mu_2 = -\mu_3$  so that  $n_2 = -n_3$ . Eq. (10.36) can be further simplified. For propagating waves in the middle layers,  $k_{2z} = -k_{3z}$  and  $\xi_3 = \xi_4 = 0$ ; thus,

$$t = \frac{1}{\cos(k_{3z}\Delta) - iY \sin(k_{3z}\Delta)} \quad (10.39)$$

where  $\Delta = d_3 - d_2$ ,  $Y = \frac{1}{2}(k_{3z}\mu_1/k_{1z}\mu_3 + k_{1z}\mu_3/k_{3z}\mu_1)$  for TE waves, and  $Y = \frac{1}{2}(k_{3z}\varepsilon_1/k_{1z}\varepsilon_3 + k_{1z}\varepsilon_3/k_{3z}\varepsilon_1)$  for TM waves. Because media 1 and 4 are made of the same material, the transmittance for propagating waves can be written as follows:

$$T'_\lambda = \frac{1}{\cos^2(k_{3z}\Delta) + Y^2 \sin^2(k_{3z}\Delta)} \quad (10.40)$$

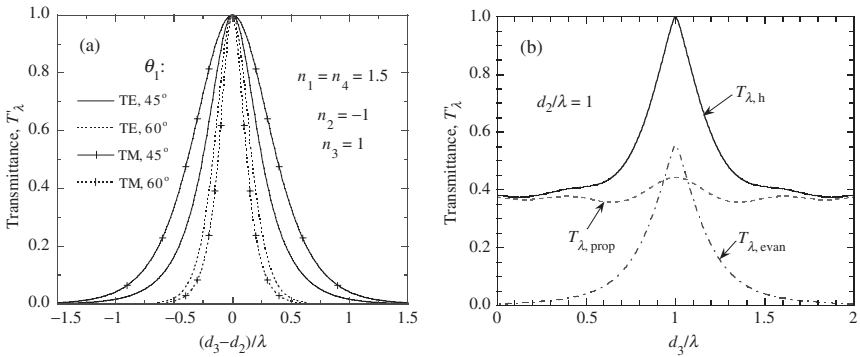
For evanescent waves, we have  $k_{2z} = k_{3z} = i\eta_3$ , where  $\eta_3 = (2\pi/\lambda) \sqrt{n_1^2 \sin^2 \theta_1 - n_3^2}$ . Now that  $\xi_1 = \xi_2 = 0$ , Eq. (10.36) can be simplified so that

$$t = \frac{1}{\cosh(\eta_3\Delta) + i \cot(\delta) \sinh(\eta_3\Delta)} \quad (10.41)$$

where  $\cot(\delta) = \frac{1}{2}(\eta_3\mu_1/k_{1z}\mu_3 - k_{1z}\mu_3/\eta_3\mu_1)$ , with  $\delta$  being the phase change upon total internal reflection from medium 1 and 2. The transmittance  $T'_\lambda = |t|^2$  is real and always decreases with increasing  $\Delta$ , the difference between the layer thicknesses. Although Eq. (10.39) and Eq. (10.41) are identical because  $\sin(ix) = i \sinh(x)$  and  $\cos(ix) = \cosh(x)$ , the use of real

variables allows us to observe the variation of transmittance with  $\Delta$  easily. When tunneling occurs, the field is highly localized near the interface between the NIM and the PIM layers, as shown in Fig. 10.10*b* for a TE wave, where the fields are sum of the forward-decaying and backward-decaying evanescent waves. The amplitude of the evanescent wave in the NIM increases in the direction of energy flow. It can be shown that the amplitude will still increase in medium 2, even though the NIM is placed in layer 3 and layer 2 is a vacuum. This corresponds to another resonance effect, which is associated with the excitation of surface electromagnetic waves or surface polaritons, to be discussed in the next section.

The directional and hemispherical transmittances for the structure shown in Fig. 10.10*a* are illustrated in Fig. 10.11 with the following parameters:  $n_1 = n_4 = 1.5$ ,  $n_2 = -1$



**FIGURE 10.11** Transmittance for a four-layer structure with one middle layer being matching-index NIM. (a) Directional transmittance. (b) Hemispherical transmittance.

( $\epsilon_2 = \mu_2 = -1$ ), and  $n_3 = 1$  (vacuum). Both the directional and hemispherical transmittances become 1 when  $d_3 = d_2$ . The hemispherical transmittance has two components due to propagating and evanescent waves. The effects of loss and dispersion have also been examined.<sup>24</sup>

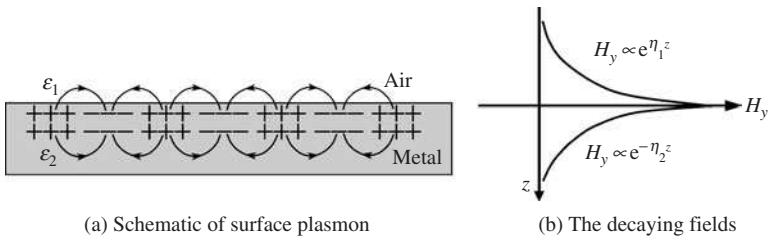
## 10.2 POLARITONS OR ELECTROMAGNETIC SURFACE WAVES

Surface plasmons, also known as *surface plasmon polaritons*, play an important role in near-field microscopy, nanophotonics, and biomolecular sensor applications.<sup>25–27</sup> Surface plasmon polaritons represent the interaction between electromagnetic waves and the oscillatory movement of free charges near the surface of metallic materials. When surface plasmons are confined to small structures, such as the tip of a scanning microscopic probe, quantum dots or nanoparticles, nanowires, or nanoapertures, they are referred to as *localized plasmons*. Surface plasmons usually occur in the electromagnetic wave spectrum in the visible or near-infrared region for highly conductive metals such as Ag, Al, and Au. In some polar dielectric materials, phonons or bound charges can also interact with the electromagnetic waves in the mid-infrared spectral region and cause resonance effects near the surface; these are called *surface phonon polaritons*, which have applications in tuning the thermal emission properties<sup>28</sup> and nanoscale nondestructive imaging.<sup>29</sup> In the following, the basic mechanisms of surface polaritons will be presented, with discussions on some

important applications. Emphasis is placed on the quantitative analysis of radiative properties for layered structures. In Sec. 10.4, the superlens concept will be introduced for imaging beyond the diffraction limit, and the energy streamline method will be presented for analyzing the energy propagation direction in the near-field regime.

### 10.2.1 Surface Plasmon and Phonon Polaritons

Plasmons are quasiparticles associated with oscillations of plasma, which is a collection of charged particles such as electrons in a metal or semiconductor. Plasmons are longitudinal excitations that can occur either in the bulk or at the interface. As shown in Fig. 10.12a, the



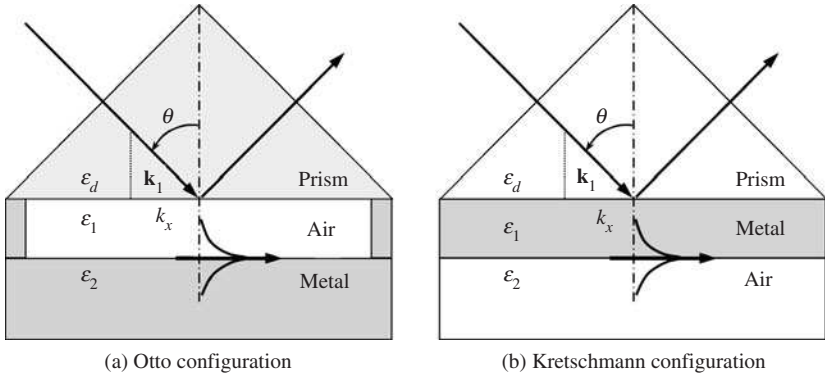
**FIGURE 10.12** Illustration of surface plasmon polariton. (a) Charge fluctuations and the magnetic field at the interface between a metal and air. (b) The exponentially decaying field amplitudes away from the interface.

charges oscillate along the surface, and such an excitation is called a *surface plasmon* or *surface plasmon polariton*. The field associated with a plasmon is localized at the surface, and the amplitude decays away from the interface, as shown in Fig. 10.12b. Such a wave propagates along the surface, and therefore, it is called a surface electromagnetic wave, similar to surface waves in fluids or the acoustic surface waves. Surface plasmons can be excited by electromagnetic waves and are important for the study of optical properties of metallic materials, especially near the plasma frequency, which usually lies in the ultraviolet. The requirement of evanescent waves on both sides of the interface prohibits the coupling of propagating waves in air to the surface plasmons. For this reason, surface waves are often regarded as nonradiative modes. The attenuated total reflectance (ATR) arrangements are commonly used to excite surface plasmons. When light is incident from the prism, it is possible for evanescent waves to occur simultaneously in the underneath metallic and air layers, as shown in Fig. 10.13, for the two typical configurations named after A. Otto (prism-air-metal) and E. Kretschmann and H. Raether (prism-metal-air). A detailed discussion with historical aspects can be found from Raether.<sup>30</sup>

In addition to the requirement of evanescent waves on both sides of the interface, the *polariton dispersion relations* must be satisfied. They are expressed as follows when both media extend to infinity in the  $z$  direction:

$$\frac{k_{1z}}{\epsilon_1} + \frac{k_{2z}}{\epsilon_2} = 0 \quad \text{for TM wave} \quad (10.42)$$

$$\frac{k_{1z}}{\mu_1} + \frac{k_{2z}}{\mu_2} = 0 \quad \text{for TE wave} \quad (10.43)$$



**FIGURE 10.13** Typical configurations for coupling electromagnetic waves with surface polaritons using attenuated total reflectance arrangements. (a) The Otto configuration (prism-air-metal). (b) The Kretschmann-Raether configuration (prism-metal-air). Note that a polar dielectric may substitute for the metal to excite surface phonon polaritons.

Let us consider lossless media first. In order for evanescent waves to occur, we must have  $k_{1z} = i\eta_1$  and  $k_{2z} = i\eta_2$  with  $\eta_1$  and  $\eta_2$  being positive, in order for the field  $e^{ik_x x - ik_{1z} z} = e^{ik_x x + \eta_1 z}$  to decay toward  $z = -\infty$  and  $e^{ik_x x + ik_{2z} z} = e^{ik_x x - \eta_2 z}$  to decay toward  $z = \infty$ . This means that the sign of permittivity must be opposite for media 1 and 2 in order to couple a surface polariton with a TM wave. On the other hand, we will need a magnetic material with negative permeability for a TE wave to be able to couple with a surface polariton. NIMs exhibit simultaneously negative permittivity and permeability in the same frequency region and are sometimes called double-negative (DNG) materials. Therefore, both TE and TM waves may excite surface plasmon polaritons with a NIM, as predicted by Ruppin.<sup>31</sup>

When compared with Fresnel's reflection coefficients, as can be seen from Eq. (10.2), the condition for the excitation of surface polaritons is that the denominator of the reflection coefficient be zero. A pole in the reflection coefficient is an indication of a resonance. Very often, the surface plasmon polariton is referred in the literature as a surface plasmon resonance. Taking a TM wave for example, since  $k_{1z}^2 = \mu_1 \epsilon_1 \omega^2 / c^2 - k_x^2$  and  $k_{2z}^2 = \mu_2 \epsilon_2 \omega^2 / c^2 - k_x^2$  from Eq. (10.1), we can solve Eq. (10.42) to obtain

$$k_x = \frac{\omega}{c} \sqrt{\frac{\mu_1 / \epsilon_1 - \mu_2 / \epsilon_2}{1 / \epsilon_1^2 - 1 / \epsilon_2^2}} \quad (10.44)$$

Equation (10.44) relates the frequency with the parallel component of the wavevector and is another form of the polariton dispersion relation. It should be noted that solutions of this equation are for both  $k_{1z} / \epsilon_1 + k_{2z} / \epsilon_2 = 0$  and  $k_{1z} / \epsilon_1 - k_{2z} / \epsilon_2 = 0$ , i.e., not only the poles but also the zeros of the Fresnel reflection coefficient are included. For nonmagnetic materials, Eq. (10.44) becomes

$$k_x = \frac{\omega}{c} \sqrt{\frac{\epsilon_1 \epsilon_2}{\epsilon_1 + \epsilon_2}} \quad (10.45)$$

One should bear in mind that the permittivities are in general functions of the frequency. For a metal with a negative real permittivity, the normal component of the wavevector is purely imaginary for any real  $k_x$  because  $\mu \epsilon \omega^2 / c^2 < 0$ . Thus, evanescent waves exist in metals regardless of the angle of incidence.

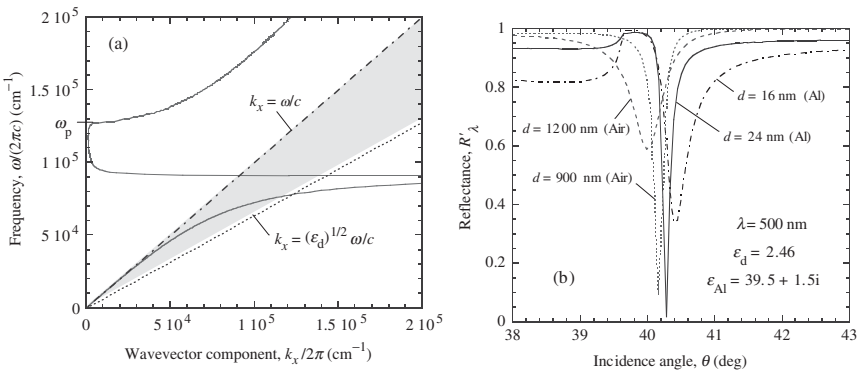
Consider either the Otto or Kretschmann-Raether configuration, and use the three-layer structure with a middle layer, medium 1, of thickness  $d$ . From Eq. (9.7), the reflection coefficient can be expressed as follows:

$$r = \frac{r_{01} + r_{12}e^{2i\beta}}{1 + r_{01}r_{12}e^{2i\beta}} = \frac{r_{01} + r_{12}e^{-2\eta_1 d}}{1 + r_{01}r_{12}e^{-2\eta_1 d}} \quad (10.46)$$

where the subscript 0 signifies the incidence medium, which is the prism, and  $\beta = k_{1z}d = i\eta_1 d$ . When  $d$  is sufficiently large,  $\exp(-2\eta_1 d) \ll 1$ , and the reflectance  $R'_\lambda = rr^* \approx r_{01}r_{01}^*$  is close to unity. When surface polaritons are excited, however,  $r_{12}$  increases dramatically and thus it is possible for  $r_{12}e^{-2\eta_1 d}$  to be of the same magnitude as  $r_{01}$ , but with an opposite phase, i.e., with a phase difference of  $\pi$ . At the condition of surface plasmon resonance, the reflectance  $R'_\lambda$  drops suddenly. Let us use an example to illustrate the polariton dispersion curves and the effect on the reflectance in ATR arrangements.

**Example 10-6.** Calculate the dispersion relation between Al and air. Calculate the reflectance versus angle of incidence for both the Otto and Kretschmann-Raether configurations at  $\lambda = 500$  nm, using Al as the metallic material. Determine the *polariton propagation length* at the wavelength  $\lambda = 500$  nm. Assume the prism is made of KBr with  $\epsilon_d = 2.46$  and the dielectric function of Al can be described by the Drude model.

**Solution.** The Drude model parameters for Al have been given in Example 8-6. Thus, we have  $\epsilon_2(\omega) = 1 - \omega_p^2/(\omega^2 + i\omega\gamma)$ , where the plasma frequency  $\omega_p = 2.4 \times 10^{16}$  rad/s and the scattering rate  $\gamma = 1.4 \times 10^{14}$  rad/s. One way to calculate the dispersion relation is to assume  $\omega$  is real and calculate  $k_x(\omega) = k'_x(\omega) + ik''_x(\omega)$ . The dispersion curves between Al and air ( $\epsilon_{\text{air}} = 1$ ) are usually plotted in a  $\omega$ - $k_x$  graph, for the real part of  $k_x$  shown in Fig. 10.14a by the solid line. At very low frequencies, the magnitude of  $\epsilon_2$  is so large that  $k_x \approx \omega/c$ . Note that the dash-dotted line with



**FIGURE 10.14** (a) The dispersion relation of surface plasmon polaritons between Al and air, where  $k_x$  is the real part solution of Eq. (10.45). (b) Reflectance in ATR arrangements, either with Al or air as the middle layer.

$k_x = \omega/c$  represents the *light line*. On the left of this line, there exist propagating waves in air; whereas on the right of the light line, evanescent waves occur in air because  $k_x > \omega/c$ . The light line can be considered as a wave travelling in air along the  $x$  direction. On the polariton dispersion curve,  $k_x$  increases quickly as  $\omega$  increases and reaches an asymptote at  $\omega = \omega_p/\sqrt{2}$ , when the real part of the dielectric function of Al approaches  $-1$ . Between  $\omega_p/\sqrt{2} < \omega < \omega_p$ , the real part of the dielectric function of Al becomes negative with an absolute value less than 1. Therefore, the solution of Eq. (10.45) has a large imaginary part, while the real part of  $k_x$  drops to near zero, as reflected by the

bending of the dispersion curve toward left and the steep rise upward. Beyond  $\omega > \omega_p$ , metal becomes transparent and the real part of the dielectric function becomes positive. Solutions beyond  $\omega > \omega_p$  correspond to zeros in the reflection coefficient and thus are not the solutions for Eq. (10.42), which are poles of the reflection coefficient. Notice that the dotted line refers to the light line of the prism. In the shaded region, there exist evanescent waves in air but propagating waves appear in the prism; as a result, surface plasmons can be coupled to propagating waves in the prism.

The reflectance is calculated from Eq. (10.46) at the wavelength  $\lambda = 500$  nm, corresponding to a wavenumber of  $20,000 \text{ cm}^{-1}$ . As can be seen from Fig. 10.14a, at this frequency, the surface polariton curve is very close to the light line in air. Therefore, the excitation of surface polariton is expected to be near the critical angle  $\theta_c \approx 39.6^\circ$  between the prism and air. The reflectance would be close to 1 at  $\theta > \theta_c$ . However, as shown in Fig. 10.14b, the reflectance drops suddenly around  $40^\circ$  due to the excitation of surface polaritons.

Furthermore, the reflectance dips are very sensitive to the thickness of the middle layer. In the Otto configuration, the air thickness of 900 nm yields a sharp dip. For the Kretschmann-Raether configuration, on the other hand, a metallic film thickness of 24 nm yields a sharp dip in the reflectance. If the Al film exceeds 50 nm, the reflectance is close to 1. The locations of the reflectance minimum and the width depend on the thickness of the middle layer.

When the surface plasmon polariton is excited, a large absorption occurs in the metal, which results in a coupling of the electromagnetic energy to a surface wave. The propagation length of the surface wave can be determined based on the imaginary part of  $k_x$ , i.e.,  $k_x''$ . Note that the field can be expressed as  $e^{ik_x x - k_x'' z}$  for surface waves propagating in the positive  $x$  direction and as  $e^{ik_x x + k_x'' z}$  for surface waves propagating in the negative  $x$  direction. The power is proportional to the square of the field amplitude, and the  $(1/e)$  power decaying length or the *polariton propagation length* is<sup>30</sup>

$$l_{\text{sp}} = 1/(2k_x'') \quad (10.47)$$

Plugging into the values in Eq. (10.45), we obtain  $l_{\text{sp}} \approx 80 \text{ } \mu\text{m}$ . Note that the Drude model somewhat underpredicts the imaginary part of the dielectric function. If  $\text{Im}(\epsilon)$  of Al were taken as 10 at  $\lambda = 500$  nm, one would obtain  $l_{\text{sp}} \approx 13 \text{ } \mu\text{m}$ , still much longer than the wavelength.

Another way to excite surface plasmon or phonon polaritons is by gratings. When light is incident onto a grating at a given  $k_x$ , the Bloch-Floquet condition given by Eq. (9.63a) in Sec. 9.4 states that the reflected and refracted waves can have different values of the parallel component of the wavevector:  $k_{x,j} = k_x + 2\pi j/\Lambda$ , where  $j$  is the diffraction order and  $\Lambda$  is the period of the gratings. For this reason, the dispersion relation can be folded into the region for  $k_x \leq \pi/\Lambda$  and surface polaritons can be excited on a grating surface. As an example, Fig. 10.15a shows the reduced dispersion relation for a binary grating made of Ag with

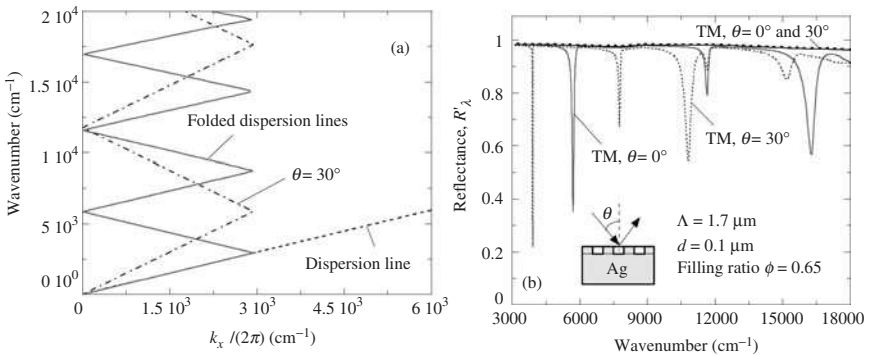


FIGURE 10.15 (a) Dispersion curves for gratings. (b) Reflectance for an Ag grating.



$\Lambda = 1.7 \mu\text{m}$ . The solid lines are the folded dispersion curves, and the dash-dotted lines, which are also folded, correspond to an incidence angle of  $30^\circ$ . The intersections identify the location where surface plasmons can be excited for a TM wave incidence, when the magnetic field is parallel to the grooves.

The reflectance of a shallow grating on Ag is calculated and plotted in Fig. 10.15b at  $\theta = 0^\circ$  and  $30^\circ$ . The grating height  $d = 100 \text{ nm}$ , and the filling ratio  $\phi = 0.65$  (see Fig. 9.18 for the grating geometry). For a TE wave, no dips exist in the reflectance because surface waves cannot be excited. The reflectance is very high for TE waves and has little difference between  $\theta = 0^\circ$  and  $\theta = 30^\circ$ . For a TM wave, the excitation of surface polaritons is responsible for the dips in the reflectance. Furthermore, the frequency locations agree well with those predicted by the dispersion curves. Note that at normal incidence, the excitation frequencies are located at the intersections between the dispersion curve and the vertical axis, as shown in Fig. 10.15a. These dips have also been known as Wood's or the Rayleigh-Wood anomalies, when a diffraction order just appears at the grazing angle; see Hessel and Oliner (*Appl. Opt.*, **4**, 1275, 1965). The actual resonance frequency may shift slightly from the frequency associated with the appearance or disappearance of a diffraction order, because the dispersion curve is not a straight line. The Rayleigh-Wood anomaly may also occur for gratings whose dielectric functions have a positive real part, i.e., not associated with surface plasmon polaritons.

It should be mentioned that many polar dielectric or semiconductor materials such as MgO, SiC, and GaAs contain a phonon absorption band, called the *reststrahlen band*, where  $\text{Re}(\epsilon)$  is negative and  $\text{Im}(\epsilon)$  is very small. The surface polariton condition described in Eq. (10.42) can be satisfied in the infrared, and the associated excitation or resonance is called a *surface phonon polariton*. In the following discussion of polaritons, the word "metal" is used to signify a material with a negative real permittivity or a negative- $\epsilon$  material.

Surface roughness is yet another way to excite surface waves because a rough surface can be considered as a Fourier expansion of multiple periodic components, each acting as a grating. Obviously, there is a large room to tune the radiative properties by surface polaritons with different geometries. The resonance behavior in nanoparticles or quantum dots has enormous applications in chemical sensing and medical diagnoses. Plasmon waveguide which is based on the resonance of nanoparticles, nanowires, and nanotips may allow electromagnetic energy transfer beyond the diffraction limit; see, for example, Maier et al. (*Nature Mater.*, **2**, 229, 2003), Dickson and Lyon (*J. Phys. Chem. B*, **104**, 6095, 2000), and Stockman (*Phys. Rev. Lett.*, **93**, 137404, 2004). Mie in 1908 developed the scattering formula to describe scattering from small absorbing particles, and expressed the *scattering coefficient* and the *absorption coefficient* in the limit of a small sphere, whose radius  $r_0$  is much smaller than the wavelength in vacuum  $\lambda$ , as

$$Q_{\text{sca},\lambda} = \frac{8}{3} \left( \frac{2\pi r_0}{\lambda} \right)^4 \frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + 2\epsilon_1} \left| \frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + 2\epsilon_1} \right|^2 \quad (10.48)$$

and

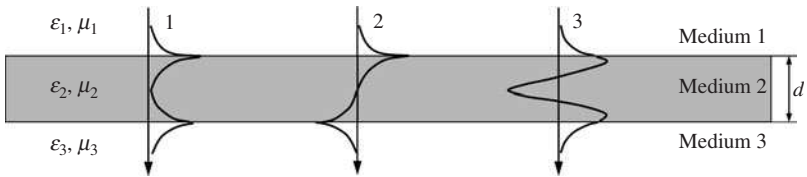
$$Q_{\text{abs},\lambda} = \frac{8\pi r_0}{\lambda} \sqrt{\epsilon_1} \text{Im} \left( \frac{\epsilon_2 - \epsilon_1}{\epsilon_2 + 2\epsilon_1} \right) \quad (10.49)$$

where  $\epsilon_1$  is the dielectric function of the surrounding dielectric medium and  $\epsilon_2$  is that of the absorbing sphere.<sup>32</sup> While Eq. (10.48) has the same form as the expression of Rayleigh scattering with the  $1/\lambda^4$  relationship of the *scattering cross section*, defined as  $4\pi r_0^2 Q_{\text{sca},\lambda}$ , the scattering of metallic spheres is distinctly different from that of dielectric spheres because the dielectric function of metals is complex and depends strongly on the wavelength. The scattering cross section is usually a very complex function of the wavelength. This is especially true when the resonance condition  $\epsilon_2 = -2\epsilon_1$  is satisfied. This resonance

is associated with the *localized surface plasmon polaritons*. Geometric optics completely failed to describe scattering and absorption of small particles. The scattering cross section can be much greater than the actual surface area. Furthermore, the absorbed energy can exceed that of a blackbody of the same size. In fact, the *blackbody* concept is misleading in the subwavelength regime. The actual resonance condition may be complicated for different geometries and coatings, as well as for clusters of particles or nanoparticle aggregates. Detailed discussion about resonance in metallic and polar dielectric materials in the absorption band can be found from Bohren and Huffman;<sup>32</sup> also see Yang et al. (*J. Chem. Phys.*, **102**, 869, 1995), Link et al. (*J. Phys. Chem. B*, **103**, 3073, 1999), Jin et al. (*Science*, **294**, 1901, 2001), and Kottmann et al. (*Phys. Rev. B*, **64**, 235402, 2001). Resonance phenomena in small particles have been applied to surface-enhanced Raman scattering microscopy and surface-enhanced fluorescence microscopy for single-molecule detection. The study of resonance phenomena in small particles continues to be an active research area because of the applications in biological imaging and molecular sensing; for details, refer to Moskovits (*Rev. Mod. Phys.*, **57**, 783, 1985), Chen et al. (*Nano Lett.*, **5**, 473, 2005), Johansson et al. (*Phys. Rev. B*, **72**, 035427, 2005), and Pustovit and Shahbazyan (*Phys. Rev. B*, **73**, 085408, 2006). Surface wave scattering has been used as a technique to characterize metallic nanoparticles.<sup>33</sup>

## 10.2.2 Coupled Surface Polaritons and Bulk Polaritons

Polaritons can exist on both surfaces of a thin film, resulting in a standing wave inside the film, as shown in Fig. 10.16. Economou performed a detailed investigation of different



**FIGURE 10.16** Illustration of polaritons in a slab. 1—symmetric mode coupled surface polaritons; 2—antisymmetric mode coupled surface polaritons; and 3—bulk polariton.

configurations of a thin-film structure;<sup>34</sup> while Kovacs and Scott (*Phys. Rev. B*, **16**, 1297, 1977) studied the optical excitation of surface plasma waves in layered structures. An essential requirement for coupled surface polaritons to occur is the existence of evanescent waves that decay in both media 1 and 3. Such a method was used in Sec. 10.1.2 for obtaining the mode equation for waveguides. A more convenient method to derive the polariton relations is to set the denominator of the reflection coefficient to zero. From Eq. (10.46), we can see that for the configuration shown in Fig. 10.16,  $r = (r_{12} + r_{23}e^{2ik_z d}) / (1 + r_{12}r_{23}e^{2ik_z d})$ , which has poles at  $1 + r_{12}r_{23}e^{2ik_z d} = 0$ . This can be expressed as follows:

$$\tanh(ik_z d) \left( \frac{k_{2z}^2}{\epsilon_2^2} + \frac{k_{1z}k_{3z}}{\epsilon_1\epsilon_3} \right) = \frac{k_{2z}}{\epsilon_2} \left( \frac{k_{1z}}{\epsilon_1} + \frac{k_{3z}}{\epsilon_3} \right) \quad (10.50)$$

which is the polariton dispersion relation for a slab sandwiched between two semi-infinite media. Because  $\tanh(ik_z d) = i \tan(k_z d)$ , Eq. (10.50) is identical to the mode equation of

a planar waveguide given in Eq. (10.16), when medium 3 is identical to medium 1. Attention should also be paid to the different meanings of the subscripts in Eq. (10.16) and Eq. (10.50). For the coupled surface polariton, however,  $k_{2z}$  is purely imaginary if loss is neglected. In the case of  $\varepsilon_3 = \varepsilon_1$  and  $\mu_3 = \mu_1$ , Eq. (10.50) can be rewritten into two equations:<sup>30</sup>

$$\frac{k_{1z}}{\varepsilon_1} + \frac{k_{2z}}{\varepsilon_2} \coth\left(\frac{k_{2z}d}{2i}\right) = 0 \quad (10.51a)$$

$$\frac{k_{1z}}{\varepsilon_1} + \frac{k_{2z}}{\varepsilon_2} \tanh\left(\frac{k_{2z}d}{2i}\right) = 0 \quad (10.51b)$$

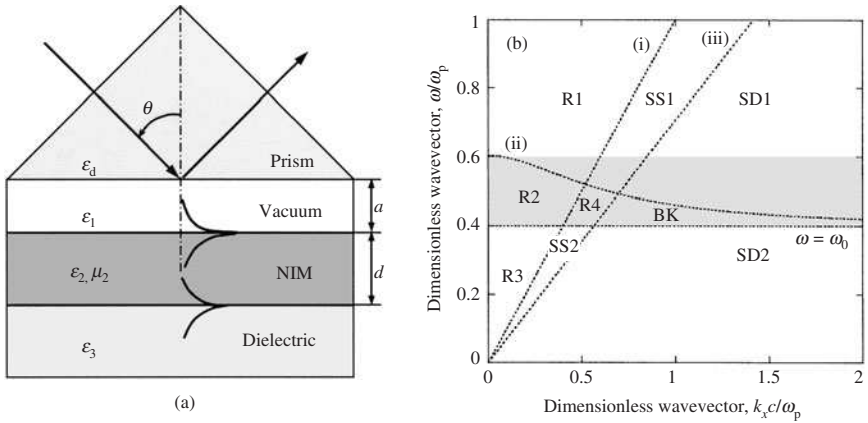
Each of them gives a dispersion curve, and the field distribution can be illustrated in Fig. 10.16 for case 1: a lower-frequency symmetric mode, where the surface charges are symmetric and the magnetic fields at the interfaces are in phase, and case 2: a higher-frequency antisymmetric mode, where the surface charges are asymmetric with respect to the middle plane and the magnetic fields at the interfaces are out of phase. Due to the coupling of surface waves, the wave inside medium 2 resembles a standing wave. It should also be noted that when  $d \rightarrow \infty$ , both Eq. (10.51a) and Eq. (10.51b) reduce to the surface polariton equation between two semi-infinite media. The discussion is also applicable to TE waves. The only change is to exchange  $\varepsilon$ 's and  $\mu$ 's in Fresnel's reflection coefficients and hence the dispersion relations. If medium 2 is a metal with a negative real permittivity ( $\varepsilon_2 < 0$ ) and media 1 and 3 are dielectric, evanescent waves must exist in the dielectric and coupled surface polaritons can interact only with TM waves for  $k_x > \max(\sqrt{\varepsilon_1}\omega/c, \sqrt{\varepsilon_3}\omega/c)$ . If medium 2 is a NIM ( $\varepsilon_2 < 0, \mu_2 < 0$ ), both TE and TM waves can excite coupled surface polaritons.

If a dielectric is placed as medium 2 between two metallic media 1 and 3, resonance is possible, even though  $k_{2z}$  is real, since  $k_{1z}$  and  $k_{3z}$  are imaginary in media 1 and 3. A standing wave is formed in medium 2, which is a guided mode discussed earlier. The guided mode is a kind of polariton, called *bulk polariton* that exists inside the material, i.e., the dielectric slab. The field distribution for a bulk polariton is illustrated in Fig. 10.16 as case 3. Both TE and TM waves can excite bulk polaritons, even at normal incidence when  $k_x = 0$ . Furthermore, several polariton modes may exist if the thickness  $d$  is large enough, corresponding to each order of the waveguide modes. Note that as in dielectric waveguides, the metal cladding can be replaced by a dielectric material of smaller refractive index. When evanescent waves exist inside media 1 and 3 and propagating waves exist in medium 2, only bulk polaritons can occur for both TE and TM waves but no surface polaritons exist.

Park et al. developed a regime for a NIM slab sandwiched between two different dielectrics, one of which is a vacuum, as shown in Fig. 10.17a.<sup>35</sup> The NIM is represented by the permittivity and permeability functions given in Eq. (8.121) and Eq. (8.122). For developing the dispersion curves, the damping terms can be assumed to be zero; therefore,

$$\varepsilon_2(\omega) = 1 - \frac{\omega_p^2}{\omega^2} \quad \text{and} \quad \mu_2(\omega) = 1 - \frac{F\omega^2}{\omega^2 - \omega_0^2} \quad (10.52)$$

Figure 10.17b represents the regimes with  $F = 0.56$  and  $\omega_0 = 0.4\omega_p$  shown in the  $\omega$ - $k_x$  graph, where both  $k_x$  and  $\omega$  are normalized with respect to  $\omega_p$ . Note that no polaritons can be excited for  $\omega > \omega_p$  because both  $\varepsilon_2$  and  $\mu_2$  are positive. In the shaded region for  $0.4\omega_p < \omega < 0.6\omega_p$ , both  $\varepsilon_2$  and  $\mu_2$  are negative, and this entails a NIM region. Four dotted lines (i), (ii), (iii), and  $\omega = \omega_0$  separate nine different regions. Lines (i), (ii), and (iii) correspond to  $\omega = k_x c / \sqrt{\varepsilon\mu}$  for media 1, 2, and 3, respectively. If the two dielectrics are identical, lines (i) and (iii) will merge and the regions in between will be eliminated. Notice that the condition for  $\omega = k_x c / \sqrt{\varepsilon\mu}$  corresponds to  $k_z = 0$  in any given medium. In the



**FIGURE 10.17** Illustration of polaritons in a NIM slab. (a) ATR arrangement. (b) Regimes of surface and bulk polaritons.<sup>35</sup>

regions on the left of line (i),  $k_x$  is too small to excite any evanescent waves in media 1 and 3; hence, no polaritons can exist in regions R1, R2, and R3. In regions between lines (i) and (iii), an evanescent wave appears in medium 1 whilst a propagating wave exists in medium 3. A surface polariton may exist only at the interface between media 1 and 2, and energy may be transmitted from the prism into medium 3. This arrangement is similar to the double-prism configuration if the dielectric is made of the same material as that of the prism. In regions on the right of line (iii), evanescent waves emerge in both media 1 and 3; hence, surface polaritons may exist at dual boundaries, and several bulk polaritons may also exist.

In the upper regions of line (ii), evanescent waves exist in the NIM layer. In the shaded area, surface polaritons may be observed in region SS1 at a single boundary and in region SD1 at dual boundaries of the NIM slab, for both polarizations. Surface polaritons may also exist in regions SS1 and SD1 above the shaded area only for TM waves. On the other hand, in regions between the lines  $\omega = \omega_0$  and (ii), propagating waves exist in the NIM layer because  $k_{z_2} > 0$ . Therefore, no polaritons may exist in region R4, whereas bulk polaritons can occur in region BK. Below the line  $\omega = \omega_0$ , medium 2 behaves like a normal metal because  $\epsilon_2 < 0$  and  $\mu_2 > 0$  surface polaritons may occur only for TM waves at a single boundary in region SS2 and both boundaries in region SD2.

