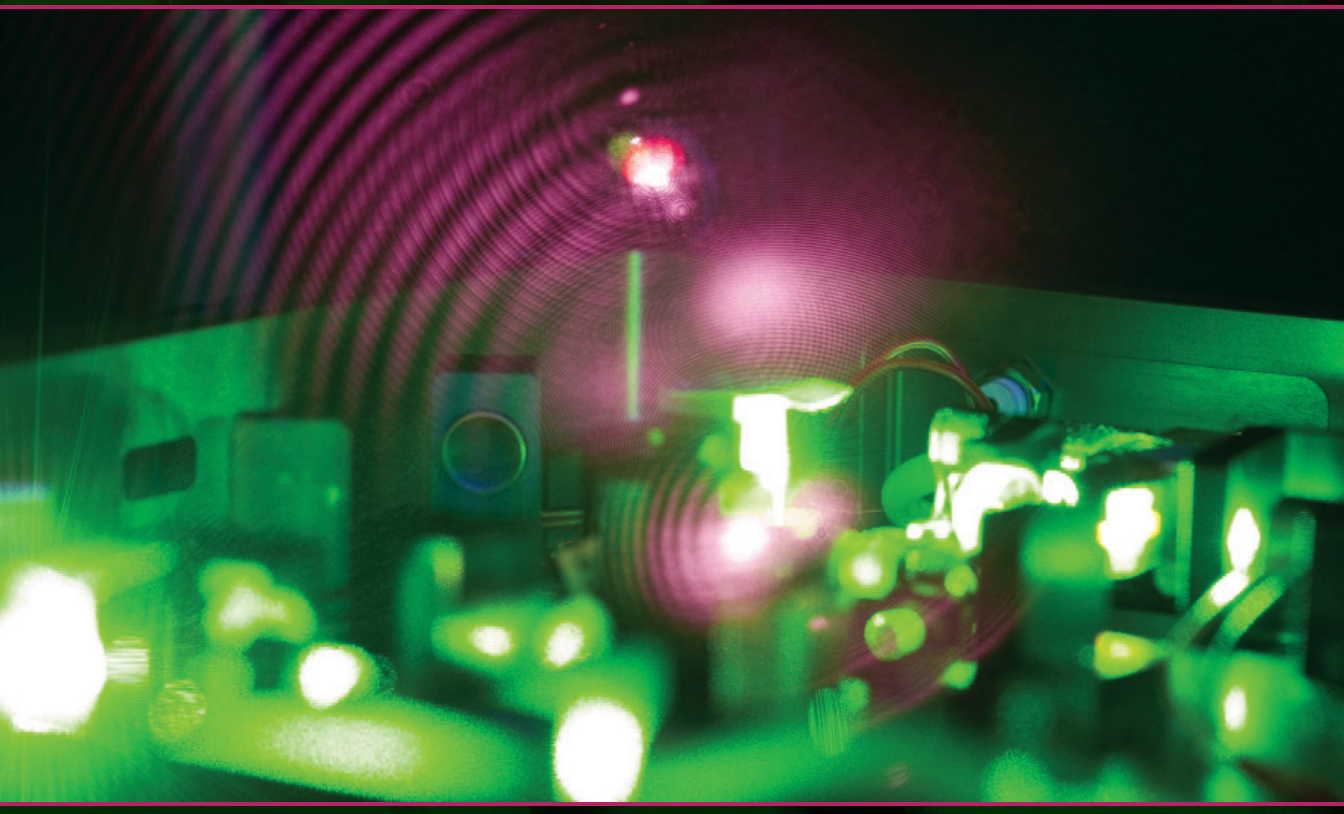
OPTICAL TECHNIQUES FOR SOLID-STATE MATERIALS CHARACTERIZATION



Edited by
Rohit P. Prasankumar
Antoinette J. Taylor

CRC Press
Taylor & Francis Group

OPTICAL TECHNIQUES FOR SOLID-STATE MATERIALS CHARACTERIZATION

OPTICAL TECHNIQUES FOR SOLID-STATE MATERIALS CHARACTERIZATION

Edited by
Rohit P. Prasankumar
Antoinette J. Taylor

CRC Press
Taylor & Francis Group
Boca Raton London New York

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

Cover Image: Volker Steger

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

Version Date: 20110705

International Standard Book Number-13: 978-1-4398-1437-6 (eBook - PDF)

To Anuradha (Rohit P. Prasankumar)

CONTENTS

PREFACE

> You can't study the darkness by flooding it with light.
>
> **Edward Abbey**

Or can you? From the earliest beginnings of humankind, darkness has evoked emotions of fear and uncertainty in nearly all of us. The control of fire and invention of the light bulb helped alleviate these feelings, but to this day, when entering a dark room, we usually respond by turning on the light. This reaction speaks to the primacy of vision among our senses, perhaps because we can immediately obtain a massive amount of information by simply opening our eyes and looking at the world around us. In much the same manner, one can rapidly learn a lot about a material by simply shining light on it. Accordingly, it is not surprising that optics has been used to study both natural and man-made phenomena from the beginning of recorded history.

In fact, light is virtually an ideal tool for studying nearly any material or phenomenon, as optical techniques have several notable advantages over other methods. Paramount is their noncontact and nondestructive nature, along with the ability to broadly tune from x-ray to terahertz (THz) frequencies, generate attosecond (10^{-18} s) duration pulses, and probe nanoscale dimensions (<10 nm). Therefore, by using one or more of the myriad optical techniques that have been developed over the last century for characterizing materials, substantial insight into phenomena ranging from high-temperature superconductivity to protein folding can be obtained.

The purpose of this book is thus to describe both basic and advanced experimental optical techniques that are commonly used to study materials. To the best of our knowledge, a single volume that describes the essential experimental techniques for optically characterizing materials does not exist, and relatively few references discuss any of these methods in detail. This then presents an opportunity to address these issues by describing established optical experiments in the same volume as more recently developed temporally and spatially resolved methods. More specifically, we aim to describe these techniques in enough detail for researchers with different levels of experience to build and/or use a working setup, acquire data on both simple and complex materials, and analyze and interpret these data to obtain insight into fundamental material properties with minimal reliance on other sources.

To accomplish this, we solicited contributions from experts who have pioneered these techniques and extensively applied them to characterize complex materials in their own laboratories. Although we have attempted to include the most widely known optical techniques for characterizing materials, this book is not meant to be exhaustive; several important methods have undoubtedly been left out, either inadvertently or due to space limitations. In addition, although our intention is to describe each of these techniques in extensive detail as mentioned

above, in a book of this nature we are bound to have overlooked some potentially important issues, for which we apologize in advance. It is worth pointing out that all of the optical techniques described here can be used, with some (usually minor) variations, to characterize non-solid-state materials (e.g., soft matter, gases); we chose to concentrate on solid-state materials to avoid diluting the content and remain within our primary area of expertise.

Part I provides background information on light–matter interactions (Chapter 1), semiconductors (Chapter 2), and metals (Chapter 3). This knowledge will be critical in understanding the experimental techniques and results described later in the book. We assume that the reader has an undergraduate-level understanding of optics, electromagnetics, quantum mechanics, and solid-state physics, along with some knowledge of basic components used in experimental optics (e.g., lenses, wave plates, filters). However, many readers may be experts in one of these areas but only somewhat familiar with other subjects, particularly regarding the information that is most relevant for understanding data obtained with different optical probes. Our aim in this part is therefore to provide a broad knowledge base for understanding the experiments and results discussed later in the book, with additional, more specialized details contained in the appropriate chapters.

Part II focuses on linear, time-integrated optical experiments used to measure basic optical properties such as transmission, reflectivity, and luminescence. These experiments can provide essential information on characteristic physical quantities, such as the dielectric function and optical conductivity, that are intimately linked to a material's microscopic properties (e.g., the electronic and vibrational structure). They can also reveal the properties of elementary excitations including excitons, phonons, and magnons. To this end, Chapter 4 discusses standard methods for obtaining the optical constants of a material (e.g., the refractive index and absorption coefficient), focusing on the techniques of visible-ultraviolet (UV) spectroscopy and Fourier transform infrared spectroscopy (FT-IR). Space limitations prevented us from a detailed description of ellipsometry, another established technique for measuring optical constants, but several references are cited for the interested reader.

Chapter 5 discusses methods for obtaining the optical constants after continuous wave (cw) laser excitation, focusing on photoluminescence (PL), photoluminescence excitation (PLE), and photocurrent (PC) spectroscopies. These measurements reveal the transitions involved in light absorption and emission, as well as energy transfer between different electronic levels. Finally, Chapter 6 treats Raman scattering, describing how one can obtain symmetry, energy, and lifetime information on a wide range of elementary excitations, even under extreme conditions (e.g., high pressure and magnetic fields).

Time-resolved optical spectroscopy is discussed in Part III, with a focus on optical techniques that can characterize the dynamic properties of materials with femtosecond (or better) time resolution. To a large extent, each of these techniques is capable of temporally resolving photoinduced changes in one or more of the properties discussed in Part II (e.g., absorption, optical conductivity) while providing insight on properties that can only be accessed with dynamic measurements (e.g., dephasing and population relaxation times). All of these methods depend on ultrashort pulsed lasers to achieve femtosecond time resolution, and therefore the basics of ultrashort pulse generation and measurement are covered in Chapter 7, with a particular focus on understanding the principles of femtosecond laser operation and practical knowledge necessary for using these lasers in the laboratory. In addition, the disparate phenomena experienced by quasiparticles after femtosecond photoexcitation (carrier–carrier/carrier–phonon scattering, radiative/nonradiative recombination, defect/surface trapping, etc.) depend on the material being studied, but are independent of the technique used to probe them. Chapter 8 describes these processes in semiconductors and metals, along with some examples in more complex materials, to provide a basis for the experimental results discussed later in this part.

The remaining chapters in this part provide detailed descriptions of the most common time-resolved optical techniques, each capable of probing different phenomena. Chapter 9 describes pump–probe spectroscopy, arguably the simplest time-resolved technique, that can temporally resolve photoinduced transmission and reflection changes and relate these changes back to dynamical material properties (e.g., carrier population and transient electronic temperature). Four-wave mixing spectroscopy, capable of measuring dephasing processes and many-body interactions between different excitations, is discussed in Chapter 10. Chapter 11 then describes terahertz spectroscopy, which can probe both the static and dynamic properties of low-energy excitations. Time-resolved photoluminescence spectroscopy, discussed in Chapter 12, is used to measure radiative recombination and energy transfer processes. Spin dynamics in complex materials can be measured using time-resolved magneto-optical spectroscopy, described in Chapter 13. Finally, Chapter 14 discusses time-resolved Raman scattering, which can directly address the dynamical properties of low-energy excitations such as phonons and magnons, in a manner complementary to the other techniques described in this part.

The ability to perform optical measurements with high spatial resolution is important when characterizing materials, particularly due to the predominance of both naturally occurring and man-made materials with nanoscale features among materials of great current interest (e.g., quantum-confined nanostructures, metamaterials, and transition metal oxides). Part IV thus contains three chapters on conventional optical microscopy, micro-optical techniques, and near-field scanning optical microscopy. Chapter 15, on conventional optical microscopy, somewhat differs from the other chapters in this book, since a multitude of references already exist on the practical aspects of optical microscopy. Therefore, this chapter begins with a detailed description of resolution and image formation, followed by an overview of different microscopic techniques and ending with a detailed description of imaging interferometric microscopy. Micro-optical techniques, in which diffraction-limited spatial resolution can provide additional information on a given sample beyond the measurements discussed in Part II, are discussed in Chapter 16. Finally, Chapter 17 describes near-field scanning optical microscopy, a powerful tool for characterizing a variety of materials with nanometer spatial resolution.

Part V deals with more advanced optical techniques for characterizing materials that are, arguably, insufficiently developed to merit their own chapters, but worthy of inclusion in this volume, especially since they will likely develop into more established methods in the coming years. This part contains only one chapter, which describes advanced time-resolved optical techniques for resolving ultrafast photoinduced changes in a material's lattice and electronic structure, and then discusses wide-field and scanning optical techniques that combine high temporal and spatial resolution in a single experiment. It is worth noting that Chapter 15 gives an overview of advanced spatially resolved techniques, which therefore are not discussed in this part.

Certain chapters contain information useful for performing nearly any optical experiment. Chapters 4, 5, 6, and 16 discuss (in varying levels of detail) common light sources, optical components, detectors, spectrometers, and cryostats used in time-integrated optical experiments, which is useful background for almost all of the experiments discussed in this book. There is some redundancy between these chapters; however, our overall approach was to include all of the information essential for understanding and implementing a given technique in the relevant chapter, with references to other chapters or other sources that give more detailed information on a given aspect of each method. In general, we thought it was best to err on the side of including too much rather than too little information.

Chapter 4 provides a brief procedure for polishing samples, which is relevant for optical measurements on samples with rough surfaces. Table 4.1 also summarizes the spectral ranges covered by different IR sources, beam splitters, detectors, and optical window materials. Chapter 5 includes a discussion of laser safety practices, important reading for anyone

working with lasers. The use and properties of metal contacts on a sample are also described in the context of photocurrent measurements in this chapter. Finally, it also discusses how to perform a simple knife edge scan for determining the width of a Gaussian beam, which is then discussed in more detail in Chapter 16. Chapter 6 includes a description of Raman scattering measurements under high pressure and magnetic field, which, to a large extent, can be adapted for many of the other optical experiments discussed here.

Chapter 9 is worth examining for any reader unfamiliar with time-resolved experiments, since it contains a great deal of information that is common to all of the techniques discussed in Part III; in some respects, the other experiments described in this part are all variations and extensions of the basic pump–probe technique described in Chapter 9. Finally, in Part IV, Chapter 16 includes a short procedure for aligning optical beams, which is critical for all of the experiments discussed in this book.