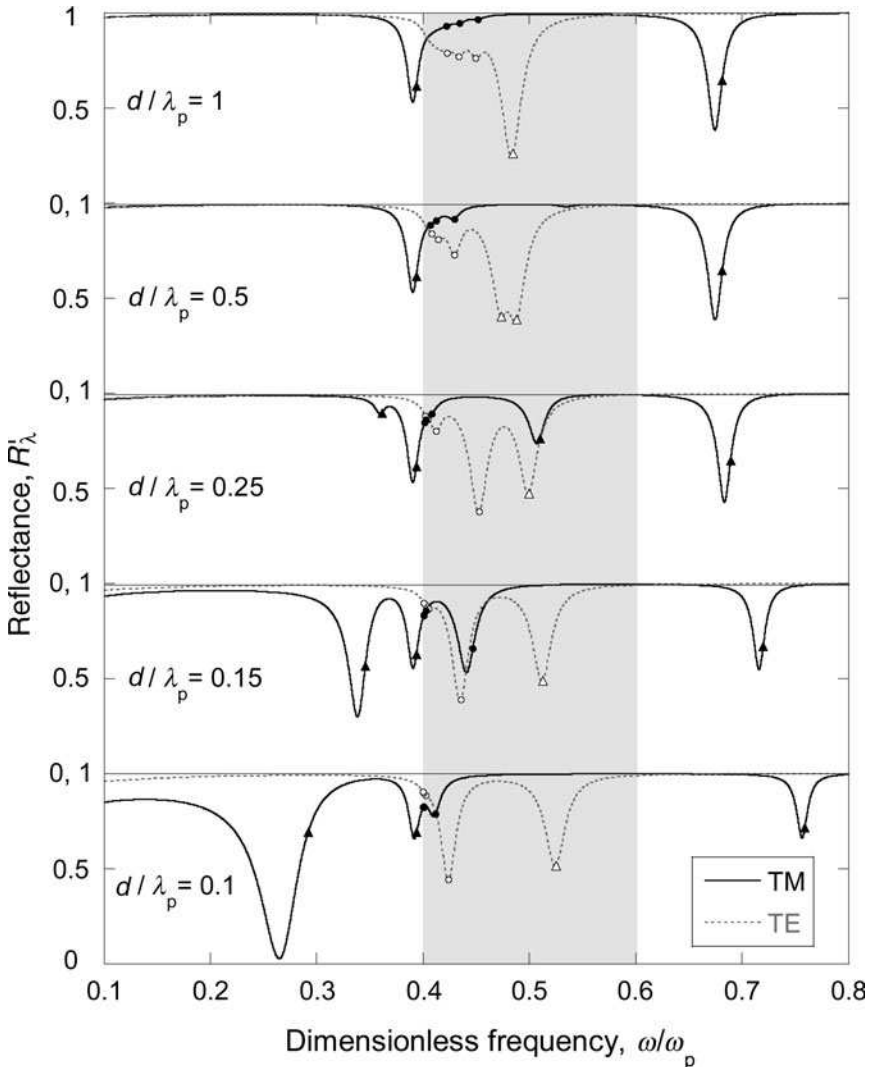
The reflection coefficient for the four-layer structure shown in Fig. 10.17a can be expressed as follows:

$$r = \frac{Y_{01}X_{12}X_{23}e^{-i\phi_1} + X_{01}X_{12}Y_{23}e^{i\phi_1} + Y_{01}Y_{12}Y_{23}e^{-i\phi_2} + X_{01}Y_{12}X_{23}e^{i\phi_2}}{X_{01}X_{12}X_{23}e^{-i\phi_1} + Y_{01}X_{12}Y_{23}e^{i\phi_1} + X_{01}Y_{12}Y_{23}e^{-i\phi_2} + Y_{01}Y_{12}X_{23}e^{i\phi_2}} \quad (10.53)$$

where  $X_{ij} = 1 + k_{jz}\epsilon_i/k_{iz}\epsilon_j$  and  $Y_{ij} = 1 - k_{jz}\epsilon_i/k_{iz}\epsilon_j$  for TM waves,  $X_{ij} = 1 + k_{jz}\mu_i/k_{iz}\mu_j$  and  $Y_{ij} = 1 - k_{jz}\mu_i/k_{iz}\mu_j$  for TE waves, and  $\phi_1 = k_{1z}a + k_{2z}d$  and  $\phi_2 = k_{1z}a - k_{2z}d$  are the phase terms. This analytical expression may be used as an alternate to the matrix formulation for the calculation of  $r$ , and subsequently, the reflectance  $R = rr^*$  of the four-layer structure. The denominator is exactly the same as that in Eq. (10.36) if the different notations are taken into consideration. Sometimes, it is more convenient to use subscript 0 for the first medium, while in other times, it is easier to use subscript 1 instead. Some precaution is necessary to identify the configuration with appropriate equations. Both Eq. (10.36)

and Eq. (10.53) are applicable for dissipative media, and can be combined to calculate the absorbance or emittance of the four-layer structure.

Figure 10.18 shows the calculated reflectance spectra for different NIM layer thicknesses, normalized to  $\lambda_p = 2\pi c/\omega_p$ , for both TM wave (solid curves) and TE wave (dotted curves). The permittivity and the permeability of the NIM are modeled with  $F = 0.56$ ,  $\omega_0 = 0.4\omega_p$ , and damping coefficients  $\gamma_e = \gamma_m = 0.012\omega_p$  using Eq. (8.121) and Eq. (8.122). The thickness of the vacuum layer is assumed to be  $a = 0.25\lambda_p$ . For the prism,  $\epsilon_d = 6$ , and for the dielectric,  $\epsilon_3 = 2$ . The incidence angle is set to  $\theta = 60^\circ$  so that only evanescent waves



**FIGURE 10.18** Reflectance of NIM slab in the ATR arrangement shown in Fig. 10.17a for both TM and TE waves at  $\theta = 60^\circ$ .<sup>35</sup>

exist in medium 3. The corresponding regions are SD1, BK, and SD2 in Fig. 10.17*b*. The shaded region corresponds to the frequencies where the refractive index of the NIM is negative. Several dips, due to surface and bulk polaritons, can be clearly seen in the reflectance spectra. Triangular and circular marks (filled for TM wave and unfilled for TE wave) represent surface and bulk polariton resonance frequencies that are obtained from the polariton dispersion relations in the lossless case. While damping terms affect the width of the dips, it is the vacuum gap distance  $a$  that affects the location of the reflectance dips strongly. For  $d/\lambda_p = 0.5$  and TE waves, there are three bulk polaritons in  $0.4 < \omega/\omega_p < 0.45$  and two surface polaritons in  $0.45 < \omega/\omega_p < 0.5$ . When the NIM layer thickness is reduced, the surface polariton of the lower frequency, in the pass band, is converted into a bulk polariton, while the other bulk polaritons are compressed to the vicinity of  $\omega_0$  and have little effect on the reflectance. The transition from a surface polariton to a bulk polariton occurs at  $d/\lambda_p$  between 0.25 and 0.5 for TE waves, and between 0.15 and 0.25 for TM waves. It is clear that both surface and bulk polaritons affect the radiative properties significantly. More detailed discussions can be found from Park et al.<sup>35</sup>

As mentioned earlier, waveguide modes are fundamentally the same as bulk polaritons with a standing wave inside the guided region that propagates along a fiber or a waveguide. Here, we have used a rather general definition of bulk polaritons. Consequently, many optical resonance phenomena can be explained with the unified theory of polaritons, including dielectrics, polar materials, metals, semiconductors, and photonic crystals. For dielectric waveguides, evanescent waves are required outside the guided region where the medium has a lower refractive index. The cladding can be made of metallic materials with a high conductivity to prevent any propagating waves from leaking outside the guided region. Similarly, a photonic crystal (PC) waveguide uses the evanescent wave in the forbidden band to guide the electromagnetic wave through holey fibers when the waveguide mode or bulk polariton condition is satisfied. A Fabry-Perot resonator cavity made of two metal films sandwiching a dielectric slab can also be explained by bulk polaritons. This structure is called metal-dielectric-metal configuration in which bulk polaritons can be excited even at normal incidence ( $k_x = 0$ ). One could argue that a thin slab in air will cause interference fringes and can be viewed as a type of Fabry-Perot resonator. The transmittance spectrum of a dielectric thin film oscillates solely because of interference of propagating waves without any evanescent wave. However, such a resonator usually does not exhibit sharp peaks. Another example of bulk polaritons is the whispering gallery mode in dielectric spheres or disks, as well as many 2-D and 3-D microcavities using dielectric or metallic PCs. These can be considered as 2-D or 3-D bulk polaritons similar to the 1-D bulk polaritons, whose resonance conditions are standing waves inside the cavity and evanescent waves outside the cavity. Additional examples of polariton-enhanced transmission, including resonance transmission and absorption, will be discussed next.

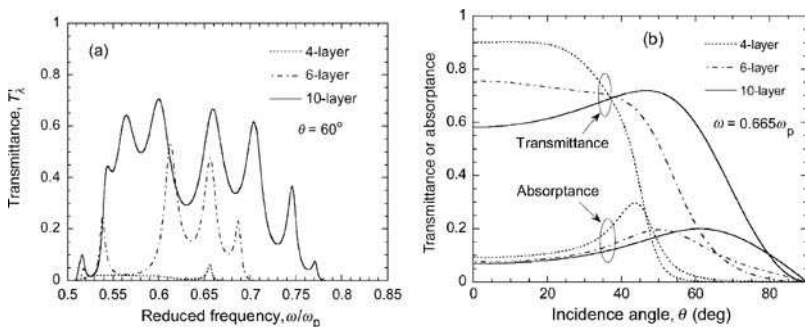
### 10.2.3 Polariton-Enhanced Transmission of Layered Structures

Surface polaritons, coupled surface polaritons, and bulk polaritons can enhance transmission in layered structures. Resonance tunneling through two or more barriers is an example of bulk polaritons because propagating waves exist in the denser dielectric medium (see Fig. 10.9 in Sec. 10.1.5). In this case, the parallel component of the wavevector  $k_x$  must lie between  $n_2\omega/c$  and  $n_1\omega/c$ . In Sec. 10.1.6, we discussed photon tunneling with a NIM layer. It can be seen that the surface plasmon polariton conditions given in Eq. (10.42) and Eq. (10.43) are always satisfied at the interface between vacuum and a medium with  $\epsilon = \mu = -1$ , or  $n = -1$ . When the phase shifts in the two layers cancel each other, complete tunneling occurs at incidence angles greater than the critical angle with a large vacuum gap distance as shown in Fig. 10.11. In reality, dispersion and dissipation cannot be avoided, and some examples of the transmittance through NIM layers will be given later in this section. Let us ask another question first:

Can polaritons enhance the transmittance of a metal film? The answer is positive, and there are different configurations for this to occur. Note that any discussion that is applicable to metals is also applicable to polar materials in the absorption band, in terms of polaritons, despite the different strengths, frequencies, widths, and mechanisms.

The prism-air-metal-prism structure can be used to excite surface polaritons at the interface between air and the metal to enable a larger tunneling transmittance for TM waves. Furthermore, air can be placed on both sides of the metal film to form a prism-air-metal-air-prism configuration that will enhance the transmittance by coupled surface polaritons. Of course, air can be replaced by a dielectric with a lower refractive index as discussed in Dragila et al. (*Phys. Rev. Lett.*, **55**, 1117, 1985). It is possible to achieve a sharp transmittance peak, even though the thickness of the metal exceeds the penetration depth (see Problems 10.19 and 10.20). Another configuration is possible by using a metal-dielectric-metal structure for exciting bulk polaritons without using a prism. The excitation of bulk polaritons can enhance the transmittance/absorption of the metal-dielectric-metal structure even in air for both polarizations, as well as at normal incidence; see Deych et al. (*Phys. Rev. E*, **57**, 7254, 1998), Villa et al. (*Phys. Rev. B*, **63**, 165103, 2001), and Wang (*Appl. Phys. Lett.*, **82**, 4385, 2003). Numerous studies have investigated the optical properties of metal-dielectric multilayers and semiconductor-semiconductor multiple quantum wells, where bulk polaritons dominate and enhance the transmittance; see for example, Scalora et al. (*J. Appl. Phys.*, **83**, 2377, 1998), Kee et al. (*J. Opt. A: Pure Appl. Opt.*, **6**, 22, 2004), Feng et al. (*Phys. Rev. B*, **72**, 085117, 2005), Schubert et al. (*Phys. Rev. B*, **71**, 035324, 2005), and Eremenchouk et al. (*Phys. Rev. B*, **71**, 235335, 2005). This field of research is multidisciplinary and continues to be of great interest to scientists and engineers in various fields. Two examples are given next for layers with NIMs and with a paired negative- $\epsilon$  and negative- $\mu$  composite.

Fu et al. calculated the tunneling transmittance of multilayer structures with alternating vacuum and NIM gaps.<sup>24</sup> The four-layer structure looks similar to the double-prism configuration, shown in Fig. 10.10, except that dispersion and loss are considered using the functional relations given in Eq. (8.121) and Eq. (8.122). Figure 10.19 shows the calculated transmittance with the following parameters:  $\omega_0 = 0.5\omega_p$ ,  $F = 0.785$ , and  $\gamma_e = \gamma_m = 0.0025\omega_p$ . For the 6-layer structure, there are 2 vacuum and 2 NIM layers between two dielectric prisms of  $\epsilon_d = 2.25$ , while there are 4 vacuum and 4 NIM layers in the middle for the 10-layer structure. The total thickness of the NIM is fixed to  $0.85\lambda_p = 1.7\pi c/\omega_p$ . The thickness of the vacuum layer is exactly the same as that of the NIM layer in the same setup. The transmittance spectra for a TE wave at an incidence angle of  $60^\circ$  are shown in Fig. 10.19a. The tunneling transmittance is greatly enhanced by reducing the individual layer thicknesses while maintaining the same total thickness. The



**FIGURE 10.19** Radiative properties of multilayer structures with NIMs for TE wave.<sup>24</sup> (a) Transmittance spectra at incidence angle  $\theta = 60^\circ$ . (b) Transmittance and absorbance as a function of incidence angle at  $\omega = 0.665\omega_p$ .

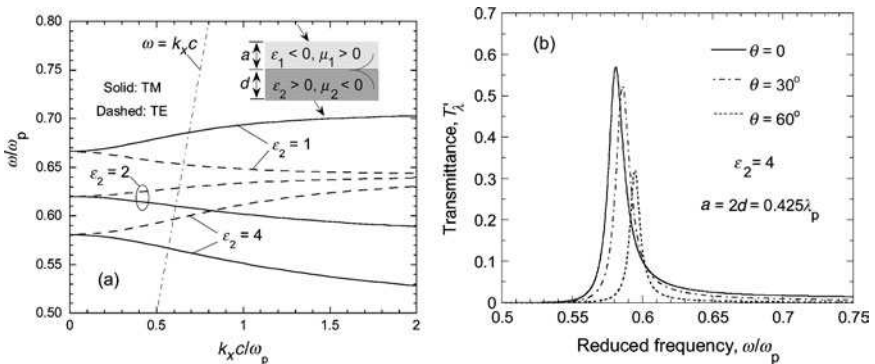
enhanced transmittance is believed to be caused by the coupled surface polaritons as well as bulk polaritons. Figure 10.19*b* illustrates the transmittance and the absorptance as functions of the incidence angle. Note that the critical angle between the prism and vacuum is  $41.8^\circ$ . While the transmittance is slightly reduced with the increased layers for propagating waves in vacuum, the tunneling transmittance is greatly enhanced. At large incidence angles, the absorptance also increases as the number of layers increases. Therefore, the enhanced transmittance is due to a reduction in the reflectance.

Some studies have dealt with a paired negative- $\varepsilon$  and negative- $\mu$  bilayer composite and demonstrated unique transmission and emission properties.<sup>36,37</sup> Consider the surface polaritons at the interface of such a structure without loss. Then,  $\varepsilon_1\varepsilon_2 < 0$  and  $\mu_1\mu_2 < 0$ ; furthermore,  $k_{1z}$  and  $k_{2z}$  are purely imaginary regardless of  $k_x$ . For simplicity, let us model the electric and magnetic properties of these two materials by

$$\varepsilon_1(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\gamma_e} \quad \text{and} \quad \mu_2 = 1 \quad (10.54)$$

and 
$$\varepsilon_2(\omega) = \varepsilon_2 \quad \text{and} \quad \mu_2(\omega) = 1 - \frac{F\omega^2}{\omega^2 - \omega_0^2 + i\omega\gamma_m} \quad (10.55)$$

where  $\varepsilon_2$  is real positive. The dispersion relations for  $\varepsilon_2 = 1, 2$ , and  $4$  are shown in Fig. 10.20*a*, for both polarizations, assuming  $\omega_0 = 0.5\omega_p$ ,  $F = 0.785$ , and  $\gamma_e = \gamma_m = 0$ . It can be seen that polaritons can be coupled with propagating waves in air, even at normal



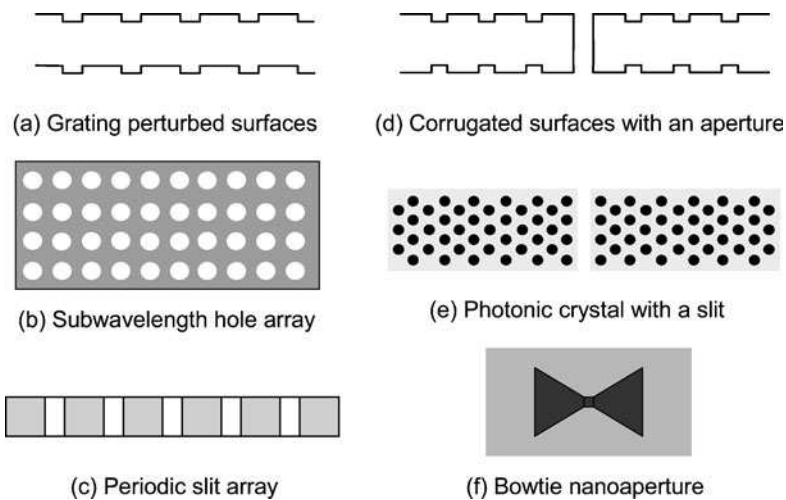
**FIGURE 10.20** Dispersion relations and transmittance of a paired negative- $\varepsilon$  and negative- $\mu$  composite in air. (a) Dispersion relations for both polarizations between two semi-infinite lossless media. (b) Spectral transmittance for a TE wave at different angles of incidence.

incidence. The transmittance for TE wave incidence for such a bilayer in air is shown in Fig. 10.20*b*, considering losses, using  $\varepsilon_2 = 4$  and  $\gamma_e = \gamma_m = 0.0025\omega_p$ . The thicknesses are assumed to be  $a = 2d = 0.425\lambda_p$ . Sharp transmission peaks occur near the surface polariton resonance frequency. Furthermore, if each individual layer is used, the transmittance is very small. The calculation of the transmittance for a TM wave is left as an exercise. Jiang et al. (*J. Appl. Phys.*, **98**, 013101, 2005) discussed the resonance transmission of a PC, made of alternating layers of negative- $\varepsilon$  and negative- $\mu$  materials, for potential application of high- $Q$  filters. The application of paired single-negative (SNG) materials for coherent emission will be discussed in Sec. 10.3.2.



### 10.2.4 Radiation Transmission through Nanostructures

The cross coupling of surface plasmon polaritons between corrugated metal films has been studied since the 1970s and employed to enhance light emission from tunnel junctions, light-emitting diode, and organic photoluminescence; see for example, Pockrand (*Opt. Commun.*, **13**, 311, 1975), Theis et al. (*Phys. Rev. Lett.*, **50**, 750, 1983), Inagaki et al. (*Phys. Rev. B*, **32**, 6238, 1985), Köck et al. (*Appl. Phys. Lett.*, **57**, 2327, 1990), Gifford and Hall (*Appl. Phys. Lett.*, **81**, 4315, 2002), and Wedge et al. (*Phys. Rev. B*, **69**, 245418, 2004). The coupled surface polaritons enable the coherent transmission of light through a narrow wavelength region in well-defined directions and thus enhance the light emission characteristics. A schematic of corrugated or grating-perturbed surfaces is shown in Fig. 10.21a, along with



**FIGURE 10.21** Various structures for transmission enhancement. (a) Grating or periodically perturbed surfaces for cross coupling of surface plasmons. (b) Subwavelength hole array. (c) 1-D periodic slit array in a metal or polar material. (d) Corrugated metallic surfaces with an aperture for directional transmission. (e) Photonic bandgap structure for beaming light. (f) Bowtie nanoaperture for near-field focusing and transmission enhancement.

some structures that have been studied intensively in recent years for the control of light transmission through nanostructures. A complete discussion of light-matter interactions in these structures is beyond the scope of this text. Therefore, only a brief review is provided here so that interested readers can find the relevant literature for further study.

The publication of enhanced transmission of metallic films perforated with subwavelength holes by Ebbesen et al. (*Nature*, **391**, 667, 1998) has raised the interest of studying transmission of light through nanostructures, including 2-D hole arrays and 1-D slit arrays, as shown in Figs. 10.21b and c, as well as annular aperture arrays. Coupled and localized surface polaritons and Fabry-Perot-type resonances are believed to be responsible for the enhancement; see Porto et al. (*Phys. Rev. Lett.*, **83**, 2845, 1999), Liu and Tsai (*Phys. Rev. B*, **65**, 155423, 2002), García-Vidal and Martín-Moreno (*Phys. Rev. B*, **66**, 155412, 2002), Marquier et al. (*Opt. Lett.*, **29**, 2178, 2004), Martín-Moreno et al. (*Phys. Rev. Lett.*, **86**, 1114, 2001), and Fan et al. (*Phys. Rev. Lett.*, **94**, 033902, 2005). It should not be surprising, though, that the location of surface polaritons may not correspond well with the actually

observed resonance behavior of the radiative properties of nanostructures, because polariton relations are obtained for infinitely extended media. Lezec and Thio (*Opt. Express*, **12**, 3629, 2004) revisited earlier measurement results and theories by comparing the enhancement and suppression of perforated metallic-type films with those of dielectric-type films; also see Thio in the January-February 2006 issue of *American Scientist* (p. 40) for some historical events that lead them to rethink the experimental and theoretical explanations. It appears that a full consideration of diffracted evanescent waves is necessary in order to explain the resonance frequency. Furthermore, the interactions of periodic 2-D and 3-D structures are similar to those of periodic layered structures, where cavity resonances and bulk polaritons may play a significant role. We may refer to these types of interactions as generalized bulk polaritons or cavity modes. More discussion will be given in Sec. 10.3.1. Laroche et al. (*Phys. Rev. B*, **71**, 155113, 2005) demonstrated resonance transmission through a 2-D PC in the forbidden band. Chan et al. (*Opt. Lett.*, **31**, 516, 2005) theoretically and experimentally studied the optical transmission through double-layer metallic sub-wavelength slit arrays, and revealed resonance transmission through coupling of evanescent fields.

Another type of the enhanced transmission configuration is an aperture in corrugated surfaces, as shown in Fig. 10.21*d* for a metallic film and Fig. 10.21*e* that uses the bandgap of PCs to beam the light. The corrugated surface serves as a funnel to guide the light into the aperture. In either case, the transmitted light becomes highly directional and the transmittance spectra exhibit sharp peaks; see Lezec et al. (*Science*, **297**, 820, 2002), Moreno et al. (*Phys. Rev. B*, **69**, 121402, 2004), Kramper et al. (*Phys. Rev. Lett.*, **92**, 113903, 2004), and Lockyear et al. (*Appl. Phys. Lett.*, **84**, 2040, 2004). These structures may be considered as periodic slits with an infinite period or distance of separation. Circularly corrugated surfaces have also been used to funnel light through a subwavelength aperture. Surface plasmon-mediated transmission through single nanoholes and double slits without corrugated surfaces has also been studied; see Yin et al. (*Appl. Phys. Lett.*, **85**, 467, 2004), Bravo-Abad et al. (*Phys. Rev. E*, **69**, 026601, 2004), Popov et al. (*Appl. Opt.*, **44**, 2332, 2005), and Schouten et al. (*Phys. Rev. Lett.*, **94**, 053901, 2005).

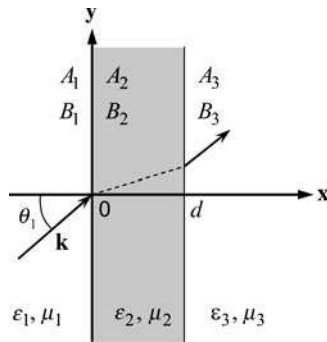
Light transmission through single nanoapertures of different shapes has been of great interest to nanolithography. The bowtie shape is illustrated in Fig. 10.21*f*. Such a nanoaperture behaves as an antenna to collect light and focus it in the near field, especially with coupled surface plasmon polaritons; see Shi et al. (*Opt. Lett.*, **28**, 1320, 2003), Matteo et al. (*Appl. Phys. Lett.*, **85**, 648, 2004), Jin and Xu (*Appl. Phys. Lett.*, **86**, 111106, 2003), Jin and Xu (*J. Quant. Spectrosc. Radiat. Transfer*, **93**, 163, 2005), Sundaramurthy et al. (*Nano Lett.*, **6**, 355, 2006), and Wang et al. (*Nano Lett.*, **6**, 361, 2006). Another type of transmission enhancement and focusing in the near field is the use of self-assembled monolayer of nanosphere arrays for lithography; see Osamu et al. (*Appl. Phys. Lett.*, **79**, 1366, 2001), Lu et al. (*Appl. Phys. Lett.*, **82**, 4143, 2003), and Li et al. (*Nanotechnology*, **15**, 333, 2004).

Numerical computations are inevitable due to the complexity of the nanostructures and the electric and magnetic properties. For simple grating structures, the rigorous coupled-wave analysis (RCWA) discussed in Sec. 9.4.1 is an effective tool for the study of the field distribution and radiative properties, especially for 1-D gratings, and has also been extended to 2-D structures. The analysis of the band structures of 1-D PCs is rather straightforward based on the 1-D matrix formulation. In some cases, a truncated 2-D PC can be viewed as a multilayered 1-D gratings. The transfer matrix method (TMM) is the most commonly used technique for calculating the dispersion relations or band structures of 2-D and 3-D PCs, as well as the transmittance and reflectance.<sup>38</sup> On the other hand, finite-difference time-domain (FDTD) is a commonly used technique for the study of radiative properties of nanostructures. FDTD is a central difference scheme in both time and space domains with the second-order accuracy.<sup>39</sup> Finite-element method (FEM) and boundary-element method (BEM) are other common numerical techniques used for diffraction optical devices.<sup>40,41</sup>

### 10.2.5 Superlens for Perfect Imaging and the Energy Streamlines

As discussed in Chap. 8, Sec. 8.4.6, a NIM or a double-negative material (DNG) forms a *flat lens* that can focus light (see Fig. 8.18). Pendry (*Phys. Rev. Lett.*, **85**, 3966, 2000) predicted that a DNG flat lens not only focuses the propagating waves but also allows complete transmission of evanescent waves because of an amplifying effect of the evanescent wave amplitude. Furthermore, a single-negative material (SNG) like a Ag film also exhibits focusing properties in the closest proximity. Such a lens is thereafter called a *perfect lens* or *superlens*. Many researchers are working on the fabrication of micro/nanostructures with tailored electric and magnetic properties. Photonic crystals have also been realized with focusing properties for electromagnetic waves (photons); see Luo et al. (*Appl. Phys. Lett.*, **81**, 2352, 2002) and Lu et al. (*Phys. Rev. Lett.*, **95**, 153901, 2005); as well as for acoustic waves (phonons); see Yang et al. (*Phys. Rev. Lett.*, **93**, 024301, 2004) and Li et al. (*Phys. Rev. B*, **73**, 054302, 2006). Researchers have also experimentally demonstrated that flat Ag lens can focus light at nanoscale distances for nanolithographic applications.<sup>42,43</sup>

While the electromagnetic wave theory describes the tunneling phenomenon and surface polaritons elegantly, the energy ray concept meets a difficulty for coupled evanescent waves because the parallel component of the wavevector for an evanescent wave is so large that no polar angle within the real space can be defined. On the other hand, the Poynting vector can always be defined, and by following the traces, the energy streamline method appears to be a promising technique for analyzing the energy flow directions in the near field. The basic concept developed in a recent study by Zhang and Lee is described next.<sup>44</sup> For convenience, let us consider the layered medium to be oriented along the  $x$  direction as shown in Fig. 10.22, where media 1 and 3 are semi-infinite. If the incident wave is a TM wave with an angular frequency  $\omega$ , the magnetic field in each region is given by



**FIGURE 10.22** Schematic of a three-layer structure, where  $A_j$  and  $B_j$  ( $j = 1, 2,$  and  $3$ ) are the coefficients of forward and backward waves at the nearest interface.

$$H_z(x,y) = \left[ A(x)e^{ik_x x} + B(x)e^{-ik_x x} \right] e^{ik_y y} \quad (10.56)$$

Here,  $A$  and  $B$  are the coefficients of forward and backward waves at the interface;  $x$  is relative to the origin in media 1 and 2, while in medium 3,  $x$  is relative to  $d$ ; and  $k_x$  and  $k_y$  are the  $x$  (normal) and  $y$  (parallel) components of the wavevector. Note that  $k_x^2 + k_y^2 = \epsilon\mu\omega^2/c^2$  for this geometry. The components of the time-averaged Poynting vector can be expressed as follows:

$$\langle S_x \rangle = \frac{1}{2\omega\epsilon_0} \operatorname{Re} \left( \frac{k_x}{\epsilon} \right) \left[ |A|^2 e^{-2k_x' x} - |B|^2 e^{2k_x' x} \right] - \frac{1}{\omega\epsilon_0} \operatorname{Im} \left( \frac{k_x}{\epsilon} \right) \operatorname{Im} \left( AB^* e^{2ik_x' x} \right) \quad (10.57)$$

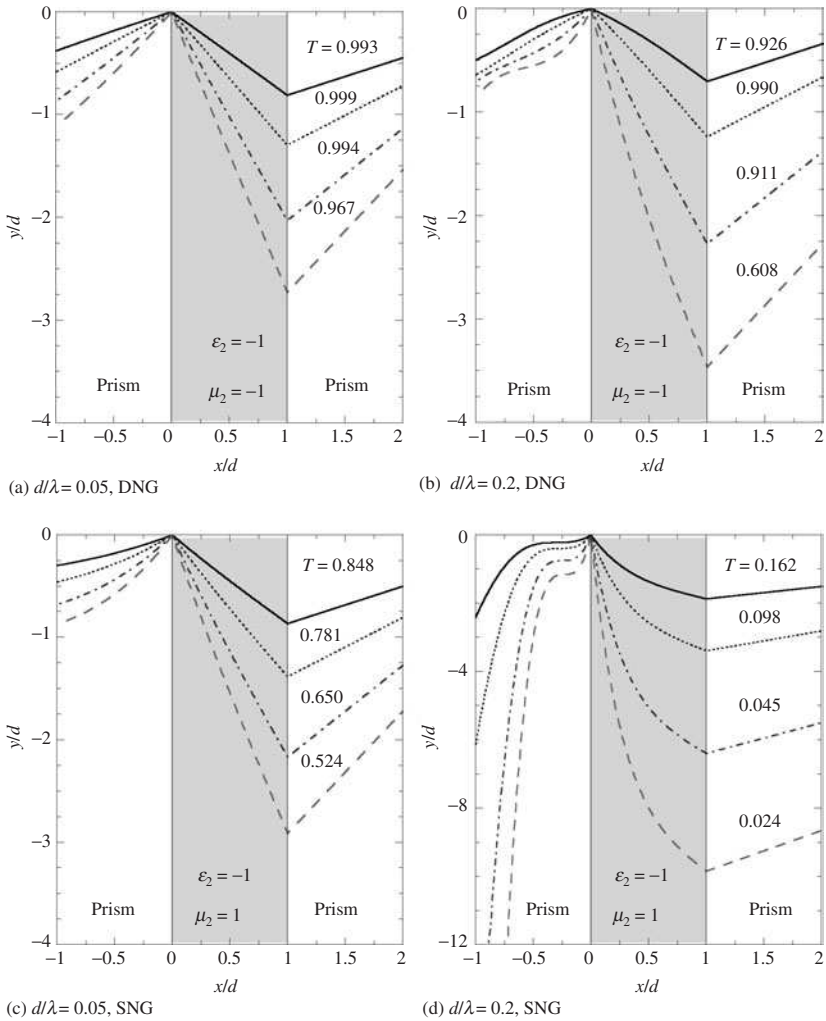
$$\langle S_y \rangle = \frac{k_y}{2\omega\epsilon_0} \operatorname{Re} \left( \frac{1}{\epsilon} \right) \left[ |A|^2 e^{-2k_x' x} + |B|^2 e^{2k_x' x} \right] + \frac{k_y}{\omega\epsilon_0} \operatorname{Re} \left( \frac{1}{\epsilon} \right) \operatorname{Re} \left( AB^* e^{2ik_x' x} \right) \quad (10.58)$$

Here,  $k_x = k_x' + ik_x''$  is the normal component of the wavevector. Note that the present section uses a slightly different notation from that of preceding sections. The last terms in Eq. (10.57) and Eq. (10.58) arise from the coupling between the forward and backward waves. The direction of  $\langle \mathbf{S} \rangle$  of the combined wave can always be defined by a polar angle  $\phi = \arctan(\langle S_y \rangle / \langle S_x \rangle)$ ; in contrast, it is not always possible to define the angle of incidence or refraction  $\theta = \arctan(k_y / k_x)$  in the real space. The trajectory of  $\langle \mathbf{S} \rangle$  for given  $\omega$  and  $k_y$  is an energy streamline, which defines the path of the net energy flow. The matrix formulation can be used to evaluate  $A$  and  $B$  in each layer by setting  $A_1 = 1$  and  $B_3 = 0$ . Note that the dependence of  $\mu$  is implicit in Eq. (10.57) and Eq. (10.58), since  $k_x$  is a function of  $\mu$ , and furthermore,  $A_j$  and  $B_j$  depend on  $k_x$ 's. For TE waves, the magnetic field can be replaced by the electric field in Eq. (10.56), and  $\epsilon$  should be replaced by  $\mu$  in Eq. (10.57) and Eq. (10.58).

The energy streamlines in the prism-DNG-prism and prism-SNG-prism configurations are shown in Fig. 10.23, for different incidence angles or  $k_y$ 's. The energy transport is from the left to the right, and the trajectory of the Poynting vector in the three regions forms a zigzag path, especially when  $d \ll \lambda$ . The  $x$ - and  $y$ -axes are normalized to the slab thickness  $d$ . For simplicity, let us use ZLs to abbreviate these zigzag energy streamlines. All ZLs are for positive  $k_y$  values and pass through the origin. With the dielectric prism ( $\epsilon = 2.25$ ), the critical angle is  $\theta_c = 41.81^\circ$ .

Causality requires that  $\langle S_x \rangle$  be positive; furthermore, when loss is neglected,  $\langle S_x \rangle$  is independent of  $x$ . Note that  $\langle S_y \rangle = (k_y / 2\omega\epsilon_0) \operatorname{Re}(1/\epsilon) |H_z|^2$  is opposite to  $k_y$  when  $\operatorname{Re}(\epsilon) < 0$ , as in the DNG layer (medium 2). At  $\theta_1 = \theta_c$ , when the phase refraction angle  $\theta_2 = 90^\circ$ , the energy refraction angle  $\phi_2$  is much less than  $90^\circ$ . In order to remove the singularity, the computation for  $\theta_1 = \theta_c$  can be approximated by using an angle that is either slightly greater or slightly smaller than  $\theta_c$ . Furthermore, the dash-dotted line in the slab separates the propagating-wave ZLs (inside the cone) from the evanescent-wave ZLs (outside the cone). The observation that the energy paths of propagating and evanescent waves are separated by a cone provides a new explanation of the photon tunneling phenomenon based on wave optics. Note that the tunneling phenomena have been extensively studied in quantum mechanics on the time delay and beam shift, but the results are somewhat controversial.<sup>21</sup> The energy transmittance through the slab, calculated by  $T = |A_3/A_1|^2$ , is labeled for each ZL. The tunneling transmittance decreases rapidly as  $d$  increases, and the ZLs are curved when  $d = \lambda/5$ . Figures 10.23c and d are for a negative- $\epsilon$  but positive- $\mu$  slab (such as a metal, but lossless). In this case, only evanescent waves exist in the slab because  $k_x$  is purely imaginary even at normal incidence. Energy is carried through medium 2 by coupled evanescent waves, whose path can be completely described by a ZL. The transmittance with a SNG slab is much smaller than that with a DNG slab, and the beam shift in the  $y$  direction becomes very large, as illustrated in Fig. 10.23d. Nevertheless, Fig. 10.23a and Fig. 10.23c look alike. When  $d \ll \lambda$ , the propagating waves and evanescent waves are similar because both the sinusoidal and hyperbolic functions are the same under the small-argument approximation.<sup>36</sup> Assume that only propagating waves exist in medium 1, and both  $\epsilon_1$  and  $\epsilon_2$  are real. The following approximation can be obtained for the energy incidence and refraction angles in the limit  $d/\lambda \rightarrow 0$ :

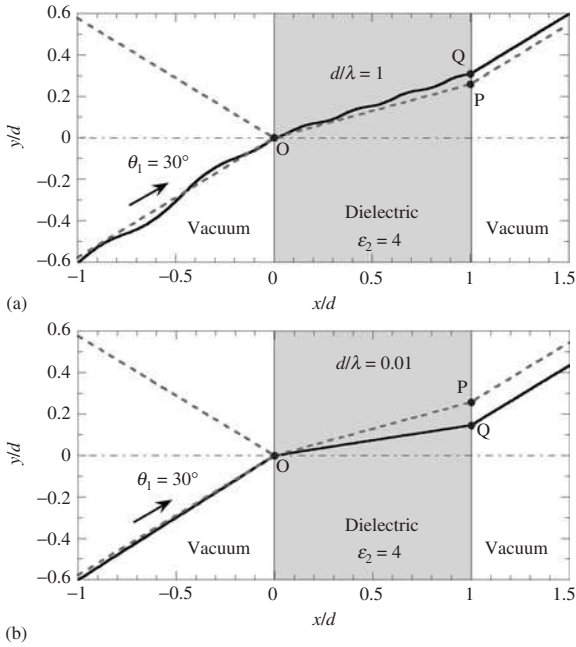
$$\phi_1 = \theta_1 \quad \text{and} \quad \tan \phi_2 = (\epsilon_1/\epsilon_2) \tan \phi_1 \quad (10.59)$$



**FIGURE 10.23** ZLs for prism-DNG-prism and prism-SNG-prism configurations at various incidence angles:<sup>44</sup>  $\theta_1 = 20^\circ$  (solid),  $30^\circ$  (dotted),  $41.81^\circ$  (dash-dotted), and  $50^\circ$  (dashed). The prism has  $\varepsilon = 2.25$  and  $\mu = 1$ , so  $\theta_1 = 41.81^\circ$  corresponds to the critical angle for DNG in (a) and (b). Only evanescent waves exist in medium 2 for SNG in (c) and (d). The transmittance  $T$  from medium 1 to 3 is shown for each incidence angle.

Note that  $\mu_2$  does not affect the TM wave results in the *electrostatic limit*, when the distance is much shorter than the wavelength. However, the effect of  $\mu_2$  becomes significant when  $d/\lambda > 0.1$ .

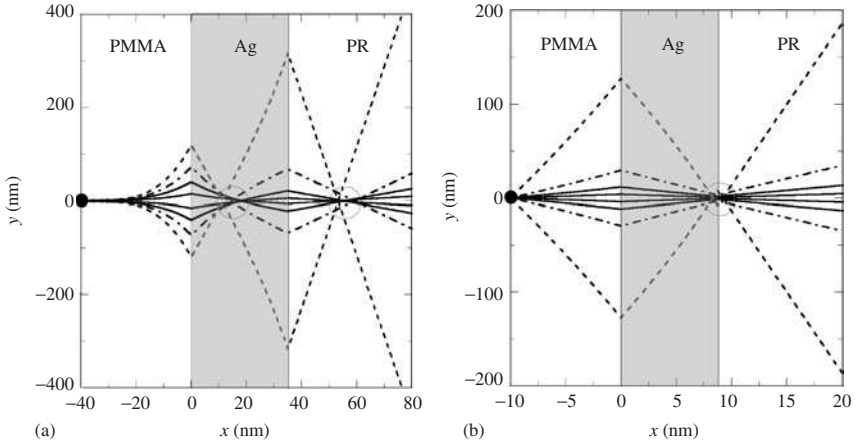
Both positive and negative phase-time shifts were noticed by Li (*Phys. Rev. Lett.*, **91**, 133903, 2003) for an optically dense dielectric slab in air without evanescent waves. It is worthwhile to take a look at the ZLs for the vacuum-dielectric-vacuum configuration. For propagating waves, because the second term in Eq. (10.58) depends on  $x$ , the ZL exhibits wavelike features for  $d = \lambda$ , as can be seen from Fig. 10.24a, where the solid curve is the



**FIGURE 10.24** The ZL for vacuum-dielectric-vacuum configuration at  $\theta_1 = 30^\circ$  when (a)  $d/\lambda = 1$  and (b)  $d/\lambda = 0.01$ .<sup>44</sup> Solid curves are ZLs, and dashed lines are the wavevector direction.

ZL and the dashed lines are the traces of the wavevector. The lateral shift of the energy line is determined by the point  $Q$  rather than  $P$ . When  $d/\lambda$  is reduced to 0.01 as shown in Fig. 10.24b, the ZL is almost a straight line in each medium. However, point  $Q$  becomes closer to the  $x$ -axis than  $P$ , contrary to Fig. 10.24a. When  $d/\lambda \ll 1$ , Snell's law determines  $\theta_2$  and Eq. (10.59) determines  $\phi_2$ . The shift of  $Q$  with respect to  $P$  depends on the incidence angle, which can be positive or negative. Perhaps the lateral shift of the energy path can be understood by the energy flow parallel to the film as a result of the combined field, similar to the Goos-Hänchen shift. The difference here is due to the fact that a plane wave of infinite width is used to calculate the lateral shift of transmission through a thin film, as well as tunneling. While Poynting vector traces have been presented for transmission through nanoslits as well as for scattering around nanoparticles, the application of the streamline method to planar layers reveals some fundamental and counterintuitive behavior; see Bohren and Huffman,<sup>32</sup> and Bashevov et al. (*Opt. Express*, **13**, 8372, 2005). It appears to be more natural for the thermal engineers to deal with energy streamlines rather than evanescent waves. This method allows the visualization of energy flow in the optical near field.

Understanding the energy transport in the subwavelength region has an enormous impact on near-field optics and nanolithography. Figure 10.25 shows the ZLs for the three-layer structure made by Fang et al.<sup>43</sup> A 35-nm-thick Ag film was evaporated over a poly-methyl methacrylate (PMMA) followed by a photoresist (PR) coating. The source is assumed to be at  $x = -40$  nm and  $y = 0$  inside the PMMA. The properties are taken from Fang et al., and accordingly,  $\epsilon_1 = 2.30 + 0.0014i$  for the PMMA,  $\epsilon_2 = -2.40 + 0.25i$  for Ag, and  $\epsilon_3 = 2.59 + 0.01i$  for the PR. The solid lines are for propagating waves in the



**FIGURE 10.25** ZLs for a three-layer structure, showing the imaging features for a silver lens with a thickness of (a)  $d = 35$  nm and (b)  $d = 8.75$  nm, at the wavelength  $\lambda = 365$  nm. The dot represents the source, and circles are for foci.

PMMA at  $\theta_1 = 20^\circ$  and  $50^\circ$ . The dash-dotted lines correspond to  $\theta_1 = 90^\circ$  or  $k_y = \text{Re}(k_1) = 2\pi\text{Re}(\sqrt{\epsilon_1})/\lambda$ , where  $\epsilon_1$  is the dielectric function of the PMMA. Outside the cone, defined by the dash-dotted lines with  $k_y = 1.06\text{Re}(k_1)$ , evanescent waves exist inside the PMMA. Note that evanescent waves exist in vacuum for  $\theta_1 > 41.25^\circ$ . The use of PMMA allows evanescent waves of the light source with  $k_y$  much greater than  $\omega/c$  to be transmitted through. In the calculations, both the PMMA and the PR are assumed semi-infinite, and this assumption should have little effect on the imaging properties.

The ZLs shown in Fig. 10.25 are curved (i.e.,  $\phi_1 \neq \theta_1$ ). The ZL graph clearly reveals two foci, one inside the Ag film and the other at about 20 nm outside the Ag film in the PR. It should be noticed that the foci are somewhat blurred due to losses. The actual structure fabricated by Fang et al. was more complicated and may require an integration over the wavevector space to fully understand the imaging properties.<sup>43</sup> To examine the proximity limit, the thickness of the Ag film and the distance between the source and the Ag film are fourfold reduced without changing other conditions. As shown in Fig. 10.25b, a single focus is formed near the Ag-PR interface, and the ZLs are nearly straight lines in each medium. Because of the loss in the Ag film, the energy refraction angle in Ag depends on  $k_y$  and is slightly greater than that calculated from Eq. (10.59) based on the real parts of  $\epsilon$ 's. The ZL method presented here provides information on the paths of light energy and can be used to study lateral beam shifts in photon tunneling and to construct near-field images inside and outside flat lenses made of a NIM or a silver film.

### 10.3 SPECTRAL AND DIRECTIONAL CONTROL OF THERMAL RADIATION

Thermal emission is a spontaneous emission process that occurs from any objects. It is misleading to think that thermal emission arises only from heated objects like an oven, a fire, or the sun. In a vacuum environment, radiation is the only mode of heat transfer, such as in space exploration and cryogenic systems. Measurements of the far-infrared and microwave

emission spectrum of deep space have revealed that the cosmic background has an effective temperature of 2.7 K. Of course, thermal radiation is very important in combustion systems as well as in industrial furnaces for materials processing. Another application of thermal emission is the incandescent lamp or light bulb invented by Thomas Edison in 1879.

Radiation emission is a reverse process of absorption when the transition occurs from a higher energy level to a lower energy level. It should be noted that transition from a higher to a lower energy level is not necessarily associated with the emission of photon, because it can also release one or more phonons (i.e., lattice waves) as well as cause other transitions. Transitions that give out photons are called *radiative transitions*; otherwise, they are called *nonradiative transitions*. The absorption processes in solids were extensively studied in Chap. 8 (see Sec. 8.4). Thermal radiation emitted from solids is generally manifested by a broad spectrum and quasi-isotropic angular behavior, just like absorption and reflection. By introducing thin-film coatings and multilayer structures, the emission spectrum can be significantly modified, and wavelength selective coatings have been developed since the 1960s for space application and solar collectors. Gratings can also modify the emission properties. These approaches can be generalized to multidimensional complex microstructures, including photonic crystals, for wavelength and directional control of spontaneous emission. There are a number of applications that require spectral and directional selection of thermal radiation. Besides space application and solar energy, thermophotovoltaic devices utilize a heating source or an emitter around 1500 K to generate electricity based on the photovoltaic principle. The efficiency is often limited by the large portion of long-wavelength photons that cannot create electron-hole pairs in the photovoltaic cell. Other applications may include nanoelectronics thermal control and remote sensing technologies.

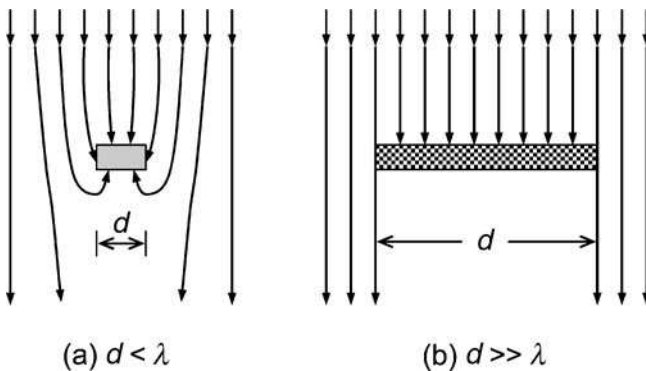
Cavity quantum electrodynamics (QED) is a field that was initiated in the 1980s to study the spontaneous emission of atoms inside a subwavelength cavity; see Haroche and Kleppner (*Physics Today*, **42**, 24, January 1989). Both enhancement and inhibition of spontaneous emission have been theoretically and experimentally demonstrated; see Yablonovitch et al. (*Phys. Rev. Lett.*, **61**, 2546, 1988), Lai and Hinds (*Phys. Rev. Lett.*, **81**, 2671, 1998), Bayer et al. (*Phys. Rev. B*, **86**, 3168, 2001), Solomon et al. (*Phys. Rev. Lett.*, **86**, 3903, 2001), and Larkin et al. (*Phys. Rev. B*, **69**, 121403R, 2004). In recent years, there have been a large number of publications dealing with spontaneous emission of microstructures through some controversial experiments as well as theoretical predictions; see Blanco and García de Abajo and references therein.<sup>45</sup> It seems quite clear that surface plasmons may enhance transmission, suppress transmission but enhance absorption, or enhance both transmission and absorption in nanostructures at the same time, depending on the coupling with the resonance and boundary conditions. There is no doubt that spontaneous emission can be suppressed at certain wavelengths using microstructures. The scientific community has yet to come up with an acceptable answer to the question: "Whether spontaneous emission can ever exceed that of blackbody radiation?" This is actually a rather ambiguous question because it did not specify where the spontaneous emission comes from, where the emission is detected, what the boundary conditions are, and what time duration is involved. Let us make it more specific: If an object is at thermal equilibrium by itself, can it emit out more energy (in any spectral range, polarization, and solid angle) to free space (far field) than a blackbody with the identical shape and size at the same temperature? Well, one may say that such a blackbody cannot exist if the object is smaller than the wavelength of interest. Imagine such a blackbody could exist and emit according to the blackbody intensity  $I_{b,\lambda}(\lambda, T)$  in all directions. The question is "Whether the intensity leaving the object can ever exceed the blackbody intensity at any particular wavelength and angle of emission?" By using intensity, we are clearly talking about the far field, not the near field.

The answer is definitely "yes" if the overall structure is less than or comparable with the wavelength, and definitely "no" if the overall structure is much greater than the wavelength, even though the structure is made by subwavelength features. Rather than



considering spontaneous emission toward an empty space, let us consider the thermodynamic equilibrium in an enclosure, where the object is placed inside and is in thermodynamic equilibrium with the enclosure. Generally speaking, stimulated emission is much smaller than stimulated absorption (see Chap. 3, Sec. 3.6), and we can treat the net absorption as stimulated absorption subtracted by stimulated emission. At thermodynamic equilibrium, the net absorption of energy must be the same as the spontaneously emitted energy of any objects inside the cavity. The density of states inside any medium is modified by its electric and magnetic properties. In the simple dielectric case, Planck's distribution is modified by the square of the refractive index, as can be seen from Eq. (10.30a) and Eq. (10.30b). If the refractive index depends on wavelength, the group velocity will be different from the phase velocity and, hence, the equilibrium distribution will further deviate from Planck's law. If absorption is also considered, the equilibrium distribution inside the medium will be completely different. However, *Planck's distribution is always observed in the evacuated region, as long as the location is away from either the object or the walls of the enclosure.* This condition or restriction implies that the enclosure must have enough room for the evacuated region to be much greater than the characteristic wavelength. It is impractical to have a blackbody with a size less than the wavelength, as noted by Planck himself many years ago.<sup>46</sup> It has been known for some time that a large field enhancement exists near the surface when surface polaritons are excited.<sup>30</sup> The enhancement also exists around subwavelength structures.<sup>32</sup> However, the energy density in an evacuated enclosure at thermal equilibrium is the same as Planck's distribution, except at close vicinity of the objects, including the walls of the enclosure.

Spontaneous emission can be viewed as a coupling of the field inside the material with that outside the material. A small object can couple with the electromagnetic field by bending the energy streamlines or the Poynting vectors (due to coupling of the incident and emitted fields) toward it, and hence, the object will absorb more energy than a *blackbody* of the same size.<sup>32</sup> Figure 10.26 illustrates qualitatively and somewhat exaggerated



**FIGURE 10.26** Schematic drawing of the energy streamlines for an incident plane wave, showing the Poynting vectors of the incident field and the cross coupling between the incident and scattered fields. (a) A small object with resonance absorption. (b) A large object with subwavelength structures.

interactions of the incident field with a small object and a large object. A small object can perturb the incoming energy streamlines by creating an additional term in the Poynting vector that arises from the coupling between the incident and scattered fields. Therefore, the absorptance and spontaneous emission can be enhanced at the resonance wavelength, which is structure dependent, as can be seen from Eq. (10.49) for the spherical case. On the other hand, for a large structure, the incoming energy is limited by the projected area without any

geometric enhancement even with surface or volume micro/nanostructures. Because of reflection and transmission, the net absorbed energy is always smaller than the energy incident on the object. Hence, it is not possible for spontaneous emission from a large object or composite to exceed the blackbody intensity in the far field. Several recent publications support this argument with detailed calculations and careful experiments; see Pigeat et al. (*Phys. Rev. B*, **57**, 9293, 1998), Luo et al. (*Phys. Rev. Lett.*, **93**, 213905, 2004), Seager et al. (*Appl. Phys. Lett.*, **86**, 244105, 2005), and Chow (*Phys. Rev. A*, **73**, 013821, 2006).

Next, we discuss some unique features when the emission/absorption spectra or angular distributions are modified by micro/nanostructures, especially by surface electromagnetic waves and resonance cavities. Kirchhoff's law is valid because the overall structure is much greater than the wavelength. In all cases, the emissivity can be calculated indirectly from the calculation of reflectance. In some cases, the distribution of electromagnetic fields has been explored to better understand the underlying mechanisms.

### 10.3.1 Gratings and Microcavities

It has been known for a long time that radiative properties, especially the directional and spectral properties, can be modified by surface roughness and structures. An example is that the cavities formed on the surface of the moon yield retroreflection of the visible light, i.e., the reflected rays are nearly antiparallel to the direction of the incoming rays from the sun. Earlier studies of surface microstructure effect on thermal radiative properties can be found, e.g., from Perlmutter and Howell (*J. Heat Transfer*, **85**, 282, 1963), Birkebak and Eckert (*J. Heat Transfer*, **87**, 85, 1965), Zipin (*Appl. Opt.*, **5**, 1954, 1966), and Torrance and Sparrow (*J. Heat Transfer*, **88**, 223, 1966). Most of the earlier studies dealt with rather simple geometries and did not consider diffraction. The diffraction by subwavelength periodic gratings was investigated under a completely different field for spectroscopic applications. The emergence of microfabrication and the increased computing capabilities have led to more systematic investigations of the effect of microstructures and material properties on thermal emission and absorption characteristics. Hesketh et al. published a series of studies on the thermal emission from periodically grooved micromachined silicon surfaces;<sup>47</sup> also see Hesketh et al. (*Phys. Rev. B*, **37**, 10795, 1998; 10803, 1998). The grooves were 45  $\mu\text{m}$  deep with straight ridges etched on heavily doped *p*-type Si wafers, with a grating period  $\Lambda$  ranging from 10 to 22  $\mu\text{m}$ . Thermal emission was measured at temperatures between 300 and 400°C in the mid-infrared wavelengths ranging from 3 to 14  $\mu\text{m}$ . Compared with smooth Si wafers, the grooved surfaces increased the spectral emittance, whereas the observed enhancement was polarization dependent even at normal incidence. Resonance features were observed in the emission spectra, but their location and dependence on the grating period were significantly different from those for TE and TM waves. The observed emittance enhancement was explained by *organ pipe* resonant modes.<sup>47</sup> Geometric optics models largely failed to predict the observed behavior. Wang and Zemel (*Infrared Phys.*, **32**, 477, 1991; *Appl. Opt.*, **31**, 732, 1992; *Appl. Opt.*, **32**, 2021, 1993) extended this work by studying the spectral emittance of gratings made of undoped Si. Several theories were examined, including the Bloch wave, coupled-mode, effective medium, and waveguide methods. It was found that the emission oscillates with wavelength for deep gratings similar to a Fabry-Perot cavity resonator. The effective medium theory could explain the observed directional and spectral variation of the measured emittance. Auslender and Hava used the RCWA and performed a parametric study to identify the conditions for nearly 100% absorptance for TE waves.<sup>48</sup> Buckius and coworkers performed rigorous calculations using the integral equation method as well as experiments for the reflection of 1-D and 2-D microstructured surfaces.<sup>49</sup>

As illustrated in Fig. 10.15, surface plasmons can reduce reflectance and enhance absorption for TM waves. Narrowband emission mediated by surface plasmons was demonstrated by Kreiter et al. (*Opt. Commun.*, **168**, 117, 1999) in the near-infrared with gold diffraction

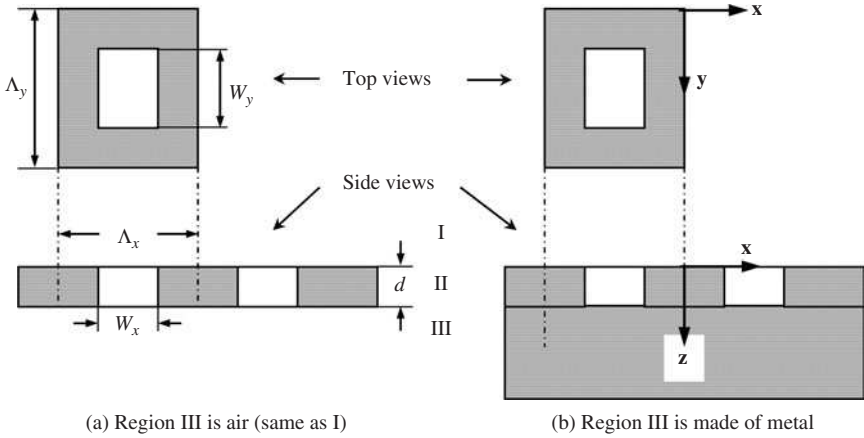
gratings. The emission is also direction dependent. Greffet and coworkers showed a strongly coherent thermal emission mediated by surface phonon polaritons.<sup>28</sup> The SiC grating, with a period of 6.25  $\mu\text{m}$  and a height of 0.28  $\mu\text{m}$ , emits thermal radiation near 11  $\mu\text{m}$  in two narrow lobes, when heated to 800 K. The emission spectrum contains a narrowband peak at the polariton resonance frequency for TM waves. Spectrally coherent thermal emission means that the emission is confined in a narrow wavelength region for any given direction; this is also referred to as *temporal coherence*, because coherence time and coherence length are interrelated. A nearly monochromatic radiation will have a very long coherence length and time. When the emission at a given wavelength is confined to a narrow angular range, it is referred to as *spatial coherence*, like a collimated beam whose wavefronts will not alter significantly as it travels. At the resonance conditions, surface plasmon or phonon polaritons are coupled with spontaneous emission due to randomly fluctuating charges, called dipoles in the thermal field; consequently, thermal emission is enhanced at a particular wavelength and direction. At wavelengths where the surface mode is not excited, the radiation emitted inside the material is either absorbed by the neighboring atoms or reflected back by the material-air interface, yielding a very low emissivity that is typical for metallic materials in the infrared. It should be noted that the bandwidth for spontaneous emission is far larger than that of lasers, which operate under the principle of stimulated emission. Nevertheless, a much longer coherence length than that of blackbody radiation or emission from plain solids has been achieved. Coherent thermal emission is expected to have applications in remote sensing, space thermal control, thermophotovoltaic devices, and nanoelectronics thermal management, especially when coupled with nanoscale heat transfer.

Heinzel et al. fabricated 2-D arrays of tungsten circular pillars as near-infrared emitters and hole arrays on gold films as wavelength-selective filters, for applications in thermophotovoltaic systems.<sup>50</sup> For these structures, the lateral period was between 1 and 2  $\mu\text{m}$ , and the thickness was between 200 and 300 nm. Theoretical modeling based on a 2-D RCWA was performed and compared with experiments. Surface plasmons and cavity resonances were believed to play a role in the wavelength selection. The emittance was measured in vacuum at temperatures up to 1700 K, with a Fourier-transform spectrometer. Maruyama et al. developed 2-D microcavities using Cr-coated Si surfaces and demonstrated discrete thermal emission peaks from these structures as a result of cavity resonances and the enhanced density of states.<sup>51</sup> Sai et al. gave a comprehensive review of their theoretical and experimental developments of 2-D microcavities on tungsten surfaces for thermophotovoltaic emitters.<sup>51</sup> By using square microcavities with a period around 1  $\mu\text{m}$ , a width around 0.8  $\mu\text{m}$ , and a depth of 1.6  $\mu\text{m}$ , high emittance was achieved at  $\lambda < 2 \mu\text{m}$ . This is the desired feature for near-infrared thermophotovoltaic applications. Kusunoki et al. studied the emission spectra from microcavities made by dry etching on tantalum and tungsten surfaces, and identified surface plasmon modes and cavity resonance modes in the mid-infrared emission spectra.<sup>52</sup> Chen and Zhang (*Opt. Commun.*, **269**, 411, 2007) recently proposed complex gratings to tailor the emission spectrum of tungsten for thermophotovoltaic radiators.

Before moving into other configurations, let us further explore the resonance conditions with the assistance of Fig. 10.27. Readers should refer to Fig. 9.18 and the discussion in Sec. 9.4.1 for the general theory of 1-D gratings, and Sec. 10.1.2 for guided modes in a waveguide. Consider the two 2-D rectangular grating structures, where region II is the binary grating with a thickness  $d$  and is made of air holes or slits inside the metal. The air portion in between the metal ridges will be referred to as cavities hereafter. Region III can either be air (Fig. 10.27a) or a metal (Fig. 10.27b) for the filter/lens or emitter/absorber type of applications. The Bloch wave conditions are

$$k_{x,p} = k_x + \frac{2\pi}{\Lambda_x} p \text{ and } k_{y,q} = k_y + \frac{2\pi}{\Lambda_y} q \quad (10.60)$$

where  $p, q = 0, \pm 1, \pm 2, \dots$  are the diffraction orders,  $k_x$  and  $k_y$  are the wavevector components of the incident plane wave, and  $\Lambda_x$  and  $\Lambda_y$  are the periods in the  $x$  and  $y$  directions,



**FIGURE 10.27** Illustration of metal gratings. (a) Both sides are open to air (hole or slit array). (b) Gratings are made on the metal.

respectively. Usually, it is the lower diffraction orders that are significant. However, a large number of diffraction orders may need to be considered; see Chen et al. (*Int. J. Thermophys.*, **25**, 1235, 2004). Because region I is air, we have  $k_x^2 + k_y^2 + k_z^2 = k^2 = \omega^2/c^2$ . Inside the cavities, we have

$$k_{x,p}^2 + k_{y,p}^2 + k_{z,pq}^2 = k^2 \tag{10.61}$$

which determines the  $z$  component of the diffraction order defined by  $(p,q)$ . Note that for 1-D gratings with grooves parallel to  $y$ ,  $\Lambda_y \rightarrow \infty$ , and we can simply replace  $k_{y,q}$  by  $k_y$  because there is no diffraction in the  $y$  direction. In Chap. 9, we used subscript  $j$  as the diffraction order in the case of 1-D gratings.

A Rayleigh-Wood anomaly mode corresponds to  $k_{z,pq} = 0$ ; therefore,

$$\left(\frac{2\pi}{\lambda_{pq}}\right)^2 = k_{pq}^2 = \left(k_x + \frac{2\pi}{\Lambda_x}p\right)^2 + \left(k_y + \frac{2\pi}{\Lambda_y}q\right)^2 \tag{10.62}$$

On the other hand, surface polariton conditions are determined when  $k_z$  matches the dispersion relation, Eq. (10.42). Because the dispersion curve, as shown in Figs. 10.14a and 10.15a, almost merges with the light line, the Rayleigh-Wood anomaly and surface polariton modes overlap, causing sharp drops in the reflectance spectra for shallow gratings, as can be seen from Fig. 10.15b. The problem is more complex for deep gratings because one needs to decide whether the interface between regions I and II or the interface between regions II and III are responsible. When the cavities are open to both ends, the surface polaritons are coupled and may divide into two modes, as discussed previously. The dielectric function of region II, however, is not the same as that of the metal, unless the filling ratio is very high (i.e., with very narrow slits or small holes). Therefore, the surface polariton resonance frequencies are highly sensitive to both the filling ratio and height of the gratings. Numerous papers have dealt with this situation, but a systematic parametric study is not yet available.

Attention is turned to cavity resonances, which are important when  $d$  is on the order of or greater than  $W_x$  or  $W_y$ . Standing waves, or *effective* standing waves, need to exist in the cavities in order to couple the incident radiation field with the gratings or through the gratings. The resonance conditions are

$$k_{x,m} = \frac{m\pi - \delta}{W_x}, m = 0, 1, 2, \dots \quad (10.63)$$

$$k_{y,n} = \frac{n\pi - \delta}{W_y}, n = 0, 1, 2, \dots \quad (10.64)$$

where the phase shift upon reflection  $\delta$  may be neglected for highly reflecting metals and

$$k_{z,l} = \frac{\pi}{d_{\text{eff}}}l, l = 0, 1, 2, \dots \quad (10.65a)$$

when both sides of the cavities are open as shown in Fig. 10.27a. Here,  $d_{\text{eff}}$  is an effective cavity length that should be on the same order as the grating height  $d$ . Because grating does not have any physical boundaries or symmetric boundaries, it is not expected that  $d_{\text{eff}}$  will be identical to  $d$ . When  $l = 0$ , it is also a Rayleigh-Wood mode ( $k_{z,l} = 0$ ), which can be coupled to the resonance mode if Eq. (10.64) and Eq. (10.65) are satisfied. For gratings whose cavities are open on one side, as shown in Fig. 10.27b, assuming a symmetric boundary condition near the top of the grating region, we have

$$k_{z,l} = \frac{\pi}{d_{\text{eff}}}\left(l + \frac{1}{2}\right), l = 0, 1, 2, \dots \quad (10.65b)$$

The resonance wavelength for a mode  $(m, n, l)$  is simply  $\lambda_{m,n,l} = 2\pi/k_{m,n,l}$  with  $k_{m,n,l}^2 = k_{x,m}^2 + k_{y,n}^2 + k_{z,l}^2$ . When resonance occurs, there will be a strong electromagnetic field inside the cavity; it is this confined and enhanced field that subsequently enhances absorption and may enhance or suppress transmission (for slit or hole arrays). Intuitively, one would match a diffracted wave with the resonance mode for the excitation. Two simple matching scenarios are (1)  $W_x = \Lambda_x/2$  and  $W_y = \Lambda_y/2$  such that the modes  $m = |p|$  and  $n = |q|$  at normal incidence and (2)  $W_x \approx \Lambda_x$  and  $W_y \approx \Lambda_y$  such that the modes  $m = 2|p|$  and  $n = 2|q|$  at normal incidence. Experiments seem to agree with these predictions, considering that metal is not a perfect conductor and the fabricated gratings may not be ideal.<sup>51,52</sup> The case is complicated if, say,  $W_x = 0.47\Lambda_x$  and  $W_y = 0.53\Lambda_y$ . It is impossible to find small numbers of  $(p, q)$  to match with small numbers of  $(m, n, l)$ . Furthermore, the resonance condition given in Eq. (10.65b) inherently requires that a propagating wave exist in the  $z$  direction. This requires that  $k_{z,pq}$  in Eq. (10.61) be real. At normal incidence, a real  $k_{z,pq}$  implies that  $\lambda < \Lambda_x$ , assuming  $\Lambda_x \leq \Lambda_y$ . However, resonance at  $\lambda > \Lambda_x$  has been observed for square cavities ( $\Lambda_x = \Lambda_y$ ) as well as with very deep gratings. Note that the diffracted field is a Fourier series of all diffraction orders, and it is not necessary for a single individual mode to match with the cavity mode. It is the field of the combined diffracted waves that must match the cavity conditions for resonance to occur. In fact, only when  $W_x = \Lambda_x/2$  and  $W_y = \Lambda_y/2$ , all diffracted modes individually matches the cavity modes. Otherwise, cavity modes rule over Bloch conditions in terms of the excitation frequency. Equation (10.63) through Eq. (10.65) should be combined to determine the wavelength of resonance modes. For a square cavity with  $\Lambda_x = \Lambda_y = 5 \mu\text{m}$ ,  $W_x = W_y = 3 \mu\text{m}$ , and  $d = 3.7 \mu\text{m}$ , Kusunoki et al. experimentally observed the modes  $(m, n, l) = (1, 0, 0)$ ,  $(1, 0, 1)$ , and  $(1, 0, 2)$  or  $(0, 1, 0)$ ,  $(0, 1, 1)$ , and  $(0, 1, 2)$  in the emission spectrum.<sup>52</sup> Note the symmetry between  $m$  and  $n$ . Their measured spectrum also exhibited a strong peak near  $\lambda = 6 \mu\text{m}$ , suggesting the existence of a cavity resonance mode with  $(1, 0, -\frac{1}{2})$  or  $(0, 1, -\frac{1}{2})$ . These modes will result in strong absorption near the top of the grating region, regardless of the boundary conditions at the bottom of the cavity. When  $W_x \leq W_y$ ,  $\lambda = 2W_x$  corresponds to the fundamental mode, and no resonance modes exist at longer wavelengths, unless the grating is very deep, as will be discussed next.

For a narrow-slit array or a small-hole array, it is possible to have resonance modes with  $l = 1, 2, \dots$  and  $m = n = 0$  (*organ pipe mode*). In an example shown by Marquier et al. (*Opt. Express*, **13**, 70, 2005) for a 1-D Ag slit array with  $\Lambda = 500 \text{ nm}$ ,  $W = 50 \text{ nm}$ , and  $d = 400 \text{ nm}$ , the organ pipe resonance mode is responsible for the large transmittance at

$l = 1$  and  $2$ . However,  $d_{\text{eff}}$  is highly sensitive to the structure as well as the order of the mode, making it difficult to predict the peak wavelengths in a simple formulation. For 1-D gratings, these resonances can occur for TM waves only, because the boundary conditions require that the parallel components of the electric field ( $E_y$  and  $E_z$ ) vanish at the left and right walls of the slit. It should be emphasized that the resonance mode at  $\lambda > \Lambda$  is formed by the Fourier series of all diffracted evanescent waves, together with the zeroth-order diffraction mode ( $k_{x,0} = k_x$  and  $k_{y,0} = k_y$ ), which is a propagating wave with the same characteristics as the incident wave. More quantitative research is needed to develop reliable regimes to characterize the absorption and transmission enhancement or suppression for each polarization. The explanation that works for 1-D surfaces may be different from that for 2-D surfaces. Mathematical solutions for circular and annular geometries are different from that for a rectangular geometry. For cylindrical holes arranged in a square array, the matching between the two coordinate systems also needs to be considered.

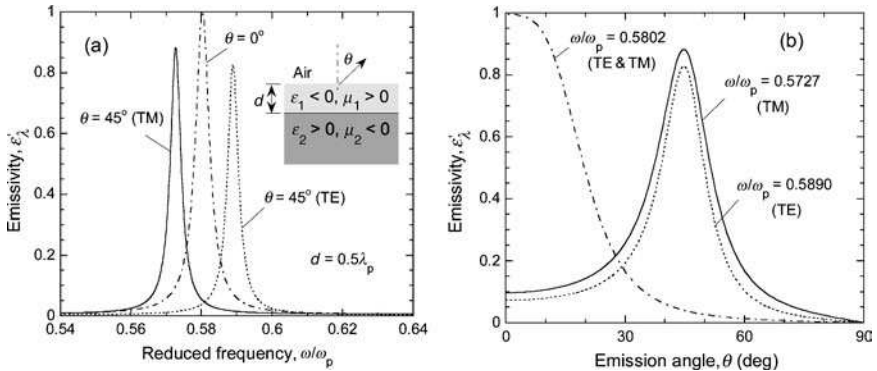
### 10.3.2 Metamaterials

While a thin film can exhibit spectral oscillations and angular lobes in the radiative emission spectra, as shown by Kollyukh et al. (*Opt. Commun.*, **225**, 349, 2003), the resonance features are not sharp enough. The use of a 1-D dielectric Fabry-Perot resonator may introduce sharp features, as suggested by Ben-Abdallah (*J. Opt. Soc. Am. A*, **21**, 1368, 2004). A dielectric layer can be formed on a highly reflecting metallic substrate or an opaque metallic film coated onto a dielectric substrate. On top of the dielectric, a partially transmitting mirror can be used so that an asymmetric Fabry-Perot structure can be built for emission toward the partially transmitting mirror. Candidates of the reflecting mirror are metallic films, photonic crystals, heavily doped Si, and SiC. For more details, see Schubert et al. (*Appl. Phys. Lett.*, **63**, 2603, 1993), Celanovic et al. (*Phys. Rev. B*, **72**, 075127, 2005), and Laroche et al. (*Opt. Commun.*, **250**, 316, 2005). Metamaterials have also been proposed for selective emitters; see Enoch et al. (*Phys. Rev. Lett.*, **89**, 213902, 2002) and Zhou et al. (*Appl. Phys. Lett.*, **86**, 101101, 2005). Fu et al. proposed to use the paired negative- $\epsilon$  and negative- $\mu$  bilayer to achieve coherent emission through the excitation of surface polaritons at all angles, for both TE and TM waves.<sup>37</sup> The following example illustrates this concept, which is promising as the development of metamaterials continues to push toward high frequencies.

**Example 10-7.** For a thin metallic-type film, with  $\text{Re}(\epsilon_1) < 0$  and  $\text{Re}(\mu_1) > 0$ , of thickness  $d$ , on an opaque magnetic material, with  $\text{Re}(\mu_2) < 0$  and  $\text{Re}(\epsilon_2) > 0$ , calculate the emissivity, using the functions given in Eq. (10.54) and Eq. (10.55). Assume the parameters are  $F = 0.785$ ,  $\omega_0 = 0.5\omega_p$ ,  $d = 0.5\lambda_p = \pi c/\omega_p$ ,  $\epsilon_2 = 4$ , and  $\gamma_c = \gamma_m = 0.002\omega_p$ .

**Solution.** Under the lossless conditions, the polariton dispersion relations are the same as shown in Fig. 10.20 for two semi-infinite media. The directional-spectral emissivity can be calculated by  $\epsilon'_\lambda = 1 - R'_\lambda$ , because the magnetic medium is semi-infinite, where  $R'_\lambda$  can be evaluated using Eq. (10.46) for each polarization. The calculation results are shown in Fig. 10.28a at normal direction as well as at  $\theta = 45^\circ$  for either TE or TM wave incidence. Reduced frequency is used again. It can be seen that the peak shifts toward lower frequencies for the TM wave and higher frequencies for the TE wave as  $\theta$  increases and the center frequency of the peak  $\omega_c$  is in good agreement with the polariton dispersion curves, shown in Fig. 10.20a. The  $Q$ -factor, defined as  $Q = \omega/\delta\omega$  with  $\delta\omega$  being the FWHM, is around 100. Figure 10.28b shows the angular distribution of the emission at the center frequencies shown on the left figure. The emission is not diffuse but rather direction selective.

Fu et al. further proposed to use a three-layer structure with a negative- $\epsilon$  film and a negative- $\mu$  film onto a negative- $\epsilon$  substrate to achieve a higher  $Q$  and a spatially coherent source. In such a case, surface polaritons at both sides of the negative- $\mu$  medium can be coupled. A temporally coherent diffuse emitter was also predicted.<sup>37</sup>



**FIGURE 10.28** Emissivity of a negative- $\epsilon$  layer of thickness  $d$  on a negative- $\mu$  layer (semi-infinite). (a) Frequency dependence at fixed angles. (b) Angular dependence at fixed frequencies.

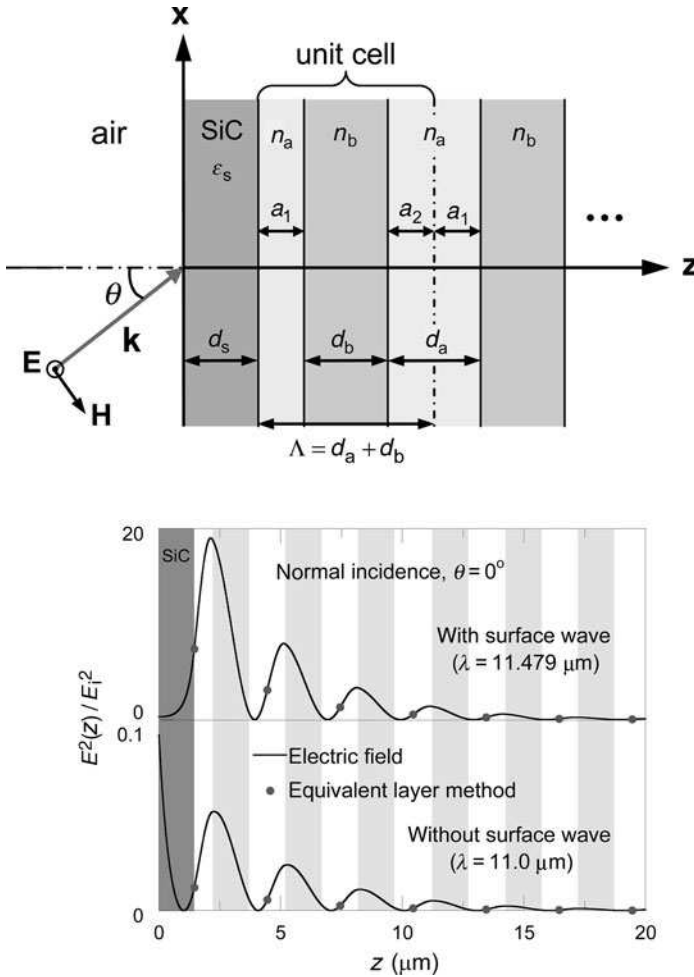
### 10.3.3 Modified Photonic Crystals for Coherent Thermal Emission

Numerous recent studies utilize the unique features of modulated microstructures (i.e., photonic crystals) and nanoparticle arrays to control and improve the optical and radiative properties for specific applications. For example, see Cornelius and Dowling (*Phys. Rev. A*, **59**, 4736, 1999), Pralle et al. (*Appl. Phys. Lett.*, **81**, 4685, 2002), Lin et al. (*Opt. Lett.*, **20**, 1909, 2003), Fleming (*Appl. Phys. Lett.*, **86**, 249902, 2005), Puscasu et al. (*J. Appl. Phys.*, **98**, 013531, 2005), Fujita et al. (*Science*, **308**, 1296, 2005), Ecoch et al. (*Appl. Phys. Lett.*, **86**, 261101, 2005), Ben-Abdallah and Ni (*J. Appl. Phys.*, **97**, 104910, 2005), Florescu et al. (*Phys. Rev. A*, **72**, 033821, 2005), Laroche et al. (*Phys. Rev. Lett.*, **96**, 123903, 2006), and Yannopoulos (*Phys. Rev. B*, **73**, 113108, 2006). In this subsection, coverage is given to the surface electromagnetic waves coupled with a PC and the resulting coherent emission characteristics. Yeh et al. showed that a PC can support surface modes or surface waves for both the TM and TE waves in the stop band.<sup>22</sup> If a metallic layer is coated on a 1-D PC, surface waves can be excited by a propagating wave in air; this will result in a strong reduction in the reflectance at the resonance frequency.<sup>53</sup> Lee et al. predicted coherent thermal emission based on a modified 1-D PC coated with a thin film of SiC.<sup>54</sup> When the thicknesses and dielectric properties are adjusted, surface waves can be excited in the stop band of the PC by radiative waves propagating in air, for either polarization. Subsequently, the emission from the proposed structure contains sharp peaks within a narrow spectral band and toward well-defined directions. The geometry and the electric field distribution are illustrated in Fig. 10.29, and will be discussed in detail next.