Optics has long been one of the first techniques applied to study new systems exhibiting novel properties, and we expect that the use of optical techniques to characterize materials will only grow in importance in the coming years. To this end, our goal was to put together a comprehensive reference for anyone using optics to experimentally characterize both simple and complex materials. We wish to empower a wide range of researchers, from experts in experimental optics who want to implement a new technique in their laboratories to those outside of optics who want to understand the basics of a particular optical technique that has been applied to a material of interest, to use and implement any of the techniques described here. Therefore, we hope that this book contributes to the vast body of literature demonstrating that, contrary to the quotation at the beginning of this preface, in fact you *can* study the darkness by simply turning on the light!

Rohit P. Prasankumar

Antoinette J. Taylor

ACKNOWLEDGMENTS

Any book of this nature, by definition, requires the collaboration of many people, all of whom have extremely busy schedules. Therefore, we deeply appreciate their efforts in contributing to this project amid their other deadlines and commitments, and apologize in advance to anyone that we unintentionally omit here. This book would not have been possible without the dedicated effort of the chapter authors, who have without exception been a pleasure to work with; their patience and perseverance during the long process of putting together this book has made it much more enjoyable. We would also like to thank Luna Han, our editor at Taylor & Francis, who helped formulate the initial idea for this book and has been extremely encouraging, responsive, and considerate throughout the whole process. We also appreciate the efforts of Kari Budyk and others at Taylor & Francis in the production stages of this project.

We have been very fortunate to work with several postdoctoral researchers who have been understanding and supportive of our efforts to put together this book, despite the fact that it limited the time we could spend advising their research. We would also like to thank our colleagues around the world and at Los Alamos National Laboratory, particularly within the Center for Integrated Nanotechnologies and Laboratory for Ultrafast Materials and Optical Science, for stimulating discussions and collaborations that helped us pursue exciting new scientific directions while also guiding the content of this book.

Finally, we would most of all like to thank our families. They have uncomplainingly put up with the amount of time that this project required, particularly in the countless nights and weekends spent in front of a computer screen instead of with them. This book would not have been possible without their love, support, and understanding.

Rohit P. Prasankumar

Antoinette J. Taylor

EDITORS

Rohit P. Prasankumar received his BS in electrical engineering from the University of Texas at Austin (1997) and his MS (1999) and PhD (2003) in electrical engineering from the Massachusetts Institute of Technology. He joined Los Alamos National Laboratory as a postdoctoral research associate in August 2003, focusing on the measurement of ultrafast mid-to-far-infrared dynamics in semiconductor nanostructures and strongly correlated compounds. He then became a technical staff member at the Center for Integrated Nanotechnologies at Los Alamos National Laboratory in February 2006, with research interests principally directed toward the measurement of dynamics in complex functional materials with high temporal and spatial resolution over a broad spectral range. He has coauthored over 30 publications, including 2 invited review papers.

Antoinette (Toni) J. Taylor is the leader of the Materials Physics and Applications Division at Los Alamos National Laboratory (LANL). Prior to this position, she was director of the Center for Integrated Nanotechnologies, a joint Sandia/LANL Nanoscience Research Center funded through the Department of Energy, Office of Basic Energy Sciences. Her research interests include the investigation of ultrafast dynamical nanoscale processes in materials and the development of novel optics-based measurement techniques for the understanding of new phenomena. She has published over 250 papers in these areas, written 2 book chapters, and edited 3 books. She has been a director-at-large of the Optical Society of America, a topical editor of the *Journal of the Optical Society B: Optical Physics*, and a member of the Solid State Science Committee, Board of Physics and Astronomy, and the National Academies; she has also chaired the National Academies' Committee on Nanophotonics Applicability and Accessibility. Currently, she is a member-at-large of the Division of Laser Science of the American Physical Society and the Optical Society of America's representative to the Joint Council of Quantum Electronics. She is an LANL fellow and a fellow of the American Physical Society, the Optical Society of America, and the American Association for the Advancement of Science. In 2003, Toni won the inaugural Los Alamos Fellow's Prize for Outstanding Leadership in Science and Engineering.

CONTRIBUTORS

Peter Abbamonte
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Marc Achermann
Department of Physics
University of Massachusetts
Amherst, Massachusetts

Richard D. Averitt
Department of Physics
Boston University
Boston, Massachusetts

Harini Barath
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Steven R.J. Brueck
Center for High Technology Materials
University of New Mexico
Albuquerque, New Mexico

Diego Casa
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Xiaoqian Chen
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Cesar E. Chialvo
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

S. Lance Cooper
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Steven T. Cundiff
JILA
University of Colorado

and

National Institute for Standards and Technology
Boulder, Colorado

Marcelo Davanço
Center for Nanoscale Science and Technology
National Institute of Standards and Technology
Gaithersburg, Maryland

and

Maryland Nanocenter
University of Maryland
College Park, Maryland

Jeffrey Davis
Centre for Atom Optics and Ultrafast Spectroscopy
Swinburne University of Technology
Victoria, Australia

Thomas Dekorsy
Department of Physics
University of Konstanz
Konstanz, Germany

Jure Demsar
Department of Physics
University of Konstanz
Konstanz, Germany

Daniele Fausti
Department of Physics
University of Trieste
Trieste, Italy

Y. Gan
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Jordan Gerton
Department of Physics and Astronomy
University of Utah
Salt Lake City, Utah

David J. Hilton
Department of Physics
University of Alabama–Birmingham
Birmingham, Alabama

Chennupati Jagadish
Department of Electronic Materials Engineering
Australian National University
Canberra, Australia

Young-Il Joe
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Jessica Johnston
Department of Physics and Astronomy
University of Utah
Salt Lake City, Utah

Robert A. Kaindl
Materials Sciences Division
Lawrence Berkeley National Laboratory
Berkeley, California

John F. Karpus
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Minjung Kim
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Andrew Kowalevicz
Technology Solutions
BAE Systems
Columbia, Maryland

Yuliya Kuznetsova
Center for High Technology Materials
University of New Mexico
Albuquerque, New Mexico

Paul H.M. van Loosdrecht
Zernike Institute for Advanced Materials
University of Groningen
Groningen, the Netherlands

Ben Mangum
Department of Physics and Astronomy
University of Utah
Salt Lake City, Utah

Nadya Mason
Department of Physics
University of Illinois at Urbana-Champaign
Urbana, Illinois

Alexander Neumann
Center for High Technology Materials
University of New Mexico
Albuquerque, New Mexico

Hidekazu Okamura
Department of Physics
Kobe University
Kobe, Japan

Willie J. Padilla
Department of Physics
Boston College
Chestnut Hill, Massachusetts

Rohit P. Prasankumar
Center for Integrated Nanotechnologies
Los Alamos National Laboratory
Los Alamos, New Mexico

Matthew T. Rakher
Center for Nanoscale Science and Technology
National Institute of Standards and Technology
Gaithersburg, Maryland

James P. Reed
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Sajan Saini
Department of Physics
Queens College of The City University of New York
Flushing, New York

Eyal Shafran
Department of Physics and Astronomy
University of Utah
Salt Lake City, Utah

C.S. Snow
Department of Physics
University of Illinois at Urbana–Champaign
Urbana, Illinois

Kartik Srinivasan
Center for Nanoscale Science and Technology
National Institute of Standards and Technology
Gaithersburg, Maryland

Antoinette J. Taylor
Materials Physics and Applications
Los Alamos National Laboratory
Los Alamos, New Mexico

Jigang Wang
Ames Laboratory and Department of Physics and Astronomy
Iowa State University
Ames, Iowa

LIST OF ABBREVIATIONS

Time	
Femtosecond	fs
Picosecond	ps
Nanosecond	ns
Distance	
Nanometer	nm
Micrometer (micron)	µm
Millimeter	mm
Centimeter	cm
Meter	m
Energy	
Femtojoule	fJ
Picojoule	pJ
Nanojoule	nJ
Microjoule	µJ
Millijoule	mJ
Milli-electron volt	meV
Electron volt	eV
Frequency	
Kilohertz	kHz
Megahertz	MHz
Gigahertz	GHz
Terahertz	THz
Ultraviolet	UV
Visible	VIS
Infrared	IR
Alternating current	AC
Direct current	DC

Physical constants

Speed of light in a vacuum	$c = 2.99792458 \times 10^{8}\,\text{m/s}$
Permittivity of free space	$\varepsilon_0 = 8.8542 \times 10^{-12}\,\text{F/m}$
Permeability of free space	$\mu_0 = 1.2566 \times 10^{-6}\,\text{H/m}$
Electron charge	$e = 1.602 \times 10^{-19}\,\text{C}$
Electron mass	$m_e = 9.01 \times 10^{-31}\,\text{kg}$
Planck's constant	$h = 6.626 \times 10^{-34}\,\text{J s}$
Boltzmann constant	$k_B = 1.38066 \times 10^{-23}\,\text{J/K}$

PART

BACKGROUND

Chapter 1

LIGHT–MATTER INTERACTIONS

Willie J. Padilla

CONTENTS

This chapter overviews the interactions of electromagnetic waves with matter—the study of which has a long and illustrious history. Inquiries into the nature of light began thousands of years ago and seem, not surprisingly, to have been first during the Classical and Hellenistic periods in ancient Greece (Pomeroy 1999). For example, the Greek philosopher Pythagoras (570–495 BC) was the first to propose a fifth element (in addition to earth, water, air, and fire), which he titled "aether," to describe the properties of light. Democritus, Empedocles, Plato, Aristotle, and others also developed several theories of light. In Euclid's book *Catoptrics* (300 BC),

he described both the rectilinear propagation and reflection of light. Heron of Alexandria (AD 10–70), a Greek mathematician, described both of these phenomena by explaining that light traverses the shortest allowed path between two points.

We do not intend to overview the entire history of light–matter interactions here, but would like to simply point out that, although ever-increasing insight regarding light and its interaction with matter has taken place over the last several thousand years, the theory of electromagnetic waves was placed on firm theoretical ground during the nineteenth century with the development of Maxwell's equations. In fact, all classical descriptions of the interactions between light and matter can be described by Maxwell's equations and the constitutive relations. Thus, we begin this chapter by describing the relationship between what is known—Maxwell's equations—and what can be measured.

The most experimentally accessible optical quantities of measure are the frequency-dependent reflectance $R(\omega)$ and transmittance $T(\omega)$. On the other hand, quantities most directly related to the electronic and magnetic structure of materials are the electrical permittivity $\varepsilon(\omega)$ and the magnetic permeability $\mu(\omega)$. Maxwell's equations facilitate an understanding of the relationship between the measured quantities and the electromagnetic structure of materials, that is, $[R,T] \leftrightarrow [\varepsilon,\mu]$. Maxwell's equations are really a collection of theories put forward by several scientists, namely, Gauss, Ampere, and Lorentz. They were unified by Maxwell in 1865 (Maxwell 1865) and put into their modern form by Gibbs and Wilson (1929) and Heaviside (1892).

Before we begin, we pause to note that throughout this chapter we use the *Système International* (SI) units, also called the rationalized meter–kilogram–second (mks) system. There are, however, four other common systems of electromagnetic units in use, all connected to the centimeter–gram–second (cgs). These four are the Gaussian system, Heaviside–Lorentz system, cgs electrostatic system, and cgs electromagnetic system. Historically the cgs electrostatic (esu) and cgs electromagnetic (emu) systems were developed first—during the middle of the nineteenth century—and were the systems of choice for the development of Maxwell's equations. Toward the end of the nineteenth century, the Heaviside–Lorentz system and the Gaussian system were developed, followed shortly by the rationalized mks system.

There can be ambiguity and confusion when converting back and forth between different systems of units. These difficulties seem to arise due to the number of fundamental dimensions the systems are based on. During the development of the electromagnetic theory in the nineteenth century, the esu, emu, and Gaussian systems were naturally based on the three fundamental dimensions of *length*, *mass*, and *time*. The rationalized mks system of units, however, recognizes a fourth fundamental "electromagnetic" dimension of *charge*, and a process of "rationalization" eliminates factors of 4π from Maxwell's equations (Cohen 2001). The rationalized mks system of units is the most common one employed today for electromagnetism and is used in nearly every introductory text.

1.1 MAXWELL'S EQUATIONS

1.1.1 Microscopic and Macroscopic Forms

Maxwell's equations describe the properties of electromagnetic waves and their interactions with matter. In free space, they take the form

$$\bar{\nabla} \cdot \bar{\mathbf{E}} = \frac{1}{\varepsilon_0}\rho, \tag{1.1}$$

$$\bar{\nabla} \cdot \bar{\mathbf{B}} = 0, \tag{1.2}$$

$$\bar{\nabla} \times \bar{\mathbf{E}} = -\frac{\partial \bar{\mathbf{B}}}{\partial t}, \tag{1.3}$$

$$\bar{\nabla} \times \bar{\mathbf{B}} = \mu_0 \bar{\mathbf{J}} + \mu_0 \varepsilon_0 \frac{\partial \bar{\mathbf{E}}}{\partial t}, \tag{1.4}$$

where

$\bar{\mathbf{E}}$ is called the electric field strength
$\bar{\mathbf{B}}$ is the magnetic flux density
ρ is the charge density
$\bar{\mathbf{J}}$ is the current density
ε_0 and μ_0 are the permittivity and permeability of free space

The permeability is defined to be exactly $\mu_0 = 4\pi \times 10^{-7}$ H/m, and thus the permittivity has a value of $\varepsilon_0 = 10^7/4\pi c^2$ F/m, that is, the uncertainty in ε_0 is the uncertainty in the velocity of light. Equation 1.1 is known as Gauss's law for electric fields, (1.2) as Gauss's law for magnetic fields, (1.3) as Faraday's induction law, and (1.4) as Ampere's law. Maxwell modified Ampere's law by adding the $\mu_0\varepsilon_0\, \partial\vec{\mathbf{E}}/\partial t$ term, and thus the particular forms of (1.1) through (1.4) are known as Maxwell's microscopic equations.