A PC is a heterogeneous structure, and for the PC discussed previously,  $\epsilon_a = n_a^2$ ,  $\epsilon_b = n_b^2$ , and  $\mu_a = \mu_b = 1$  (nonmagnetic). Hence, it is inappropriate to define the equivalent  $\epsilon$  and  $\mu$  of the PC separately by considering it as a homogeneous medium. However, surface waves can be excited at the stop band of the PC because there exists in the PC an *effective evanescent wave*, which is an oscillating field whose amplitude gradually decays to zero as  $z$  approaches infinity. The effective evanescent wave does not carry energy into a semi-infinite PC. Since the wavelength range corresponding to stop bands of the PC can be scaled by changing the thickness of the unit cell,  $\Lambda$  is chosen to be  $3 \mu\text{m}$  in order to approximately match the wavelengths corresponding to the first bandgap of the 1-D PC, shown in Fig. 9.16, with the phonon absorption band of SiC. Surface waves can be excited at the SiC-PC interface within the SiC phonon absorption band for both polarizations. By using the equivalent layer method<sup>53</sup> or the supercell method proposed by Ramos-Mendieta

and Halevi (*J. Opt. Soc. Am. B*, **14**, 370, 1997), it is possible to obtain dispersion relations of surface waves between a PC and another medium similar to Eq. (10.42) and Eq. (10.43).

Figure 10.29 also shows the square of the electric field, normalized to the incident, inside the SiC-PC structure for  $\theta = 0$ . The real part of the complex electric field is used to show the actual field inside the structure. The solid line represents the field calculated from the matrix formulation described in Sec. 9.2.2. An oscillating field exists inside the



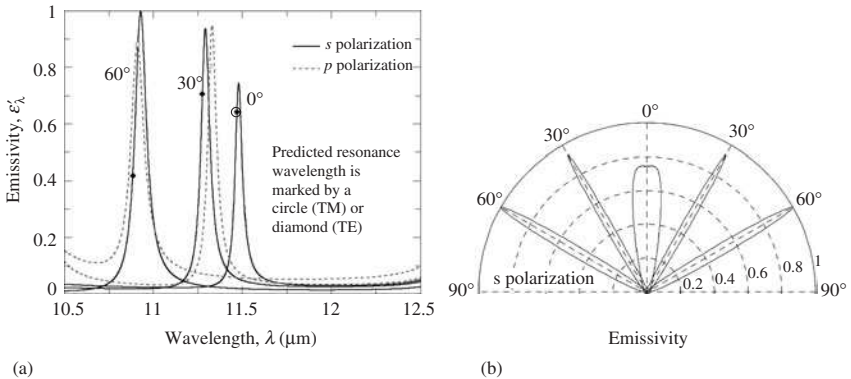
**FIGURE 10.29** Schematic of the SiC-coated 1-D PC (upper) and the field distributions (lower) for a TE wave incident from air.<sup>54</sup>

PC, and the amplitude of the oscillating field decays gradually toward larger  $z$ . The dots represent the electric field obtained using the equivalent layer method, which matches the matrix solutions at the boundaries of each unit cell.<sup>53</sup> The upper panel corresponds to the wavelength ( $\lambda = \lambda_c = 11.479 \mu\text{m}$  when a surface wave is excited, and the lower panel



corresponds to ( $\lambda = 11.0 \mu\text{m}$  without a surface wave. The field strength at the boundary between SiC and the PC is enhanced by more than an order of magnitude due to excitation of the surface wave. When a surface wave is excited, the incident energy is resonantly transferred to the surface wave, which causes a large absorption in SiC. Because SiC is the only material in the structure that can absorb the incident energy, it is also responsible for the emission of radiation from the SiC-PC structure. It is interesting to note that the maximum electric field is slightly off from the interface between SiC and the PC, which has been observed previously. If a smooth curve connects all the dots, the magnitude of the electric field will be maximum at the SiC-PC interface and decay gradually deep into the PC. Furthermore, the Poynting vector or the energy flux toward the positive  $z$  direction is zero inside the PC at the stop band. Therefore, the effective field inside the PC at the stop band resembles an evanescent wave in a semi-infinite medium. The fact that the field near the SiC-PC interface is greatly enhanced confirms the existence of a surface wave. Further, surface waves at the interface between SiC and the PC can be excited at any angle of incidence and for both polarizations.

Figure 10.30a shows the spectral-directional emissivity spectra in the wavelengths between 10.5 and 12.5  $\mu\text{m}$  at  $\theta = 0^\circ, 30^\circ,$  and  $60^\circ$  for both polarizations.<sup>54</sup> Notice that



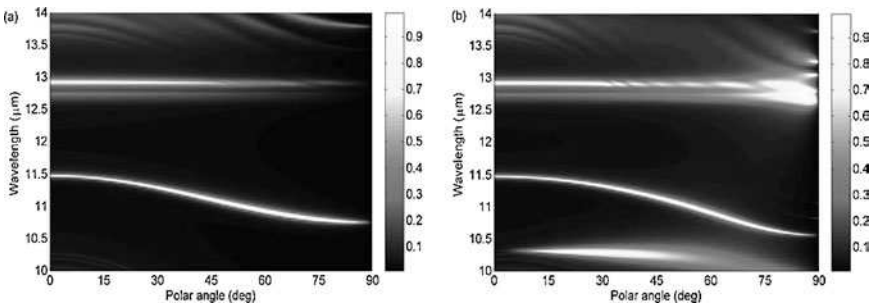
**FIGURE 10.30** The directional-spectral emissivity of the SiC-PC structure when surface wave is excited.<sup>54</sup> (a) Spectral dependence at different polar angles for both polarizations. (b) Polar plot of showing the angular distribution of the emissivity at  $\lambda_c = 11.479, 11.293,$  and  $10.929 \mu\text{m}$  for a TE wave ( $s$  polarization).

since the emission peak values depend on the thickness of SiC,  $d_c$  can be tuned to maximize the emissivity for any given emission angle and polarization states. Here, the thickness of SiC is set to be  $d_c = 1.45 \mu\text{m}$ , which results in a near-unity emissivity at  $\theta = 60^\circ$  for  $s$  polarization and slightly lower emission peaks at other conditions. For  $s$  polarization, as an example, very large  $\epsilon'_\lambda$  can be seen in a narrow wavelength band centered at  $\lambda_c = 11.479, 11.293,$  and  $10.929 \mu\text{m}$ , for emission angles of  $0^\circ, 30^\circ,$  and  $60^\circ$ , respectively. The spectral emission peaks clearly indicate temporal coherence of the thermal emission. The corresponding quality factor  $Q = \lambda_c/\delta\lambda$  are 230, 185, and 133, respectively, which are comparable to those for SiC gratings.<sup>28</sup> From the solution of the surface wave dispersion relation, assuming no absorption in SiC, the resonance wavelength can be predicted for the given emission angle. These values are also marked as diamonds and circles for TE and TM waves, respectively. The deviation in the emission peaks from the

predicted resonance frequency is due to the fact that the thickness of SiC layer is so thin that the radiation damping effect causes a considerable shift in the emissivity peak positions from the predicted solution.<sup>30,34,35</sup>

The spatial coherence of the proposed emission source can be seen from the angular distributions of the emissivity, shown in Fig. 10.30b at the three peak wavelengths for the TE wave. It is important to note that the emissivity is plotted as a polar plot to clearly show the angular lobe into a well-defined direction. However, if one considers the actual source with finite dimensions, due to the axial symmetry of the planar structure, the coherent emission from the SiC-PC structure exhibits circular patterns, in contrast to the antenna shape for the grating surfaces. The emissivity at each  $\lambda_c$  is confined in a very narrow angular region, although the angular spread corresponding to the peak at  $\theta = 0^\circ$  is larger than the other two peaks.

In order to examine the resonance modes fully, the spectral-directional emissivity is illustrated by the contour plot in Fig. 10.31, for both polarizations, as a function of wavelength and emission angle. Large emissivity values can be seen in a certain range of



**FIGURE 10.31** Contour plot of the directional-spectral emissivity of the SiC-PC structure.<sup>54</sup> (a) TE wave. (b) TM wave.

wavelengths and emission angles. There exist three different mechanisms for the enhanced emissivity in the SiC-PC structure. In addition to surface waves, cavity resonance can occur for both polarizations at wavelengths near 13  $\mu\text{m}$ . The emission band is flat, suggesting that a nearly diffuse emitter can be formed with the cavity resonance mode. There also exists a Brewster mode for the TM wave due to the small reflection coefficient near the Brewster angle, which is located at wavelengths ranging from 10 and 10.3  $\mu\text{m}$  at  $\theta > 10^\circ$ . More recently, Lee and Zhang (*J. Appl. Phys.*, **100**, 063529, 2006) demonstrate spectral coherence near the wavelength of 1  $\mu\text{m}$  using truncated 1-D PC on Ag, which was deposited on a silicon substrate.

## 10.4 RADIATION HEAT TRANSFER AT NANOMETER DISTANCES

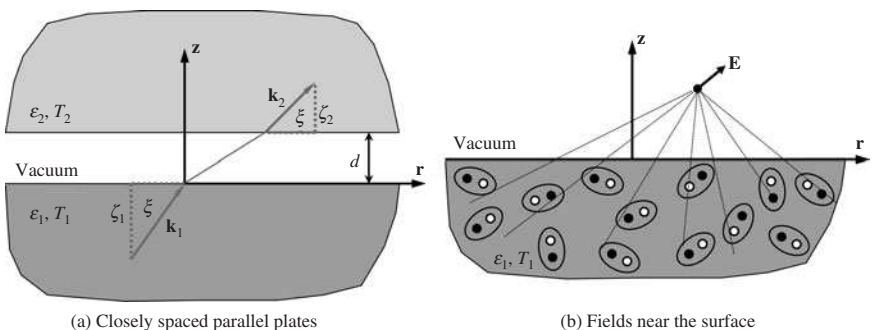
Heat transfer between surfaces placed at extremely short distances has important applications in near-field scanning thermal microscopy.<sup>55–58</sup> The concept of microscale thermophotovoltaic devices has been proposed to improve the energy conversion efficiency, by bringing the hot source very close to the receiving surface so that photon tunneling can

enhance the net radiant power flux.<sup>18</sup> Negative index materials can be used to enhance photon tunneling through longer distances.<sup>23,24</sup> The calculation of near-field radiation heat transfer between dielectric materials is rather straightforward and has already been described in Sec. 10.1.4. Nanoscale radiation heat transfer can be enhanced by several orders of magnitude when absorption is considered. While many metals support surface waves through surface plasmon polaritons, the plasma frequencies are usually much higher than the characteristic frequencies of thermal sources. Consequently, the near-field enhancement of thermal radiation is not very large for good conductors. On the other hand, semiconductors and semimetals, with smaller electric conductivities, may greatly enhance radiation heat flux at nanometer scales; see Polder and van Hove (*Phys. Rev. B*, **4**, 3303, 1971) and Loomis and Maris (*Phys. Rev. B*, **50**, 18517, 1994). Most of the theoretical works were centered on the prediction of the net heat flux between two parallel metallic plates, using a simple Drude model for the dielectric function. Several studies also considered the nanoscale energy transfer between a sphere and a surface or between two spheres.<sup>19,59</sup> The use of SiC allows surface phonon polaritons to be excited, resulting in large near-field radiation heat transfer that is concentrated in a very narrow wavelength band.<sup>19,59,60</sup>

This section introduces fluctuational electrodynamics, originally developed by Rytov in the 1950s, based on the fluctuation-dissipation theorem.<sup>61</sup> Detailed discussions will be given on the calculation of the near-field thermal radiation between two parallel plates, with an example based on doped silicon. The fluctuation-dissipation theorem has applications in the study of thermal conductivity of nanostructures and has also been used to study the van der Waals forces and noncontact friction at nanometer distances; see Volokitin and Persson (*Phys. Rev. Lett.*, **91**, 106101, 2003) and Zurita-Sanchez et al. (*Phys. Rev. A*, **69**, 022902, 2004). Another application of fluctuational electrodynamics is that it provides the first-principle calculation of thermal emission as well as emissivity. The method has been used to predict the emissivity from 1-D metallodielectric photonic crystals.<sup>60</sup>

### 10.4.1 The Fluctuational Electrodynamics

Consider the geometry shown in Fig. 10.32a, where two media, each at equilibrium but with different temperatures  $T_1$  and  $T_2$ , are separated by a vacuum gap of width  $d$ , ranging from several tens of micrometers down to 1 nm. For nonmagnetic, homogeneous, and isotropic media, the complex dielectric function or relative permittivity is the only property needed to fully characterize the optical behavior. The foundation of fluctuational electrodynamics is the fluctuation-dissipation theorem, under which thermal radiation is assumed



**FIGURE 10.32** Schematic drawings for the study of near-field thermal radiation in the cylindrical coordinates. (a) Radiation heat transfer between two parallel plates separated by a vacuum gap. (b) The electric field near the surface due to thermally induced charge fluctuations.

to arise from the random movement of charges inside the medium at temperatures exceeding 0 K. The fluctuated charge movement causes fluctuating electric currents that will result in distributions of electromagnetic fields in space and time. The movement of charges from an equilibrium position can also be viewed as dipoles illustrated in Fig. 10.32*b*. The electromagnetic field at any location is a superposition of contributions from all point sources in the radiating region. The electromagnetic waves deep inside the medium will attenuate due to absorption (i.e., dissipation) inside the medium. The following discussion is based on the work of Fu and Zhang.<sup>20</sup> The basic assumptions are as follows: (a) Each medium is semi-infinite and at a thermal equilibrium, presumably due to a sufficiently large thermal conductivity of the solid. (b) Both media are nonmagnetic, isotropic, and homogeneous, so that the frequency-dependent complex dielectric function (relative permittivity)  $\varepsilon_1$  or  $\varepsilon_2$  is the only material property that characterizes the electrodynamic response and thermally excited dipole emission of medium 1 or 2. (c) Each surface is perfectly smooth, and the two surfaces are parallel to each other.

Because of axial symmetry, cylindrical coordinates can be used so that the space variable  $\mathbf{x} = \mathbf{r} + \mathbf{z} = r\hat{\mathbf{r}} + z\hat{\mathbf{z}}$ . Consider a monochromatic electromagnetic wave propagating from medium 1 to 2. The complex wavevectors in media 1 and 2 are  $\mathbf{k}_1$  and  $\mathbf{k}_2$ , respectively, with  $k_1^2 = \varepsilon_1 k_0^2$  and  $k_2^2 = \varepsilon_2 k_0^2$ , where  $k_0 = \omega/c = 2\pi/\lambda$  is the magnitude of the wavevector in vacuum. Because  $\varepsilon_1$  and  $\varepsilon_2$  are in general complex,  $k_1$  and  $k_2$  should be viewed as complex variables of  $\omega$ . Only real and positive  $\omega$ 's are considered so that  $k_0$  is always real. The monochromatic plane wave can be expressed in terms of a time- and frequency-dependent field  $\exp(i\mathbf{k}_j \cdot \mathbf{x} - i\omega t)$ , where  $j = 0, 1, \text{ or } 2$  refers to vacuum, medium 1, or medium 2, respectively. The phase-matching condition requires the parallel components of all three wavevectors to be the same. To simplify the notation, let us use  $\xi$  for the parallel component and  $\zeta_j$  for the normal component of the wavevector  $\mathbf{k}_j$ . Thus,  $\mathbf{k}_j = \xi\hat{\mathbf{r}} + \zeta_j\hat{\mathbf{z}}$  and  $\zeta_j = \sqrt{k_j^2 - \xi^2}$ . The spatial dependence of the field in vacuum can be expressed as  $\exp(i\xi r + i\zeta_0 z)$ . Because its amplitude must not change along the  $r$  direction,  $\xi$  must be real. Keep in mind that both  $r$  and  $\xi$  are positive in the cylindrical coordinates. The normal component of the wavevector in vacuum  $\zeta_0 = \sqrt{k_0^2 - \xi^2}$  will be real when  $0 \leq \xi \leq \omega/c$  and purely imaginary when  $\xi > \omega/c$ . Thus, an evanescent wave exists in vacuum when  $\xi > \omega/c$ . Note that  $\zeta_1$  and  $\zeta_2$  are in general complex.

The random thermal fluctuations produce a spatial-time-dependent electric current density  $\mathbf{j}(\mathbf{x}, t)$  inside the medium whose time average is zero. The current density can be decomposed into the frequency domain using the Fourier transform, which gives  $\mathbf{j}(\mathbf{x}, \omega)$ . With the assistance of the dyadic Green function  $\mathbf{G}(\mathbf{x}, \mathbf{x}', \omega)$ , the induced electric field in the frequency domain can be expressed as a volume integration:

$$\mathbf{E}(\mathbf{x}, \omega) = i\omega\mu_0 \int_V \mathbf{G}(\mathbf{x}, \mathbf{x}', \omega) \cdot \mathbf{j}(\mathbf{x}', \omega) d\mathbf{x}' \quad (10.66)$$

where  $\mu_0$  is the magnetic permeability of vacuum, and the integral is over the region  $V$  that contains the fluctuating sources. The physical significance of the Green function is that it is a transfer function for a current source  $\mathbf{j}$  at a location  $\mathbf{x}'$  and the resultant electric field  $\mathbf{E}$  at  $\mathbf{x}$ . Mathematically, the dyadic Green function satisfies the vector Helmholtz equation:

$$\nabla \times \nabla \times \mathbf{G}(\mathbf{x}, \mathbf{x}', \omega) - k^2 \mathbf{G}(\mathbf{x}, \mathbf{x}', \omega) = \mathbf{I} \delta(\mathbf{x} - \mathbf{x}') \quad (10.67)$$

where  $k$  is the amplitude of the wavevector at  $\mathbf{x}$ , and  $\mathbf{I}$  is a unit dyadic. The corresponding magnetic field  $\mathbf{H}(\mathbf{x}, \omega)$  can be obtained from the Maxwell equation:  $\mathbf{H}(\mathbf{x}, \omega) = (i\omega\mu_0)^{-1} \nabla \times \mathbf{E}(\mathbf{x}, \omega)$ . The spectral energy density of the thermally emitted

electromagnetic field in vacuum can be calculated from Eq. (8.19) based on the ensemble average. Therefore,

$$u_{\omega}(\mathbf{x}, \omega) = \frac{\varepsilon_0}{4} \langle \mathbf{E}(\mathbf{x}, \omega) \cdot \mathbf{E}^*(\mathbf{x}, \omega) \rangle + \frac{\mu_0}{4} \langle \mathbf{H}(\mathbf{x}, \omega) \cdot \mathbf{H}^*(\mathbf{x}, \omega) \rangle \quad (10.68)$$

where “ $\langle \rangle$ ” denotes the ensemble average of the random currents. The emitted energy flux can be expressed by the ensemble average of the Poynting vector, i.e.,

$$\langle \mathbf{S}(\mathbf{x}, \omega) \rangle = \frac{1}{2} \langle \text{Re}[\mathbf{E}(\mathbf{x}, \omega) \times \mathbf{H}^*(\mathbf{x}, \omega)] \rangle \quad (10.69)$$

To evaluate the ensemble average, the required spatial correlation function between the fluctuating currents at two locations  $\mathbf{x}'$  and  $\mathbf{x}''$  inside the emitting medium is given as<sup>20</sup>

$$\langle j_m(\mathbf{x}', \omega) j_n^*(\mathbf{x}'', \omega) \rangle = \frac{4\omega\varepsilon_0 \text{Im}(\varepsilon)\Theta(\omega, T)}{\pi} \delta_{mn} \delta(\mathbf{x}' - \mathbf{x}'') \quad (10.70)$$

where  $j_m$  ( $m = 1, 2, \text{ or } 3$ ) stands for the  $x, y, \text{ or } z$  component of  $\mathbf{j}$ ,  $\delta_{mn}$  is the Kronecker delta function, and  $\delta(\mathbf{x}' - \mathbf{x}'')$  is the Dirac delta function. In Eq. (10.70),  $\Theta(\omega, T)$  is the mean energy of a Planck oscillator at the frequency  $\omega$  in thermal equilibrium and is given by

$$\Theta(\omega, T) = \frac{\hbar\omega}{\exp(\hbar\omega/k_B T) - 1} \quad (10.71)$$

In Eq. (10.71), the term  $\frac{1}{2}\hbar\omega$  that accounts for vacuum fluctuation is omitted since it does not affect the net radiation heat flux; see Milonni and Shih (*Am. J. Phys.*, **59**, 684, 1991) for a detailed discussion about the vacuum fluctuation or zero-point energy. The calculated energy density should be regarded as being relative to the vacuum ground energy density.<sup>61</sup> A factor of 4 has been included in Eq. (10.70) to be consistent with the conventional definitions of the spectral energy density and the Poynting vector expressed in Eq. (10.68) and Eq. (10.69), respectively, since only positive values of frequencies are considered here.<sup>20</sup> The local density of states or density of modes is defined by the following relation:<sup>61,62</sup>

$$u_{\omega}(z, \omega) = D(z, \omega)\Theta(\omega, T) \quad (10.72)$$

The energy density and density of states are independent of  $r$  because of the infinite-plate assumption. The physical significance of  $D(z, \omega)$  [ $1/(\text{m}^3 \cdot \text{rad/s})$ ] is the number of modes per unit frequency interval per unit volume. Equation (10.72) assumes that the contribution is only from the medium and did not consider the contribution from free space as well as that reflected by the interface. This omission is justifiable in the near-field regimes because the contribution from free space may be orders of magnitude smaller than that from the medium.

### 10.4.2 Heat Transfer between Parallel Plates

The Green function depends on the geometry of the physical system, and for two parallel semi-infinite media sketched in Fig. 10.32a, it takes the following form:

$$\mathbf{G}(\mathbf{x}, \mathbf{x}', \omega) = \int_0^{\infty} \frac{i}{4\pi\zeta_1} (\hat{\mathbf{s}}_t \hat{\mathbf{s}} + \hat{\mathbf{p}}_t \hat{\mathbf{p}}_p) e^{i(\xi z - \xi' z')} e^{i\xi(r-r')} \xi d\xi \quad (10.73)$$

where  $\mathbf{x} = r\hat{\mathbf{r}} + z\hat{\mathbf{z}}$  and  $\mathbf{x}' = r'\hat{\mathbf{r}} + z'\hat{\mathbf{z}}$ .<sup>19,59</sup> Note that  $t_s$  and  $t_p$  are the transmission coefficients from medium 1 to medium 2 for  $s$  and  $p$  polarizations, respectively, and can be calculated using Airy's formula given in Eq. (9.8). The unit vectors are  $\hat{\mathbf{s}} = \hat{\mathbf{r}} \times \hat{\mathbf{z}}$ ,  $\hat{\mathbf{p}}_1 = (\xi\hat{\mathbf{z}} - \zeta_1\hat{\mathbf{r}})/k_1$ , and  $\hat{\mathbf{p}}_2 = (\xi\hat{\mathbf{z}} - \zeta_2\hat{\mathbf{r}})/k_2$ . If the interest is to calculate the radiation

field from a medium to vacuum,  $t_s$  and  $t_p$  can be replaced by the Fresnel transmission coefficients between the medium and vacuum. The electric field can be calculated by substituting Eq. (10.70) and Eq. (10.73) into Eq. (10.66). The energy density, the energy flux, and the local density of states can then be calculated. The local density of states in vacuum near the surface of medium 1 can be expressed in two terms as follows:<sup>19,20</sup>

$$D(z, \omega) = D_{\text{prop}}(\omega) + D_{\text{evan}}(z, \omega) \quad (10.74)$$

$$\text{where} \quad D_{\text{prop}}(\omega) = \int_0^{\omega/c} \frac{\omega}{2\pi^2 c^2 \zeta_0} \left(2 - \rho_{01}^s - \rho_{01}^p\right) \xi d\xi \quad (10.75a)$$

$$\text{and} \quad D_{\text{evan}}(z, \omega) = \int_0^{\omega/c} \frac{e^{-2z\text{Im}(\zeta_0)}}{2\pi^2 \omega |\zeta_0|} \left[ \text{Im}(r_{01}^s) + \text{Im}(r_{01}^p) \right] \xi^3 d\xi \quad (10.75b)$$

Here,  $r_{01}$  is the Fresnel reflection coefficient and  $\rho_{01} = |r_{01}|^2$  is the (far-field) reflectivity at the interface between vacuum and medium 1, the superscripts  $s$  and  $p$  signify  $s$  polarization and  $p$  polarization, respectively. Note that  $r_{01}^s = (\zeta_0 - \zeta_1)/(\zeta_0 + \zeta_1)$  and  $r_{01}^p = (\zeta_0 - \zeta_1/\epsilon_1)/(\zeta_0 + \zeta_1/\epsilon_1)$ . It should be mentioned that, in deriving Eq. (10.74), the imaginary part of the permittivity of medium 1 in Eq. (10.70) has been combined with other terms. No matter how small  $\text{Im}(\epsilon_1)$  may be, such as for a dielectric, it must not be zero for the semi-infinite assumption to hold. The contribution of propagating waves given by Eq. (10.75a) is independent of  $z$  and exists in both near and far fields; whereas the contribution of evanescent waves decreases with increasing  $z$ . In the far-field limit, the contribution of the propagating waves is responsible for thermal emission, and one can see the directional-spectral emissivity terms, i.e.,  $\epsilon_{\omega,1}^s = 1 - \rho_{01}^s$  and  $\epsilon_{\omega,1}^p = 1 - \rho_{01}^p$  from Eq. (10.75a). As it gets closer and closer to the surface, the contribution of evanescent waves near the surface may dominate when  $\text{Im}(r_{01}^p)$  is large especially in the case of surface phonon polaritons. Subsequently, very large energy densities exist near the surface at that particular frequency.<sup>19,62</sup>

The spectral energy flux from medium 1 to medium 2 is calculated by projecting the time-averaged Poynting vector from Eq. (10.69) into the  $z$  direction; thus,

$$q''_{\omega,1-2} = \frac{\Theta(\omega, T_1)}{\pi^2} \int_0^{\omega/c} Z_{12}(\omega, \xi) \xi d\xi \quad (10.76)$$

where

$$Z_{12}(\omega, \xi) = \frac{4\text{Re}(\zeta_1)\text{Re}(\zeta_2)|\zeta_0^2 e^{2i\zeta_0 d}|}{|(\zeta_0 + \zeta_1)(\zeta_0 + \zeta_2)(1 - r_{01}^s r_{02}^s e^{2i\zeta_0 d})|^2} + \frac{4\text{Re}(\epsilon_1 \zeta_1^*)\text{Re}(\epsilon_2 \zeta_2^*)|\zeta_0^2 e^{2i\zeta_0 d}|}{|(\epsilon_1 \zeta_0 + \zeta_1)(\epsilon_2 \zeta_0 + \zeta_2)(1 - r_{01}^p r_{02}^p e^{2i\zeta_0 d})|^2}$$

Here,  $Z_{12}(\omega, \xi)$  can be regarded as an *exchange function*, which provides information on the contribution to the spectral energy flux at a given  $\xi$ . Equation (10.76) includes the contributions from both propagating and evanescent waves. The expression of  $q''_{\omega,2-1}$  is readily obtained by replacing  $\Theta(\omega, T_1)$  with  $\Theta(\omega, T_2)$  since the exchange function is reciprocal, namely,  $Z_{12}(\omega, \xi) = Z_{21}(\omega, \xi)$ . The net total energy flux is the integration of  $q''_{\omega,1-2} - q''_{\omega,2-1}$  over all frequencies, viz.,

$$q''_{\text{net}} = \int_0^{\infty} (q''_{\omega,1-2} - q''_{\omega,2-1}) d\omega = \frac{1}{\pi^2} \int_0^{\infty} [\Theta(\omega, T_1) - \Theta(\omega, T_2)] \int_0^{\omega/c} Z_{12}(\omega, \xi) \xi d\xi d\omega \quad (10.77)$$

which provides an *ab initio* calculation of the thermal radiation that is applicable for both the near- and far-field heat transfer. The contribution of evanescent waves with imaginary  $\zeta_0$  (for  $\xi > \omega/c$ ) reduces as  $d$  increases and is negligible when  $d$  is on the order of the wavelength. The energy transfer can also be separated into contributions of propagating waves and coupled evanescent waves (i.e., photon tunneling).

The exchange function  $Z$  can be rewritten using the Fresnel coefficients and reflectivity for propagating waves as

$$Z_{\text{prop}}(\omega, \xi) = \frac{(1 - \rho_{01}^s)(1 - \rho_{02}^s)}{4|1 - r_{01}^s r_{02}^s e^{-2i\zeta_0 d}|^2} + \frac{(1 - \rho_{01}^p)(1 - \rho_{02}^p)}{4|1 - r_{01}^p r_{02}^p e^{-2i\zeta_0 d}|^2} \quad \text{for } \xi < \omega/c \quad (10.78)$$

Substituting Eq. (10.78) into Eq. (10.76) and noting that  $\xi = (\omega/c) \sin \theta$ , where  $\theta$  is the polar angle in vacuum, we can evaluate the integration from  $\theta = 0$  to  $\pi/2$ , by averaging the oscillation terms to obtain the far-field and incoherent limit ( $d \gg \lambda$ ):  $|1 - r_{01}^s r_{02}^s e^{-2i\zeta_0 d}|^2 \rightarrow (1 - \rho_{01}^s \rho_{02}^s)$ . It can also be shown that  $(1 - \rho_{01}^s \rho_{02}^s)/[(1 - \rho_{01}^s)(1 - \rho_{02}^s)] = 1/\epsilon_{\omega,1}' + 1/\epsilon_{\omega,2}' - 1$ , which also holds for  $p$  polarization. The total energy flux in the far-field limit becomes

$$q''_{\text{net, far}} = \frac{1}{4\pi^2 c^2} \int_0^\infty [\Theta(\omega, T_1) - \Theta(\omega, T_2)] \omega^2 d\omega \int_0^{\pi/2} \cos \theta \sin \theta d\theta \quad (10.79)$$

$$\times \left( \frac{1}{1/\epsilon_{\omega,1}' + 1/\epsilon_{\omega,2}' - 1} + \frac{1}{1/\epsilon_{\omega,1}^p + 1/\epsilon_{\omega,2}^p - 1} \right)$$

which is similar to the equation found in radiation heat transfer texts, except that angular frequency is used here instead of wavelength. The wavelength integration can be obtained by converting blackbody intensity from  $\hbar\omega^3/[4\pi^3 c^2(e^{\hbar\omega/k_B T} - 1)]d\omega$  to  $2hc^2/[\lambda^5(e^{hc/\lambda k_B T} - 1)]d\lambda$ . While the energy flux includes the contributions by both polarizations, one should integrate for the two polarizations separately according to Eq. (10.79). The expression of  $Z$  for the contribution of evanescent waves is

$$Z_{\text{evan}}(\omega, \xi) = \frac{\text{Im}(r_{01}^s)\text{Im}(r_{02}^s)e^{-2\text{Im}(\zeta_0)d}}{|1 - r_{01}^s r_{02}^s e^{-2\text{Im}(\zeta_0)d}|^2} + \frac{\text{Im}(r_{01}^p)\text{Im}(r_{02}^p)e^{-2\text{Im}(\zeta_0)d}}{|1 - r_{01}^p r_{02}^p e^{-2\text{Im}(\zeta_0)d}|^2} \quad \text{for } \xi > \omega/c \quad (10.80)$$

Clearly, the exchange function decays exponentially as the distance of separation  $d$  increases.

### 10.4.3 Asymptotic Formulation

At the nanometer scale, when near-field radiation dominates, especially for metallic media, doped silicon, or polar materials in the absorption band, the exchange factor from Eq. (10.80) can be expressed with an approximate formula. Note that at  $\xi \gg \omega/c$ , we have  $\zeta_1 \approx \zeta_2 \approx \zeta_0 \approx i\xi$  so that  $r_{01}^s$  and  $r_{02}^s$  are negligibly small, and the contribution of TE waves can be ignored. Furthermore,  $r_{01}^p \approx \frac{\epsilon_1 - 1}{\epsilon_1 + 1}$  and  $r_{02}^p \approx \frac{\epsilon_2 - 1}{\epsilon_2 + 1}$  are independent of  $\xi$ ; therefore,

$$Z_{\text{evan}}(\omega, \xi) \approx \frac{\text{Im}(r_{01}^p)\text{Im}(r_{02}^p)e^{-2\xi d}}{|1 - r_{01}^p r_{02}^p e^{-2\xi d}|^2} \quad (10.81)$$

Using the relation:  $\text{Im}\left(\frac{\varepsilon - 1}{\varepsilon + 1}\right) = \frac{2\text{Im}(\varepsilon)}{|\varepsilon + 1|^2}$ , the spectral heat flux from 1 to 2 in the limit  $d \rightarrow 0$  can then be expressed as

$$q''_{\omega,1-2} \approx \frac{\Theta(\omega, T_1)}{\pi^2 d^2} \frac{\text{Im}(\varepsilon_1)\text{Im}(\varepsilon_2)}{|\varepsilon_1 + 1)(\varepsilon_2 + 1)|^2} \int_{x_0}^{\infty} \left| 1 - \frac{(\varepsilon_1 - 1)(\varepsilon_2 - 1)}{(\varepsilon_1 + 1)(\varepsilon_2 + 1)} e^{-x} \right|^{-2} x e^{-x} dx$$

where  $x_0 = 2d\omega/c$ . The heat flux will be inversely proportional to  $d^2$  in the proximity limit. The integral approaches 1 when  $\left| \frac{(\varepsilon_1 - 1)(\varepsilon_2 - 1)}{(\varepsilon_1 + 1)(\varepsilon_2 + 1)} \right| \ll 1$ ; consequently, the net spectral flux becomes

$$q''_{\omega,1-2} - q''_{\omega,2-1} \approx \frac{1}{\pi^2 d^2} \frac{\text{Im}(\varepsilon_1)\text{Im}(\varepsilon_2)}{|\varepsilon_1 + 1)(\varepsilon_2 + 1)|^2} \left[ \Theta(\omega, T_1) - \Theta(\omega, T_2) \right] \quad (10.82)$$

When  $\xi \gg \omega/c$ , Eq. (10.75b) reduces to  $D_{\text{evan}}(z, \omega) \approx \frac{1}{\pi^2 \omega} \frac{\text{Im}(\varepsilon_1)}{|\varepsilon_1 + 1|^2} \int_{\omega/c}^{\infty} e^{-2\xi z} \xi^2 d\xi$ . By evaluating the integration and keeping the highest-order terms only, one obtains the following asymptotical expression for  $z \rightarrow 0$  as

$$D_{\text{evan}}(z, \omega) \approx \frac{1}{4\pi^2 \omega z^3} \frac{\text{Im}(\varepsilon_1)}{|\varepsilon_1 + 1|^2} \quad (10.83)$$

This equation suggests that, as  $z$  decreases, the near-field density of states increases with  $z^{-3}$  and is localized at the surface.

#### 10.4.4 Nanoscale Radiation Heat Transfer between Doped Silicon

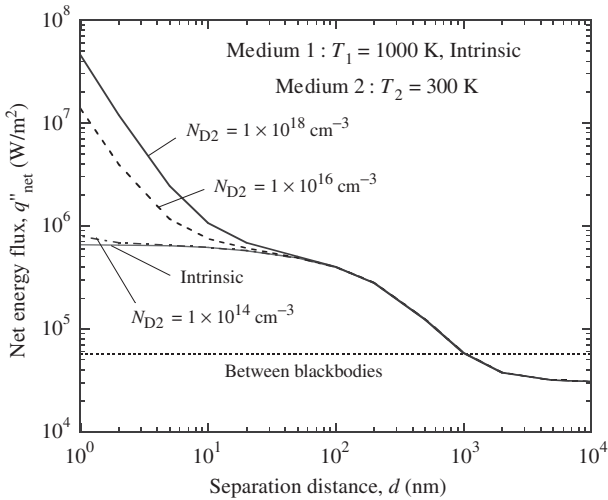
Radiation heat transfer may be important when their characteristic dimensions are on the nanometer scales. AFM cantilevers with integrated heaters and with nanoscale sharp tips made of doped silicon have been developed for thermal writing and reading.<sup>58</sup> These heated cantilever tips may provide local heating for the study of radiative energy transfer between two objects separated by a few nanometers. It is critical to quantitatively predict the near-field radiation heat flux between doped silicon. The dielectric function of doped silicon can be described by the Drude model, considering the effects of temperature and doping level on the concentrations and scattering times of electrons and holes, as described in Sec. 8.4.4 of Chap. 8.