Matter may be polarized in the presence of electric and/or magnetic fields. If we assume a linear response and isotropic media, the electric and magnetic polarizations are taken to be proportional to the external electric and magnetic fields. Additionally, an explicit factor of ε_0 and/or μ_0 appears. The polarization fields plus the incident fields together form the total fields, called the electric displacement $\bar{\mathbf{D}}$ and the magnetic field strength $\bar{\mathbf{H}}$. These relations are derived explicitly below. The electric displacement is defined as

$$\bar{\mathbf{D}} \equiv \varepsilon_0 \bar{\mathbf{E}} + \bar{\mathbf{P}}, \tag{1.5}$$

where

$$\bar{\mathbf{P}} \equiv \varepsilon_0 \chi_e \bar{\mathbf{E}}, \tag{1.6}$$

with $\bar{\mathbf{P}}$ the polarization and χ_e the electric susceptibility. The electric displacement can thus be written as

$$\bar{\mathbf{D}} = \varepsilon \bar{\mathbf{E}}, \tag{1.7}$$

where

$$\varepsilon \equiv \varepsilon_0 (1 + \chi_e) = \varepsilon_0 \varepsilon_r. \tag{1.8}$$

The magnetic field strength is defined as

$$\bar{\mathbf{H}} \equiv \frac{1}{\mu_0} \bar{\mathbf{B}} - \bar{\mathbf{M}}, \tag{1.9}$$

where

$$\bar{\mathbf{M}} \equiv \chi_m \bar{\mathbf{H}}, \tag{1.10}$$

$\bar{\mathbf{M}}$ is the magnetization
χ_m is the magnetic susceptibility

The magnetic flux density can thus be written as

$$\bar{\mathbf{B}} = \mu \bar{\mathbf{H}}, \tag{1.11}$$

where

$$\mu \equiv \mu_0(1 + \chi_m) = \mu_0 \mu_r. \tag{1.12}$$

Using the definitions of $\bar{\mathbf{D}}$ and $\bar{\mathbf{B}}$ above, Maxwell's equations are rewritten as shown below.

$$\bar{\nabla} \cdot \bar{\mathbf{D}} = \rho. \tag{1.13}$$

$$\bar{\nabla} \cdot \bar{\mathbf{B}} = 0. \tag{1.14}$$

$$\bar{\nabla} \times \bar{\mathbf{E}} = -\frac{\partial \bar{\mathbf{B}}}{\partial t}. \tag{1.15}$$

$$\bar{\nabla} \times \bar{\mathbf{H}} = \bar{\mathbf{J}} + \frac{\partial \bar{\mathbf{D}}}{\partial t}. \tag{1.16}$$

Equations 1.13 through 1.16 are termed Maxwell's macroscopic equations, in contrast to (1.1) through (1.4), the microscopic form.

The microscopic Maxwell's equations completely describe the properties of electromagnetic fields; that is, for specified sources ρ and $\bar{\mathbf{J}}$ Maxwell's microscopic equations give the fields $\bar{\mathbf{E}}$ and $\bar{\mathbf{B}}$ everywhere in space. If this is the case, then why do we need the form of Maxwell's equations presented in (1.13) through (1.16)? We can answer this by considering electromagnetic wave propagation in matter. In this case, the number of sources—electrons—may be on the order of 10^{23}. Solving for the electric and magnetic fields in matter is thus of such complexity that is clearly not feasible. However, by introduction of the optical constants, via Maxwell's macroscopic equations, we may describe the average of fields and sources over scales much larger than individual atomic sizes (Jackson 1998). Thus, the optical constants facilitate the description of interactions between electromagnetic waves and matter.

Before proceeding, we pause to specify one other form of Maxwell's equations that is particularly useful for studying harmonic phenomena. We consider regions of space in which there are no charges or currents ($\rho = \bar{\mathbf{J}} = 0$), and that all fields in (1.13) through (1.16) vary as $\exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$, where $\bar{\mathbf{k}}$ is called the wave vector or the propagation vector, ω is the angular frequency, $\bar{\mathbf{r}}$ is the position vector, and t is time. In this case we find that Maxwell's equations can be written as

$$\bar{\mathbf{k}} \cdot \varepsilon \bar{\mathbf{E}} = 0. \tag{1.17}$$

$$\bar{\mathbf{k}} \cdot \mu \bar{\mathbf{H}} = 0. \tag{1.18}$$

$$\bar{\mathbf{k}} \times \bar{\mathbf{E}} = \omega\mu\bar{\mathbf{H}}. \tag{1.19}$$

$$\bar{\mathbf{k}} \times \bar{\mathbf{H}} = -\omega\varepsilon\bar{\mathbf{E}}. \tag{1.20}$$

Thus $\bar{\mathbf{E}}$ and $\bar{\mathbf{H}}$ are mutually perpendicular and perpendicular to $\bar{\mathbf{k}}$.

1.1.2 Boundary Conditions

Maxwell's equations may be used to derive the boundary conditions—that is, the relationship between the fields at an interface between two different types of media described by their respective optical constants. We will not do this explicitly here, and instead simply list the results. The interested reader is referred to Jackson (1998), who has derived these:

$$\bar{\mathbf{n}} \cdot \left(\bar{\mathbf{D}}^{(2)} - \bar{\mathbf{D}}^{(1)}\right) = 0, \tag{1.21}$$

$$\bar{\mathbf{n}} \cdot \left(\bar{\mathbf{B}}^{(2)} - \bar{\mathbf{B}}^{(1)}\right) = 0, \tag{1.22}$$

$$\bar{\mathbf{n}} \times \left(\bar{\mathbf{E}}^{(2)} - \bar{\mathbf{E}}^{(1)}\right) = 0, \tag{1.23}$$

$$\bar{\mathbf{n}} \times \left(\bar{\mathbf{H}}^{(2)} - \bar{\mathbf{H}}^{(1)}\right) = 0, \tag{1.24}$$

where $\bar{\mathbf{n}}$ is a unit vector perpendicular to the interface and the superscripts (1) and (2) refer to the two different media which form the interface. The specific form of the boundary conditions shown above is valid in regions of space in which there are no charges or currents ($\rho = \mathbf{J} = 0$). Furthermore, (1.21) through (1.24) allow us to explicitly state the relationships between tangential and normal components of the field at the interface. For example, (1.21) and (1.22) allow us to specify the relationship between the normal components of $\bar{\mathbf{D}}$ and $\bar{\mathbf{B}}$ as

$$\bar{\mathbf{D}}_{\mathrm{n}}^{(1)} = \bar{\mathbf{D}}_{\mathrm{n}}^{(2)} \quad \text{and} \quad \bar{\mathbf{B}}_{\mathrm{n}}^{(1)} = \bar{\mathbf{B}}_{\mathrm{n}}^{(2)} \tag{1.25}$$

and (1.23) and (1.24) allow us to specify the relationship between the tangential components of $\bar{\mathbf{E}}$ and $\bar{\mathbf{H}}$ as

$$\bar{\mathbf{E}}_{\mathrm{t}}^{(1)} = \bar{\mathbf{E}}_{\mathrm{t}}^{(2)} \quad \text{and} \quad \bar{\mathbf{H}}_{\mathrm{t}}^{(1)} = \bar{\mathbf{H}}_{\mathrm{t}}^{(2)}. \tag{1.26}$$

where the subscripted n's in (1.25) stand for normal components and the subscripted t's in (1.26) stand for the tangential components.

Thus we may summarize the boundary conditions by stating that the normal components of the electric displacement and the magnetic flux density are continuous at the interface, and that the tangential components of the electric field strength and the magnetic field strength are continuous at the interface.

1.1.3 Optical Constants

The relation between the microscopic and macroscopic forms of Maxwell's equations was first investigated by Lorentz (Choy 1999, Sihvola 1999) and is a subject detailed in various condensed matter texts (Ashcroft and Mermin 1976, Kittel 1996). However, an earlier connection between the atomic polarizability and the electric permittivity was developed by Mossotti in 1850, and independently by Clausius in 1879. This relationship, known as the Clausius–Mossotti equation, established that for any given material, the dielectric constant ε should be proportional to the density of atoms

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{1}{3\varepsilon_0} \sum N_j \alpha_j, \tag{1.27}$$

where

N_j is the concentration
α_j is the atomic polarizability

Equation 1.27 thus provides an important connection between macroscopic and microscopic theories. The atomic polarizability would be calculated by a microscopic theory, which could then be connected to a dielectric function, which would describe the response of the ions to the local field acting on them.

The dielectric constant (also called the electric permittivity) ε and the magnetic permeability μ are called the optical constants. Although in some cases it may be suitable to treat them as constant, they are, however, functions of frequency and, in general, may also depend on the wave vector, that is, $\varepsilon = \varepsilon(\omega, \mathbf{k})$, and $\mu = \mu(\omega, \mathbf{k})$. Here we denote angular frequency (cycles per unit time) as ω, which has units of rad/s, and the wave vector (wavelengths in a length of 2π) as **k**, which has units of 1/m. The relationship between the angular frequency and frequency is $\omega = 2\pi f$. It is typical to refer to ω as "frequency," as it is encountered more often than *f*.

As they have been derived here the optical constants are linear response functions that obey causality. Thus we may write them as complex functions $\tilde{\varepsilon} = \varepsilon_1 + i\varepsilon_2$ and $\tilde{\mu} = \mu_1 + i\mu_2$, where the subscripts 1 and 2 denote the real and imaginary parts, respectively. The real part of the complex optical constants describes the amplitude "response," and imaginary portions describe the damping or loss, due to external electric or magnetic fields.

Notice that ε and μ are the only two parameters in Maxwell's macroscopic equations which describe interactions of electromagnetic waves and matter. Other quantities that describe electromagnetic–matter interactions, called material parameters, have simple algebraic relations to the optical constants. If we assume a time-harmonic convention $e^{-i\omega t}$ (which we use throughout), the wave impedance (*Z*), conductivity (σ), and refractive index (*n*) are related to ε and μ by

$$\tilde{n} = n_1 + in_2 = \sqrt{\tilde{\varepsilon}_r \tilde{\mu}_r}. \tag{1.28}$$

$$\tilde{Z} = Z_1 + iZ_2 = \sqrt{\frac{\tilde{\mu}}{\tilde{\varepsilon}}}. \tag{1.29}$$

$$\tilde{\sigma} = \sigma_1 + i\sigma_2 = i\omega(\varepsilon_0 - \tilde{\varepsilon}). \tag{1.30}$$

The impedance (Z) of a medium to an electromagnetic wave is closely related to energy propagation. By analogy with a one-dimensional transmission line, the key parameter was shown to be the ratio of the total electric field to the total magnetic field (Schelkunoff 1938, Stratton 1941, Ramo et al. 1994). Thus, if we consider a wave propagating along the z-direction (with its electric field in the x-direction and the magnetic field in the y-direction), the load presented to the wave at any point z is given by

$$\tilde{Z}(z) = \frac{\tilde{E}_x(z)}{\tilde{H}_y(z)}, \tag{1.31}$$

which has units of resistance (Ohms). For isotropic media (1.29) and (1.31) are equal.

The ability of a material to carry an electric current is given by the real conductivity (σ_1), which has units of *siemens per meter* (S/m). Equivalently, materials are often characterized by their resistivity (ρ), which is the inverse of the conductivity $\rho = 1/\sigma_1$. The units of resistivity are ohm-meters, and thus 1 ohm = $(1\ \mathrm{S})^{-1}$.

When light enters a medium it may appear to be bent. The material parameter which describes this effect is the real part of the index of refraction (n_1). In addition to (1.28), which describes the connection between the optical constants and n_1, the index of refraction is the ratio between the speed of light c in vacuum and the speed of light v in that medium $n_1 = c/v$. We develop this idea later in this chapter.

Although (1.28) through (1.30) are a compact way of presenting the relationships between the optical constants and the material parameters, they offer little physical insight. Thus we expand each of these terms and list their explicit dependence in both real and imaginary parts in Table 1.1.

From Table 1.1 we can observe that the real (imaginary) part of the dielectric function is connected to the imaginary (real) part of the conductivity. It was previously mentioned that

TABLE 1.1
Relationships between the Optical Constants and the Material Parameters

	Dielectric Constant, $\tilde{\varepsilon}$	**Electrical Conductivity, $\tilde{\sigma}$**	**Index of Refraction and Wave Impedance, $\tilde{n}$, $\tilde{Z}$**
$\tilde{\varepsilon}$	$\tilde{\varepsilon} = \varepsilon_1 + i\varepsilon_2$	$\varepsilon_1 = \varepsilon_0 - \frac{1}{\omega}\sigma_2$	$\varepsilon_1 = c\frac{n_1 Z_1 + n_2 Z_2}{Z_2^1 + Z_2^2}$
		$\varepsilon_2 = \frac{1}{\omega}\sigma_1$	$\varepsilon_2 = c\frac{n_2 Z_1 - n_1 Z_2}{Z_1^2 + Z_2^2}$
$\tilde{\sigma}$	$\sigma_1 = \omega\varepsilon_2$	$\tilde{\sigma} = \sigma_1 + i\sigma_2$	$\sigma_1 = \omega c\frac{n_2 Z_1 - n_1 Z_2}{Z_1^2 + Z_2^2}$
	$\sigma_2 = \omega(\varepsilon_0 - \varepsilon_1)$		$\sigma_2 = \omega\left[\varepsilon_0 - c\frac{n_1 Z_1 + n_2 Z_2}{Z_1^2 + Z_2^2}\right]$
$\tilde{n}, \tilde{Z}$	$n_1 = \varepsilon_1 Z_1 - \varepsilon_2 Z_2$	$n_1 = c\left[Z_1\left(\varepsilon_0 - \frac{1}{\omega}\sigma_2\right) - Z_2\left(\frac{1}{\omega}\sigma_1\right)\right]$	$\tilde{n} = n_1 + in_2$
	$n_2 = \varepsilon_1 Z_2 + \varepsilon_2 Z_1$	$n_2 = c\left[Z_1\left(\frac{1}{\omega}\sigma_1\right) + Z_2\left(\varepsilon_0 - \frac{1}{\omega}\sigma_2\right)\right]$	$\tilde{Z} = Z_1 + iZ_2$

the imaginary portions of the optical constants characterize the loss of electromagnetic fields in a material. Thus we see that the real part of the electrical conductivity σ_1, which governs how well a material conducts electricity, is directly proportional to the quantity ε_2, which characterizes the loss of the electric component of an electromagnetic wave in a material. This connection will be made clear when we discuss the Drude model later in this chapter and in Chapter 3.

Although Table 1.1 is useful for converting between the conductivity and the dielectric function, the equations describing the relations between the index of refraction and the impedance to other material parameters and optical constants are less clear. This complexity arises due to the inclusion of the magnetic permeability. If we are in a regime of the electromagnetic spectrum where the magnetic response is weak (i.e., $\mu = \mu_0$), we are considering a material that has only an electric response, or we are at sufficiently high frequencies that magnetic response may be ignored (Landau et al. 1984)—these relations may be further simplified. Equations used to convert from the index of refraction to the dielectric function and back, with $\mu = \mu_0$, are

$$\begin{aligned} \varepsilon_1 &= \varepsilon_0(n_1^2 - n_2^2), \\ \varepsilon_2 &= 2\varepsilon_0 n_1 n_2, \end{aligned} \tag{1.32}$$

$$\begin{aligned} n_1 &= \left\{ \frac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2\varepsilon_0} \right\}^{1/2}, \\ n_2 &= \left\{ \frac{-\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2}}{2\varepsilon_0} \right\}^{1/2}, \end{aligned} \tag{1.33}$$

where, again, the subscripts 1 and 2 denote the real and imaginary portions, respectively, of the complex optical constants and material parameters.

1.1.4 Wave Equation

In the late nineteenth century, scientists believed that light needed a medium to be propagated in—the so-called luminiferous aether. In 1887, Michelson and Morley showed that the velocity of light was a constant c, which was the first strong evidence against the aether hypothesis. We now know that light is an electromagnetic wave—the speed of light emerging from Maxwell's equations is a crowning achievement of Maxwell's electromagnetic theory. We next derive the equations that govern the propagation of electromagnetic waves from Maxwell's equations.

If we assume there are no charges or currents ($\rho = \bar{\mathbf{J}} = 0$), then using the vector identity

$$\bar{\nabla} \times \left(\bar{\nabla} \times \bar{\mathbf{E}} \right) = \bar{\nabla} \left(\bar{\nabla} \cdot \bar{\mathbf{E}} \right) - \nabla^2 \bar{\mathbf{E}}, \tag{1.34}$$

we take the curl of (1.15) and, plugging in (1.13) and (1.16), we obtain

$$\nabla^2 \bar{\mathbf{E}} = \mu\varepsilon \frac{\partial^2 \bar{\mathbf{E}}}{\partial t}. \tag{1.35}$$

We may perform a similar operation for (1.16) and plug in (1.14) and (1.15) to obtain

$$\nabla^2\bar{\mathbf{B}} = \mu\varepsilon\frac{\partial^2\bar{\mathbf{B}}}{\partial t^2}. \tag{1.36}$$

Equations 1.35 and 1.36 are second-order partial differential equations, which describe the propagation of electromagnetic waves through a medium with a velocity $v = (\varepsilon\mu)^{-1/2}$, and are termed the wave equations. They have the general solution

$$\begin{Bmatrix}\bar{\mathbf{E}}\\ \bar{\mathbf{B}}\end{Bmatrix} = \begin{Bmatrix}\bar{\mathbf{E}}_0\\ \bar{\mathbf{B}}_0\end{Bmatrix}\exp i(\bar{\mathbf{k}}\cdot\bar{\mathbf{r}} - \omega t), \tag{1.37}$$

where

$\bar{\mathbf{k}}$ is called the wave vector or the propagation vector
ω is the frequency
$\bar{\mathbf{r}}$ is the position vector
$\bar{\mathbf{E}}_0$, $\bar{\mathbf{B}}_0$ are the amplitudes

In Figure 1.1 we plot an electromagnetic wave.

Upon substituting the general solution (1.37) into the wave equation (1.35) or (1.36) we find

$$k^2 = \omega^2\mu\varepsilon, \tag{1.38}$$

where k is the magnitude of the wave vector. Equation 1.38 is known as a dispersion relation and it describes the connection between the spatial dispersion and frequency dispersion of an electromagnetic wave, and its relation to the ratio of the speed of the electromagnetic wave to its speed in vacuum. We further develop the ideas of electromagnetic wave velocity next.

1.1.5 Group and Phase Velocities

The general solution (1.37) to Maxwell's equations describes a plane wave. For a particular wave vector $\bar{\mathbf{k}}$, a phase front is determined by $\bar{\mathbf{k}}\cdot\bar{\mathbf{r}}$ = constant. The phase front is a plane and is perpendicular to the wave vector. The phase velocity of an electromagnetic wave is defined to be

$$v_p = \frac{\omega}{k}. \tag{1.39}$$

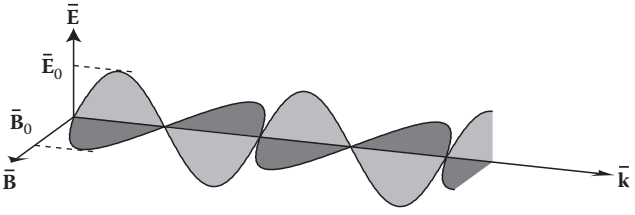


Figure 1.1 Transverse electromagnetic (TEM) wave. The wave is said to be "polarized" in the direction determined by the vector **E**.

This terminology originates from the sinusoidal dependence of waves, that is, the argument of a sinusoid is called its phase, so the velocity for which the phase is constant is called the phase velocity (Ramo et al. 1994). The group velocity is defined to be

$$v_g = \frac{d\omega}{dk}. \tag{1.40}$$

The definition of the group velocity in (1.40) is for a particular direction $\bar{\mathbf{k}}$, which we have not specified. However, (1.40) holds in Cartesian coordinates and thus we may express it in vector form as

$$\bar{\mathbf{v}}_g = \bar{\nabla}_k \omega(k). \tag{1.41}$$

We can put the group velocity (1.40) in a more useful form by writing our dispersion relation (1.38) as

$$k = \frac{n_1}{c}\omega = \frac{n_1}{\lambda}2\pi \tag{1.42}$$

where

$c = 1/\sqrt{\varepsilon_0\mu_0}$ is the speed of light in vacuum
n_1 is the real part of the index of refraction, defined in (1.28)
λ is the wavelength

If $n_1 = n_1(\omega)$, then we can write (1.42) as

$$dk = \frac{1}{c}d[n_1(\omega)\omega] = \frac{1}{c}[n_1(\omega)d\omega + \omega dn_1(\omega)]. \tag{1.43}$$

Dividing $d\omega$ by (1.43), we can write the group velocity as

$$v_g = \frac{c}{n_1(\omega) + \omega(dn_1(\omega)/d\omega)}. \tag{1.44}$$

The group velocity is the velocity with which a signal composed of a narrow band of frequency components propagates, that is, it is the transmission velocity of a wave packet (Collin 1991, Kittel 1996). For wave propagation in frequency-dispersive media, one may define a number of other velocities, for example, signal velocity and energy velocity, as well as the Sommerfeld precursor and the Brillouin first and second precursors (Sommerfeld 1950, Brillouin 1960). We will not elaborate on these other velocities and phenomena here, but refer to several texts where this is detailed (Brillouin 1960, Collin 1991). We do note, however, that the velocity of energy transfer is determined by the Poynting vector, which we discuss next.

1.1.6 Poynting Theorem

Conservation of energy is a law of physics which states that *for any closed (isolated) system, the energy is conserved over time*. Of course the implication of this is that energy cannot be created or destroyed, but instead can only be transferred back and forth between different forms.

In electricity and magnetism conservation of energy can be found from Maxwell's equations by dotting $\bar{\mathbf{H}}$ into (1.15) and subtracting $\bar{\mathbf{E}}$ dotted into (1.16). This yields

$$\bar{\mathbf{H}}\cdot\bar{\nabla}\times\bar{\mathbf{E}}-\bar{\mathbf{E}}\cdot\bar{\nabla}\times\bar{\mathbf{H}}=-\left(\bar{\mathbf{E}}\cdot\bar{\mathbf{J}}+\bar{\mathbf{H}}\cdot\frac{\partial\mathbf{B}}{\partial t}+\bar{\mathbf{E}}\cdot\frac{\partial\mathbf{D}}{\partial t}\right). \tag{1.45}$$

Using the identity $\bar{\nabla}\cdot\left(\bar{\mathbf{E}}\times\bar{\mathbf{H}}\right)=\bar{\mathbf{H}}\cdot\bar{\nabla}\times\bar{\mathbf{E}}-\bar{\mathbf{E}}\cdot\bar{\nabla}\times\bar{\mathbf{H}}$, (1.45) can be written as

$$\bar{\nabla}\cdot\left(\bar{\mathbf{E}}\times\bar{\mathbf{H}}\right)=-\left(\bar{\mathbf{E}}\cdot\bar{\mathbf{J}}+\bar{\mathbf{H}}\cdot\frac{\partial\bar{\mathbf{B}}}{\partial t}+\bar{\mathbf{E}}\cdot\frac{\partial\bar{\mathbf{D}}}{\partial t}\right). \tag{1.46}$$

Thus we observe that the conservation of energy in electricity and magnetism is determined by (1.46). One may also define the Poynting vector as

$$\bar{\mathbf{S}}=\bar{\mathbf{E}}\times\bar{\mathbf{H}}, \tag{1.47}$$

which has the units of (energy/area × time), and thus represents the power flux density. The quantity

$$\bar{\mathbf{H}}\cdot\frac{\partial\bar{\mathbf{B}}}{\partial t}+\bar{\mathbf{E}}\cdot\frac{\partial\bar{\mathbf{D}}}{\partial t}\equiv\frac{\partial u}{\partial t} \tag{1.48}$$

represents the time rate of change of the stored electric and magnetic energy ($u=u_{\mathrm{e}}+u_{\mathrm{m}}$) over a volume, and $\bar{\mathbf{E}}\cdot\bar{\mathbf{J}}$ is the power supplied by a current density $\bar{\mathbf{J}}$ over some volume. We may also define, over some volume, the electric energy $u_{\mathrm{e}}=\bar{\mathbf{E}}\cdot\bar{\mathbf{D}}$ and the magnetic energy $u_{\mathrm{m}}=\bar{\mathbf{B}}\cdot\bar{\mathbf{H}}$. Thus finally we may write (1.46) as

$$\bar{\nabla}\cdot\bar{\mathbf{S}}+\frac{\partial u}{\partial t}=-\bar{\mathbf{J}}\cdot\bar{\mathbf{E}}. \tag{1.49}$$

The meaning of (1.49) is that the energy transported by the fields per unit area per unit time, plus the time rate of change of the energy density, is equal to the negative of the total work done by the fields on the sources. It should be noted that (1.49) is a statement of the conservation of energy, but is valid only for linear responses (electric and magnetic) and is not valid for hysteretic materials.

Since the physical meaning of $\bar{\mathbf{S}}$ is that it is the energy flow, it must be real. In (1.47) we may specify that we will only consider the real parts of each vector $\bar{\mathbf{S}}$, $\bar{\mathbf{E}}$, $\bar{\mathbf{H}}$. Further, we are interested in waves that have a harmonic dependence $e^{-i\omega t}$. Thus we show, without proof, that if we define $\bar{\mathbf{S}}$ over one period, then we may write (Jackson 1998)

$$\bar{\mathbf{S}}=\frac{1}{2}\mathrm{Re}\left[\bar{\mathbf{E}}\times\bar{\mathbf{H}}^{*}\right], \tag{1.50}$$

where $\bar{\mathbf{H}}^{*}$ is the complex conjugate of the magnetic field strength. The usefulness of the Poynting vector shown in (1.50) will be made clear when we discuss the Fresnel equations later in this chapter.

The particular form of the time rate of change of the electric energy density in (1.48) permits us to consider the response of a Drude metal (discussed in more detail later and in Chapter 3), here an electron gas, to an additional electron moving within the solid. This moving electron produces a field $\bar{\mathbf{D}}$, which we assume is time varying as $\bar{\mathbf{D}} = \bar{\mathbf{D}}_0 \exp i(\bar{\mathbf{k}} \cdot \bar{\mathbf{r}} - \omega t)$, and thus the electronic energy density absorbed per unit volume is

$$\frac{\partial u_e}{\partial t} = \frac{\bar{\mathbf{D}}}{\tilde{\varepsilon}} \cdot \frac{\partial \bar{\mathbf{D}}}{\partial t} = -\frac{i\omega}{\tilde{\varepsilon}} |D_0|^2 . \tag{1.51}$$

If we consider the real and imaginary parts of (1.51), then upon equating real parts we find

$$\mathrm{Re}\left(\frac{\partial u_e}{\partial t}\right) = -\omega |D_0|^2 \;\; \mathrm{Re}\left(\frac{i}{\tilde{\varepsilon}}\right) = \omega |D_0|^2 \;\; \mathrm{Im}\left(\frac{1}{\tilde{\varepsilon}}\right). \tag{1.52}$$

Thus we find that the imaginary part of $1/\tilde{\varepsilon}$ describes the energy loss due to an electron moving through the medium. Specifically we associate the term

$$\mathrm{Im}\left(\frac{1}{\tilde{\varepsilon}}\right) = -\frac{\varepsilon_2}{\varepsilon_1^2 + \varepsilon_2^2} \tag{1.53}$$

with this loss, and thus (1.53) is called the loss function. Although we have left out its explicit frequency dependence, in general the loss function is a function of frequency. It can be observed that the loss function peaks where the real part of the dielectric function is near zero and where the imaginary part of the dielectric function (losses) is small. The loss function is important in optics, where it determines the center frequency of longitudinal optical phonons (and other longitudinal resonant phenomena) (Burns 1990), and in plasmonics where it determines the frequency of surface, geometrical, and bulk plasmons (Cottam and Tilley 1989, Kittel 1996).