To calculate the radiative energy flux, it is essential to evaluate the integration of the exchange function  $Z_{12}(\omega, \xi)$  over the wavevector  $\xi$  ranging from 0 to infinity and the integration of the spectral energy flux over all frequencies. The integration over  $\xi$  from 0 to  $\omega/c$  corresponds to radiation heat transfer by propagating waves. In this range, the integrand exhibits highly oscillatory behavior for large  $d$ . In this regard, Simpson's rule is an effective technique in dealing with oscillatory integrands. The integration for  $\xi$  from  $\omega/c$  to infinity corresponds to radiation heat transfer by evanescent waves, and the exchange factor is given as  $Z_{\text{evan}}(\omega, \xi)$ . For small  $d$  values, the upper limit  $\xi_{\text{max}}$  should be on the order of  $1/d$ ; but for large  $d$  values,  $1/d$  would be less than  $\omega/c$ . A semi-empirical criterion can be used to set  $\xi_{\text{max}}$  as  $3/d$  or  $100\omega/c$ , whichever is larger, to ensure an integration error less than 1%. An effective way to perform the integration is to break it into several parts and evaluate each part using Simpson's rule. For example, the integration can be carried out in two parts,  $\omega/c < \xi < 6\omega/c$  and  $6\omega/c < \xi < \xi_{\text{max}}$ . A relative difference of 0.1% may be used as the convergence criterion between consecutive iterations. For conventional radiation heat transfer calculations, the lower and upper bounds of the integration over frequency



(or wavelength) can be selected such that 99% of the blackbody emissive power falls between the limits. For example, 99% of blackbody radiation emissive power is concentrated between 1.2 and 25  $\mu\text{m}$  at 1000 K, and between 4 and 85  $\mu\text{m}$  at 300 K. The enhancement of near-field radiation heat transfer is generally greater at longer wavelengths, and as such, the integration should be performed over a much broader spectral region.

Figure 10.33 shows the predicted radiation heat transfer between two silicon plates. Medium 1 is intrinsic silicon at  $T_1 = 1000$  K, whereas medium 2 is at  $T_2 = 300$  K whose



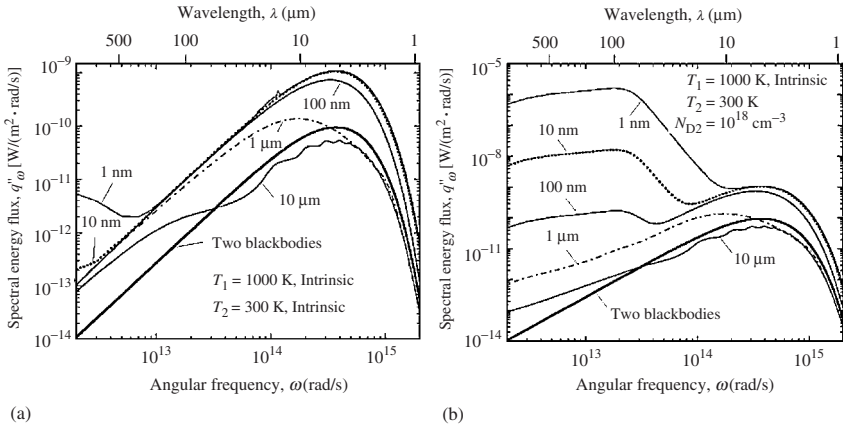
**FIGURE 10.33** Net energy flux between intrinsic silicon, medium 1 at 1000 K and medium 2 at 300 K, with different doping levels.<sup>20</sup>

doping levels vary from intrinsic to heavily doped silicon with phosphorus as the donor ( $n$ -type). In the calculations, the wavelength region was chosen in the range from 0.94 to about 1880  $\mu\text{m}$  ( $\omega$  range from  $10^{12}$  to  $2 \times 10^{15}$  rad/s). The dotted line represents the far-field radiation heat flux between two blackbodies,  $\sigma_{\text{SB}}(T_1^4 - T_2^4)$ , as predicted by the Stefan-Boltzmann law. Wien's displacement law suggests that the dominant wavelength  $\lambda_{\text{mp}}$  for the 1000-K emitter is around 3  $\mu\text{m}$ . The energy flux is essentially a constant when the distance  $d$  is greater than 10  $\mu\text{m}$ , which is the far-field regime. The net energy flux increases quickly when  $d < \lambda_{\text{mp}}$  due to photon tunneling. When medium 2 is intrinsic or lightly doped, i.e.,  $N_{\text{D}2} < 10^{15} \text{ cm}^{-3}$ , the maximum  $q''_{\text{net}}$  is achieved when  $d < 50$  nm. The maximum net energy flux is 21.3 times that of the far-field limit and 11.7 times that of blackbodies for intrinsic silicon, as predicted earlier when the silicon plates are treated as dielectrics. On the other hand,  $q''_{\text{net}}$  for  $N_{\text{D}2} > 10^{16} \text{ cm}^{-3}$  continues to increase as  $d$  is reduced and does not saturate. Note that the results for  $N_{\text{D}2}$  ranging from  $10^{17}$  to  $10^{20} \text{ cm}^{-3}$  are very similar to that for  $N_{\text{D}2} = 10^{18} \text{ cm}^{-3}$ . The heat flux at  $d = 1$  nm with  $N_{\text{D}2} = 10^{18} \text{ cm}^{-3}$  is 800 times greater than that between two blackbodies.

If one of the media is a slightly absorbing dielectric, as for silicon with a carrier concentration less than  $10^{15} \text{ cm}^{-3}$ , the Fresnel coefficients beyond the critical angle become imaginary. There is a propagating wave in the medium and an evanescent wave in vacuum (corresponding to frustrated total internal reflection). If the refractive index of the

dielectric medium is  $n$ , then  $Z_{\text{evan}}(\omega, \xi)$  will be nonzero for  $\omega/c < \xi < n\omega/c$ . However, because the extinction coefficient  $\kappa$  is negligibly small,  $Z_{\text{evan}}(\omega, \xi)$  will be very small beyond  $n\omega/c$  and will decay exponentially with increasing  $\xi$ . Therefore, for lightly doped silicon, the enhancement is limited to approximately  $(n^2 - 1)\sigma_{\text{SB}}(T_1^4 - T_2^4)$  and the near-field flux becomes  $q''_{\text{net}} \approx n^2\sigma_{\text{SB}}(T_1^4 - T_2^4)$ , as discussed previously. Because of the small difference between the refractive indices of the two media,  $n$  is used here for both media for simplicity. On the other hand, if  $\kappa$  is large, the integration over  $\xi > n\omega/c$  will have a significant contribution to the heat flux and may even dominate the heat flux when  $d$  reaches a few nanometers.

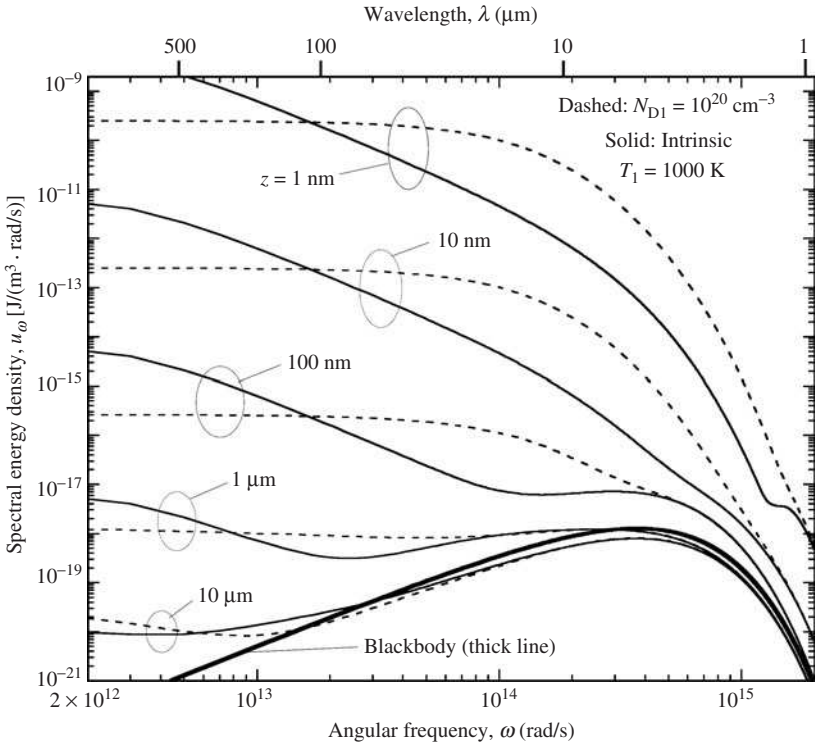
The enhancement of near-field heat transfer can be better understood from the energy flux spectra shown in Fig. 10.34. The units of  $q''_{\omega}$  is expressed as  $\text{W}/(\text{m}^2 \cdot \text{rad/s})$  rather than



**FIGURE 10.34** Spectral energy flux for different separation distances between silicon plates, where medium 1 is always intrinsic,  $T_1 = 1000 \text{ K}$ , and  $T_2 = 300 \text{ K}$ .<sup>20</sup> (a) Medium 2 is intrinsic. (b) Medium 2 is  $n$ -type silicon with a donor concentration  $N_{\text{D}2} = 10^{18} \text{ cm}^{-3}$ .

$\text{J}/(\text{m}^2 \cdot \text{rad})$  to keep the integrity of the angular frequency units, i.e.,  $\text{rad/s}$ . Notice that at  $1000 \text{ K}$ , the carrier concentration is about  $10^{18} \text{ cm}^{-3}$ . The spectral flux between two blackbodies at  $1000$  and  $300 \text{ K}$ , calculated from Planck's spectral emissivity power, is also shown for comparison. Interference becomes important at  $d = 10 \mu\text{m}$  and causes the wavy features in the spectral energy flux. When the receiver is intrinsic, as Fig. 10.34a reveals, the shape of the spectrum is similar for  $d < 100 \text{ nm}$  and scaled up with  $n^2$  ( $\approx 11.7$ ) times that of the blackbody. However, the slightly increased  $\kappa$  due to phonon absorption and, in the far-infrared, due to free carriers can result in an increase in the spectral energy flux, while the increment is not significant enough to vary the total flux. The near-field spectral flux is greatly enhanced with doping, as can be seen from Fig. 10.34b, especially in the far-infrared region. As mentioned earlier, the increased energy flux in the longer wavelengths requires the integration to be carried out much broader than that of blackbody spectrum. The mechanism underlying the nanoscale enhancement is discussed next. When Eq. (10.82) is used to calculate the spectral energy flux at  $d = 1 \text{ nm}$ , the predicted values are nearly half of those obtained by integration in the frequency region from  $10^{12}$  to  $10^{14} \text{ rad/s}$  for the case shown in Fig. 10.34b. This is because Eq. (10.81) is not applicable for  $\omega > 10^{14} \text{ rad/s}$ , where the major contribution of evanescent waves comes for  $\omega/c < \xi < n\omega/c$ , i.e., propagating waves in silicon. Therefore, care must be taken in applying the asymptotic expression given in Eq. (10.82).

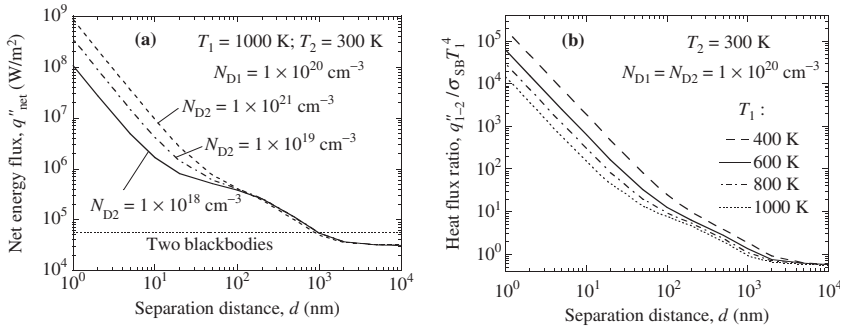
The enhanced thermal radiation can also be understood from the very large energy density in the vicinity of the surface. The spectral energy density  $u(z, \omega)$  near the surface of medium 1 at 1000 K is evaluated using Eq. (10.68), and the results are shown in Fig. 10.35



**FIGURE 10.35** Spectral energy density near semi-infinite silicon at  $T_1 = 1000$  K at different distances from the surface for intrinsic and doped silicon.<sup>20</sup>

when the medium is either intrinsic or doped silicon with  $N_{D1} = 10^{20} \text{ cm}^{-3}$ . The energy density of a blackbody enclosure is also shown for comparison. As the height  $z \leq 100 \text{ nm}$ , the energy density is greatly enhanced because the local density of states increases rapidly as  $z$  decreases toward the nanometer regime. Furthermore, the maximum appears at a different wavelength when compared to the maximum for Planck's blackbody distribution function. The dependence of  $u_\omega$  on doping can be understood by examining the term  $\text{Im}(\epsilon_1)/|\epsilon_1 + 1|^2$  in Eq. (10.83). Because the plasma frequency of silicon is pushed toward shorter wavelengths by increasing the carrier concentration,  $\text{Im}(\epsilon_1)/|\epsilon_1 + 1|^2$  increases significantly in the spectral region from 2 to 100  $\mu\text{m}$ . Although  $u_\omega$  for intrinsic silicon is larger toward longer wavelengths, the contribution to the total energy density is very small when  $\omega < 10^{13} \text{ rad/s}$ . Hence, the near-field energy flux may increase further when the emitter is heavily doped.

Figure 10.36 plots the radiation heat transfer between heavily doped silicon. In Fig. 10.36a, the 1000-K emitter is assumed to have a fixed dopant concentration of  $N_{D1} = 10^{20} \text{ cm}^{-3}$ , while the dopant concentration of the 300-K receiver varies from  $N_{D2} = 10^{18} \text{ cm}^{-3}$  to  $N_{D2} = 10^{21} \text{ cm}^{-3}$ . The result for  $N_{D2} = 10^{20} \text{ cm}^{-3}$  (not shown) is



**FIGURE 10.36** Net energy flux between heavily doped silicon.<sup>20</sup> (a)  $q''_{\text{net}}$  versus  $d$  for several  $N_{D2}$  values when  $N_{D1} = 10^{20}$  cm<sup>-3</sup>, at  $T_1 = 1000$  K and  $T_2 = 300$  K. (b) The energy flux from medium 1 to medium 2, normalized by  $\sigma T_1^4$  at different  $T_1$  for  $T_2 = 300$  K.

slightly lower but very close to that for  $N_{D2} = 10^{21}$  cm<sup>-3</sup>. As expected, the nanoscale thermal radiation is enormous. For example, at  $d = 1$  nm,  $q''_{\text{net}} \sim 10^9$  W/m<sup>2</sup>, which is 15,000 times greater than that between two blackbodies. The effect of source temperature on the net energy flux is investigated by varying  $T_1$ , while  $T_2$  is fixed at 300 K. The normalized energy flux from medium 1 to medium 2, with respect to that of blackbody emissive power at  $T_1$ , is shown in Fig. 10.36b with dopant concentrations  $N_{D1} = N_{D2} = 10^{20}$  cm<sup>-3</sup>. The enhancement is greater when  $T_1$  is closer to 300 K and reaches  $\approx 2 \times 10^5$  when  $T_1 = 400$  K, at  $d = 1$  nm. Even at  $d = 10$  nm, the enhancement is significant for a number of applications, such as enhanced heating and cooling for thermal control and energy conversion near room temperature. Using different parameters in the Drude model of heavily doped silicon, Marquier et al. (*Opt. Commun.*, **237**, 379, 2004) predicted similar enhancement of near-field radiation heat transfer for heavily doped silicon. Measurements have rarely been performed on the infrared properties of heavily doped silicon.

For SiC, the enhanced near-field radiation has been attributed to surface phonon polaritons.<sup>19,59,60</sup> For doped silicon, the plasma frequency is in the infrared region. Because of the large scattering rate  $1/\tau$ ,  $\text{Re}(\epsilon)$  may never be negative in the infrared; even though it becomes negative at some frequencies, the magnitude would be much smaller than  $\text{Im}(\epsilon)$ . Therefore, the enhancement of nanoscale radiation may be understood from the large values of the exchange function around the plasma frequency where  $\text{Im}(\epsilon)/|1 + \epsilon|^2$  is large in both media. Although surface plasmon resonance condition is not satisfied, evanescent waves are present in vacuum as well as in the media for sufficiently large  $\xi$  values. The near-field energy flux spectrum for Si exhibits a broader peak when the doping concentration is less than  $10^{18}$  cm<sup>-3</sup>, as can be seen from Fig. 10.34. The spectral width of the heat flux peak decreases with increasing doping level but is still much broader than that for SiC.

**Example 10-8.** At what distance  $d$ , would the nanoscale thermal radiation between two plates at  $T_1 = 400$  K and  $T_2 = 300$  K, exceed that of heat conduction by air at the pressure  $P = 1$  atm?

**Solution.** When  $d$  is much smaller than the mean free path, which is about 70 nm at standard atmospheric conditions, boundary scattering or ballistic scattering dominates gas conduction. The thermal conductivity decreases linearly as  $d$  decreases, whereas the heat flux is independent of  $d$  in this regime. Assuming a thermal accommodation coefficient of 1, the heat transfer by gas conduction can be estimated from the theory in Chap. 4, Eq. (4.93), as

$$q''_{\text{cond}} = \frac{c_v(\gamma + 1)P}{(8\pi RT_m)^{1/2}}(T_1 - T_2)$$

where  $R$  is the ideal gas constant,  $P$  is the pressure,  $T_m = 4T_1T_2/(\sqrt{T_1} + \sqrt{T_2})^2$  is a mean temperature, and  $c_v$  is the specific heat at constant volume evaluated at  $T_m$ . The resulting  $q''_{\text{cond}}$  for air at a pressure  $P = 1$  atm is approximately  $1.1 \times 10^7$  W/m<sup>2</sup> for  $T_1 = 400$  K and  $T_2 = 300$  K. The calculated near-field net energy transfer by radiation is at the same level when  $d = 3$  nm with heavily doped silicon. At  $d = 1$  nm, the near-field radiation heat transfer can be an order of magnitude greater than the heat transfer by air conduction at the atmospheric pressure. Because the conduction heat flux further decreases as the pressure is reduced, nanoscale thermal radiation may dominate the heat transfer process for scanning thermal probes and heated cantilever tips using heavily doped silicon.

The radiation heat transfer coefficient can be defined as  $h_r = q''_{\text{net}}/(T_1 - T_2)$  in analogy to Newton's law of cooling. It can be seen from Fig. 10.36a that for heavily doped silicon,  $h_r \sim 10^6$  W/(m<sup>2</sup> · K) at  $d = 1$  nm and  $h_r \sim 10^4$  W/(m<sup>2</sup> · K) at  $d = 10$  nm. It is important to verify whether the local-equilibrium assumption is valid. Assume that the near-field radiation penetration depth is 100 nm and the thermal conductivity for doped silicon is 100 W/(m · K). For a heat flux of  $10^9$  W/m<sup>2</sup>, the temperature drop would be 1 K within the radiation penetration depth. Therefore, the local-equilibrium assumption should still be valid. However, for a wafer of 100- $\mu$ m thickness, the temperature drop would be 1000 K. The preceding calculations suggest that indeed near-field radiation can be an effective way of heating and cooling. As an alternative to the parallel-plate configuration, it is possible to pattern one of the silicon wafers with a 2-D array of truncated cones or pyramids to remove heat locally for thermal control in nanoelectronics, for example.

Experimental investigations of near-field radiative energy transfer are very limited. Tien and coworkers<sup>63</sup> and Hargreaves<sup>64</sup> were among the first to measure the energy flux of two parallel plates at cryogenic temperatures. Kutateladze et al. performed similar measurements.<sup>65</sup> The smallest separation distance was about 1.5  $\mu$ m. Xu et al. used an STM stage with an indium needle that has a flat tip surface of 100- $\mu$ m diameter.<sup>56</sup> Müller-Hirsch et al. investigated the heat transfer between a tungsten tip and a planar thermocouple in the substrate by cooling the substrate.<sup>57</sup> While the proximity effect was observed at distances down to 10 nm or so, it was difficult to quantitatively determine the absolute heat flux between the tip and the substrate, as well as to accurately measure the temperatures of the tip and the substrate.<sup>56,57</sup> Further research is needed to quantitatively demonstrate near-field heat transfer enhancement with different materials, including doped silicon, at the nanoscale distances. Another interesting question is the coupling between near-field radiation and thermionic and field-emission effects for cooling and direct energy conversion.

## 10.5 SUMMARY

This last chapter of the book, while a little bit long, described a field that is clearly associated with radiation heat transfer but with its foundations deeply within physical optics, electrodynamics, and perhaps quantum electrodynamics and quantum mechanics. The research has so far largely been performed in the areas of nanooptics, nanophotonics, nanomaterials, and nanooptoelectronics. The aspects of fabrication and specific applications were not sufficiently covered, because the focus has been on the fundamental mechanisms that are generally applicable to any given devices or systems. Nevertheless, through the references cited at the end of the chapter, together with those cited in the texts, readers will be able to access the literature database to obtain further information about any specific topic. It is hoped that this chapter will bridge gaps between different disciplines and provide a solid foundation for the readers to appreciate this exciting and dynamic field, which is believed to be a new frontier in nanoscale heat transfer.

## REFERENCES

1. D. W. Pohl, W. Denk, and M. Lanz, "Optical stethoscopy: Image recording with resolution  $\lambda/20$ ," *Appl. Phys. Lett.*, **44**, 651–653, 1984.
2. A. Lewis, M. Isaacson, A. Harootunian, and A. Muray, "Development of a 500 Å spatial resolution light microscope," *Ultramicroscopy*, **13**, 227–232, 1984.
3. E. Betzig and R. J. Chichester, "Single molecules observed by near-field scanning optical microscopy," *Science*, **262**, 1422–1425, 1993; E. Betzig, J. K. Trautman, R. Wolfe, et al., "Near-field magneto-optics and high density storage," *Appl. Phys. Lett.*, **61**, 142–144, 1992.
4. Y. F. Lu, B. Hu, Z. H. Mai, W. J. Wang, W. K. Chim, and T. C. Chong, "Laser-scanning probe microscope based nanoprocessing of electronics materials," *Jpn. J. Appl. Phys.*, **40**, 4395–4398, 2001.
5. A. Chimmalgi, G. P. Grigoropoulos, and K. Komvopoulos, "Surface nanostructuring by nano/femtosecond laser-assisted scanning force microscopy," *J. Appl. Phys.*, **97**, 104319, 2005.
6. H. A. Haus, *Waves and Fields in Optoelectronics*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
7. J. A. Kong, *Electromagnetic Wave Theory*, 2nd ed., Wiley, New York, 1990.
8. H. K. V. Lotsch, "Beam displacement at total reflection: The Goos-Hänchen effect, I, II, III, IV," *Optik*, **32**, (2) 116–137; (3) 189–204; (4) 299–319; (6) 553–569, 1970–1971.
9. A. Puri and J. L. Birman, "Goos-Hänchen beam shift at total internal reflection with application to spatially dispersive media," *J. Opt. Soc. Am. A*, **3**, 543–549, 1986.
10. D.-K. Qing and G. Chen, "Goos-Hänchen shifts at the interfaces between left- and right-handed media," *Opt. Lett.*, **29**, 872–874, 2004; X. L. Hu, Y. D. Huang, W. Zhang, D.-K. Qing, and J. D. Peng, "Opposite Goos-Hänchen shifts for transverse-electric and transverse-magnetic beams at the interface associated with single-negative materials," *Opt. Lett.*, **30**, 899–901, 2005.
11. H. M. Lai, F. C. Cheng, and W. K. Tang, "Goos-Hänchen effect around and off the critical angle," *J. Opt. Soc. Am. A*, **3**, 550–557, 1986.
12. I. V. Shadrivov, A. A. Zharov, and Y. S. Kivshar, "Giant Goos-Hänchen effect at the reflection from left-handed metamaterials," *Appl. Phys. Lett.*, **83**, 2713–2715, 2003.
13. X. Chen and C.-F. Li, "Lateral shift of the transmitted light beam through a left-handed slab," *Phys. Rev. E*, **69**, 066617, 2004.
14. K. J. Vahala, "Optical microcavities," *Nature*, **424**, 839–846, 2003.
15. Z. Guo and H. Quan, "Energy transfer to optical microcavities with waveguides," *J. Heat Transfer*, **129**, 44–52, 2007.
16. R. F. Cregan, B. J. Mangan, J. C. Knight, et al., "Single-mode photonic band gap guidance of light in air," *Science*, **285**, 1537–1539, 1999; P. St. J. Russell, "Photonic crystal fibers," *Science*, **299**, 358–362, 2003.
17. E. G. Cravalho, C. L. Tien, and R. P. Caren, "Effect of small spacing on radiative transfer between two dielectrics," *J. Heat Transfer*, **89**, 351–358, 1967; C. L. Tien and G. R. Cunnington, "Cryogenic insulation heat transfer," *Adv. Heat Transfer*, **9**, 349–417, 1973.
18. M. D. Whale and E. G. Cravalho, "Modeling and performance of microscale thermophotovoltaic energy conversion devices," *IEEE Trans. Energy Conversion*, **17**, 130–142, 2002.
19. J.-P. Mulet, K. Joulain, R. Carminati, and J.-J. Greffet, "Nanoscale radiative heat transfer between a small particle and a plane surface," *Appl. Phys. Lett.*, **78**, 2931–2933, 2001; J.-P. Mulet, K. Joulain, R. Carminati, and J.-J. Greffet, "Enhanced radiative heat transfer at nanometric distance," *Microscale Thermophys. Eng.*, **6**, 209–222, 2002; G. Dominges, S. Volz, K. Joulain, and J.-J. Greffet, "Heat transfer between two nanoparticles through near field interaction," *Phys. Rev. Lett.*, **94**, 085901, 2005.
20. C. J. Fu and Z. M. Zhang, "Nanoscale radiation heat transfer for silicon at different doping levels," *Int. J. Heat Mass Transfer*, **49**, 1703–1718, 2006.
21. R. Y. Chiao and A. M. Steinberg, "Tunneling times and superluminality," *Progress in Optics*, **37**, 345–405, 1997.
22. P. Yeh, "Resonant tunneling of electromagnetic radiation in superlattice structures," *J. Opt. Soc. Am. A*, **2**, 568–571, 1985; P. Yeh, *Optical Waves in Layered Media*, Wiley, New York, 1988.

23. Z. M. Zhang and C. J. Fu, "Unusual photon tunneling in the presence of a layer with a negative refractive index," *Appl. Phys. Lett.*, **80**, 1097–1099, 2002; C. J. Fu and Z. M. Zhang, "Transmission enhancement using a negative-refraction layer," *Microscale Thermophys. Eng.*, **7**, 221–234, 2003.
24. C. J. Fu, Z. M. Zhang, and D. B. Tanner, "Energy transmission by photon tunneling in multilayer structures including negative index materials," *J. Heat Transfer*, **127**, 1046–1052, 2005.
25. S. Kawata (ed.), *Near-Field Optics and Surface Plasmon Polaritons*, Springer, Berlin, 2001.
26. J. Tominaga and D.P. Tsai (eds.), *Optical Nanotechnologies – The Manipulation of Surface and Local Plasmons*, Springer, Berlin, 2003.
27. J. Homola, S. S. Yee, and G. Gauglitz, "Surface plasmon resonance sensors: Review," *Sensors and Actuators B*, **54**, 3–15, 1999.
28. J.-J. Greffet, R. Carminati, K. Joulain, J.-P. Mulet, S. Mainguy, and Y. Chen, "Coherent emission of light by thermal sources," *Nature*, **416**, 61–64, 2002; F. Marquier, K. Joulain, J.-P. Mulet, R. Carminati, J.-J. Greffet, and Y. Chen, "Coherent spontaneous emission of light by thermal sources," *Phys. Rev. B*, **69**, 155412, 2004.
29. R. Hillenbrand, T. Taubner, and F. Kellmann, "Phonon-enhanced light-matter interaction at the nanometer scale," *Nature*, **418**, 159–162, 2002; R. Hillenbrand, "Towards phonon photonics: Scattering-type near-field optical microscopy reveals phonon-enhanced near-field interaction," *Ultramicroscopy*, **100**, 421–427, 2004.
30. H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer-Verlag, Berlin, 1988.
31. R. Rupin, "Surface polaritons of a left-handed medium," *Phys. Lett. A*, **277**, 61–64, 2000; R. Rupin, "Surface polaritons of a left-handed material slab," *J. Phys.: Condens. Matter*, **13**, 1811–1819, 2001.
32. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, New York, 1983.
33. G. Videen, M. M. Aslan, and M. P. Mengüç, "Characterization of metallic nano-particles via surface wave scattering: A. Theoretical framework and formulation," *J. Quant. Spectrosc. Radiat. Transfer*, **93**, 195–206, 2005; M. M. Aslan, M. P. Mengüç, and G. Videen, "Characterization of metallic nano-particles via surface wave scattering: B. Physical concept and numerical experiments," *idem*, **93**, 207–217, 2005; P. P. Venkata, M. M. Aslan, M. P. Mengüç, and G. Videen, "Surface plasmon scattering by gold nanoparticles and two-dimensional agglomerates," *J. Heat Transfer*, **129**, 60–70, 2007.
34. E. N. Economou, "Surface plasmons in thin films," *Phys. Rev.*, **182**, 539–554, 1969.
35. K. Park, B. J. Lee, C. J. Fu, and Z. M. Zhang, "Study of the surface and bulk polaritons with a negative index metamaterial," *J. Opt. Soc. Am. B*, **22**, 1016–1023, 2005.
36. A. Alu and N. Engheta, "Pairing an epsilon-negative slab with a mu-negative slab: Resonance, tunneling and transparency," *IEEE Trans. Antennas Propag.*, **51**, 2558–2571, 2003.
37. C. J. Fu, Z. M. Zhang, and D. B. Tanner, "Planar heterogeneous structures for coherent emission of radiation," *Opt. Lett.*, **30**, 1873–1875, 2005.
38. J. B. Pendry and A. MacKinnon, "Calculation of photon dispersion relations," *Phys. Rev. Lett.*, **69**, 2772–2775, 1992; J. B. Pendry, "Photonic band structures," *J. Mod. Opt.*, **41**, 209–229, 1994; J. B. Pendry, "Calculating photonic band structures," *J. Phys.: Condens. Matter*, **8**, 1085–1108, 1996.
39. A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd ed., Artech House, Boston, MA, 2005.
40. B. Lichtenberg and N. C. Gallagher, "Numerical modeling of diffractive devices using the finite element method," *Opt. Eng.*, **33**, 3518–3526, 1994.
41. D. W. Prather, M. S. Mirotznik, and J. N. Mait, "Boundary integral methods applied to the analysis of diffractive optical elements," *J. Opt. Soc. Am. A*, **14**, 34–43, 1997; D. W. Prather, J. N. Mait, M. S. Mirotznik, and J. P. Collins, "Vector-based synthesis of finite aperiodic subwavelength diffraction optical elements," *J. Opt. Soc. Am. A*, **15**, 1599–1607, 1998.
42. D. O. S. Melville, R. J. Blaikie, and C. R. Wolf, "Submicron imaging with a planar silver lens," *Appl. Phys. Lett.*, **84**, 4403–4405, 2004.

43. N. Fang, H. Lee, C. Sun, and X. Zhang, "Sub-diffraction-limited optical imaging with a silver superlens," *Science*, **308**, 534–537, 2005.
44. Z. M. Zhang and B. J. Lee, "Lateral shift in photon tunneling studied by the energy streamline method," *Opt. Express*, **14**, 9963–9970, 2006.
45. L. A. Blanco and F. J. García de Abajo, "Spontaneous light emission in complex nanostructures," *Phys. Rev. B*, **69**, 205414, 2004.
46. M. Planck, *The Theory of Heat Radiation*, Dover Publications, New York, 1959.
47. P. J. Hesketh, J. N. Zemel, and B. Gebhart, "Organ pipe radiant modes of periodic micromachined silicon surfaces," *Nature*, **324**, 549–551, 1986; P. J. Hesketh, B. Gebhart, and J. N. Zemel, "Measurements of the spectral and directional emission from microgrooved silicon surfaces," *J. Heat Transfer*, **110**, 680–686, 1998.
48. M. Auslender and S. Hava, "Zero infrared reflectance anomaly in doped silicon lamellar gratings. I. From antireflection to total absorption," *Infrared Phys. Technol.*, **36**, 1077–1088, 1995.
49. R. A. Dimenna and R. O. Buckius, "Electromagnetic theory predictions of the directional scattering from triangular surfaces" *J. Heat Transfer*, **116**, 639–645, 1994; D. W. Cohn, K. Tang, and R. O. Buckius, "Comparison of theory and experiments for reflection from microcontoured surfaces," *Int. J. Heat Mass Transfer*, **40**, 3233–3235, 1997; K. Tang and R. O. Buckius, "Bi-directional reflection measurements from two-dimensional microcontoured metallic surfaces," *Microscale Thermophys. Eng.*, **2**, 245–260, 1998.
50. A. Heinzel, V. Boerner, A. Gombert, B. Bläsi, V. Wittwer, and J. Luther, "Radiation filters and emitters for the NIR based on periodically structured metal surfaces," *J. Mod. Opt.*, **47**, 2399–2419, 2000.
51. S. Maruyama, T. Kashiwa, H. Yugami, and M. Esashi, "Thermal radiation from two-dimensionally confined modes in microcavities," *Appl. Phys. Lett.*, **79**, 1393–1395, 2001; H. Sai, Y. Kanamori, and H. Yugami, "Tuning of the thermal radiation spectrum in the near-infrared region by metallic surface microstructures," *J. Micromech. Microeng.*, **15**, S243–S249, 2005.
52. F. Kusunoki, J. Takahara, and T. Kobayashi, "Quantitative change of resonant peaks in thermal radiation from periodic array of microcavities," *Electron. Lett.*, **39**, 23–24, 2003; F. Kusunoki, T. Kohama, T. Hiroshima, S. Fukumoto, J. Takahara, and T. Kobayashi, "Narrow-band thermal radiation with low directivity by resonant modes inside tungsten microcavities," *Jpn. J. Appl. Phys.*, **43**, 5253–5258, 2004.
53. J. A. Gaspar-Armenta and F. Villa, "Photonic surface-wave excitation: Photonic crystal-metal interface," *J. Opt. Soc. Am. B*, **20**, 2349–2354, 2003.
54. B. J. Lee, C. J. Fu, and Z. M. Zhang, "Coherent thermal emission from one-dimensional photonic crystals," *Appl. Phys. Lett.*, **87**, 071904, 2005; B. J. Lee and Z. M. Zhang, "Coherent thermal emission from modified periodic multilayer structures," *J. Heat Transfer*, **129**, 17–26, 2007.
55. C. C. Williams and H. K. Wickramasinghe, "Scanning thermal profiler," *Appl. Phys. Lett.*, **49**, 1587–1589, 1986; H. F. Hamann, Y. C. Martin, and H. K. Wickramasinghe, "Thermally assisted recording beyond traditional limits," *Appl. Phys. Lett.*, **84**, 810–812, 2004.
56. J.-B. Xu, K. Läger, R. Möller, K. Dransfeld, and I. H. Wilson, "Heat transfer between two metallic surfaces at small distances," *J. Appl. Phys.*, **76**, 7209–7216, 1994; J.-B. Xu, K. Läger, K. Dransfeld, and I. H. Wilson, "Thermal sensors for investigation of heat transfer in scanning probe microscopy," *Rev. Sci. Instrum.*, **65**, 2262–2266, 1994.
57. W. Müller-Hirsch, A. Kraft, M. T. Hirsch, J. Parisi, and A. Kittel, "Heat transfer in ultrahigh vacuum scanning thermal microscopy," *J. Vac. Sci. Technol. A*, **17**, 1205–1210, 1999.
58. W. P. King, T. W. Kenny, K. E. Goodson, et al., "Atomic force microscope cantilevers for combined thermomechanical data writing and reading," *Appl. Phys. Lett.*, **78**, 1300–1302, 2001; W. P. King and K. E. Goodson, "Thermomechanical formation and thermal detection of polymer nanostructures," in *Heat Transfer and Fluid Flow in Microscale and Nanoscale Structures*, M. Faghri and B. Sunden (eds.), WIT Press, Southampton, pp. 131–171, 2003.
59. A. I. Volokitin and B. N. J. Persson, "Radiative heat transfer between nanostructures," *Phys. Rev. B*, **63**, 205404, 2001; A. I. Volokitin and B. N. J. Persson, "Resonance phonon tunneling of the radiative heat transfer," *Phys. Rev. B*, **69**, 045417, 2003.
60. A. Narayanaswamy and G. Chen, "Surface modes for near field thermophotovoltaics," *Appl. Phys. Lett.*, **82**, 3544–3546, 2003; A. Narayanaswamy and G. Chen, "Thermal emission control



- with one-dimensional metallodielectric photonic crystals," *Phys. Rev. B*, **70**, 125101, 2005; A. Narayanaswamy and G. Chen, "Thermal radiation in 1D photonic crystals," *J. Quant. Spectrosc. Radiat. Transfer*, **93**, 175–183, 2005.
61. S. M. Rytov, "Correlation theory of thermal fluctuations in an isotropic medium," *Sov. Phys. JETP*, **6**, 130–140, 1958; S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics III: Elements of Random Fields*, Vol. 3 (Chapter 3), Springer-Verlag, Berlin, 1987.
  62. K. Joulain, R. Carminati, J.-P. Mulet, and J.-J. Greffet, "Definition and measurement of the local density of electromagnetic states close to an interface," *Phys. Rev. B*, **68**, 245405, 2003; K. Joulain, J.-P. Mulet, F. Marquier, R. Carminati, and J.-J. Greffet, "Surface electromagnetic waves thermally excited: Radiative heat transfer, coherence properties and Casimir forces revisited in the near field," *Surf. Sci. Rep.*, **57**, 59–112, 2005.
  63. E. G. Cravalho, G. A. Domoto, and C. L. Tien, "Measurements of thermal radiation of solids at liquid helium temperatures," in *Progress in Aeronautics and Astronautics*, J. T. Bevans (ed.), **21**, 531–542, 1968; G. A. Domoto, R. F. Boehm, and C. L. Tien, "Experimental investigation of radiative transfer between metallic surfaces at cryogenic temperatures," *J. Heat Transfer*, **92**, 412–417, 1970.
  64. C. M. Hargreaves, "Anomalous radiative transfer between closely-spaced bodies," *Phys. Lett.*, **30A**, 491–492, 1969; C. M. Hargreaves, "Radiative transfer between closely-spaced bodies," *Philips Res. Rep. Supp.*, No. 5, pp. 1–80, 1973.
  65. S. S. Kutateladze, N. A. Rubtsov, and Ya. A. Bal'tsevich, "Effect of magnitude of gap between metal plates on their thermal interaction at cryogenic temperatures," *Sov. Phys. Doklady*, **8**, 577–578, 1979.