1.2 CONSTITUTIVE RELATIONS

Maxwell's macroscopic equations, as shown in (1.13) through (1.16) are a set of four equations involving the components of the four fields $\bar{\mathbf{E}}$, $\bar{\mathbf{B}}$, $\bar{\mathbf{D}}$, $\bar{\mathbf{H}}$. As we have shown in a simply linear case, a medium may be "polarized" by incident $\bar{\mathbf{E}}$, $\bar{\mathbf{H}}$ fields, resulting in a total $\bar{\mathbf{D}}$, $\bar{\mathbf{B}}$ field. Equations that govern the relationships between the incident and total fields are called constitutive equations. It is useful to characterize materials by the specific form of the constitutive relations, and the symmetry they posses. In a very general way one may express the constitutive relations as

$$\bar{\mathbf{D}} = \bar{\mathbf{D}}\left[\bar{\mathbf{E}}, \bar{\mathbf{H}}\right], \tag{1.54}$$

$$\bar{\mathbf{B}} = \bar{\mathbf{B}}\left[\bar{\mathbf{E}}, \bar{\mathbf{H}}\right], \tag{1.55}$$

where the square brackets indicate that the relations between the given fields and the resultant fields are not necessarily simple, and may depend upon the entire past history of the sample (hysteresis), may be nonlinear, or may be bianisotropic, etc. (Jackson 1998).

1.2.1 Magneto-Optical Permittivities

For isotropic, anisotropic, or more complicated types of media, we may express the constitutive relations generally as (Cheng and Kong 1968a,b, Kong 1990)

$$\begin{pmatrix} \bar{\mathbf{D}} \\ \bar{\mathbf{B}} \end{pmatrix} = \begin{pmatrix} \bar{\bar{\varepsilon}} & \bar{\bar{\xi}} \\ \bar{\bar{\zeta}} & \bar{\bar{\mu}} \end{pmatrix} \cdot \begin{pmatrix} \bar{\mathbf{E}} \\ \bar{\mathbf{H}} \end{pmatrix}. \tag{1.56}$$

The quantities $\bar{\bar{\varepsilon}}$ and $\bar{\bar{\mu}}$, the dielectric permittivity and the magnetic permeability, have already been introduced. The double bar above each indicates that it is a tensor quantity, in this particular case of rank 2. Thus, for example, we may write the dielectric permittivity as

$$\bar{\bar{\varepsilon}} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{pmatrix}. \tag{1.57}$$

A similar form may be written for the other quantities $\left(\bar{\bar{\mu}}, \bar{\bar{\xi}}, \bar{\bar{\zeta}}\right)$ in (1.56). It should be stated that all of the quantities in (1.56) may depend on frequency (ω) and wave vector ($\boldsymbol{k}$) but, for simplification, we drop this explicit dependence. Typically in crystal optics, the magnetic permeability does not exhibit frequency dependence and its value may be assumed to be equal to that of free space. Under these conditions, one may further distinguish between different types of crystals by considering the symmetry of the dielectric tensor. For example, usually the off-diagonal terms in (1.57) are zero such that

$$\bar{\bar{\varepsilon}} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}. \tag{1.58}$$

In the case of (1.58), if we have the specific relationship $\varepsilon_{xx} = \varepsilon_{yy} \neq \varepsilon_{zz}$ between diagonal components of the permittivity, then the crystal is termed "uniaxial." For the specific case where the relationship is $\varepsilon_{xx} \neq \varepsilon_{yy} \neq \varepsilon_{zz}$, the crystal is called "biaxial." By considering the simple relationship between the dielectric permittivity and the index of refraction (1.28), notice that for a uniaxial crystal there exist two different refractive indices. These two different refractive indices are termed the "ordinary index" (n_o) if light is polarized in the $\hat{x}$ or $\hat{y}$ direction, and the "extraordinary index" (n_e) for light polarized in the $\hat{z}$ direction.

For the more complicated case where we cannot ignore the magnetic response of materials we must consider the full constitutive equations (1.56). Beyond having just a magnetic and electric response we also allow a magneto-electric response. In order to describe such responses it is useful to define the terms $\bar{\bar{\xi}}$ and $\bar{\bar{\zeta}}$, which are called the magneto-optical permittivities. They describe coupling of the magnetic-to-electric response $\bar{\bar{\xi}}$ and electric-to-magnetic response $\bar{\bar{\zeta}}$. Multiferroics are one example of a class of materials in which (1.56) facilitates a description of their electromagnetic response (Eerenstein et al. 2006). Just as in the case of the symmetry of the dielectric permittivity tensor, we may describe various media based on the symmetry of the magneto-optical permittivity tensors, shown in (1.56). This is summarized in Table 1.2.

TABLE 1.2
Description of Materials Based on the Symmetry of the Magneto-Optical Permittivity Tensor (1.56)

Material Type	$\bar{\bar{\varepsilon}}, \bar{\bar{\mu}}$	$\bar{\bar{\xi}}, \bar{\bar{\zeta}}$
Isotropic	$\propto \mathbf{I}$	$\bar{\bar{\xi}}, \bar{\bar{\zeta}} = 0$
Anisotropic	$\bar{\bar{\varepsilon}}$ and/or $\bar{\bar{\mu}}$ not $\propto \mathbf{I}$	$\bar{\bar{\xi}}, \bar{\bar{\zeta}} = 0$
Bi-isotropic	$\propto \mathbf{I}$	$\propto \mathbf{I}$
Bi-anisotropic	All other cases	All other cases

1.3 FRESNEL EQUATIONS

If an isotropic opaque material may be described by the macroscopic form of Maxwell's equations (1.13 through 1.16), we may calculate the reflectance of the intensity of a plane wave from the material with the Fresnel equations. We begin by deriving the appropriate equations for two different types of polarization—the so-called transverse electric (TE) and the transverse magnetic (TM) cases.

1.3.1 Reflection and Transmission at an Interface

In Figure 1.2 we show a plane wave incident on the interface between a medium with optical constants $\tilde{\varepsilon}^{(1)}$, $\tilde{\mu}^{(1)}$ and a medium (shaded gray region) described by the optical constants $\tilde{\varepsilon}^{(2)}$, $\tilde{\mu}^{(2)}$, where the wave vector forms an angle θ_i with the surface normal and the superscripted (1) and (2) refer to the first and second media, respectively. $\bar{\mathbf{E}}$ represents the electric field, $\bar{\mathbf{H}}$ the magnetic field, and $\bar{\mathbf{k}}$ the wave vector of the incident wave. The subscripted i, r, and t on each of the fields and wave vectors stand for the incident, reflected, and transmitted waves, respectively. The incoming and outgoing wave vectors form the "plane of incidence," and TE and TM denote the relation of the electric field vector and magnetic field vector with respect to the plane of incidence. TE means that the electric field vector is perpendicular to the plane

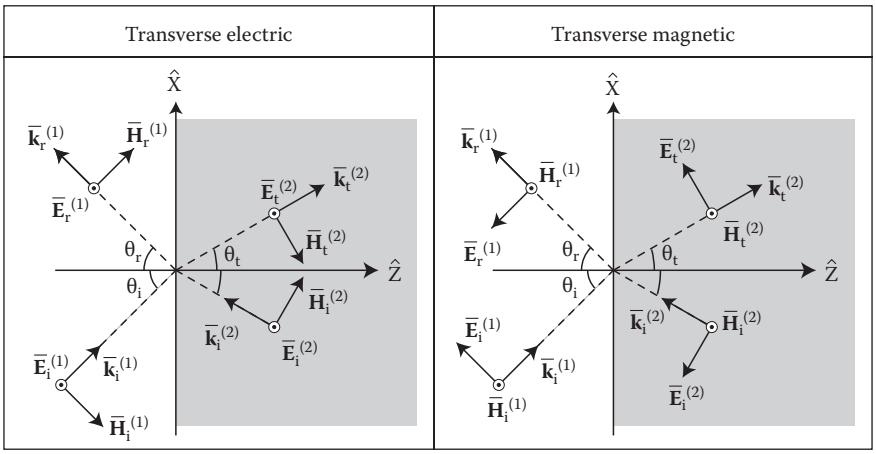


Figure 1.2 Schematic depicting the transmission and reflection of an electromagnetic wave at an interface. The left and right panels show the TE and TM polarization cases, respectively. In both cases, waves are incident from both sides of the interface.

of incidence (with the magnetic field vector parallel to the plane of incidence). TM is the opposite, with the magnetic field vector perpendicular and the electric field vector parallel to the plane of incidence.

We first discuss the TE case, as depicted on the left portion of Figure 1.2. The interface lies in the *xy*-plane and thus the electric field is polarized only in the *y*-direction. The magnetic field, however, has two components lying in the *x*- and *z*-directions. The boundary conditions (1.21) through (1.24) permit us to specify the relationship of the electric and magnetic fields across the interface. For example, (1.26) dictates that the parallel (tangential) components of the electric field must be the same on either side of the interface, that is,

$$E_{\mathrm{i}}^{(1)} + E_{\mathrm{r}}^{(1)} = E_{\mathrm{t}}^{(2)} + E_{\mathrm{i}}^{(2)}. \tag{1.59}$$

Likewise the same must hold for the magnetic field,

$$H_{\mathrm{ix}}^{(1)} + H_{\mathrm{rx}}^{(1)} = H_{\mathrm{tx}}^{(2)} + H_{\mathrm{ix}}^{(2)}, \tag{1.60}$$

where we have added a subscripted *x* to denote that we are only considering the components parallel to the interface. We can use Maxwell's equation (1.19) to write the magnetic field in terms of the electric field, that is,

$$-k_{\mathrm{iz}}^{(1)} E_{\mathrm{i}}^{(1)} = \omega\mu^{(1)} H_{\mathrm{ix}}^{(1)}; \quad k_{\mathrm{rz}}^{(1)} E_{\mathrm{r}}^{(1)} = \omega\mu^{(1)} H_{\mathrm{rx}}^{(1)}; \quad -k_{\mathrm{tz}}^{(2)} E_{\mathrm{t}}^{(2)} = \omega\mu^{(2)} H_{\mathrm{tx}}^{(2)}. \tag{1.61}$$

The first equation in (1.61) is for the incident components (i subscripts), the second for the reflected components (r subscripts), the third for the transmitted components (t subscripts), and there is a similar equation for the wave incident on the interface from medium (2). Plugging in (1.61) into our boundary condition for the magnetic field (1.60) we see

$$-\frac{k_z^{(1)}}{\mu^{(1)}} E_{\mathrm{i}}^{(1)} + \frac{k_z^{(1)}}{\mu^{(1)}} E_{\mathrm{r}}^{(1)} = -\frac{k_z^{(2)}}{\mu^{(2)}} E_{\mathrm{t}}^{(2)} + \frac{k_z^{(2)}}{\mu^{(2)}} E_{\mathrm{i}}^{(2)}. \tag{1.62}$$

Our boundary conditions are now all expressed in terms of the electric field. The system of Equations 1.59 and 1.62 may thus be expressed in matrix form as

$$\begin{pmatrix} 1 & 1 \\ -\dfrac{k_z^{(1)}}{\mu^{(1)}} & \dfrac{k_z^{(1)}}{\mu^{(1)}} \end{pmatrix} \begin{pmatrix} E_{\mathrm{i}}^{(1)} \\ E_{\mathrm{r}}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -\dfrac{k_z^{(2)}}{\mu^{(2)}} & \dfrac{k_z^{(2)}}{\mu^{(2)}} \end{pmatrix} \begin{pmatrix} E_{\mathrm{t}}^{(2)} \\ E_{\mathrm{i}}^{(2)} \end{pmatrix}. \tag{1.63}$$

We may express (1.63) in terms of the electric fields on one side of the interface in terms of the other by

$$\begin{pmatrix} E_{\mathrm{t}}^{(2)} \\ E_{\mathrm{i}}^{(2)} \end{pmatrix} = \mathbf{M}^{\mathrm{TE}} \begin{pmatrix} E_{\mathrm{i}}^{(1)} \\ E_{\mathrm{r}}^{(1)} \end{pmatrix}, \tag{1.64}$$

where $\mathbf{M}^{\mathrm{TE}}$ is called the transfer matrix and has the explicit form

$$\mathbf{M}^{\mathrm{TE}} = \frac{1}{2} \begin{pmatrix} 1 + \dfrac{\mu^{(2)}}{\mu^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} & 1 - \dfrac{\mu^{(2)}}{\mu^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} \\ 1 - \dfrac{\mu^{(2)}}{\mu^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} & 1 + \dfrac{\mu^{(2)}}{\mu^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} \end{pmatrix}. \tag{1.65}$$

Equation 1.65 is the general form for the transmission matrix in the TE polarization case when waves are incident from both sides of the interface. In the case that waves are only incident from the left side of the interface we have $E_{\mathrm{i}}^{(2)} = 0$, and we can write (1.64) as

$$\begin{aligned} E_{\mathrm{t}}^{(2)} &= M_{11}^{\mathrm{TE}} E_{\mathrm{i}}^{(1)} + M_{12}^{\mathrm{TE}} E_{\mathrm{r}}^{(1)} \\ E_{\mathrm{i}}^{(2)} &= M_{21}^{\mathrm{TE}} E_{\mathrm{i}}^{(1)} + M_{22}^{\mathrm{TE}} E_{\mathrm{r}}^{(1)} = 0. \end{aligned} \tag{1.66}$$

The transmission and reflection coefficients for the electric field are defined as

$$t_{\mathrm{TE}} = \frac{E_{\mathrm{t}}^{(2)}}{E_{\mathrm{i}}^{(1)}} \quad \text{and} \quad r_{\mathrm{TE}} = \frac{E_{\mathrm{r}}^{(1)}}{E_{\mathrm{i}}^{(1)}}. \tag{1.67}$$

Using (1.64) we may solve the system of equations (1.66) to give

$$t_{\mathrm{TE}} = \frac{\det \mathbf{M}^{\mathrm{TE}}}{M_{22}^{\mathrm{TE}}} = \frac{2\mu^{(2)} k_z^{(1)}}{\mu^{(1)} k_z^{(2)} + \mu^{(2)} k_z^{(1)}} \tag{1.68}$$

and

$$r_{\mathrm{TE}} = -\frac{M_{21}^{\mathrm{TE}}}{M_{22}^{\mathrm{TE}}} = \frac{\mu^{(2)} k_z^{(1)} - \mu^{(1)} k_z^{(2)}}{\mu^{(2)} k_z^{(1)} - \mu^{(1)} k_z^{(2)}}. \tag{1.69}$$