## PROBLEMS

---

- 10.1.** For incidence from glass with  $n = 1.5$  to air, calculate the Goos-Hänchen phase shift  $\delta$  for both TE and TM waves. Plot  $\delta$  as a function of the incidence angle  $\theta_1$ .
- 10.2.** Show that the normal component of the time-averaged Poynting vector is zero in both media when total internal reflection occurs. Prove Eq. (10.8).
- 10.3.** Calculate the Goos-Hänchen lateral shift upon total internal reflection from a dielectric with  $n = 2$  to air. Plot the lateral shift for both TE and TM waves as a function of  $\theta_1$ . Discuss the cause and the physical significance of the lateral beam shift.
- 10.4.** A perfect conductor can be understood based on the Drude free-electron model by neglecting the collision term. The dielectric function becomes  $\epsilon(\omega) = 1 - \omega_p^2/\omega^2$ , where  $\omega_p$  is the plasma frequency. For radiation incident from air to a perfect conductor, calculate the phase shift when  $\omega = \omega_p/2$  for TE and TM waves as a function of the incidence angle. Use Eq. (10.9) to calculate the lateral beam shift for a TM wave and modify it for a TE wave. Do you expect a sign difference between the TE and TM waves?
- 10.5.** For a planar waveguide with  $n_1 = 1.54$  and  $n_2 = 1.23$ , with a thickness of  $d = 200$  nm, how many total modes are there in the waveguide at  $\lambda = 635$  nm,  $\lambda = 1.55$   $\mu\text{m}$ , and  $\lambda = 3.2$   $\mu\text{m}$ ?
- 10.6.** Derive Eq. (10.16) by setting the determinant of the characteristic  $4 \times 4$  matrix to be zero. Sometimes, it is desirable to plot the solutions of Eq. (10.16) in curves that relate  $\omega$  to  $k_x$ . These curves are called waveguide dispersion relations. Given  $n_1 = 1.6$ ,  $n_2 = 1.3$ , and  $d = 300$   $\mu\text{m}$ , plot the dispersion curves for the first four TE modes. Explain why the group velocities are different for different modes, even though the refractive indices are independent of wavelength.
- 10.7.** In an asymmetric dielectric waveguide, the guided region (refractive index  $n_1 = 3.5$ ) is sandwiched between two different materials ( $n_2 = 1.5$  and  $n_3 = 2.5$ ). Show that the mode equation can be expressed as  $2k_{1z}d + \delta_2 + \delta_3 = 2m\pi$ , for  $m = 0, 1, 2, \dots$ , where  $\delta_2$  and  $\delta_3$  are the phase angles upon total internal reflection by media 2 and 3, respectively. If the thickness of the guided region is  $d = 3$   $\mu\text{m}$ , find the wavelength region where the fiber is a single-mode fiber (TE<sub>0</sub> only). Find the wavelength region where the fiber allows only TE<sub>0</sub> and TM<sub>0</sub> modes to be guided, i.e., single mode for each polarization.
- 10.8.** Using Eq. (10.20b) and Eq. (10.25) to show that the normal component of the Poynting vector in medium 2 is not a function of  $z$  in the case of photon tunneling (see Fig. 10.7a). Prove that the  $\langle S \rangle_z$  in medium 3 is the same as that in medium 2. Can you separate the incident power from the reflected power at the interface between media 2 and 3?

**10.9.** For thermal radiation originated from medium 1 with a refractive index of  $n_1 = 2$  to air, what is the critical angle? If medium 3 is close to medium 1 to form an air gap of width  $d$ , plot the directional-spectral transmittance at different angles as a function of  $d/\lambda$ . Calculate the hemispherical transmittance for  $s$  polarization, for both propagating and evanescent waves, and plot it against  $d/\lambda$ .

**10.10.** Two dielectric materials 1 and 3 are placed in close vicinity at cryogenic temperatures, separated by a vacuum gap of thickness  $d$ . Given  $n_1 = n_3 = 4$ ,  $T_1 = 4$  K, and  $T_3 = 2$  K, calculate the net radiative energy transfer when  $d = 1$  mm. Plot the radiative energy transfer from 1 to 3 and that from 3 to 1, as a function of  $d$ .

**10.11.** In the resonance tunneling setup discussed in Sec. 10.1.5 and shown in Fig. 10.9a, show that for  $N = 2$  the transmittance can be expressed as  $T'_\lambda = \sin^2(\delta)/[\sin^2(\delta) + 4p^2 \sinh(\eta b)]$ , where  $p = \cos(k_{1z} a) \cosh(\eta b) + \cot(\delta) \sin(k_{1z} a) \sinh(\eta b)$ . For  $n_1 = 3$ ,  $n_2 = 2$ , and  $a = b$ , find the wavelengths where resonance tunneling occurs. Plot the transmittance spectra for the TM wave, and determine the FWHM of each peak. [Discussion: It is interesting to find out the field distribution and localization in the three middle layers. The amplitude of the evanescent wave may either increase or decrease in the forward direction. One can use the matrix formulation to solve the field distribution to demonstrate the growth of evanescent waves in this arrangement. Discuss the lateral beam shift of the transmitted beam due to the parallel energy flow in the central layer.]

**10.12.** Derive Eq. (10.39) through Eq. (10.41), with layer 3 being a NIM and layer 2 being a PIM, with the same absolute values of refractive index. How will the field distribution in Fig. 10.10b change if the two middle layers switch positions?

**10.13.** Refer to photon tunneling with negative index layers. Consider two dielectric prisms of refractive index  $n_1 = n_5 = 1.5$ , sandwiching three middle layers of thicknesses  $d_2$ ,  $d_3$ , and  $d_4$ . Media 2 and 4 are vacuum with  $n_2 = n_4 = 1$ , while the middle layer, medium 3, is made of a NIM with  $n_3 = -1$ , i.e.,  $\epsilon_3 = \mu_3 = -1$ . Show that when the incidence angle is greater than the critical angle, the transmission coefficient can be expressed as follows:  $t = [\coth(\eta_2 \Delta) + i \cot(\delta) \sinh(\eta_2 \Delta)]^{-1}$ , where  $\eta_2 = \sqrt{k_x^2 - n_2^2 \omega^2 / c^2}$ ,  $\delta$  is the phase angle upon total internal reflection from medium 1 to 2, and  $\Delta = d_2 + d_4 - d_3$ . Plot the transmittance as a function of  $\Delta/\lambda$ , at incidence angles of  $45^\circ$  and  $60^\circ$  for each polarization. Derive the expression for the transmittance of propagating waves. Calculate the hemispherical transmittance for a chosen polarization, and plot it as a function of  $\Delta/\lambda$  for both the propagating and evanescent waves in vacuum.

**10.14.** Calculate the real and imaginary parts of  $k_x$  based on the surface plasmon polariton relation given in Eq. (10.45) for Al, as in Example 10-6. What is  $k'_x$  for  $\lambda = 400$  nm? Assuming that the prism has an index of refraction  $n_d = 1.53$ , find the incidence angle that would yield  $k_x = k'_x$ . Calculate the reflectance for Al in the ATR arrangements at  $\lambda = 400$  nm. Discuss whether the obtained reflectance dip in the angular distribution of the reflectance agrees with that predicted by the surface plasmon polariton. Calculate the polariton propagation length at this wavelength.

**10.15.** Studies suggest that the surface plasmon dispersion relation described in Eq. (10.45) can be solved by assuming that  $k_x$  is real but  $\omega = \omega' + i\omega''$ . The real part of  $\omega$  corresponds to the surface polariton resonance frequency, while the imaginary part corresponds to the bandwidth. Develop a computer program to solve  $\omega'(k_x)$  and  $\omega''(k_x)$  for Al. Assume that the Al film of thickness 24 nm is adjacent to the prism with  $\epsilon_d = 2.46$ . For  $\theta = 40.23^\circ$ , calculate the reflectance spectrum near the resonance frequency of the surface polariton, and compare the bandwidth with the calculated  $\omega''$ .

**10.16.** Examine Fig. 10.15 to confirm whether the surface polariton resonance frequencies predicted by the dispersion relation agree with the reflectance dips for a TM wave incident on a grating. Note that one of the dips in the dotted line ( $\theta = 30^\circ$ ) overlaps with that of the solid line ( $\theta = 0^\circ$ ) near  $12,000 \text{ cm}^{-1}$ . Hence, there are three notable dips in the reflectance at  $\theta = 0^\circ$  and five notable dips at  $\theta = 30^\circ$ .

**10.17.** Discuss why a nanoparticle can absorb more energy than a blackbody of the same size. Is it possible for a nanoparticle of radius  $r_0$  to emit more energy than  $4\pi r_0^2 \sigma_{\text{SB}} T^4$ , where  $T$  is the temperature of the spherical particle? Furthermore, is it possible for a nanoaperture to transmit more energy than the product of the incident energy flux (i.e., radiance) times its area? Why or why not?

**10.18.** Reproduce some cases in Fig. 10.18 under the same conditions for  $a = 0.25\lambda_p$  and  $d = 0.25\lambda_p$ . To examine the effect of  $a$ , recalculate the reflectance spectra with  $a = 0.15\lambda_p$  and  $0.1\lambda_p$  for the same  $d$ . Compare your results with those of Park et al.<sup>35</sup>

**10.19.** Based on the dielectric function model of SiC, at  $\lambda = 11 \text{ }\mu\text{m}$ ,  $\epsilon_s = -3.256 + 0.208i$  and  $n_e = 0.059 + 1.953i$ , which correspond to a radiation penetration depth of  $0.448 \text{ }\mu\text{m}$ . If a film of SiC with a thickness of  $d = 1.8 \text{ }\mu\text{m}$  is sandwiched between two prisms of the same dielectric constant  $\epsilon_d = 2.89$ , calculate the transmittance as a function of the incidence angle. Considering a

prism-air-SiC-prism arrangement, where the width of the air gap is  $a = 5 \mu\text{m}$ , calculate the transmittance again. You should see a peak near  $43^\circ$  with a transmittance around 0.3 for a TM wave. Verify that the transmittance enhancement is due to surface plasmon excitation, by calculating the angle-dependent reflectance.

**10.20.** Consider a prism-air-Al-prism configuration with  $\epsilon_d = 2.45$  for both prisms, air gap width  $a = 120 \text{ nm}$ , and aluminum thickness  $d = 30 \text{ nm}$ . Use the dielectric function of Al from Example 8-6 to calculate the transmittance and the reflectance at  $\lambda = 180 \text{ nm}$  for a TM wave at the incidence angle  $\theta = 47^\circ$ . Discuss the effect of the air gap width.

**10.21.** First reproduce Fig. 10.20*b*, i.e., the transmittance of SNG bilayer for a TE wave. Then, calculate the transmittance for a TM wave under the same conditions.

**10.22.** Consider two thin films with negative- $\epsilon$  (thickness  $a$ ) and negative- $\mu$  (thickness  $d$ ) onto a negative- $\epsilon$  substrate which is opaque. Such a layered structure may exhibit coherent emission to air. The structure is basically air-NGE-NGM-NGE, where NG stands for negative, E for permittivity, and M for permeability. The electric and magnetic properties can be modeled with Eq. (10.54) and Eq. (10.55), using  $F = 0.785$ ,  $\omega_0 = 0.5\omega_p$ ,  $\gamma_e = \gamma_m = 0.0025$ , and  $\epsilon_2 = 4$ .

(a) Let  $d = 0.1\lambda_p$  and  $a = 0.55\lambda_p$ . For  $p$  polarization, calculate the emissivity of this structure at  $\theta = 30^\circ$  for  $\omega/\omega_p$  between 0.5 and 0.9. Then calculate the angular distribution of the emissivity at  $\omega/\omega_p = 0.8383$ .

(b) Let  $d = 0.1\lambda_p$  and  $a = 0.3\lambda_p$ . Repeat the calculation of the emissivity spectrum at  $\theta = 30^\circ$  for  $p$  polarization. Then, calculate the angular distribution of the emissivity at  $\omega/\omega_p = 0.5425$ .

**10.23.** Reproduce Fig. 10.24, and discuss the features of energy streamlines for the radiative transfer through a dielectric film. Switch the vacuum and the dielectric regions so that the structure becomes dielectric-vacuum-dielectric, with  $d/\lambda = 0.1$  and  $0.01$ . Show the ZLs for both propagation and evanescent waves in air.

**10.24.** For the transmittance through the bilayer structure, shown in Fig. 10.20*b*, develop the energy streamlines at  $30^\circ$  and  $60^\circ$  incidence at the frequencies corresponding to the transmittance peaks.

**10.25.** A 2-D grating is made of microcavities with the following parameters:  $\Lambda_x = \Lambda_y = 6 \mu\text{m}$ ,  $W_x = W_y = 4.5 \mu\text{m}$ , and  $d = 5 \mu\text{m}$ , on a perfectly conducting metal. What is the largest wavelength at which the cavity resonance mode can be excited? List other possible modes, i.e., excitation wavelengths. Discuss the dependence on polarization, as well as potential anisotropic emission and absorption between  $\phi = 0^\circ$  and  $\phi = 45^\circ$ , where  $\phi$  is the azimuthal angle. Optional: Use FDTD or BEM software to confirm your results.

**10.26.** Calculate nanoscale heat transfer between two SiC plates with a vacuum gap as a function of the gap width. Assume that  $T_1 = 600 \text{ K}$  and  $T_2 = 300 \text{ K}$ , and use the dielectric function of SiC at room temperature. Plot and discuss the spectral energy transfer in the near field.

**10.27.** Team Project: It is proposed to use an asymmetric Fabry-Perot cavity for coherent thermal emission in the near-infrared, by depositing a  $1.5\text{-}\mu\text{m}$  SiO<sub>2</sub> coating onto a smooth Al substrate and then depositing a  $15\text{-nm}$  Al film atop the SiO<sub>2</sub> later. Assume that the refractive index of SiO<sub>2</sub> is  $n = 1.5$  and is independent of frequency. Calculate the emissivity for a TM wave at  $\theta = 30^\circ$  and  $60^\circ$  in the frequency range from  $5000$  to  $12,000 \text{ cm}^{-1}$ . Locate some of the peaks, and calculate the angle-dependent emissivity. If possible, plot the directional-spectral emissivity in a contour, showing the dependence on wavenumber and emission angle. See Lee and Zhang (*J. Appl. Phys.*, **100**, 063529, 2006).

**10.28.** Team Project: Develop the matrix formulation for 1-D multilayer structures, and use it to reproduce Figs. 10.30 and 10.31.

**10.29.** Team Project: Consider the radiation heat transfer between two plates at  $T_1 = 800 \text{ K}$  and  $T_2 = 300 \text{ K}$ , separated by a vacuum gap of width  $d$ . The dielectric function of the plates can be modeled as a Drude model:  $\epsilon_1(\omega) = \epsilon_2(\omega) = 1 - \omega_p^2/(\omega^2 + i\omega\gamma)$ . Choose different values of  $\omega_p$  and  $\gamma$  to calculate the near-field and far-field radiation heat transfer. Comment on the effect of each parameter. [Hint: You probably want to set  $\omega_p$  in the near-infrared, say, at  $8000 \text{ cm}^{-1}$ , and  $\gamma \approx 0.01\omega_p$  to start with.]

---

# APPENDIX A

---

## PHYSICAL CONSTANTS, CONVERSION FACTORS, AND SI PREFIXES

---

### PHYSICAL CONSTANTS

---

Avogadro's constant	$N_A$	$6.022 \times 10^{26} \text{ kmol}^{-1}$
Boltzmann's constant	$k_B$	$1.381 \times 10^{-23} \text{ J/K}$
Electric permittivity (vacuum)	$\epsilon_0$	$1/\mu_0 c_0^2 = 8.854 \times 10^{-12} \text{ C}^2/(\text{N} \cdot \text{m}^2)$
Electron charge (absolute value)	$e$	$1.602 \times 10^{-19} \text{ C (coulomb)}$
Electron mass	$m_e$	$9.109 \times 10^{-31} \text{ kg}$
Magnetic permeability (vacuum)	$\mu_0$	$4\pi \times 10^{-7} \text{ N/A}^2 \text{ (exact)}$
Planck's constant	$h$	$6.626 \times 10^{-34} \text{ J} \cdot \text{s}$
Proton mass	$m_p$	$1.673 \times 10^{-27} \text{ kg}$
Speed of light in vacuum	$c_0$	$2.998 \times 10^8 \text{ m/s (299,792,458 m/s, exact)}$
Standard acceleration of gravity	$g_n$	$9.80665 \text{ m/s}^2 \text{ (exact)}$
Stefan-Boltzmann constant	$\sigma_{SB}$	$5.670 \times 10^{-8} \text{ W}/(\text{m}^2 \cdot \text{K}^4)$
Universal gas constant	$\bar{R}$	$8.314 \text{ kJ}/(\text{kmol} \cdot \text{K})$

### CONVERSION FACTORS

---

1 atm = 760 mmHg = 101.325 kPa (standard atmosphere, exact)

1 eV =  $1.602 \times 10^{-19} \text{ J}$  (electron volt)

### SI PREFIXES

---

Power	$10^{-21}$	$10^{-18}$	$10^{-15}$	$10^{-12}$	$10^{-9}$	$10^0$	$10^9$	$10^{12}$	$10^{15}$	$10^{18}$	$10^{21}$
Prefix	zepto	atto	femto	pico	nano	—	giga	tera	peta	exa	zetta
Symbol	z	a	f	p	n	—	G	T	P	E	Z

---

Reference: <http://physics.nist.gov/cuu/index.html>

*This page intentionally left blank*

---

# APPENDIX B

---

## MATHEMATICAL BACKGROUND

---

### B.1 SOME USEFUL FORMULAE

---

#### B.1.1 Series and Integrals

Binary equation:

$$\begin{aligned}
 (a + b)^N &= b^N + Nab^{N-1} + \frac{N(N-1)}{2!}a^2b^{N-2} + \frac{N!}{3!(N-3)!}a^3b^{N-3} \\
 &+ \dots + Na^{N-1}b + a^N = \sum_{M=0}^N \frac{N!}{M!(N-M)!} a^M b^{N-M}
 \end{aligned}
 \tag{B.1}$$

Geometric series:

$$1 + e^{-x} + e^{-2x} + e^{-3x} + \dots = \frac{1}{1 - e^{-x}} \quad (x > 0) \tag{B.2}$$

Using the Taylor expansion, we can write

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \tag{B.3}$$

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad (-1 < x < 1) \tag{B.4}$$

Integrate  $\int_{-\infty}^{\infty} e^{-x^2} dx$ . This integral may be evaluated by a transformation from Cartesian coordinates to polar coordinates:

$$\begin{aligned}
 \left( \int_{-\infty}^{\infty} e^{-x^2} dx \right) \left( \int_{-\infty}^{\infty} e^{-y^2} dy \right) &= \iint_{\substack{-\infty < x < \infty \\ -\infty < y < \infty}} e^{-(x^2+y^2)} dx dy \\
 &= \iint_{\substack{0 < r < \infty \\ 0 < \phi < 2\pi}} e^{-r^2} r dr d\phi = \int_0^{\infty} e^{-r^2} 2\pi r dr = \pi \int_0^{\infty} e^{-t} dt = \pi
 \end{aligned}$$

Therefore, 
$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \quad (\text{B.5})$$

It can be seen that  $\int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\pi/a}$ . It should be noticed that  $\int_{-\infty}^{\infty} xe^{-ax^2} dx = 0$ , but

$$\int_0^{\infty} xe^{-ax^2} dx = \frac{1}{2a}$$

Furthermore, 
$$\int_0^{\infty} x^{n+2} e^{-ax^2} dx = \frac{n+1}{2a} \int_0^{\infty} x^n e^{-ax^2} dx \quad (n = 0, 1, 2, \dots) \quad (\text{B.6})$$

Another type of important integral equation is the following:

$$\int_0^{\infty} \frac{x^n e^x}{(e^x - 1)^2} dx = n \int_0^{\infty} \frac{x^{n-1}}{e^x - 1} dx \quad (\text{B.7})$$

where 
$$\int_0^{\infty} \frac{x^{n-1}}{e^x - 1} dx = (n-1)! \zeta(n) \quad (\text{B.8})$$

Here,  $\zeta(n)$  is the Riemann zeta function defined as

$$\zeta(n) = 1 + \frac{1}{2^n} + \frac{1}{3^n} + \frac{1}{4^n} + \dots \quad (\text{B.9})$$

The values of  $\zeta(n)$  are given in the following table for several  $n$  values:

$n$	1	2	3	4	5	6	7	8
$\zeta(n)$	$\infty$	$\frac{\pi^2}{6}$	1.202...	$\frac{\pi^4}{90}$	1.037...	$\frac{\pi^6}{945}$	1.008...	$\frac{\pi^8}{9450}$

Examples are  $\int_0^{\infty} \frac{x}{e^x - 1} dx = \frac{\pi^2}{6}$ ,  $\int_0^{\infty} \frac{x^2}{e^x - 1} dx = 2.404\dots$ , and  $\int_0^{\infty} \frac{x^3}{e^x - 1} dx = \frac{\pi^4}{15}$

### B.1.2 The Error Function

The error function is defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-x^2} dx \quad (\text{B.10})$$

The complementary error function is  $\operatorname{erfc}(x) \equiv 1 - \operatorname{erf}(x)$ . The error function can only be evaluated numerically. As shown in the following table,  $\operatorname{erf}(x)$  changes with  $x$  almost linearly for  $x < 0.5$  but approaches to unity rapidly as  $x$  increases.

$x$	0	0.01	0.1	0.2	0.5	1	2	3	$\infty$
$\operatorname{erf}(x)$	0	0.0113	0.1125	0.2227	0.5205	0.8427	0.9953	0.99998	1

### B.1.3 Stirling’s Formula

Stirling’s formula is an approximation of the logarithm of a factorial for large numbers. Note that

$$\begin{aligned} \ln x! &= \ln 1 + \ln 2 + \ln 3 + \dots + \ln x \\ &= \sum_{n=1}^x \ln n \approx \int_1^x \ln x dx = x \ln x - x + 1 \approx x \ln x - x \end{aligned}$$

More complicated analysis results in the same approximation for large  $x$ . Stirling’s formula is then

$$\ln x! \approx x \ln x - x \tag{B.11}$$

The relative error of this approximation is 13.8% for  $x = 10$  and less than 1% for  $x > 100$ . Therefore, it is applicable for very large  $x$ .

## B.2 THE METHOD OF LAGRANGE MULTIPLIERS

The method of Lagrange multipliers is a procedure for determining the maximum/minimum point in a continuous function subject to one or more constraints. Consider a continuous function  $f(x_1, x_2, \dots, x_n)$ . At the maximum/minimum point,

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i = 0 \tag{B.12}$$

Therefore, if  $x_{i_s}$  are independent, we must have

$$\frac{\partial f}{\partial x_i} = 0, \quad i = 1, 2, \dots, n \tag{B.13}$$

If they are dependent and related by  $m$  ( $m < n$ ) constraint equations (or constraints), then

$$\psi_j(x_1, x_2, \dots, x_n) = 0, \quad j = 1, 2, \dots, m \tag{B.14}$$

that is,

$$d\psi_j = \sum_{i=1}^n \frac{\partial \psi_j}{\partial x_i} dx_i = 0, \quad j = 1, 2, \dots, m \tag{B.15}$$

Multiplying  $\beta_j$  to the  $j$ th equation in Eq. (B.15) and adding them to Eq. (B.12), we obtain

$$\sum_{i=1}^n \left( \frac{\partial f}{\partial x_i} + \sum_{j=1}^m \beta_j \frac{\partial \psi_j}{\partial x_i} \right) dx_i = 0 \tag{B.16}$$



where  $\beta_{j,s}$  are called Lagrangian multipliers. For Eq. (B.16) to hold, we must have

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^m \beta_j \frac{\partial \psi_j}{\partial x_i} = 0, \quad i = 1, 2, \dots, n \quad (\text{B.17})$$

The  $n$  equations allow the determination of  $m$   $\beta_{j,s}$  and  $n-m$  independent variables.

**Example B-1.** Determine the positive values of  $x$ ,  $y$ , and  $z$  that will maximize the function  $f(x, y, z) = 8xyz$ , subject to the constraint  $(x/a)^2 + (y/b)^2 + (z/c)^2 = 1$ , where  $a$ ,  $b$ , and  $c$  are positive constants.

**Solution.** The constraint equation may be rewritten as  $\psi(x, y, z) = (x/a)^2 + (y/b)^2 + (z/c)^2 - 1 = 0$ .

$$df = 8yzdx + 8xzdy + 8xydz = 0$$

$$\beta d\psi = \frac{2\beta x}{a^2} dx + \frac{2\beta y}{b^2} dy + \frac{2\beta z}{c^2} dz = 0$$

Adding the preceding two equations and setting the coefficients to zero, we have  $8yz + 2\beta x/a^2 = 0$ ,  $8xz + 2\beta y/b^2 = 0$ , and  $8xy + 2\beta z/c^2 = 0$ ; that is,  $\beta = -4a^2yz/x$ ,  $\beta = -4b^2xz/y$ , and  $\beta = -4c^2xy/z$ . Dividing the product of the three equations,  $\beta^3 = -64a^2b^2c^2xyz$ , by each equation gives  $\beta^2 = 16 a^2b^2c^2 (x^2/a^2)$ ,  $\beta^2 = 16a^2b^2c^2 (y^2/b^2)$ , and  $\beta^2 = 16a^2b^2c^2 (z^2/c^2)$ . Solving for  $\beta$  and substituting  $x^2/a^2$ ,  $y^2/b^2$ , and  $z^2/c^2$  into the constraint equation, we obtain  $\beta = -4abc/\sqrt{3}$ . Therefore,  $x = a/\sqrt{3}$ ,  $y = b/\sqrt{3}$ , and  $z = c/\sqrt{3}$ . Thus, the maximum of the given function under the specified constraint is  $f_{\max} = 8abc/3\sqrt{3} \approx 1.54abc$ .

### B.3 PERMUTATION AND COMBINATION

This section discusses several permutation and combination problems that are directly related to the derivation of equilibrium distributions of different types of particles, such as molecular gases, electrons in a conductor, electrons and holes in semiconductors, photons in a thermodynamic equilibrium, and phonons in crystalline solids.

Case 1. *How many ways are there to arrange  $N$  distinguishable objects in a row?*

There are  $N$  objects to select for the first place,  $N - 1$  for the second,  $N - 2$  for the third, and so on. The number of permutations of  $N$  objects is therefore given by

$${}_N P_N = N! \quad (\text{B.18})$$

Case 2. *How many ways are there to choose and then arrange a subset of  $N$  objects out from a group of  $g$  distinguishable objects ( $N \leq g$ )?*

An equivalent problem is *How many ways are there to put  $N$  distinguishable objects into  $g$  distinguishable boxes with a limit that each box can at most have one object ( $N \leq g$ )?* There are  $g$  ways of placing the first object,  $g - 1$  ways of placing the second,  $g - 2$  ways of placing the third, . . . , and  $g - N + 1$  ways of placing the  $N$ th object. Therefore, the number of permutations of  $g$  objects taken  $N$  at a time is given by

$${}_g P_N = g(g - 1)(g - 2)\cdots(g - N + 2)(g - N + 1) = \frac{g!}{(g - N)!} \quad (\text{B.19})$$

Case 3. *How many ways are there to put  $N$  distinguishable objects into  $g$  distinguishable boxes (without regard to order within the boxes)?*

Because each box can contain any number of objects, there are  $g$  ways of placing each object. Hence, the number of ways is

$$g^N \tag{B.20}$$

Here,  $g$  can be smaller than, equal to, or greater than  $N$ . Note that this is equivalent to the permutation problem with repetition: *How many ways are there to arrange  $N$  objects taken from  $g$  types of objects (each type has more than  $N$  identical objects) by allowing repetition?*

**Example B-2**

- (a) How many 4-digit integers can be made from the numbers 1, 2, . . . , and 9, allowing no repeated usage of any number?
- (b) Same as (a) but the number can be repeatedly used.
- (c) Same as (b) with the possible inclusion of zero.

**Solution.** (a) There are  $9 \times 8 \times 7 \times 6 = 3024$ . (b) There are  $9 \times 9 \times 9 \times 9 = 6561$ . (c) There are  $9 \times 10 \times 10 \times 10 = 9000$ , because the first digit must be nonzero in a 4-digit integer.

**Example B-3**

- (a) How many ways are there to place 3 different books on 5 shelves without considering their order on each shelf?
- (b) Same as (a) but each shelf cannot have more than one book.

**Solution.** (a) Since each shelf can have any number of books and each book can go to any shelf, the ways to put the books are  $5 \times 5 \times 5 = 125$ . (b) In this case, there are  $5 \times 4 \times 3 = 60$  ways only.

Case 4. *How many ways are there to choose  $N$  objects from  $g$  distinguishable objects without caring about their order ( $N \leq g$ )?*

This is a combination problem. Because the order to arrange the objects is not considered, the number of combinations of  $N$  objects taken from a group of  $g$  objects is then given by

$${}_g C_N = \frac{g!}{N!(g - N)!} \tag{B.21}$$

It can be noted that the product of Eq. (B.21) and Eq. (B.18) gives Eq. (B.19). An equivalent problem is *How many ways are there to put  $N$  indistinguishable objects into  $g$  distinguishable boxes with a limit of at most one object in each box?* We learned from Case 2 that there are  $g!/(g - N)!$  ways of placing  $N$  distinguishable objects in  $g$  boxes. Now that the  $N$  objects are indistinguishable, the number of ways is reduced by a factor of  $N!$ .

Case 5. *How many ways are there to place  $N$  distinguishable objects into  $r$  distinguishable boxes such that there are  $N_1$  objects in the first box,  $N_2$  in the second, . . . , and  $N_r$  in the  $r$ th box?*

Because the order within each box is not considered, we must divide the total number of arrangements  $N!$  by the number of arrangements in each box, keeping in mind that  $N_1 + N_2 + \dots + N_r = N$ . Therefore, the number of ways is

$$\frac{N!}{N_1!N_2! \cdots N_r!} = \frac{N!}{\prod_{i=1}^r N_i!} \tag{B.22}$$

Case 6. *How many ways are there to put  $N$  indistinguishable objects into  $g$  distinguishable boxes without limiting the number of objects in each box?*

The answer to this problem is not so straightforward as compared with previous cases. The order within each box does not matter since the objects are indistinguishable. Let us use a dot for each object and use  $g - 1$  slashes to separate them into  $g$  groups such that:

$$\bullet\bullet/\bullet\bullet\bullet\bullet/\bullet/\bullet\bullet\bullet \dots \dots //\bullet\bullet\bullet$$

Each arrangement corresponds to one way of placing  $N$  indistinguishable objects in  $g$  distinguishable boxes. Although the slashes are identical, their order makes the “boxes” distinguishable. Note that the dot and slash are symbols: each occupies one location. The question becomes *How many ways are there to select  $g - 1$  slash locations out from  $N + g - 1$  total locations?* Said differently, *How many ways are there to select  $N$  dot locations out from  $N + g - 1$  total locations?* The answer is equivalent to the combination problem given in Case 4, except that there are  $N + g - 1$  total locations, i.e.,

$$\frac{(N + g - 1)!}{N!(g - 1)!} \quad (\text{B.23})$$

## B.4 EVENTS AND PROBABILITIES

If an evenly cast coin is tossed, the probability of ending up with a head or tail would each be 0.5. Denoting the occurrence of head as event  $A$  and that of tail as event  $B$ , we can write the probability of each event as  $p(A) = 0.5$  and  $p(B) = 0.5$ . In general, the probability of any event is between 0 and 1, i.e.,

$$0 \leq p(A) \leq 1 \quad (\text{B.24})$$

If  $p(A) = 0$ , it is an impossible event, and if  $p(A) = 1$ , it is a certain event. If  $A^*$  is used for anything but  $A$ , then  $p(A) + p(A^*) = 1$ . Two events may be dependent or independent. If one tosses the coin twice, the result of the second toss is independent of that of the first. Similarly, if one throws two dice, the result of each die is independent of that of the other. On the other hand, if two balls are drawn sequentially from a box containing three red and four yellow balls, the probability of the second ball being red depends upon whether the first ball is red or yellow. If  $A$  and  $B$  are independent events, then the probability for both  $A$  and  $B$  to happen is

$$p(A \text{ and } B) = p(A) \times p(B) \quad (\text{B.25})$$

while the probability of either  $A$  or  $B$  to happen is

$$p(A \text{ or } B) = p(A) + p(B) - p(A) \times p(B) = 1 - p(A^*) \times p(B^*) \quad (\text{B.26})$$

**Example B-4.** What is the probability for the sum of the numbers on the faces to be 7 if two dice are thrown?

**Solution.** The numbers on the six faces of each die are 1, 2, 3, 4, 5, and 6. Therefore the total number of combinations is 36. The combinations that yield 7 are (1,6), (2,5), (3,4), (4,3), (5,2), and (6,1). Thus, there are 6 out of 36 combinations that will give a sum of 7. The probability of getting 7 as the sum of the numbers on the faces is then  $p(7) = 1/6$ . It can be shown that the probability of getting 8 is  $p(8) = 5/36$ .

Consider an experiment for which the probability of event  $A$  to occur is  $\phi$ . For a single trial, the probability is  $\phi$  for event  $A$  and  $1 - \phi$  for anything but  $A$ . For  $N$  trials, the probability for event  $A$  to occur  $M$  times is given by the following equation:

$$p(M) = {}_N C_M \phi^M (1 - \phi)^{N-M} = \frac{N!}{M!(N - M)!} \phi^M (1 - \phi)^{N-M} \quad (\text{B.27})$$

which is equal to the corresponding coefficient of the binomial equation, Eq. (B.1), by setting  $a = \phi$  and  $b = 1 - \phi$ .

**Example B-5.** Toss three coins; what are the probabilities for getting all tails, one head and two tails, two heads and one tail, and all heads?

**Solution.** Here  $\phi = 0.5$  and  $1 - \phi = 0.5$ . Notice that

$$\left(\frac{1}{2} + \frac{1}{2}\right)^3 = \left(\frac{1}{2}\right)^3 + 3\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right) + 3\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 = \frac{1}{8} + \frac{3}{8} + \frac{3}{8} + \frac{1}{8}$$

We have  $p(0) = 0.125$ ,  $p(1) = 0.375$ ,  $p(2) = 0.375$ , and  $p(3) = 0.125$ .

**Example B-6.** Calculate the probability for the number 4 to show up on the face more than twice in 6 tosses of a fairly weighted die.

**Solution.** The probability for the number 4 to appear in any single toss is  $\phi = 1/6$ . Using Eq. (B.27), we have

$$\begin{aligned} p(0) &= 1 \times \left(\frac{1}{6}\right)^0 \times \left(\frac{5}{6}\right)^6 \approx 0.3349 \\ p(1) &= 6 \times \left(\frac{1}{6}\right)^1 \times \left(\frac{5}{6}\right)^5 \approx 0.4019 \\ p(2) &= 15 \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^4 \approx 0.2009 \end{aligned}$$

Therefore,  $p(> 2) = 1 - p(0) - p(1) - p(2) \approx 0.0623$ .

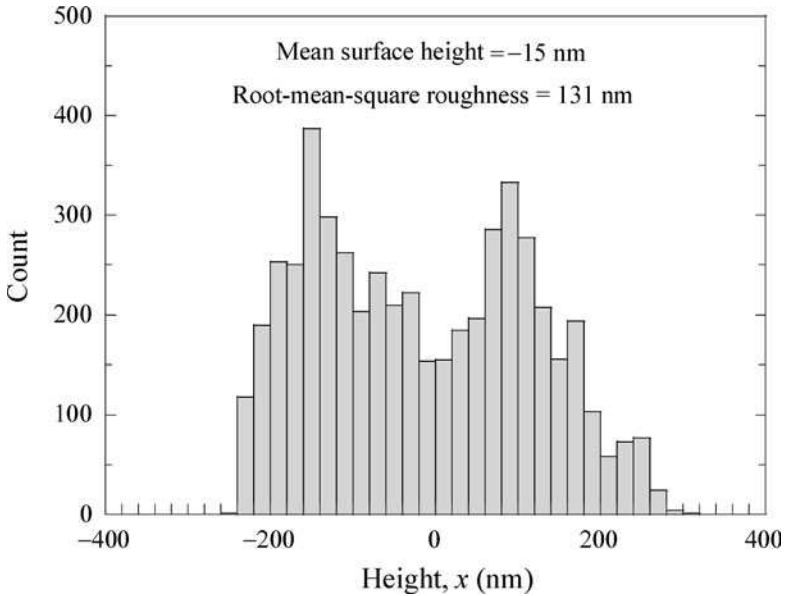
### B.5 DISTRIBUTION FUNCTIONS AND THE PROBABILITY DENSITY FUNCTION

Figure B.1 shows a plot of a surface roughness distribution (histogram) measured by using an atomic force microscope (AFM) for an unpolished silicon wafer within a  $50 \mu\text{m} \times 1 \mu\text{m}$  area with a total of  $512 \times 10 = 5120$  data points. The vertical axis records the number of points with height between  $x_{i-1}$  and  $x_i$ . Let  $N$  be the total number of data points and  $N_i$  the number of points with a height greater or equal to  $x_{i-1}$  but less than  $x_i$ . Then,  $N = \sum_i N_i$ , and *average* and *variance* (mean-square deviation) are obtained, respectively, from

$$\bar{x}_i = \frac{1}{N} \sum_i x_i N_i \quad \text{and} \quad u_{\text{var}} = \frac{1}{N} \sum_i (x_i - \bar{x}_i)^2 N_i \quad (\text{B.28})$$

The average is the mean surface height, and the square root of the variance is the *root-mean-square* (rms) roughness. The rms value associated with a set of measurements is called the *standard deviation*. If we randomly pick a point, the probability for it to have a height between  $x_{i-1}$  and  $x_i$  is

$$p(x_{i-1}, x_i) = N_i/N \quad (\text{B.29})$$



**FIGURE B.1** Histogram of surface roughness for a silicon surface measured using an AFM.

For large  $N$ , we may expect a continuous distribution function,

$$f(x) = \lim_{\Delta x_i \rightarrow 0} \left( \frac{N_i}{\Delta x_i} \right)$$

where  $\Delta x_i = x_i - x_{i-1}$ , and  $f(x)$  is called a *distribution function*. By definition,

$$\int_{x_{i-1}}^{x_i} f(x) dx = N_i \quad \text{and} \quad \int_{-\infty}^{\infty} f(x) dx = N \quad (\text{B.30})$$

The average and variance of the distribution can then be expressed as

$$\bar{x} = \frac{1}{N} \int_{-\infty}^{\infty} x f(x) dx \quad \text{and} \quad u_{\text{var}} = \frac{1}{N} \int_{-\infty}^{\infty} (x - \bar{x})^2 f(x) dx \quad (\text{B.31})$$

The average of  $x^2$ ,  $\bar{x^2}$ , is in general different from  $u_{\text{var}}$ , and is given as

$$\bar{x^2} = \frac{1}{N} \int_{-\infty}^{\infty} x^2 f(x) dx \quad (\text{B.32})$$

The distribution function  $f(x)$  may be normalized by dividing  $N$  to obtain

$$F(x) \equiv \frac{f(x)}{N} \quad (\text{B.33})$$

where  $F(x)$ , the normalized distribution function, is called the *probability density function (PDF)*. It is related to the probability by

$$p(x_1, x_2) = \int_{x_1}^{x_2} F(x)dx, \quad p(-\infty, x) = \int_{-\infty}^x F(x)dx, \quad \text{and} \quad \frac{d}{dx}p(-\infty, x) = F(x) \quad (\text{B.34})$$

Furthermore, it can be shown that

$$\int_{-\infty}^{\infty} F(x)dx = 1, \quad \int_{-\infty}^{\infty} xF(x)dx = \bar{x}, \quad \text{and} \quad \int_{-\infty}^{\infty} (x - \bar{x})^2F(x)dx = u_{\text{var}} \quad (\text{B.35})$$

**Example B-7.** Under certain conditions, the  $x$ -component velocity  $U$  of  $N$  particles in a fixed volume obeys the following distribution (*the Gaussian distribution or normal distribution*):

$$f(U) = A \exp\left(-\frac{U^2}{2\sigma^2}\right)$$

where  $U \in (-\infty, \infty)$ , and  $A$  and  $\sigma$  are positive constants. Determine the following: (a) the number of particles  $N$  in the volume; (b) the probability density function  $F(U)$ ; (c) the average velocity  $\bar{U}$ ; (d) the variance  $u_{\text{var}}$ ; and (e) the average of  $U^2$ .

**Solution.** Using the definitions and formulations given earlier, we have

$$(a) \quad N = \int_{-\infty}^{\infty} f(U)dU = \int_{-\infty}^{\infty} A \exp\left(-\frac{U^2}{2\sigma^2}\right) dU = A\sqrt{2\pi}\sigma$$

$$(b) \quad F(U) = \frac{f(U)}{N} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{U^2}{2\sigma^2}\right) \quad (c) \quad \bar{U} = \int_{-\infty}^{\infty} UF(U)dU = 0$$

$$(d) \quad u_{\text{var}} = \int_{-\infty}^{\infty} U^2F(U)dU = \sigma^2 \quad (e) \quad \bar{U}^2 = \sigma^2 = u_{\text{var}} \text{ because } \bar{U} = 0$$

**Discussion.** The general form of the Gaussian probability density function is

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

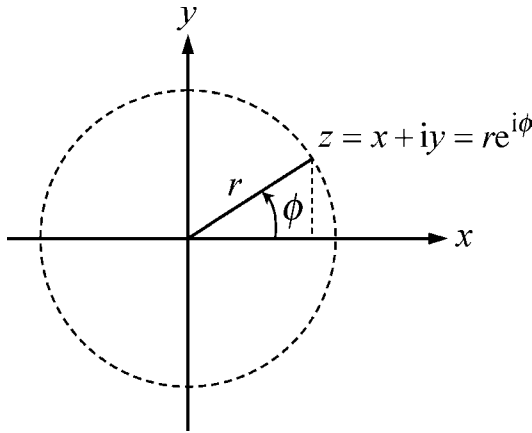
It is a bell-shaped graph centered around  $\bar{x} = \mu$  with  $u_{\text{var}} = \sigma^2$ . It has two inflection points at  $x = \mu \pm \sigma$ , at which the second-order derivative becomes zero. If the Gaussian statistics is used to describe the variations of a set of experimental measurements, the standard deviation  $\sigma$  is called the *standard uncertainty*. The probability for a measurement to fall within  $|x - \mu| < \sigma$  is 68%, and increases to 95% within  $|x - \mu| < 2\sigma$ . The *expanded uncertainty* is usually defined based on the 95% confidence interval, which is approximately  $2\sigma$  for Gaussian statistics.

## B.6 COMPLEX VARIABLES

A complex quantity  $z$  may be expressed in terms of a real component  $x = \text{Re}(z)$  and an imaginary component  $y = \text{Im}(z)$  so that

$$z = x + iy \quad (\text{B.36})$$

where  $i = \sqrt{-1}$ , and  $x$  and  $y$  are both real. The most convenient way to understand a complex variable is to use the complex plane shown in Fig. B.2. The expression of a complex



**FIGURE B.2** Illustration of the complex plane and complex quantity.

number is very similar to a 2-D vector. Notice that  $r = |z| = \sqrt{x^2 + y^2}$  is the magnitude or *complex modulus* and  $\phi = \arg|z| = \tan^{-1}(y/x)$  is the phase or *complex argument* of  $z$ . It is obvious that  $x = r \cos \phi$  and  $y = r \sin \phi$ . By defining

$$e^{i\phi} = \cos \phi + i \sin \phi \quad (\text{B.37})$$

we can also express the complex quantity in terms of its magnitude and phase as follows:

$$z = r e^{i\phi} \quad (\text{B.38})$$

The complex conjugate is defined as

$$z^* = x - iy = r e^{-i\phi} \quad (\text{B.39})$$

Hence,

$$z z^* = x^2 + y^2 = r^2 = |z|^2 \quad (\text{B.40})$$

Most of the algebra for real variables can be readily transformed to complex algebra. For example, if  $A = A' + iA'' = r_A e^{i\phi_A}$  and  $B = B' + iB'' = r_B e^{i\phi_B}$ , then

$$A \pm B = (A' \pm B') + i(A'' \pm B'') \text{ and } AB = r_A r_B e^{i(\phi_A + \phi_B)} \quad (\text{B.41})$$

It can be shown that

$$(A \pm B)^* = A^* \pm B^* \text{ and } (AB)^* = A^*B^* \tag{B.42}$$

Furthermore,

$$A^n = r^n e^{in\phi} = r^n [\cos(n\phi) + i \sin(n\phi)] \tag{B.43}$$

**Example B-8.** Suppose  $z = -1 + i\delta$ , where the real number  $\delta \ll 1$ . First, evaluate  $y = z^2$ , and then,  $x = y^{1/2}$ .

**Solution:** Clearly,  $z$  is in the second quadrant of the complex plane and  $y = 1 - i2\delta - \delta^2 = (1 - \delta^2) - i2\delta$  is in the fourth quadrant. Alternatively, we can write  $z = \sqrt{1 + \delta^2}e^{i\phi}$ , where  $\phi = \tan^{-1}(-\delta) \approx \pi - \delta$  for small  $\delta$ . Hence,  $y = (1 + \delta^2)e^{i2\phi} \approx (1 + \delta^2)e^{i(2\pi - 2\delta)} = (1 + \delta^2)e^{-i2\delta}$ . Finally,  $x = y^{1/2} = \sqrt{1 + \delta^2}e^{-i\delta} \approx 1 - i\delta = -z$ . However, if we use  $y \approx (1 + \delta^2)e^{i(2\pi - 2\delta)}$ , we will end up with  $x = \sqrt{z^2} = z$ . This example shows that multiple solutions often exist in complex algebra. Which solution should be accepted depends on the particular physical problem. Care must be taken when using a computer to do complex calculations to ensure that the final solution is physically meaningful.

Sometimes, we may deal with problems involving a complex quantity  $z$  with a complex magnitude  $\alpha = \alpha' + i\alpha''$  and a complex phase  $\beta = \beta' + i\beta''$  such that

$$z = \alpha e^{i\beta} \tag{B.44}$$

It can be considered as the multiplication of two complex quantities such that  $|z| = e^{-\beta''} \sqrt{\alpha'^2 + \alpha''^2}$ , and  $\arg(z) = \arg(\alpha) + \beta'$ . Alternatively, we can write  $\text{Re}(z) = \alpha' e^{-\beta''} \cos\beta' - \alpha'' e^{-\beta''} \sin\beta'$  and  $\text{Im}(z) = \alpha' e^{-\beta''} \sin\beta' + \alpha'' e^{-\beta''} \cos\beta'$ . Note that  $|e^{i\beta}| = e^{-\beta''}$ , which is not equal to 1 unless  $\beta''$  is 0.

Complex functions  $f = f(z)$  can be defined when  $z$  is a complex variable. The derivative and the integration can also be performed. In addition to the difficulty in dealing with multiple solutions, singularities are frequently involved.

## B.7 THE PLANE WAVE SOLUTION

---

The wave equation is a hyperbolic equation. In the 1-D case, it is given as

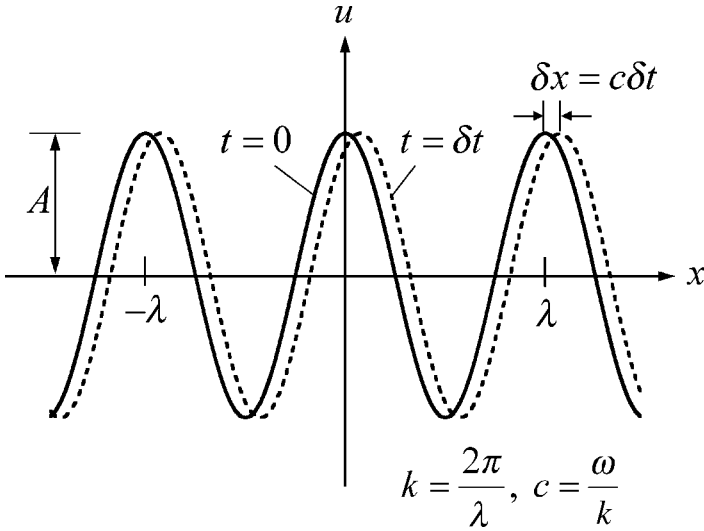
$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \tag{B.45}$$

where  $x$  is the spatial coordinate and  $t$  is the time. It can be verified that the following is a solution of the wave equation:

$$u(x,t) = A \cos(kx - \omega t) \tag{B.46}$$

as long as  $k = \omega/c$ . Equation (B.46) is only one solution, not a general expression of the solution of Eq. (B.45), that we choose to illustrate the nature of the wave equation. Let us further simplify the problem by taking only positive values of  $A$ ,  $k$ ,  $\omega$ , and  $c$ . Figure B.3 shows





**FIGURE B.3** Illustration of a wavefunction and the phase speed.

the spatial dependence of a wavefunction at  $t = 0$  and  $t = \delta t$ . Clearly,  $A$  is the amplitude of the wave. The period in space, which is the wavelength  $\lambda$ , is related to  $k$  by

$$k = \frac{2\pi}{\lambda} \quad (\text{B.47})$$

Therefore,  $k$  is called the wavevector because it is a vector in the 3-D coordinates with a magnitude  $k$ . The wavenumber is defined as the number of waves per unit length, i.e.,  $\bar{\nu} = 1/\lambda$ . From the time dependence, we can see that the period  $T = 2\pi/\omega$ . The frequency is the number of periods (cycles) per unit time; hence,  $\nu = 1/T$ , with a unit Hz. Therefore,  $\omega = 2\pi\nu$  is called the angular frequency with units rad/s. Notice that  $\beta = kx - \omega t$  is called the phase. The speed of propagation is determined by the movement of the constant phase plane, i.e.,

$$v_p = \left( \frac{dx}{dt} \right)_\beta = \frac{\omega}{k} = c \quad (\text{B.48})$$

We have just shown that  $c$  is the speed of propagation of the wave or the phase speed. In a 3-D case, the wave equation is written as

$$\nabla^2 u = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \quad (\text{B.49})$$

where  $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ . The solutions for a given frequency is

$$u(\mathbf{r}, t) = A e^{i\mathbf{k} \cdot \mathbf{r} - i\omega t} \quad (\text{B.50})$$

where  $\mathbf{k} = k_x \hat{\mathbf{x}} + k_y \hat{\mathbf{y}} + k_z \hat{\mathbf{z}}$  is called the wavevector and  $k = 2\pi/\lambda = (k_x^2 + k_y^2 + k_z^2)^{1/2}$ . It should be noted that from  $\omega = kc$ , we get  $\lambda\nu = c$ . It can be shown that Eq. (B.50)

represents a plane wave whose constant phase plane is always perpendicular to  $\mathbf{k}$ , and this wave propagates in the  $\mathbf{k}$  direction. Furthermore,

$$\frac{\partial}{\partial x}u = ik_x u, \quad \frac{\partial}{\partial y}u = ik_y u, \quad \text{and} \quad \frac{\partial}{\partial z}u = ik_z u$$

or the gradient 
$$\nabla u = \left( \hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z} \right) u = i\mathbf{k}u \quad (\text{B.51})$$

Similarly, 
$$\nabla^2 u = -k^2 u \quad (\text{B.52})$$

In the preceding discussion,  $u$  is treated as a scalar. Frequently, we will need to deal with a vector as the function, such as the electric field  $\mathbf{E}$ . Then, the wave equation can be written as

$$\nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (\text{B.53})$$

Its solution can be expressed as

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{A} \exp(i\mathbf{k} \cdot \mathbf{r} - i\omega t) \quad (\text{B.54})$$

where the amplitude  $\mathbf{A}$  is a vector. It can be shown that

$$\frac{\partial}{\partial x}E_x = ik_x E_x, \quad \frac{\partial}{\partial y}E_y = ik_y E_y, \quad \text{and} \quad \frac{\partial}{\partial z}E_z = ik_z E_z$$

Thus, the divergence is 
$$\nabla \cdot \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = i\mathbf{k} \cdot \mathbf{E} \quad (\text{B.55})$$

the curl is 
$$\nabla \times \mathbf{E} = \begin{pmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{pmatrix} = i \begin{pmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ k_x & k_y & k_z \\ E_x & E_y & E_z \end{pmatrix} = i\mathbf{k} \times \mathbf{E} \quad (\text{B.56})$$

and 
$$\nabla^2 \mathbf{E} = -k^2 \mathbf{E} \quad (\text{B.57})$$

Equation (B.54) can be called the solution of a monochromatic wave (single frequency). When multiple frequencies are involved, the speed is frequency dependent in a dispersive medium. In such a case, waves of different frequency will travel with a different speed. A wave group or wave packet contains waves of more than one frequency. The group velocity of the wave packet represents the velocity of energy carried by the wave packet and is given by

$$\mathbf{v}_g = \frac{d\omega}{d\mathbf{k}} = \hat{\mathbf{x}} \frac{d\omega}{dk_x} + \hat{\mathbf{y}} \frac{d\omega}{dk_y} + \hat{\mathbf{z}} \frac{d\omega}{dk_z} \quad (\text{B.58})$$

The functional relation  $\omega = \omega(\mathbf{k})$  is called a *dispersion relation*. In the 1-D case or an isotropic medium, we have

$$v_g = d\omega/dk \quad (\text{B.59})$$

If the phase speed  $c = \omega/k$  is constant, i.e., the dispersion relation is linear, the group velocity is the same as the phase velocity because  $v_g = d\omega/dk = c = \omega/k = v_p$ .

**Example B-9.** Consider light propagating in a glass whose refractive index  $n = 1.5 + \alpha\omega^2$ , where  $\alpha$  is a very small coefficient. Find the dispersion relation, the phase speed, and the group speed as functions of  $\omega$ .

**Solution.** The speed of light in a medium  $c = c_0/n$ , where  $n$  is the refractive index. Therefore,  $v_p(\omega) = \omega/k = c = c_0/(1.5 + \alpha\omega^2)$ . The dispersion relation is given by  $1.5\omega + \alpha\omega^3 = c_0k$ . The group speed  $v_g(\omega) = (dk/d\omega)^{-1} = c_0/(1.5 + 3\alpha\omega^2) = c_0/n_g$ , where  $n_g = n + \omega dn/d\omega$  is called the group index.

Some useful vector operators and identities are given for convenience as follows:

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A} \quad (\text{B.60})$$

$$\mathbf{A} \times \mathbf{B} = -\mathbf{B} \times \mathbf{A} \quad (\text{B.61})$$

$$\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = (\mathbf{C} \times \mathbf{B}) \times \mathbf{A} = (\mathbf{A} \cdot \mathbf{C})\mathbf{B} - (\mathbf{A} \cdot \mathbf{B})\mathbf{C} \quad (\text{B.62})$$

$$\mathbf{A} \cdot (\mathbf{B} \times \mathbf{C}) = \mathbf{B} \cdot (\mathbf{C} \times \mathbf{A}) = \mathbf{C} \cdot (\mathbf{A} \times \mathbf{B}) = \det(\mathbf{A} \ \mathbf{B} \ \mathbf{C}) \quad (\text{B.63})$$

$$\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \nabla^2 \mathbf{A} \quad (\text{B.64})$$

$$\nabla \cdot (\phi \mathbf{A}) = \phi \nabla \cdot \mathbf{A} + \mathbf{A} \cdot \nabla \phi \quad (\text{B.65})$$

$$\nabla \times (\phi \mathbf{A}) = \phi \nabla \times \mathbf{A} + \nabla \phi \times \mathbf{A} \quad (\text{B.66})$$

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = \mathbf{B} \cdot (\nabla \times \mathbf{A}) - \mathbf{A} \cdot (\nabla \times \mathbf{B}) \quad (\text{B.67})$$

$$\nabla \times (\mathbf{A} \times \mathbf{B}) = (\mathbf{B} \cdot \nabla)\mathbf{A} - \mathbf{B}(\nabla \cdot \mathbf{A}) + \mathbf{A}(\nabla \cdot \mathbf{B}) - (\mathbf{A} \cdot \nabla)\mathbf{B} \quad (\text{B.68})$$

$$\text{and } \nabla(\mathbf{A} \cdot \mathbf{B}) = \mathbf{A} \times (\nabla \times \mathbf{B}) + (\mathbf{A} \cdot \nabla)\mathbf{B} + \mathbf{B} \times (\nabla \times \mathbf{A}) + (\mathbf{B} \cdot \nabla)\mathbf{A} \quad (\text{B.69})$$

If  $\mathbf{A}$  is a constant matrix, say  $\mathbf{A} = \mathbf{K}$ , then Eq. (B.68) and Eq. (B.69) reduce respectively to

$$\nabla \times (\mathbf{K} \times \mathbf{B}) = \mathbf{K}(\nabla \cdot \mathbf{B}) - (\mathbf{K} \cdot \nabla)\mathbf{B}$$

$$\nabla(\mathbf{K} \cdot \mathbf{B}) = \mathbf{K} \times (\nabla \times \mathbf{B}) + (\mathbf{K} \cdot \nabla)\mathbf{B}$$

The divergence theorem or Gauss's theorem is expressed as

$$\iiint_V \nabla \cdot \mathbf{E} \, dV = \iint_A \mathbf{E} \cdot \mathbf{n} \, dA \quad (\text{B.70})$$

which states that *the integral of the divergence over the entire volume is equal to the surface integral over the enclosed surface*. The curl theorem or Green's theorem states that

$$\iint_A (\nabla \times \mathbf{E}) \cdot \mathbf{n} \, dA = \int_C \mathbf{E} \cdot d\mathbf{r} \quad (\text{B.71})$$

In this equation, it is assumed that  $C$  is a closed, piecewise smooth curve that bounds the surface area  $A$ . The equation converts a surface integration of the curl of a vector to a line integration of the vector. Both the divergence theorem and the curl theorem can be considered special cases of the Stokes theorem.

### B.8 THE SOMMERFELD EXPANSION

In the free-electron theory discussed in Chap. 5, the integration often includes the Fermi-Dirac function

$$f_{\text{FD}}(\varepsilon, T) = \frac{1}{e^{(\varepsilon - \mu)/k_{\text{B}}T} + 1} \tag{B.72}$$

where  $\varepsilon$  and  $\mu$  are the electron energy and chemical potential, respectively,  $k_{\text{B}}$  is the Boltzmann constant, and  $T$  is absolute temperature. Unless the temperature is very high,  $k_{\text{B}}T \ll \mu$ , where  $\mu$  itself is a weak function of temperature, i.e.,  $\mu = \mu(T)$ . At  $T \rightarrow 0$  K, the chemical potential is called the Fermi energy  $\mu_{\text{F}} = \mu(0)$ . However, the chemical potential  $\mu$  is often called Fermi level or Fermi energy as well in many texts. It can be seen that at very low temperatures,  $f_{\text{FD}}(\varepsilon, 0) \equiv f_{\text{FD}}(\varepsilon, T \rightarrow 0) = 1$ , when  $\varepsilon < \mu_{\text{F}}$ , and  $f_{\text{FD}}(\varepsilon, 0) = 0$ , when  $\varepsilon > \mu_{\text{F}}$ , as illustrated in Fig. 5.5a. Thus,

$$\int_0^{\infty} G(\varepsilon) f_{\text{FD}}(\varepsilon, 0) d\varepsilon = \int_0^{\mu_{\text{F}}} G(\varepsilon) d\varepsilon \tag{B.73}$$

When  $k_{\text{B}}T \ll \mu$ ,  $f_{\text{FD}}(\varepsilon, T)$  is essentially the same as  $f_{\text{FD}}(\varepsilon, 0)$ , except when  $|\varepsilon - \mu| < k_{\text{B}}T$ . The following approximation is often used since when  $k_{\text{B}}T \ll \mu$ ,

$$\int_0^{\mu} G(\varepsilon) d\varepsilon = \int_0^{\mu_{\text{F}}} G(\varepsilon) d\varepsilon + (\mu - \mu_{\text{F}})G(\mu_{\text{F}}) + \dots \tag{B.74}$$

Let us now consider the derivative

$$\frac{\partial}{\partial \varepsilon} f_{\text{FD}}(\varepsilon, T) = -\frac{1}{k_{\text{B}}T} \frac{e^{(\varepsilon - \mu)/k_{\text{B}}T}}{[e^{(\varepsilon - \mu)/k_{\text{B}}T} + 1]^2} \tag{B.75}$$

The derivative is nonzero only when  $|\varepsilon - \mu| < k_{\text{B}}T$ . When  $T \rightarrow 0$ , the peak at  $\varepsilon = \mu$  goes to infinity. Note that

$$\int_0^{\infty} \frac{\partial f_{\text{FD}}}{\partial \varepsilon} d\varepsilon = \int_0^{\infty} d(f_{\text{FD}}) = f_{\text{FD}} \Big|_0^{\infty} = 0 - 1 = -1$$

Therefore,  $\partial f_{\text{FD}}/\partial \varepsilon$  is a Dirac delta function, i.e.,

$$\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \approx -\delta(\varepsilon - \mu), \quad k_{\text{B}}T \ll \mu \tag{B.76}$$

Hence, 
$$\int_0^{\infty} G(\varepsilon) \frac{\partial f_{\text{FD}}}{\partial \varepsilon} d\varepsilon \approx -G(\mu), \quad k_{\text{B}}T \ll \mu \tag{B.77}$$

The preceding equation is exact only at absolute zero temperature. A difficulty arises when the integrand contains terms such as  $(\varepsilon - \mu)$  or  $(\varepsilon - \mu)^2$ . When this is the case, higher-order terms must be retained. Sommerfeld in 1927 developed an expansion to handle the integral. A detailed discussion can be found from the work of McDougall and Stoner

(*Phil. Trans. Roy. Soc. London A*, **237**, 67, 1938). The approximations necessary for the free-electron model of metals are discussed next. When  $T > 0$  K, Eq. (B.73) can be written in terms of an expansion as follows:

$$\int_0^\infty G(\varepsilon) f_{\text{FD}} d\varepsilon = \int_0^\mu G(\varepsilon) d\varepsilon + \frac{\pi^2(k_B T)^2}{6} G'(\mu) + \frac{7\pi^4(k_B T)^4}{360} G'''(\mu) + \dots \quad (\text{B.78})$$

where  $G'(\mu) = \left. \frac{dG}{d\varepsilon} \right|_{\varepsilon=\mu}$  and  $G'''(\mu) = \left. \frac{d^3G}{d\varepsilon^3} \right|_{\varepsilon=\mu}$

**Example B-10.** When  $G(0) = 0$ , show that

$$\int_0^\infty G(\varepsilon)(\varepsilon - \mu) \left( -\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \right) d\varepsilon \approx \frac{\pi^2(k_B T)^2}{3} G'(\mu) \quad (\text{B.79})$$

and  $\int_0^\infty G(\varepsilon)(\varepsilon - \mu)^2 \left( -\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \right) d\varepsilon \approx \frac{\pi^2(k_B T)^2}{3} G(\mu)$  (B.80)

**Solution.** We will use Eq. (B.78) by dropping the term with  $(k_B T)^4 G'''(\mu)$ . Therefore,

$$\begin{aligned} & \int_0^\infty G(\varepsilon)(\varepsilon - \mu) \left( -\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \right) d\varepsilon \\ &= -f_{\text{FD}} G(\varepsilon)(\varepsilon - \mu) \Big|_0^\infty + \int_0^\infty G(\varepsilon) f_{\text{FD}} d\varepsilon + \int_0^\infty (\varepsilon - \mu) G'(\varepsilon) f_{\text{FD}} d\varepsilon \\ &= \int_0^\mu G(\varepsilon) d\varepsilon + \frac{\pi^2(k_B T)^2}{6} G'(\mu) + \int_0^\mu (\varepsilon - \mu) G'(\varepsilon) d\varepsilon + \frac{\pi^2(k_B T)^2}{6} G'(\mu) \\ &= \frac{\pi^2(k_B T)^2}{3} G'(\mu) \end{aligned}$$

since  $\int_0^\mu (\varepsilon - \mu) G'(\varepsilon) d\varepsilon = (\varepsilon - \mu) G(\varepsilon) \Big|_0^\mu - \int_0^\mu G(\varepsilon) d\varepsilon = -\int_0^\mu G(\varepsilon) d\varepsilon$ . The proof of Eq. (B.80) is similar, and it is left as an exercise.

Another useful equation is

$$\frac{\partial f_{\text{FD}}}{\partial T} = \frac{e^{(\varepsilon-\mu)/k_B T}}{[e^{(\varepsilon-\mu)/k_B T} + 1]^2} \frac{1}{k_B T} \left( -\frac{\varepsilon - \mu}{T} - \frac{d\mu}{dT} \right) = -\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \left( \frac{\varepsilon - \mu}{T} + \frac{d\mu}{dT} \right) \quad (\text{B.81})$$

If we neglect  $d\mu/dT$ , then

$$\frac{\partial f_{\text{FD}}}{\partial T} \approx -\frac{\partial f_{\text{FD}}}{\partial \varepsilon} \frac{\varepsilon - \mu}{T} \quad (\text{B.82})$$

# INDEX

---

*This page intentionally left blank*

Absorbing, dissipative, or lossy medium, 290, 308  
 Absorptance or absorptivity, 48, 308–309, 335  
 Absorption, 69, 80, 92, 215  
   by free carriers, 315, 323, 325  
   fundamental or interband, 236, 314, 322  
   intra-band or intersubband, 322  
   local absorption distribution, 352  
   phonons or lattice vibrations, 166, 318, 323  
 Absorption bands, 320, 323  
 Absorption coefficients, 50, 293, 400  
 Absorption edge, 314, 321  
 Absorption spectra of gases, 80  
 Acceptance cone, 383  
 Acceptors, 199, 233  
 Accommodation coefficients, 125, 176, 178  
   energy or thermal, 125, 130, 436  
   momentum or velocity, 125  
 Acoustic mismatch model (AMM), 220, 271, 274  
 Acoustic waves, 139, 162, 165  
 Acoustically thick or thin limits, 264  
 Active medium, 94  
 Adiabatic availability, 32  
 Adiabatic process, 28  
 Advection, 44  
 Affinity, 172  
 Airy's formulae, 337  
 Ampere law, 286  
 Angle of incidence, 306  
 Angular frequency, 287, 298, 306, 379  
 Anharmonic vibration, 91, 147  
 Anisotropy, 177, 286, 367, 371  
 Anomalous dispersion, 319  
 Anomalous skin effect, 318  
 Antireflection coating, 335, 340, 350  
 Anti-Stokes, 224, 225  
 Aperture, near field, 377, 408  
 Aperture, numerical, 383  
 Apertureless NSOM, 378  
 Atomic binding, 199  
   covalent bonds, 200  
   hydrogen bonds, 200  
   ionic bonds, 199  
   metallic bonds, 200  
   molecular bonds, 200  
 Atomic emission, 88  
 Atomic force microscope (AFM), 14, 132, 277, 371, 431  
   artifacts, 365, 371  
   contact mode, 367  
   heated cantilever, 12, 277  
   tapping mode, 367  
 Atomic theory, 5, 195  
 Atomistic simulation, 21, 182, 185, 200, 253  
 Atomistic smoothness, 186, 271, 273  
 Atoms, 4, 5, 196  
 Attenuated total reflectance (ATR), 396, 403  
 Aufbau principle, 196  
 Autocorrelation function, 364, 367  
 Autocorrelation length, 274, 345, 363  
 Autocovariance function, 345  
 Average collision distance, 108  
 Average phonon speed, 141, 162, 258  
 Averages  
   ensemble, 102, 428  
   local, 102



- Averages (*Cont.*):  
 spectral, 339  
 time, 290–291
- Avogadro's constant, 38, 443
- Ballistic, 123, 131, 165, 184, 247, 258, 276
- Ballistic-diffusion approximation, 269
- Balmer series, 89
- Band structures  
 allowable band, 197  
 conduction band, 197  
 extended-zone scheme, 214  
 forbidden band, 197  
 of photonic crystals, 352, 356  
 reduced-zone scheme, 214  
 valence band, 197
- Bandgap, 197, 207, 213  
 direct or indirect, 207, 216, 224, 322
- Bandgap absorption, 198, 322  
*See also* absorption, interband
- Beam divergence, 334, 342
- Bidirectional reflectance, 312, 362
- Bidirectional reflectance distribution  
 function (BRDF), 312, 362–372  
 measurements, 368–372  
 microfacet slope method (MSM), 364  
 Monte Carlo method, 364  
 out-of-plane, 371–372  
 reciprocity, 313  
 surface generation method (SGM), 364
- Binding energy, 199–200, 227
- Biomolecule imaging, 17
- Birefringence, 386
- Blackbody  
 cavity, 294, 299  
 concept, 295  
 enclosure, 294
- Blackbody radiation  
 cosmic background, 299  
 dilute, 303  
 solar radiation, 297  
 spectral distribution, 296
- Bloch-Floquet condition, 358, 399
- Bloch formula, 157
- Bloch theorem, 210
- Bloch (wave) condition, 353, 418
- Bloch wavevector, 354, 391
- Body or volume forces, 3, 46, 116, 121
- Bohr radius, 88–90
- Bolometer, 236, 284
- Boltzmann constant, 39, 67, 68, 75
- Boltzmann transport equation (BTE),  
 116, 178  
 Chapman-Enskog method, 117  
 Collision term, 117  
*See* relaxation-time approximation
- Boltzons, 66
- Born–von Kármán periodic  
 conditions, 148
- Bose-Einstein condensate, 63, 69
- Bose-Einstein distribution function, 64,  
 140, 295
- Bose-Einstein (BE) statistics, 62, 66
- Bosons, 66, 69
- Boundary conditions in electrodynamics,  
 307, 354, 360, 379, 421
- Boundary conditions, periodic, 87, 148
- Boundary element method (BEM),  
 357, 409
- Boundary layers, 44, 114, 122  
 thermal, 44  
 velocity, 44
- Boundary scattering, 156, 163, 174
- Boyle's law, 105
- Bragg reflectors, 9, 327, 355
- Bravais lattices, 201–203  
 conventional unit cells, 201  
 non-Bravais lattice, 206  
 translational symmetry, 201  
 types of, 202
- Brewster angle, 266
- Brewster mode, 425
- Brewster window, 250, 257
- Brightness temperature, 300
- Brillouin zone, first, 209
- Brownian motion, 94, 121
- Buckminsterfullerene, 11
- Built-in potential, 238
- Cantilever, *See* atomic force microscope
- Carbon nanotubes (CNTs), 2, 12, 15  
 field emission, 231  
 multi-walled—(MWNT), 154, 186  
 single-walled—(SWNT), 12,  
 154, 186

- Carbon nanotubes (CNTs) (*Cont.*):
  - specific heat, 153
  - structure, 208
  - thermal conductivity, 185, 186
- Casimir force, 440
- Casimir limit, 266
- Cattaneo equation, 250, 253
- Causality, principle of, 248, 262, 292, 314, 411
- Cavity resonance, 327, 417, 425
- Central equation in electronic band theory, 211
- Characteristic functions in thermodynamics, 33
- Characteristic lengths, 3, 122, 247, 276
- Characteristic temperatures
  - Einstein temperature, 138
  - for rotation, 76, 79
  - for translation, 76
  - for vibration, 76, 80
  - See* Debye temperature
- Characteristic wavelengths, 61, 166, 275, 297
  - See* de Broglie wavelength
  - See* most probable wavelength
  - See also* thermal wavelength
- Charge neutrality, 233
- Chemical bonds, *See* atomic binding
- Chemical etching, 7, 11, 368, 418
- Chemical potential, 28, 34, 68, 161, 184
- Chemical vapor deposition (CVD), 10, 12, 333, 346
- Christiansen wavelength, 320
- Classifications of solids, 195–201
  - amorphous, 199
  - conductors, 197
  - dielectric, 291
  - insulators, 197
  - metals (alkali, noble, or transition), 147, 198
  - See* crystals
  - See* semiconductors
- Coherence, 342–344
  - degree of, 341
  - function, 342
  - spatial, 417–418
  - spectral width, 342
  - temporal, 417–418
- Coherent emission, 92
  - coherent thermal emission, 418, 421, 422
  - for spontaneous radiation, 415
- Collision, 106, 117
- Collision frequency (or rate), 106
- Collision time, 106
- Complementary-metal-oxide-semiconductor (CMOS), 8, 360
- Complex conductivity, 292
- Complex planes, 327, 454
- Complex refractive index, 291–292, 327
- Complex variables, 454
- Conductance quantization, *See* quantum conductance
- Conduction band, 197, 215, 232, 322
- Conductors, 197, 199
- Conservative equations, 45–46, 117–119, 221–225, 286, 290, 323
- Constitutive equations, 118, 254, 286, 319
- Constraints, 25, 64, 447
- Contact resistance, thermal, 44, 271
- Continuity equation, 45, 119, 286
- Continuum assumption, 1, 121
- Continuum regime, 122
- Corpuscular theory, 5
- Correlation function, spatial, 428
- Cosmic background radiation, 298, 331, 415
- Coulomb's force, 88
- Coupled transport processes, 172
- Creation or annihilation reactions, 95
- Critical angle, 274, 310, 383, 387
- Critical point, 35
  - critical pressure, 36
  - critical temperature, 36
- Crystal momentum, 220
- Crystal structures
  - benzene-ring structure, 154
  - body-centered cubic (bcc), 201, 204
  - cesium chloride structure, 206, 207
  - diamond structure, 206, 207
  - face-centered cubic (fcc), 201, 204
  - hexagonal close-packed (hcp), 203, 204
  - sodium chloride structure, 206, 207
  - zincblende structure, 206, 207
  - See also* Bravais lattices

- Crystal, types of, 199–200
  - covalent, 200
  - ionic, 199
  - molecular, 200
  - polycrystalline, 200
- Current density, electrical, 155
- Damping coefficients, 155, 316, 319, 328
- Dark current, 241
- de Broglie wavelength, 61, 85, 146
- Debye model, 139–143
- Debye temperature, 140, 264, 268, 273
- Defect (or impurity) scattering, 156, 163, 223, 235
- Degeneracy, 61, 66, 80, 144, 154
- Degrees of freedom, 75, 79
  - diatomic gases, 75
  - polyatomic gases, 75
  - vibrational modes, 75
- Density of modes, 428
- Density of states (DOS), 120
  - for electrons, 145, 160
  - local DOS, 428, 434
  - in 1-D or 2-D solid, 150
  - for phonons, 141, 149, 165
  - quantization, 154, 158, 163
  - in semiconductors, 232
- Dielectric functions, 291, 314–329
  - insulators, 319
  - metals, 315
  - metamaterials, 326
  - semiconductors, 321
  - superconductors, 325
- Diffraction elements, 356
- Diffraction grating, *See* gratings, 48
- Diffraction limit, 17, 146, 377, 396
- Diffraction order, 358, 399
- Diffuse emitter, 48
- Diffuse-gray surface, 48, 267
- Diffuse mismatch model (DMM), 274
- Diffuse surface, 48, 266, 313
- Diffusion, 108, 238
  - electrons, 238
  - heat, *See* heat diffusion, 110
  - mass, 112
  - momentum, 109
- Diffusion coefficient, binary, 112, 118
- Diffusion length, 238
  - thermal, 238
- Diffusivity (mass, momentum, or thermal), 45, 114
- Digital voltmeter/multimeter (DVM), 168
- Dimensionality for solids, 151
- Dipole moment, 319
- Dipoles, 310
  - electric dipoles, 310, 314
  - induced dipoles, 310
  - magnetic dipoles, 310
  - thermally excited dipoles, 418, 427
- Dirac delta function, 161, 313, 326, 428, 459
- Direct simulation Monte Carlo, 115, 124
- Discharge glow by ionization of gas molecules, 229
- Discrete ordinates method ( $S_N$  approximation), 50, 269
- Dispersion relation, 149, 457
- Dissipative, 290, 308
- Dissipative structure, *See* nonequilibrium thermodynamics
- Distribution functions, 66, 102, 116, 451
  - equilibrium, 105, 117
  - free-path, 107–108
  - Gaussian or normal, 73–74, 364, 376, 453
  - isotropic, 104
  - molecular, 102, 106
  - nonequilibrium, 119, 257, 264
  - normalized, 97, 452
  - Planck's, 295
- Donor, 198, 233
- Doping concentration, 233
- Doppler shift, 69
- Double negative (DNG) materials, 327, 397
- Drift velocity, 155, 194, 234, 235
- Drude model for free carriers, 155, 315, 323, 326, 328
- Drude-Lorentz theory, 155–156
- Dual-phase-lag model, 254–255, 257
- Duality between electric and magnetic quantities, 309, 402
- Dulong-Petit law, 138, 265