The intensity transmittance is given by the ratio of the energy flow in the two media that make up the interface. Using our equation for the description of energy flow for harmonic phenomena (1.50) we have

$$S^{(1)} = \frac{1}{2} \mathrm{Re}\left[E_{\mathrm{i}}^{(1)} H_{\mathrm{ix}}^{(1)*} \right] = \frac{1}{2\omega} \mathrm{Re}\left[\frac{k_z^{(1)}}{\mu^{(1)}} \right] \left| E_{\mathrm{i}}^{(1)} \right|^2 \tag{1.70}$$

$$S^{(2)} = \frac{1}{2} \mathrm{Re}\left[E_{\mathrm{t}}^{(2)} H_{\mathrm{tx}}^{(2)*} \right] = \frac{1}{2\omega} \mathrm{Re}\left[\frac{k_z^{(2)}}{\mu^{(2)}} \right] \left| E_{\mathrm{t}}^{(2)} \right|^2 \tag{1.71}$$

where $S^{(1)}$, $S^{(2)}$ are the energy flows in the first and second media, respectively. Since $E_{\mathrm{t}}^{(2)} = t_{\mathrm{TE}} E_{\mathrm{i}}^{(1)}$ we may finally write the transmittance (or transmissivity) and reflectance (or reflectivity) as

$$T_{\mathrm{TE}} = \frac{S^{(2)}}{S^{(1)}} = \frac{\mathrm{Re}\left[k_z^{(2)} / \mu^{(2)} \right]}{\mathrm{Re}\left[k_z^{(1)} / \mu^{(1)} \right]} \left| t_{\mathrm{TE}} \right|^2 \tag{1.72}$$

and

$$R_{\mathrm{TE}} = \left| r_{\mathrm{TE}} \right|^2. \tag{1.73}$$

Next we discuss the TM case, as depicted on the right portion of Figure 1.2. The interface lies in the *xy*-plane and thus the magnetic field is polarized only in the *y*-direction. The electric

field, however, has two components lying in the x- and z-directions. Using the same boundary conditions as for the TE case, (1.21) through (1.24), we have

$$H_{\mathrm{i}}^{(1)} + H_{\mathrm{r}}^{(1)} = H_{\mathrm{t}}^{(2)} + H_{\mathrm{i}}^{(2)}. \tag{1.74}$$

Likewise the same must hold for the electric field

$$E_{\mathrm{i}x}^{(1)} + E_{\mathrm{r}x}^{(1)} = E_{\mathrm{t}x}^{(2)} + E_{\mathrm{i}x}^{(2)}, \tag{1.75}$$

where we have added a subscripted x to denote that we are only considering the components parallel to the interface. We can use Maxwell's equation (1.20) to write the electric field in terms of the magnetic field, that is,

$$k_z^{(1)} H_{\mathrm{t}}^{(1)} = \omega\varepsilon^{(1)} E_{\mathrm{i}x}^{(1)}; \quad -k_z^{(1)} H_{\mathrm{r}}^{(1)} = \omega\varepsilon^{(1)} E_{\mathrm{r}x}^{(1)}; \quad -k_z^{(2)} H_{\mathrm{t}}^{(2)} = \omega\varepsilon^{(2)} E_{\mathrm{t}x}^{(2)}. \tag{1.76}$$

We obtain, in matrix form, the relationship

$$\begin{pmatrix} 1 & 1 \\ \dfrac{k_z^{(1)}}{\varepsilon^{(1)}} & -\dfrac{k_z^{(1)}}{\varepsilon^{(1)}} \end{pmatrix} \begin{pmatrix} H_{\mathrm{i}}^{(1)} \\ H_{\mathrm{r}}^{(1)} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -\dfrac{k_z^{(2)}}{\varepsilon^{(2)}} & \dfrac{k_z^{(2)}}{\varepsilon^{(2)}} \end{pmatrix} \begin{pmatrix} H_{\mathrm{t}}^{(2)} \\ H_{\mathrm{i}}^{(2)} \end{pmatrix}. \tag{1.77}$$

We may express (1.77) in terms of the magnetic fields on one side of the interface in terms of the other by

$$\begin{pmatrix} H_{\mathrm{t}}^{(2)} \\ H_{\mathrm{i}}^{(2)} \end{pmatrix} = \mathbf{M}^{\mathrm{TM}} \begin{pmatrix} H_{\mathrm{i}}^{(1)} \\ H_{\mathrm{r}}^{(1)} \end{pmatrix} \tag{1.78}$$

where $\mathbf{M}^{\mathrm{TM}}$ has the explicit form,

$$\mathbf{M}^{\mathrm{TM}} = \frac{1}{2} \begin{pmatrix} 1 + \dfrac{\varepsilon^{(2)}}{\varepsilon^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} & 1 - \dfrac{\varepsilon^{(2)}}{\varepsilon^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} \\ 1 - \dfrac{\varepsilon^{(2)}}{\varepsilon^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} & 1 + \dfrac{\varepsilon^{(2)}}{\varepsilon^{(1)}} \dfrac{k_z^{(1)}}{k_z^{(2)}} \end{pmatrix} \tag{1.79}$$

Equation 1.79 is the general form for the transmission matrix in the TM polarization case when waves are incident from both sides of the interface. In the case that waves are only incident from the left side of the interface we have $H_{\mathrm{i}}^{(2)} = 0$, and we can write the transmission and reflection coefficients for the magnetic field as

$$t_{\mathrm{TM}} = \frac{\det \mathbf{M}^{\mathrm{TM}}}{M_{22}^{\mathrm{TM}}} = \frac{2\varepsilon^{(2)} k_z^{(1)}}{\varepsilon^{(1)} k_z^{(2)} + \varepsilon^{(2)} k_z^{(1)}} \tag{1.80}$$

and

$$r_{\mathrm{TM}} = -\frac{M_{21}^{\mathrm{TM}}}{M_{22}^{\mathrm{TM}}} = \frac{\varepsilon^{(2)} k_z^{(1)} - \varepsilon^{(1)} k_z^{(2)}}{\varepsilon^{(2)} k_z^{(1)} - \varepsilon^{(1)} k_z^{(2)}}. \tag{1.81}$$

As with the TE case, the intensity transmittance is given by the ratio of the energy flow in the two media that make up the interface, and thus we have

$$S^{(1)} = \frac{1}{2}\mathrm{Re}\left[E_{ix}^{(1)}H_{\mathrm{i}}^{(1)*}\right] = \frac{1}{2\omega}\mathrm{Re}\left[\frac{k_z^{(1)}}{\varepsilon^{(1)}}\right]\left|H_{\mathrm{i}}^{(1)}\right|^2 \tag{1.82}$$

$$S^{(2)} = \frac{1}{2}\mathrm{Re}\left[E_{\mathrm{tx}}^{(2)}H_{\mathrm{t}}^{(2)*}\right] = \frac{1}{2\omega}\mathrm{Re}\left[\frac{k_z^{(2)}}{\varepsilon^{(2)}}\right]\left|H_{\mathrm{t}}^{(2)}\right|^2 \tag{1.83}$$

We may write the transmittance and reflectance for TM waves as

$$T_{\mathrm{TM}} = \frac{S^{(2)}}{S^{(1)}} = \frac{\mathrm{Re}\left[k_z^{(2)}/\varepsilon^{(2)}\right]}{\mathrm{Re}\left[k_z^{(1)}/\varepsilon^{(1)}\right]}\left|t_{\mathrm{TM}}\right|^2 \tag{1.84}$$

and

$$R_{\mathrm{TM}} = \left|r_{\mathrm{TM}}\right|^2 . \tag{1.85}$$

The reflectance for both TE, (1.72) through (1.73), and TM, (1.84) and (1.85), waves can easily be expressed in more common forms found in many textbooks. For example, the z-component of our incident and transmitted wave vectors may be expressed in terms of the incident and transmitted angles by

$$k_z^{(1)} = k^{(1)}\cos\theta_{\mathrm{i}} \quad \text{and} \quad k_z^{(2)} = k^{(2)}\cos\theta_{\mathrm{t}}. \tag{1.86}$$

If the interface on the left is free space, then $\varepsilon^{(1)} = \varepsilon_0$, $\mu^{(1)} = \mu_0$. Further we may use (1.42) to express the wave vector in terms of the index of refraction. With these substitutions and using Snell's law, $n^{(1)} \sin\theta_1 = n^{(2)} \sin\theta_2$, we find

$$R_{\mathrm{TE}} = \left|r_{\mathrm{TE}}\right|^2 = \left|\frac{\cos\theta - \mu_{\mathrm{r}}^{-1}\sqrt{n^2 - \sin^2\theta}}{\cos\theta + \mu_{\mathrm{r}}^{-1}\sqrt{n^2 - \sin^2\theta}}\right|^2 , \tag{1.87}$$

$$R_{\mathrm{TM}} = \left|r_{\mathrm{TM}}\right|^2 = \left|\frac{\varepsilon_{\mathrm{r}}\cos\theta - \sqrt{n^2 - \sin^2\theta}}{\varepsilon_{\mathrm{r}}\cos\theta + \sqrt{n^2 - \sin^2\theta}}\right|^2 \tag{1.88}$$

where

- θ is the angle of incidence
- ε_{r}, μ_{r}, and n are the relative permittivity, relative permeability, and index of refraction of the medium, respectively

Here we assume the combination of absorption and material thickness is sufficient such that the electromagnetic wave is attenuated before reaching the other side of the material, that is,

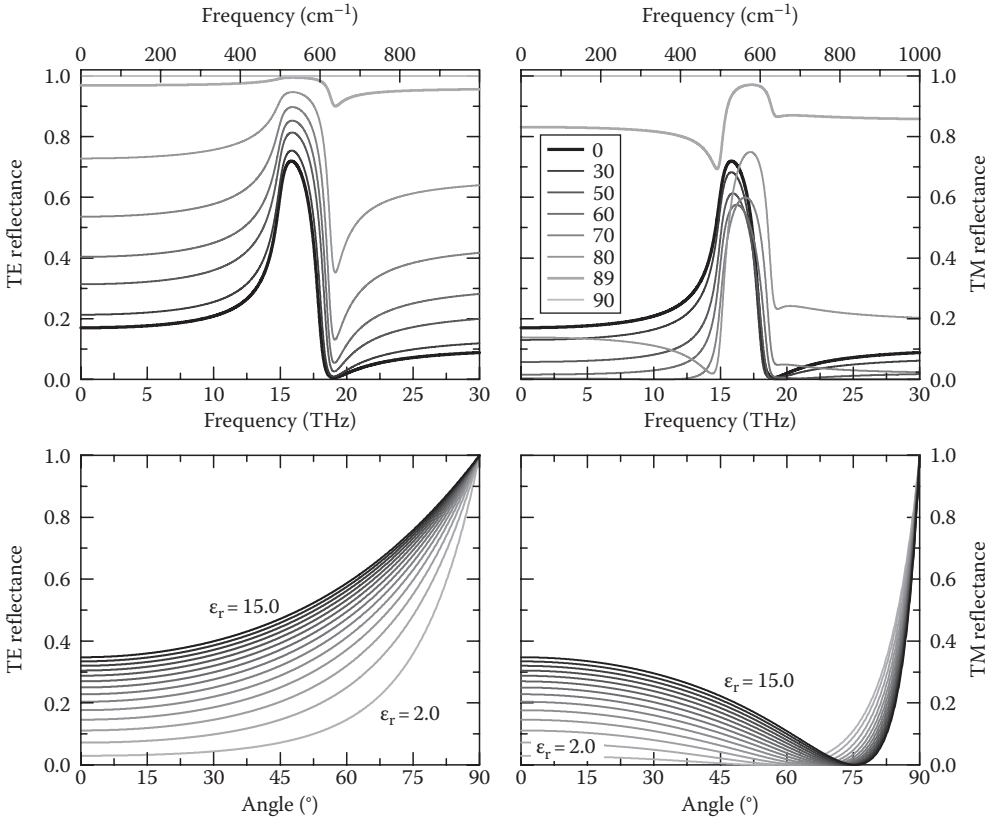


Figure 1.3 Reflection from an opaque material with a frequency-dependent dielectric function (top panels) and as a function of various constant dielectric values (bottom panels). The top panels are the reflectance (1.87) and (1.88) as a function of frequency, plotted in both THz and cm^{-1} units, and assume a single Lorentz oscillator (1.130), with parameters $\varepsilon_\infty = 4.0$, $\omega_p = 2\pi \times 20\,\text{THz}$, $\omega_0 = 2\pi \times 15\,\text{THz}$, and $\gamma = 2\pi \times 0.75\,\text{THz}$. The bottom panels show the TE (1.87) and TM (1.88) reflectance for a material of varying dielectric constant, ranging $\varepsilon_r = 2.0$–15, as a function of angle.

there is no reflection or transmission from the outgoing interface. Figure 1.3 plots (1.87) and (1.88) for both frequency-dependent optical constants (top panels) and for constant optical constants (bottom panels).