- Effective mass, 216, 232, 317, 323
- Effective medium approximation (EMA), 362
- Effective medium formulation, 357, 359
- Effective medium theory (EMT), 361
- Effective temperature, 252, 259, 266
- electrons, 259
  - nonequilibrium, 252, 265, 269
  - phonons or lattice, 259
- Effective thermal conductivity, 175–182, 270–271
- Eigenchannel, 185
- Eigenfunction, 85, 148, 360
- Eigenvalue, 83, 91, 355
- Einstein coefficients, 93
- Einstein model of specific heat, 138, 143
- Einstein relation, 238
- Electrical conductivity, 155, 161
- Electrical resistivity, 158
- Electrochemical potential, 167, 184
- Electromagnetic spectra, 4, 92, 298
- Electromagnetic surface waves,  
*See* surface waves
- Electromagnetic waves, 285–294
- Electromotive force (emf), 168
- Electron configuration, 196
- Electron microscopy, 6, 12, 146
- Electron-phonon coupling constant, 259
- Electron-phonon scattering, 156–157
- Electron spin degeneracy, 144, 184, 196
- Electron tunneling, 3, 14, 230
- Electronic band structures, 157, 209, 214
- Electronic transitions, 89, 92
- bound-bound, 92
  - bound-free, 92
  - free-free, 92
  - interband, 215, 314, 316, 322
  - intraband, 215, 314, 322
- Electrostatic force, 199, 200
- Electrostatic limit, 412
- Electrostatic potential, 167–168, 184
- Emission of photons
- atomic, 88
  - diffuse, 48
  - fluorescence, 238
  - luminescence, 237
  - phosphorescence, 238
- Emission of photon (*Cont.*):
- radiative transitions, 49, 237, 415
  - spontaneous, 92, 386
  - stimulated, 92
  - thermal, 48, 295, 426
- Emissive power, 47–48, 295–297
- Emissivity or emittance, 48, 311–312, 422, 424
- Energy, 26
- Energy density
- near-field, 427, 434
  - phonon, 265
  - photon, 93, 290, 295
- Energy levels, 58, 85, 89, 183
- Energy storage modes, 75, 80
- Energy streamlines, 410–414
- Enthalpy, definition, 32
- Entropy, 27
- definition in statistical mechanics, 67
  - entropy intensity, 301
- Equation of phonon radiative transfer (EPRT), 263–271
- generation by irreversibility, 27
  - of mixing, 54
  - radiation entropy, 301–305
- Equation of radiative transfer (ERT), 50, 269, 303
- Equation of state, 37
- Equilibrium, 27, 29
- chemical, 30
  - mechanical, 30
  - stable-equilibrium state principle, 27
  - thermal, 30
- Equipartition principle, 78, 138, 146
- Error function, 364, 446
- complementary, 249, 446
- E-S* graph, 31
- Eucken's formula, 111
- Euler relation, 34
- Evanescence waves, 293, 359, 378, 386, 409, 429
- effective evanescent wave, 422
- Ewald-Oseen extinction theorem, 310
- Exponential attenuation or decay, 107, 293, 310, 380
- Exponential integral, 51, 179, 267
- Extinction coefficient, 291
- Extinction theorem, 310

- Fabry-Perot interferometer, 347–348  
 Fabry-Perot resonator, 386, 393, 405, 417, 421  
 Faraday's law, 286  
 Fermat's least-time principle, 283  
 Fermi-Dirac function, 144–145, 232  
 Fermi-Dirac (FD) statistics, 62, 65  
 Fermi energy or level, 69, 145, 215  
 Fermi function, *See* Fermi-Dirac function  
 Fermi velocity, 146, 156  
 Fermions, 66, 69  
 Fick's law, 112, 238  
 Field emission, 229–232  
 Figure of merit ( $ZT$ ), 172  
 Filling ratio, 362  
 Finesse, 348  
   *See also*  $Q$ -factors  
 Finite element method (FEM), 357, 386  
 Finite-difference time domain (FDTD), 357, 409  
 First law of thermodynamics, 26  
 Fluctuating current, 427–428  
 Fluctuational electrodynamics, 312, 426  
 Fluctuation-dissipation theorem, 426  
 Fluxes  
   charge, 160, 166  
   definition of, 103  
   entropy, 43, 173  
   heat or energy, 42, 104, 161, 429  
   momentum, 104  
   particle, 103  
 Fourier transform infrared (FTIR) spectrometer, 80, 334, 341, 346  
 Fourier's law, 42, 110, 120  
 Free electron gas, 62, 143  
 Free molecule flow, 121, 131  
 Free spectral range, 338  
 Fresnel's coefficients, 307, 309, 337, 380  
 Fresnel's rhomb, 380  
 Friction factor, 45  
 Frustrated total internal reflection, 379  
 Fullerene, 11, 154  
 Full-width-at-half-maximum (FWHM), 348, 421  
 Fundamental absorption process, 314  
 Fundamental relation in thermodynamics, 28  
 Gain medium, 94  
 Galvanometer, 168  
 Gauss's law, 286  
 Gauss's theorem, 286, 458  
 General dielectric, 287  
 Generation of electron-hole pairs, 237  
 Geometric optics, 342, 344, 371, 381  
 Geometric optics approximation, 363  
 Giant magnetoresistive (GMR) effect, 12, 195  
 Gibbs free energy, 33  
 Gibbs relation, 28  
 Gibbs-Duhem relation, 34  
 Goos-Hänchen shift, 379–382  
 Gratings, surface relief, 357  
   complex gratings, 417  
   diffraction order of, 358, 399  
   grating equation, 358–359  
   surface plasmon excitation in, 399  
   Wood's anomaly, 400  
 Gray medium, 50, 267  
 Gray surface, 48, 266  
 Green's function, dyadic, 427–428  
 Green's theorem, 286, 458  
 Ground-state energy, 31, 77, 91  
 Group velocity, 217, 265, 288, 386, 457  
 Guided modes in a waveguide, 383  
 Hagen-Rubens equation, 316  
 Hall effect, 193  
   quantum, 195  
   semiconductor, 235  
 Hamiltonian operator, 83, 210  
 Harmonic oscillator, 77  
 Heat capacity, 36  
 Heat carriers, 110  
 Heat conduction, 42, 110  
   ballistic, 131, 184, 258, 264  
   diffusive, 42, 110  
   by electrons, 158  
   nonequilibrium, 258, 263  
   non-Fourier, 250, 254, 257, 263  
   by phonons, 162  
   regimes, 275–277

- Heat diffusion, 110
  - equation, 42, 119, 250
  - kinetic description, 110
- Heat equations
  - hyperbolic, 250, 253, 262
  - lagging, 250, 254
  - parabolic, 42, 119, 250
  - See* dual-phase-lag model
  - See* Jeffrey's equation
  - See also* two-temperature model
- Heat interaction, 28
- Heat reservoir, 37
- Heat transfer, 41
  - conduction, 42
  - convection, 44
  - evaporative cooling, 114
  - free molecule, 129
  - in microchannel, 124
  - near-field radiation, 428
  - radiation, 46
- Heat wave, 251
- Helmholtz equation, 427
- Helmholtz free energy, 33
- Hermite polynomials, 90
- Heterogeneous state, 33
- Heterogeneous structures, 6, 10, 15, 182, 229
- Heterojunction, 9, 240
- Highest entropy principle, 27, 32, 68
- Histogram, 365, 451–452
- Hot electrons, 228
- Hot-film shear-stress sensors, 121
- Hot spots (local heating), 7, 137
- Hottel's zonal method, 50
- Hot-wire anemometers, 121
- Huygens' principle, 5, 283
- Hydrodynamic equations, 46, 117
- Hydrogen atom, 88
- Hydrogen molecules, ortho- and para-, 98
- Hydrogen technologies, 16
  
- Ideal gas, 38, 102, 105
- Information technology, 6
- Infrared radiation, discovery of, 283
- Insulators, 197, 198, 291
- Integrated circuits, 4, 6, 17, 240
- Intensities
  - blackbody, 48
  - Intensities (*Cont.*):
    - entropy, 302
    - optical, 341
    - phonon, 263
    - radiation, 47
    - Raman, 225
- Interference, 284, 333, 337, 341, 387, 390
- Intermolecular forces, 57, 72, 102, 115, 134
- Intermolecular potential, 115, 124, 200, 275
- Internal energy, 27
- International Temperature Scale (ITS), 30
- Ionization energy, 196, 198, 233
- Irradiance, 297, 312
  - total solar irradiance (TSI), 297
- Irreversible thermodynamics, 172–173
  
- Jeffrey's equation, 254, 275
- Joule heating, 170, 290
  
- Kapitza resistance, *See* thermal boundary resistance
- Kinetic temperature, 105
- Kinetic theory, 57, 101, 111, 116, 155, 162, 174
- Kirchhoff's approximation, 363
- Kirchhoff's law, 49, 284
- Knudsen number, 122
- Kramers-Kronig dispersion relations, 314–315
- Kretschmann-Raether configuration, 396–398
- Kronecker delta, 104, 428
- Kronig-Penney model, 213
  - electron, 174–182
  - molecular flow, 122, 129
  - phonon, 174, 264
  
- Lab-on-a-chip, 11
- Lagging behavior, 254
- Lagrangian multipliers, 68, 448
- Laguerre polynomials, associated, 88
- Landauer's formulation, 184
- Laser ablation, 333
- Laser cooling (and) trapping, 69, 302, 305

- Lasers, 8  
 diode laser, 368  
 history, 8  
 population inversion, 94  
 principle, 94  
 quantum well, 9  
 semiconductor, 9–10, 236  
 types, 9  
 ultrafast, 3, 8, 258, 261
- Latent heat, 35–37
- Lattice Boltzmann method, 123
- Lattice, *See* Bravais lattice
- Lattice constant, 148, 202, 204, 207  
 for photonic crystals, 353
- Lattice vibrations or lattice waves, 138, 139, 217
- Lattice wavevector, 148
- Left-handed materials (LHMs),  
*See* metamaterials
- Legendre polynomials, associated, 88
- Length scales, 2, 3, 247, 275
- Lennard-Jones potential, 115
- Lewis number, 113
- Light-emitting diodes (LEDs), 10, 208, 241
- Light line, 356, 398, 403
- Liouville equation, 116
- Lithography, 7, 17, 238  
 deep-UV, 8  
 dip-pen nano-, 15  
 e-beam nano-, 11  
 focused-ion beam (FIB), 11  
 photolithography, 7, 8  
 x-ray lithography, 8
- Local equilibrium, 3, 119, 160
- Local heating, 7
- Lorentz force, 194, 235
- Lorentz number, 158, 186
- Lorentz oscillator model, 318–321, 328
- Loss or lossy medium, 290, 327
- Lowest energy principle, 32
- Lyman series, 89
- Mach number, 122
- Macroscale regime, 105, 276
- Macrostate, 59
- Magnetic materials, 292, 326
- Matthiessen's rule, 156, 163, 176
- Matrix formulation, 348, 354, 386
- Maximum kinetic energy, 227
- Maxwell equations, 285, 287, 427  
 in inhomogeneous media, 357  
*See also* electromagnetic waves
- Maxwell relations, thermodynamic, 37, 53
- Maxwell-Boltzmann distribution, 64, 232
- Maxwell-Boltzmann (MB) statistics, 62, 66
- Maxwell's displacement current, 286
- Maxwell's velocity distribution, 73
- Mayer relation, 39
- Mean free path, 3, 105  
 bulk, 174, 177  
 distribution, 107  
 effective, 156, 163, 175  
 electron, 106, 156  
 molecule, 106  
 phonon, 163  
 photon, *See* penetration depth
- Mechanistic length, 3, 174
- Melting temperature or melting point, 35, 140, 159
- Memory effect, 254
- Memory function, 255
- Metal-oxide-semiconductor field-effect transistor (MOSFET), 7, 195, 240
- Metamaterials, 292, 326–329, 393–405, 421
- Michelson interferometer, 5, 341
- Microcavity, optical, 417
- Microchannel, 11, 121, 125
- Microelectromechanical systems (MEMS), 1, 10, 121
- Microelectronics, 6–8, 238–240
- Microfabrication or micromachining, 10, 11
- Microfluidics, 121–129  
 regimes, 122
- Micro-heat pipes, 121
- Micro/nanostructures, 1, 4, 121
- Micro-particle image velocimetry, 121
- Microscale regimes, 105, 276
- Microscopy, optical, 4, 377
- Microstate, 59
- Microwave, 219, 292, 298, 316, 327
- Miller indices for crystal planes, 205

- Miniaturization, 1, 3, 6, 10
- Mobility, 234–235
- Modes
  - for energy storage in ideal gases, 75, 80
  - of interaction, 26
  - optical fiber, 384
  - optical cavity, 418–420
  - phonon, 154, 219
  - quantum confinement, 183, 186
  - waveguide, 383
- Modified Fourier equation, *See* Cattaneo equation
- Molecular beam epitaxy (MBE), 9, 333
- Molecular chaos, 102
- Molecular dynamics simulation, 57, 82, 124, 185, 277
- Molecular electronics, 17
- Molecular hypothesis, 101
- Molecular weight, 38
- Molecule, 4–5
  - C60, 11
  - diameter of, 107
  - DNA, 14, 18
- Momentum flux, 104, 109
- Momentum space, *See* velocity space
- Monochromatic radiation, 46
- Monochromatic (radiation) temperature, 266, 302
- Monochromator, 334, 371
- Monte Carlo methods, 50
  - direct simulation Monte Carlo (DSMC), 115, 124
  - for phonons, 277
  - for surface scattering, 364
- Moore's law, 7
- Most probable microstate, 59
- Most probable wavelength, 273, 297
  
- Nanoaperture, 395, 409
- Nanocontact, 185
- Nanocrystals, 4, 148, 153
- Nanoelectromechanical systems (NEMS), 1, 121, 185
- Nanoelectronics, 2, 8, 377
- Nanofluidics, 121
- Nanofluids, 1, 15, 24
  
- Nanoindentation, 15
- Nanolithography, 15, 409
  - See also* lithography
- Nanoparticles, 17, 148, 153, 400
- Nanophotonics, 10, 395
- Nanostructures, other, 4, 12, 121, 170, 305, 333, 400
- Nanotechnology, 1, 19
- Nanowires, 3, 12, 16, 151, 172, 185, 395
- National Nanotechnology Initiative (NNI), 19
- Navier-Stokes equation, 46, 118, 122
- Near-field
  - optics, 377, 413
  - radiative transfer, 3, 389, 431
- Near-field scanning optical microscopes (NSOMs), 4, 17, 377–378
- Nearly free electron model,
  - See* one-electron model
- Negative absolute temperature, 98
- Negative absorption, 93
- Negative index materials (NIM),
  - See* metamaterials
- Negative refractive index, 326, 328
- Nernst theorem, 31
- Neutron scattering, 143, 149, 219, 225
- Newton's law of cooling, 45
- Newton's law of motion, 57, 124, 155
- Newton's law of shear stress, 45, 110
- Newton's prism, 5, 283
- Nonabsorbing, nondissipative, or lossless medium, 286, 308
- Nonequilibrium thermodynamics, 174
- Nonmagnetic materials, 286, 287, 292
- Nonradiative transitions, 237, 415
- Nuclear spin degeneracy, 79, 88, 98
- Number density
  - electrons and holes, 232
  - electrons in a metal, 145
  - molecules, 39, 102
  - phonons, 141
  - photons (occupation number), 295
- Nusselt number, 45, 128
  
- Occupation number, 295
- Ohm's law, microscopic, 155, 286
  - at high frequencies, 292



- One-electron model, 210  
 Onsager reciprocity, 173  
 Onsager's theorem, 172  
 Optical communication, 10, 382–383  
 Optical constants, 291  
   *See also* complex refractive index  
 Optical fiber, 10, 382  
   applications, 10  
   holey or photonic crystal, 386  
   modes, 383  
   principle, 310, 383  
 Optical path length, 50  
 Optical properties, 292, 333, 357, 396, 406  
   complex refractive index, 291–292  
   *See* absorption  
   *See* dielectric functions  
   *See* emission  
   *See also* radiative properties  
 Optical tweezers, 301  
 Optically thick limit, 51, 264  
 Optically thin limit, 51  
 Optoelectronics, 8, 236–242, 382, 386  
 Organ pipe resonance modes, 417, 420  
 Oscillator strength, 319  
 Otto configuration, 397  
  
 Partial coherence, 340–346  
   coherent limit, 337, 342  
   complex degree of coherence, 341  
   incoherent limit, 334, 342, 383  
   mutual coherence function, 341  
 Participating media, 50  
 Partition function, 67, 75, 79, 98  
 Pauli's exclusion principle, 62, 192, 196  
 Peclet number, 128  
 Peltier effect, 168–169  
 Penetration depth, radiation, 106, 293  
   for phonons, 264  
 Perfect gas, 39  
 Perfect lens, *See* superlens  
 Periodic microstructures, 333, 352, 356, 408, 409, 417  
 Periodic potential, 209, 211  
 Permeability, magnetic, 286, 292, 379  
 Permittivity, electric, 286, 291, 379  
 Permutation, 448  
  
 Perpetual motion, 26  
 Perpetual-motion machine of the first kind (PMM1), 26  
 Perpetual-motion machine of the second kind (PMM2), 29  
 Phase diagrams, 35–36  
 Phase lag, 254–255  
 Phase-matching condition, 348, 354, 379, 427  
 Phase rule, Gibbs, 34  
 Phase shift, thin film, 337  
 Phase space, 59, 71  
 Phase velocity, 287, 456  
 Phonon branches, 165, 217  
   acoustic, 154, 166, 218  
   optical, 166, 218  
 Phonon dispersion relations, 165, 218, 219  
 Phonon modes, 141, 154, 166  
   axial, 153  
   longitudinal, 219  
   planar or surface, 153  
   transverse, 219  
   twisting or torsional, 154  
 Phonon-phonon scattering, 162, 221–222  
   four-phonon processes, 222  
   normal or  $N$ -processes, 221, 257, 262, 275  
   three-phonon processes, 221  
   umklapp or  $U$ -processes, 221, 257, 262  
 Phonons, 139, 217, 221  
   absorption or emission of, 224, 322–323  
   dispersion, 217  
   phonon gas, 137  
   polarization, 149, 165  
   radiative, 263  
   scattering, 221  
 Photoconductivity, 236  
 Photocurrent, 240–241  
 Photoelectric effect or photoemission, 5, 94, 226  
 Photon tunneling, 378, 386, 405, 425  
 Photonic crystal fibers (PCFs), 386  
 Photonic crystals, 352–356, 422  
   Bloch wavevector, 354, 391  
   effective evanescent wave, 422  
   pass band, 355  
   stop band, 355

- Photons, radiation quanta
  - momentum, 61, 93
  - photon gas, 294
- Photovoltaics, 15, 240, 241
- Piezoelectricity, 14, 121
- Planck's constant, 48
- Planck's law for blackbody radiation,
  - 48, 63, 285, 294
  - derivation, 294
  - limitations, 305, 415–416
- Plane waves, 306, 455
  - angle of incidence, 306
  - constant-amplitude planes, 293, 337
  - constant-phase planes, 293, 337
  - group front, 288
  - plane of incidence, 306
  - transverse electric (TE), 306
  - transverse magnetic (TM), 309
  - wavefront, 283, 287
- Plasma frequency, 316, 328, 435
- Plasma oscillation, 317, 396
- Plasmon, *See* surface plasmon
- $p$ - $n$  junction, 238–240
- Poiseuille flow, 126, 128
- Polariton dispersion relations, 396, 402
- Polaritons, 396
  - bulk polaritons, 404, 409
  - coupled surface polaritons, 401
  - localized surface plasmon polaritons, 395, 401
  - phonon polaritons, 395, 400
  - surface plasmon, 396–397
  - surface polaritons, 396
  - See also* surface waves
- Polarization, 288–290
  - circularly polarized, 289
  - co-polarization, 366, 367
  - cross-polarization, 366, 367
  - depolarization, 365
  - elliptically polarized, 290
  - linearly polarized, 289
  - longitudinal, 141, 154
  - parallel or  $p$ -polarized, 309
  - perpendicular or  $s$ -polarized, 306
  - propagation length, 399
  - randomly polarized, 290
  - unpolarized, 290
  - vibration ellipse, 289
- Polarization vector, 319
- Polychromatic light or radiation, 288, 339
- Polymethyl methacrylate (PMMA), 15, 413
- Positive index materials (PIM), 326
- Potential barrier, 229
- Potential well, 84
- Potentiometer, 168
- Poynting vector, 290, 308, 352, 411, 428
- Prandtl number, 45, 111, 128
- Pressure, 28, 104, 301
- Principal angle, 310
- Principal values of integral, 315
- Probability, 450–451
- Probability density function (PDF), 61, 83, 107, 453
- Properties, thermodynamic, 25, 33
  - additive, 26, 27
  - extensive, 34
  - intensive, 33
  - specific, 34
- Pseudopotential method, 214
- Pulse heating, 258
- Pump-and-probe method, 260
- $Q$ -factors, 348, 386, 421
- Quantization, 81, 85, 92
  - of conductance, 185
  - of specific heat, 153
- Quantum computing, 8, 185
- Quantum conductance, 182–187
  - thermal, 186
- Quantum confinement, 149, 172, 183
  - electron density of states, 183
  - phonon density of states, 149–151
- Quantum dots (QDs), 10, 17, 148, 183, 400
- Quantum efficiency, 237, 241
- Quantum electrodynamics (QED), cavity, 10, 386, 415
- Quantum number, 85, 89, 196
- Quantum size effect, 148, 182
  - second quantum size effect, 153
  - on specific heat, 148–154
  - See also* quantum conductance
- Quantum states, 61, 85
- Quantum theory, 58, 85

- Quantum tunneling, *See* electron tunneling and photon tunneling
- Quantum wells, 9, 85, 149, 151, 183  
multiple quantum wells, 9, 172
- Quantum wires, 183
- Radiance, 47, 300, 312  
*See also* intensities, radiation
- Radiance temperature, 300, 303
- Radiation detectors, 10, 236  
bolometer, 236, 284  
charge-coupled devices (CCD), 10  
MCT, 208, 236  
nonequilibrium or nonthermal, 236  
photoconductive, 236  
photodiode, 368  
photomultiplier tube (PMT), 225  
photovoltaic, 240  
thermal or bolometric, 236  
thermopile, 284
- Radiation jump (or slip), 269, 270
- Radiation pressure, 53, 301
- Radiation tunneling, *See* photon tunneling
- Radiative cooling, 225, 302
- Radiative equilibrium, 265, 267
- Radiative properties, 48, 308–309, 312–313, 334–340  
bidirectional, 312–313  
directional, 48  
directional-hemispherical, 48  
hemispherical, 48  
normal, 311  
spectral, 47, 48  
spectral and directional control, 414–425  
spectrally averaged, 342  
specular, 345  
total, 47, 48
- Radiative thick limit, 264, 268
- Radiative thin limit, 264, 266
- Radiometer, cryogenic, 298
- Radiosity, 267, 268
- Rad-Pro software, 324, 352
- Raman scattering, 219, 224, 401  
anti-Stokes shift, 224  
Stokes shift, 224
- Raman spectroscopy, 80, 224  
micro-Raman, 277
- Rapid thermal processing, 7, 300, 333, 362
- Rarefied gas dynamics, 121
- Rayleigh scattering, 49, 400
- Rayleigh-Jeans formula, 285, 296
- Rayleigh-Rice perturbation theory, 363
- Rayleigh-Wood anomaly, 400, 419
- Ray-tracing method, 334, 351, 364
- Reciprocal lattice, 209
- Reciprocal lattice space, 149, 209
- Reciprocal lattice vector, 209, 211
- Recombination or annihilation, 237  
Auger effect, 237  
multiphonon emission, 237  
nonradiative, 237  
radiative, 237  
recombination lifetime, 237
- Reflectance or reflectivity, 48  
in ATR configuration, 398, 403–404  
with a metallic grating, 399  
by a microfacet, 366  
for multilayer structures, 347–352  
for photonic crystals, 357  
from rough surfaces, 362–371  
between semi-infinite media, 308, 309  
for a thick film, 334–335  
for a thin film, 335–337
- Reflection, 285, 306  
diffuse, 313  
Lambertian surface, 313  
specular or mirror-like, 48, 307
- Reflection coefficient of electrons, 228
- Reflection coefficients for electromagnetic waves, 307, 309, 337, 398
- Refraction, 285, 306
- Refractive index, 288, 291
- Relativity, special theory of, 94–95
- Relaxation time, 106, 117, 156, 162, 253  
energy relaxation time, 162  
momentum relaxation time, 162  
second relaxation time, 276
- Relaxation-time approximation, 117, 119, 160, 253
- Resonance frequency, 90, 319, 328  
for a harmonic oscillator, 90  
in metamaterials, 328  
for phonon oscillators, 319  
of polaritons, 407, 418

- Resonance tunneling, 391–393, 405  
 Rest energy, 60, 94  
 Reststrahlen band, 320, 400  
 Retardation time, 254, 261  
 Retroreflection, 369, 471  
 Reynolds number, 45, 121, 128  
 Richardson constant, 228  
 Richardson–Dushman equation, 228  
 Riemann zeta function, 446  
 Rigid rotor, 86  
 Rigorous coupled-wave analysis (RCWA), 358–360, 409  
 Roughness statistics, 345, 363  
   anisotropy, 367, 371  
   autocorrelation length, 363  
   Gaussian surface, 345, 364, 376  
   height statistics, 364, 451  
   microfacets, 364  
   power spectral density (PSD), 363  
   rms roughness, 363, 451  
   shadowing function, 364  
  
 Sackur-Tetrode equation, 73  
 Saturated liquid, 35  
 Saturated vapor, 36  
 Saturation dome, 36  
 Scalar scattering theory, 345  
 Scanning electron microscope (SEM), 12  
 Scanning near-field optical microscope (SNOM), *See* NSOM  
 Scanning probe microscopes (SPMs), 1, 4  
 Scanning thermal microscope (SThM), 15, 277  
 Scanning tunneling microscope (STM), 14, 231  
 Scattering  
   albedo, 50  
   cross-section, 106, 400  
   diffuse, 176, 178  
   elastic, 180, 272  
   inelastic, 178, 305  
   phase function, 50  
   probability, 117  
   specular, 180  
 Scattering coefficient, 50, 400  
 Scattering matrix (*S*-matrix), 185  
  
 Scattering rate, 107, 156, 162, 221–223, 316  
 Scatterometer, optical, 368  
 Schmidt number, 113  
 Schrödinger equation, 61, 82  
   one-electron model, 210  
   solutions, 84, 91  
   time-dependent, 83  
 Second law of thermodynamics, 27, 252  
   Clausius statement, 29, 30  
   Kelvin-Planck statement, 29  
   *See* entropy  
 Second relaxation time, 276  
 Second sound, *See* temperature wave, 253, 257, 276  
 Seebeck effect, 167  
 Self-assembly, self-organization, 12, 18, 174  
 Semiconductor, 197  
   electrical conductivity, 234  
   extrinsic, intrinsic, 198  
   *n*-type, *p*-type, 198, 199, 238  
   wide band, 198  
 Semimetal, 197, 198, 208  
 Shear stresses, 45, 104, 109  
 Simpson's rule, 431  
 Single-negative (SNG) material, 407, 410, 412  
 Size effect  
   classical, *See* boundary scattering  
   *See* quantum conductance  
   *See* quantum size effect  
   *See* specific heat  
   *See* thermal conductivity  
 Slip boundary condition, 126  
 Slip flow, 123, 126  
 Slope distribution function (SDF), 364, 367  
 Smith shadowing function, 364, 365  
 Snell's law, 272, 307, 335  
 Solar cells, dye-sensitized, 16  
 Solid angle, 47, 265, 312, 369  
 Solid-state energy conversion  
   devices, 15, 166, 229  
 Sommerfeld constant, 147  
 Sommerfeld expansion, 145, 459  
 Specific heat, 36  
   constant pressure, 36

- Specific heat (*Cont.*):  
 constant volume, 36  
 ratio, 39
- Specific heat of nanostructures, 148  
 carbon nanotubes, 154  
 graphene, graphite, 154  
 nanocrystals, 153  
 nanoparticles, 153  
 nanowires, 151, 153  
 quantum wells, 151  
 second quantum size effect, 153  
 thin films, 151
- Specific heat models, 39–40, 137  
 electron contribution, 143  
 ideal gases, 39  
 ideal incompressible liquids, 40  
 ideal incompressible solids, 40  
 lattice contribution, 138–143
- Spectral energy density, 295, 428
- Spectral heat flux, 47, 265, 429
- Specularity (parameter), 180, 273
- Speed, 60, 74, 146  
 average, 74, 146  
 distribution, 73–74  
 root-mean-square, 74, 164  
 of sound, 53, 166, 191
- Spherical coordinates, 47
- Spherical harmonic method  
 ( $P_N$  approximation), 50, 269
- Standard conditions, 38, 57
- Standard deviation, 81, 85, 97, 368
- Standing waves, 85, 148, 383, 386,  
 401, 419
- State principle, stable-equilibrium, 27
- Statistic hypothesis, 102
- Statistical ensembles, 81  
 canonical, 81, 92  
 fluctuations, density, 81  
 grand canonical, 81  
 microcanonical, 81
- Steam table, 41
- Stefan-Boltzmann constant, 48, 296  
 measurement, 298  
 one dimensional, 186  
 for phonons, 264
- Stefan-Boltzmann law, 48, 284, 296
- Stokes' hypothesis, 46, 118
- Sublimation, 36
- Superconductivity, 63, 69, 168  
 BCS theory, 325  
 Cooper pairs, 69  
 crystal structure, 206, 208  
 high- $T_c$ , 69, 169, 325  
 magnetic resonance imaging  
 (MRI), 12  
 SQUIDs, 12  
 Thermal conductivity, 177  
 two-fluid model, 325–326
- Superfluidity, 63, 69  
 liquid helium, 69, 257  
 $\lambda$ -point or transition, 69, 257  
 second sound, 257  
 two-fluid model, 257
- Superlattices, 9, 85, 148, 172, 182
- Superlens, 410
- Supermolecule, 200
- Surface scattering, 345, 364, 368
- Surface-enhance fluorescence microscopy,  
 401
- Surface-enhance Raman microscopy  
 (SERS), 401
- Surface forces, 3, 109, 212
- Surface plasmon, 396
- Surface realization or generation, 364  
 rejection method, 365  
 spectral method, 364
- Surface roughness, 345, 367  
*See also* roughness statistics
- Surface topography, 367  
 AFM measurements, 365, 367  
 anisotropic surface, 365, 367  
 spatial resolution, 365  
 stylus profiler, 367
- Surface-to-volume ratio, 121, 153
- Surface waves, electromagnetic, 285, 378,  
 395, 422
- Suspended MEMS bridges, 186
- $T^3$  law, 143, 265
- Temperature, 28  
 absolute zero, 31  
 kinetic temperature, 105  
 scale, 30  
 of the sun, 397  
 thermodynamic, 28  
 of the universe, 299

- Temperature jump, 123–127, 269, 270
- Temperature measurement,
  - See* thermometry
- Temperature pulse, 251
- Temperature wave, 251, 257, 262
  - negative entropy generation, 252
  - speed of propagation, 251
- Thermal boundary resistance (TBR), 262, 271, 273–274, 277
- Thermal conductivity, 42, 119
  - classical size effect, 174–178
  - derivation from BTE, 119
  - effective for rarefied gases, 131
  - ideal gases, 111
  - insulators, 162–166
    - across layered structures, 262–271
    - metals, 158–162
    - nanotubes, 186
    - quantum size effect, 286
    - superlattices, 172, 182
    - along a thin film, 178
    - along a thin wire, 181
- Thermal creep, 126
- Thermal diffusion
  - average speed, 249, 257
  - infinite-speed paradox, 248–249, 281
  - See* heat diffusion
- Thermal equilibrium, 27, 30, 415
- Thermal fluctuations, 427
- Thermal metrology, 277
- Thermal radiation, 46, 283, 426
- Thermal resistance, 44, 172, 268–271
- Thermal time constant, 261
- Thermal velocity of electrons, 224, 235
- Thermal wave, 251
- Thermal wavelength of phonons, 151, 166
- Thermalization time, 261
- Thermionic emission, 227–229, 231, 436
- Thermionic refrigeration, 229
- Thermocouple, 168, 284
- Thermocouple junction, 168
- Thermodynamic cycles, 26
  - Carnot, 40, 172, 284, 302
  - Rankine, 40
- Thermodynamic equilibrium, 27, 68, 304
- Thermodynamic probability, 59, 66, 69
- Thermodynamic processes, 26
  - adiabatic, 28
  - irreversible, 26
  - quasi-equilibrium or quasi-static, 29
  - reversible, 26
  - spontaneous, 26
- Thermodynamic systems, 25
  - closed, 28
  - constituents, 25
  - environment or surroundings, 25
  - isolated, 26
  - open, 28
  - parameters, 25
- Thermoelectric devices, 15, 170
  - figure of merit, 172
  - with nanostructured materials, 172
  - thermal efficiency, 171
- Thermoelectric effect or thermoelectricity, 15, 166
  - absolute thermopower, 169
  - cooling or refrigeration, 170
  - power generation, 170
  - thermoelectric voltage, 168
- Thermometry, 30, 298
  - absolute, 298
  - lightpipe, 7
  - radiation, 298
  - Raman, 277
- Thermophotovoltaics (TPV), 15, 240, 379, 418, 425
- Thermopower, *See* thermoelectric effect
- Thermorefectance, 260, 277
- Thin films, 3–4, 9, 333
  - radiative properties, 335, 337
  - specific heat, 151
  - thermal conductivity, 174, 268
- Thin-film optics, 336, 342, 347
- Third law of thermodynamics, 31, 69
- Thompson effect, 169
- Tight-binding model, 214
- Time-harmonic, 287, 290, 357, 385
- Total internal reflection, 310, 378, 383
- Transfer matrix method (TMM), 409
- Transistors, 6, 240
  - field-effect transistor (FET), 240
  - MOSFET, 7

- Transition flow, 121
- Transmission coefficients
  - for electromagnetic waves, 307, 309, 337
  - for phonons, 272
  - in quantum conductance, 185
  - in quantum tunneling, 230–231
- Transmission electron microscope (TEM), 12
  - high-resolution, 12
- Transmission enhancement
  - with the excitation of polaritons, 405–407
  - by nanostructures, 408–409
  - by photon tunneling, 386–389
  - by resonance tunneling, 391–393
- Transmission probability, 230–231
- Transmissivity, internal, 335
- Transmittance, 48
  - including surface roughness, 344–346
  - of multilayer thin films, 347–352
  - with phonons, 272–273
  - spectrally averaged, 339, 342
  - of a thick film, 335
  - of a thin film, 337
- Transport equations, 108, 116
- Triple point, 35
- Two-phase mixture, 34
- Two-relaxation-time approximation, 257, 276
- Two-temperature model, 258, 261
- Ultra-shallow junction, 8
- Ultraviolet catastrophe, 297
- Uncertainty principle, 61, 85
- Unit cells, 201
  - basis, 201, 206
  - conventional unit cell, 201
  - lattice, *See* Bravais lattice
  - lattice points, 201
  - primitive unit cell, 201
- Universal gas constant, 38, 443
- Vacuum fluctuations,
  - See* zero-point energy
- Valence band, 197, 215, 232, 322
- Valence electrons, 196, 201, 216
- van der Waals force, 115, 200
- Velocity
  - bulk or mean, 109, 118
  - distribution, 73, 104
  - random or thermal, 109, 118
- Velocity slip, 124, 127
- Velocity space, 60
- Vertical cavity surface emitting laser (VCSEL), 9
- Very-large-scale integration (VLSI), 7
- Vibration-rotation spectrum, 49, 80
- Virial theorem, 124
- Viscosity, 45, 110
  - dynamic, 110
  - kinematic, 45
  - microscopic description, 110
- Volumetric entropy generation rate, 173, 303
- Volumetric heat capacity, 158, 256, 259, 265
- Volumetric thermal energy generation rate, 42
- von Klitzing constant, 195
- Waves
  - acoustic, 162, 165, 271, 288
  - backward, 349
  - evanescent, 293
  - forward, 349
  - homogeneous, 293
  - inhomogeneous, 293
  - packet, 457
  - propagating, 293
  - standing, 85, 148
  - surface waves, 285, 396
  - wave optics, 342, 344
  - See also* plane waves
- Wavefront, 251, 283
- Wavefunctions, 82–84, 91, 211, 212, 230
- Waveguides, 382–386
  - conducting, 386
  - dielectric, 386
  - guided mode, 383
  - plasmon, 400
- Wavenumber, 80, 298
- Wave-particle duality, 6, 60, 82, 220
- Wavevector, 221, 287
  - in electromagnetic wave, 287, 291
  - lattice wavevector, 148, 211

- Wavevector (*Cont*):  
  wavevector space or  $k$ -space, 149,  
  209, 288
- Weak-potential assumption, 211, 213
- Whispering gallery mode (WGM), 386
- Wiedemann-Franz law, 158, 162,  
  180, 182
- Wien's displacement law, 166, 297
- Wien's formula, 285, 296
- Wigner-Seitz primitive cell, 205, 209
- WKB approximation, also BWK or  
  KWB, 230
- Wood's anomaly, *See* Rayleigh-Wood  
  anomaly
- Work function, 227, 229
- Work interaction, 28, 302
- X-ray crystallography, 201
- X-ray diffraction, 245
- X-ray lithography, 8
- X-ray photoelectron spectroscopy  
  (XPS), 227
- X-ray properties, 4, 317
- Young's double-slit experiment, 5,  
  284, 341
- Zeolites, 220
- Zero absolute temperature, *See* third law of  
  thermodynamics
- Zero-entropy states, 32
- Zero-point energy, 91, 149, 428
- Zeroth law of thermodynamics, 30, 262
- ZnO nanobelts and nanowires, 12,  
  13, 16, 201
- Zone edge, 210, 212