For normal incidence these equations may be further simplified and expressed as

$$R = |r|^2 = \left|\frac{\mu_r - n}{\mu_r + n}\right|^2 = \left|\frac{Z - Z_0}{Z + Z_0}\right|^2, \tag{1.89}$$

where $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ is the impedance of free space, and we have dropped the TE and TM subscripts, since these are undefined for $\theta = 0$. The phase of the reflected wave ϕ_r is given by

$$\tan(\phi_r) = \frac{\text{Im}(r)}{\text{Re}(r)} = \frac{-2n_2\mu_r}{\mu_r^2 - n_1^2 - n_2^2} = \frac{2Z_0Z_2}{Z_1^2 + Z_2^2 - Z_0^2}, \tag{1.90}$$

where Im and Re are the imaginary and real portions of the complex reflection coefficient *r*. Likewise at normal incidence the transmittance of a layer can be simplified using (1.42) and (1.86) and expressed as

$$T = \mathrm{Re}\left[\frac{n}{\mu_r}\right]\left|\frac{2\mu_r}{n+\mu_r}\right|^2 = \mathrm{Re}\left[\frac{n}{\mu_r}\right]\left|\frac{2Z}{Z+Z_0}\right|^2 \tag{1.91}$$

1.3.2 Brewster's Angle

Notice that the reflectances given in (1.69) and (1.81) for TE and TM polarizations, respectively, both involve a numerator that is the difference between two terms. This leads us to ask whether there is a particular combination of optical constants and incident angle at which the reflectivity goes to zero. By setting each of the two terms, in the numerator of both (1.69) and (1.81), equal to each other, and using (1.42) and Snell's law we find

$$\cos^2\theta_{B,TE} = \frac{\varepsilon^{(2)}/\varepsilon^{(1)} - \mu^{(1)}/\mu^{(2)}}{\mu^{(2)}/\mu^{(1)} - \mu^{(1)}/\mu^{(2)}} \tag{1.92}$$

and

$$\cos^2\theta_{B,TE} = \frac{\mu^{(2)}/\mu^{(1)} - \varepsilon^{(1)}/\varepsilon^{(2)}}{\varepsilon^{(2)}/\varepsilon^{(1)} - \varepsilon^{(1)}/\varepsilon^{(2)}}, \tag{1.93}$$

where the subscripted B,TE and B,TM refer to Brewster's angle (also known as the polarizing angle) for TE and TM polarizations, respectively. At the particular angles specified by (1.92) and (1.93), the reflectivity of light is zero and the wave is entirely refracted. If the materials under consideration do not exhibit a magnetic response, such that $\mu^{(1)} = \mu^{(2)} = \mu_0$, then no Brewster angle exists for TE polarization; however, (1.93) for TM polarization may be expressed as

$$\tan\theta_{B,TM} = \frac{n^{(2)}}{n^{(1)}}. \tag{1.94}$$

The effect of Brewster's angle can be observed in Figure 1.3, where we plot the TM reflectance as a function of incident angle for media consisting of different dielectric constants.

1.3.3 Reflection and Transmission for a Slab

We next consider the TE transmission and reflection from a flat slab of material, as shown in Figure 1.4. The transmission matrices for the two interfaces of the slab have already been found in (1.65), and the transmission matrix that describes propagation across a material with optical constants $\varepsilon^{(2)}$, $\mu^{(2)}$ is

$$\mathbf{M}^{\mathrm{slab}} = \begin{pmatrix} e^{ik_z^{(2)}d} & 0 \\ 0 & e^{-ik_z^{(2)}d} \end{pmatrix} \tag{1.95}$$

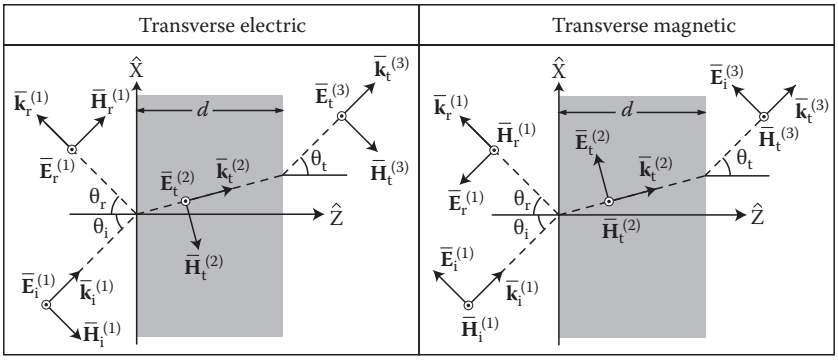


Figure 1.4 Schematic depicting the transmission and reflection of an electromagnetic wave from a slab of material. The left and right panels show the TE and TM polarization cases, respectively.

We would like to find a transmission matrix that describes the total transmission and reflection of the two interfaces plus the slab. This is given by the transmission matrix composition law (Markos and Soukoulis 2008) as

$$\mathbf{M}^{\text{total}} = \mathbf{M}_{12}\begin{pmatrix} e^{ik_z^{(2)}d} & 0 \\ 0 & e^{-ik_z^{(2)}d} \end{pmatrix}\mathbf{M}_{23}, \tag{1.96}$$

where

$\mathbf{M}_{12}$ is the transmission matrix for the interface between media (1) and (2)
$\mathbf{M}_{23}$ is the transmission matrix for the interface between media (2) and (3)

More generally if we have N total slabs and interfaces, we may write

$$\mathbf{M}^{\text{total}} = \mathbf{M}_N\mathbf{M}_{N-1}\cdots\mathbf{M}_2\mathbf{M}_1. \tag{1.97}$$

For TE polarization we may write

$$\begin{pmatrix} E_{\text{t}}^{(3)} \\ E_{\text{i}}^{(3)} \end{pmatrix} = \mathbf{M}^{\text{total}}\begin{pmatrix} E_{\text{t}}^{(1)} \\ E_{\text{i}}^{(1)} \end{pmatrix}. \tag{1.98}$$

By performing the matrix multiplication specified in (1.96) we find the specific matrix elements

$$M_{11}^{\text{total}} = \frac{1}{2}\left[1+\frac{\mu^{(3)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)+\frac{i}{2}\left[\frac{\mu^{(3)}}{\mu^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}+\frac{\mu^{(2)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.99}$$

$$M_{12}^{\text{total}} = \frac{1}{2}\left[1-\frac{\mu^{(3)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)+\frac{i}{2}\left[\frac{\mu^{(3)}}{\mu^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}-\frac{\mu^{(2)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.100}$$

$$M_{22}^{\text{total}} = \frac{1}{2}\left[1+\frac{\mu^{(3)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)-\frac{i}{2}\left[\frac{\mu^{(3)}}{\mu^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}+\frac{\mu^{(2)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.101}$$

$$M_{21}^{\text{total}} = \frac{1}{2}\left[1-\frac{\mu^{(3)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)-\frac{i}{2}\left[\frac{\mu^{(3)}}{\mu^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}-\frac{\mu^{(2)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.102}$$

As in the case of the single interface, we can obtain the transmission and reflection coefficients as

$$t_{\text{TE}} = \frac{E_{\text{t}}^{(3)}}{E_{\text{i}}^{(1)}} = \frac{\det \mathbf{M}^{\text{total}}}{M_{22}^{\text{total}}}, \tag{1.103}$$

$$r_{\text{TE}} = \frac{E_{\text{r}}^{(1)}}{E_{\text{i}}^{(1)}} = -\frac{M_{21}^{\text{total}}}{M_{22}^{\text{total}}} \tag{1.104}$$

and thus the transmittance and reflectance are

$$T = \frac{S^{(3)}}{S^{(1)}} = \frac{\mu^{(1)}}{\mu^{(3)}}\frac{\text{Re}\,k_z^{(3)}}{\text{Re}\,k_z^{(1)}}\left|t_{\text{TE}}\right|^2. \tag{1.105}$$

$$R = \left|r_{\text{TE}}\right|^2. \tag{1.106}$$

The case for TM polarization follows in a straightforward manner. We simply list here the matrix elements for the total scattering matrix

$$M_{11}^{\text{total}} = \frac{1}{2}\left[1+\frac{\varepsilon^{(3)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)+\frac{i}{2}\left[\frac{\varepsilon^{(3)}}{\varepsilon^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}+\frac{\varepsilon^{(2)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.107}$$

$$M_{12}^{\text{total}} = \frac{1}{2}\left[1-\frac{\varepsilon^{(3)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)+\frac{i}{2}\left[\frac{\varepsilon^{(3)}}{\varepsilon^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}-\frac{\varepsilon^{(2)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.108}$$

$$M_{22}^{\text{total}} = \frac{1}{2}\left[1+\frac{\varepsilon^{(3)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)-\frac{i}{2}\left[\frac{\varepsilon^{(3)}}{\varepsilon^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}+\frac{\varepsilon^{(2)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.109}$$

$$M_{21}^{\text{total}} = \frac{1}{2}\left[1-\frac{\varepsilon^{(3)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(3)}}\right]\cos\left(k_z^{(2)}d\right)-\frac{i}{2}\left[\frac{\varepsilon^{(3)}}{\varepsilon^{(2)}}\frac{k_z^{(2)}}{k_z^{(3)}}-\frac{\varepsilon^{(2)}}{\varepsilon^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.110}$$

Equations 1.99 through 1.102 and 1.107 through 1.110 completely describe the transmission and reflection of an electromagnetic wave from a material $\varepsilon^{(2)}$, $\mu^{(2)}$, where the wave originates in a material $\varepsilon^{(1)}$, $\mu^{(1)}$ and is transmitted into a different material $\varepsilon^{(3)}$, $\mu^{(3)}$.

We next consider the special case where the slab is a nonmagnetic dielectric and is embedded in free space such that $\mu^{(1)} = \mu^{(2)} = \mu^{(3)} = \mu_0$ and $\varepsilon^{(1)} = \varepsilon^{(3)} = \varepsilon_0$. Since our slab is embedded within the same type of material we may also explicitly write $k_z^{(1)} = k_z^{(3)}$. The matrix elements for the TE case become

$$M_{11}^{\text{total}} = \cos\left(k_z^{(2)}d\right) + \frac{i}{2}\left[\frac{k_z^{(2)}}{k_z^{(1)}} + \frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right) \tag{1.111}$$

$$M_{12}^{\text{total}} = \frac{i}{2}\left[\frac{k_z^{(2)}}{k_z^{(1)}} - \frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.112}$$

$$M_{22}^{\text{total}} = \cos\left(k_z^{(2)}d\right) - \frac{i}{2}\left[\frac{k_z^{(2)}}{k_z^{(1)}} + \frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.113}$$

$$M_{21}^{\text{total}} = -\frac{i}{2}\left[\frac{k_z^{(2)}}{k_z^{(1)}} - \frac{\mu^{(2)}}{\mu^{(1)}}\frac{k_z^{(1)}}{k_z^{(2)}}\right]\sin\left(k_z^{(2)}d\right). \tag{1.114}$$

We obtain the transmission and reflection coefficients as

$$t_{\text{TE}} = \frac{1}{M_{22}^{\text{total}}}, \tag{1.115}$$

$$r_{\text{TE}} = -\frac{M_{12}^{\text{total}}}{M_{22}^{\text{total}}}, \tag{1.116}$$

and thus the transmittance and reflectance are

$$T = \left|t_{\text{TE}}\right|^2, \tag{1.117}$$

$$R = \left|r_{\text{TE}}\right|^2. \tag{1.118}$$

Using (1.111) through (1.118), it is then possible to characterize the complex dielectric properties, $\tilde{\varepsilon}(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$, of single-layer materials. In this case, one must assume that the material exhibits only an electric response and that the magnetic permeability is equal to the free space value. The transmittance $T(\omega)$, reflectance $R(\omega)$, and thickness d can be measured and then used to calculate the dielectric function using least-squares nonlinear curve-fitting algorithms (Chapter 4).

1.3.4 Beer's Law

The form we have obtained for the transmittance of TE polarized light (1.117) completely describes, as we have mentioned, the transmittance. However, we would like to derive a simpler form valid for normal incidence that is easier to use for calculation.

When light passes through a slab of material, its absorption is described by Beer's law as

$$I(z) = I_0 e^{-\alpha z}, \tag{1.119}$$

where

z is the direction the wave propagates
$I(z)$ is the intensity of the wave
I_0 is the intensity of the wave at $z = 0$
α is the absorption coefficient

Beer's law describes the amount of power absorbed in a medium per unit length, that is, α has units of 1/m. We may also express the electric field of an electromagnetic wave traveling along the z-direction within a medium as

$$\overline{\mathbf{E}}(z) = \overline{\mathbf{E}}_0 e^{i(kz-\omega t)} = \overline{\mathbf{E}}_0 e^{i(k_1 z-\omega t)} e^{-k_2 z}, \tag{1.120}$$

where we have expressed the wave vector in terms of its real and imaginary parts, that is, $\tilde{k} = k_1 + ik_2$. The intensity of this electromagnetic wave is thus

$$I(z) = I_0 e^{-2k_2 z}, \tag{1.121}$$

where $I_0 = |E_0|^2$. Upon equating (1.121) through (1.119) we may write the relationship between the absorption coefficient and the imaginary part of the wave vector and imaginary index of refraction as

$$\alpha = 2k_2 = 2n_2 \frac{\omega}{c}. \tag{1.122}$$

Thus if light is incident upon a medium of thickness d, we may express the total transmittance as the transmittance of the first interface (1.91), multiplied by the power lost within the medium (1.119) or (1.121), then finally multiplied by the transmittance of the outgoing interface (1.91). (We assume the medium is reciprocal and thus the transmittance of both interfaces is identical.)

We may then express the total transmittance as

$$T^{total} = \left(\mathrm{Re}\left[\frac{n}{\mu_r} \right] \left| \frac{2Z}{Z + Z_0} \right|^2 \right)^2 e^{-2k_2 d} \tag{1.123}$$

Although this form is simpler than (1.117), it does not account for multiple reflected and transmitted paths that may occur due to the parallel interfaces of the material. Both (1.117) and (1.123) are plotted in Figure 1.5.

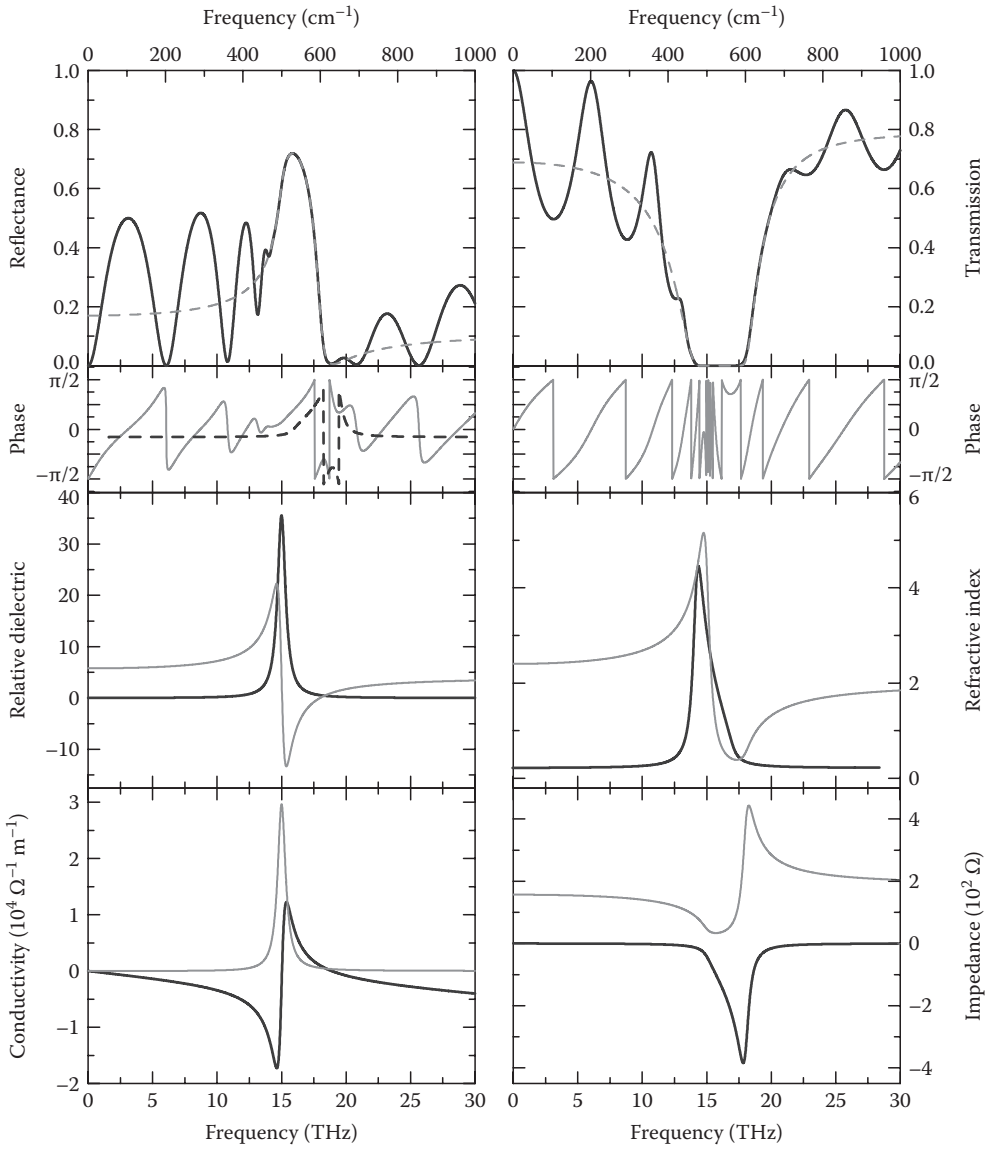


Figure 1.5 Reflection (1.118), reflected phase (ϕ_R = arctan(Im r_{TE}/Re r_{TE})), transmission (1.117), transmitted phase (ϕ_T = arctan(Im t_{TE}/Re t_{TE})), dielectric function (1.130), index of refraction (1.28), impedance (1.29), and the conductivity (1.30) for a transparent material of thickness $d = 10\,\mu m$. Here the material is modeled to consist of a single Lorentz oscillator (1.130), with parameters $\varepsilon_\infty = 4.0$, $\omega_p = 2\pi \times 20\,THz$, $\omega_0 = 2\pi \times 15\,THz$, and $\gamma = 2\pi \times 0.75\,THz$. The real and imaginary portions of the optical constants are plotted as gray and dark gray, respectively, for the bottom four panels. The reflectance and reflected phase for an opaque material with the same parameters is also plotted as a dashed light gray curve and dashed dark gray curve, respectively. The transmission is also plotted (dashed light gray curve top right) via (1.123), which neglects multiple reflections within the material.

1.4 ELECTROMAGNETIC RESPONSE OF MATERIALS

1.4.1 Drude–Lorentz Model

Much resonant phenomena in materials can be expressed in terms of Drude–Lorentz oscillators. We consider the electrons within a material to be tightly bound to the atoms, and in an externally applied electric field their motion can be described by

$$m\frac{d^2\bar{\mathbf{r}}}{dt^2} + m\gamma\frac{d\bar{\mathbf{r}}}{dt} + m\omega_0^2\bar{\mathbf{r}} = -e\bar{\mathbf{E}}, \tag{1.124}$$

where

m and $-e$ are the mass and charge of the electron
γ is the damping
$\bar{\mathbf{r}}$ is the displacement
ω_0 is the resonant frequency
$\bar{\mathbf{E}}$ is the external electric field

The loss term is general, as required by causality. Assuming an electric field space and time dependence of $e^{i(\bar{\mathbf{k}}\cdot\bar{\mathbf{r}} - \omega t)}$, we find the solution

$$\bar{\mathbf{r}} = -\frac{e\bar{\mathbf{E}}}{m}\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}. \tag{1.125}$$

Physically this means that under the action of a harmonic electric field $\bar{\mathbf{E}}$, charges are displaced from their equilibrium positions in accord with (1.125). Thus the polarization can then be expressed as

$$\bar{\mathbf{P}} = -ne\bar{\mathbf{r}} = \frac{ne^2\bar{\mathbf{E}}}{m}\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}, \tag{1.126}$$

where n is the number density (number of charges per unit volume). Using (1.6) and (1.8) we may then write the relative dielectric function as

$$\tilde{\varepsilon}_r(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i\gamma\omega}, \tag{1.127}$$

where we have defined

$$\omega_p^2 \equiv \frac{ne^2}{\varepsilon_0 m}. \tag{1.128}$$

ω_p^2 is known as the oscillator strength, or the square of the plasma frequency. Equation 1.127 is known as the Drude–Lorentz equation and describes a resonant response to a driving force, in this case, a time-varying electric field.

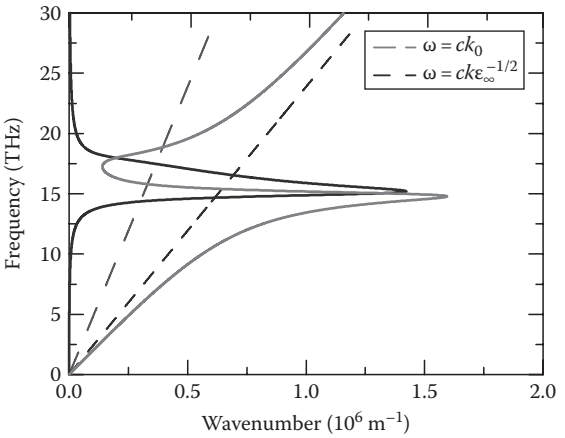


Figure 1.6 Dispersion relation (1.38) for a Lorentz oscillator (1.130), with parameters ε_∞ = 4.0, $\omega_p = 2\pi \times 20$ THz, $\omega_0 = 2\pi \times 15$ THz, and $\gamma = 2\pi \times 0.75$ THz. The real and imaginary portions are plotted as gray and dark gray, respectively. The long dashed and short dashed lines are the free space light line and the light line in the material, respectively.

If we have the case that there are n oscillators, then Equation 1.127 may be generalized to

$$\tilde{\varepsilon}_r(\omega) = 1 + \sum_n \frac{\omega_{p,n}^2}{\omega_{0,n}^2 - \omega^2 - i\gamma_n\omega}. \tag{1.129}$$

However, it is more convenient to modify (1.128) even further when considering the spectroscopy of actual materials. For example, one may characterize phonons for a solid in the far-infrared frequency regime, but contributions to the dielectric function may occur due to higher-energy modes. Thus typically (1.128) is written as

$$\tilde{\varepsilon}_r(\omega) = \varepsilon_\infty + \sum_n \frac{\omega_{p,n}^2}{\omega_{0,n}^2 - \omega^2 - i\gamma_n\omega}. \tag{1.130}$$

where ε_∞ is called "epsilon infinity" and describes contributions to the dielectric function at frequencies higher than the region of consideration (Chapter 3).

In Figure 1.5 we plot the reflection, transmission, reflected phase, transmitted phase, dielectric function, refractive index, conductivity, and impedance for a Lorentz oscillator using (1.130). One may calculate the index of refraction from the relative dielectric function through (1.28) and then further relate this to the wave vector through the dispersion relation (1.38). Figure 1.6 plots the dispersion relation—real and imaginary portions of the wave vector versus frequency—for a Lorentz oscillator (1.130).

1.4.2 Drude Model

If our resonator is centered at zero frequency $\omega_0 = 0$, (1.127) becomes

$$\tilde{\varepsilon}_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\gamma\omega}. \tag{1.131}$$

We can express this in terms of the conductivity by substituting (1.127) into (1.30) with the help of (1.8). We find

$$\tilde{\sigma}(\omega) = \varepsilon_0 \frac{\omega_p^2}{\gamma - i\omega}. \tag{1.132}$$

We can express (1.132) explicitly in real and imaginary parts to find

$$\tilde{\sigma}(\omega) = \varepsilon_0 \omega_p^2 \left[\frac{\tau}{1+\omega^2\tau^2} + i\frac{\omega\tau^2}{1+\omega^2\tau^2} \right], \tag{1.133}$$

where we have used the scattering time in place of the damping

$$\tau = \frac{1}{\gamma}. \tag{1.134}$$

The Drude–Lorentz model, although based on an ad hoc model, accurately describes the response of many dielectric and metallic materials; several examples are given in Chapter 3. The Drude model (1.132) is useful for characterizing the response of good metals to electromagnetic waves, whereas the Drude–Lorentz model (1.130) may properly describe "bound" resonant phenomena, such as optically active phonons.

1.4.3 Extended Drude Model

Although the Drude model is extremely successful in describing the electromagnetic response of metals to electromagnetic waves, there are some cases where a more complex model may elucidate the underlying physics (Timusk and Tanner 1989). For example, the response of carriers to external time-varying electric fields and the relationship to dissipative effects on carriers through scattering by phonons and defects can be described generally by a transport equation. The Drude model for metals (1.132) may be derived in a much more rigorous way from the Boltzmann transport equation (Huang 1987), with the approximation of a collision term that is described by a single collision frequency $\gamma = 1/\tau$ (Ashcroft and Mermin 1976, Kittel 1996). However, the Boltzmann transport formalism permits a much more valid description of the electromagnetic response of metals than just simply reproducing the Drude equation. The scattering rate (1.134) and carrier mass (1.128), under the Boltzmann formalism, become frequency dependent and can be described in terms of the real and imaginary parts, respectively, of a complex memory function

$$\tilde{\gamma}(\omega) = \gamma_1(\omega) + i\gamma_2(\omega). \tag{1.135}$$

Defining a dimensionless parameter

$$\lambda(\omega) = -\frac{\gamma_2(\omega)}{\omega}, \tag{1.136}$$

and plugging (1.135) through (1.136) into the Drude equation (1.132), we find

$$\tilde{\sigma}(\omega) = \varepsilon_0 \frac{\omega_p^2/[1+\lambda(\omega)]}{\gamma_1(\omega)/[1+\lambda(\omega)] - i\omega}. \tag{1.137}$$

By defining a renormalized plasma frequency

$$\omega_p^{*2}(\omega) = \frac{\omega_p^2}{1+\lambda(\omega)}, \tag{1.138}$$

and damping rate

$$\gamma^*(\omega) = \frac{\gamma_1(\omega)}{1+\lambda(\omega)}, \tag{1.139}$$

Equation 1.137 can be rewritten in a more familiar "Drude" form (compared to (1.132)):

$$\tilde{\sigma}(\omega) = \varepsilon_0 \frac{\omega_p^{*2}}{\gamma^*(\omega) - i\omega}. \tag{1.140}$$

Renormalization of the plasma frequency (1.123) occurs via introduction of an effective mass

$$\frac{m^*(\omega)}{m} = 1+\lambda(\omega). \tag{1.141}$$

By rearranging (1.136) we can describe the scattering rate and effective mass in terms of the conductivity

$$\frac{1}{\tau(\omega)} = \gamma(\omega) = \varepsilon_0 \omega_p^2 \,\mathrm{Re}\left(\frac{1}{\tilde{\sigma}(\omega)}\right), \tag{1.142}$$

$$\frac{m^*(\omega)}{m} = -\varepsilon_0 \omega_p^2 \omega^{-1} \,\mathrm{Im}\left(\frac{1}{\tilde{\sigma}(\omega)}\right). \tag{1.143}$$

The conductivity is a causal response function and thus the frequency-dependent damping and effective mass are also causal and related to each other by the Kramers–Kronig relations.

1.4.4 Kramers–Kronig Relations

The Kramers–Kronig relations are equations that connect the real and imaginary parts of any analytic complex function (Kronig 1926, Kramers 1927). Typically, in spectroscopy, they are used to infer the real (or imaginary) part of a response function, given the imaginary (or real) part. The connection between the Kramers–Kronig relations and response functions exists because causality implies analyticity, and conversely analyticity implies causality (Toll 1956). A system that responds to an external stimulus may be described by a response function and, in general, for a linear system we may write

$$X(\mathbf{r},t) = \int_{-\infty}^{\infty} G(\mathbf{r},\mathbf{r}',t,t') f(\mathbf{r}',t') d\mathbf{r}' dt'. \tag{1.144}$$

This describes the response $X(\mathbf{r}, t)$ of the system at location $\mathbf{r}$ and time t to a stimulus $f(\mathbf{r}', t')$, which acts at a time t' and a location $\mathbf{r}'$. The quantity $G(\mathbf{r}, \mathbf{r}', t, t')$ is called the response function.

In our case the stimulus will be an electromagnetic wave, which we shall assume has a sufficiently long enough wavelength that we neglect spatial dispersion, that is, we assume that the response functions do not depend on the wave vector. Thus, we make the local approximation—what happens at a particular location depends only on the fields existing at that place. Thus, our response function can be written as

$$G(\mathbf{r},\mathbf{r}',t,t') = \delta(\mathbf{r}-\mathbf{r}')G(t-t'), \tag{1.145}$$

and (1.144) becomes

$$X(t) = \int_{-\infty}^{\infty} G(t-t')f(t')dt'. \tag{1.146}$$

The requirement of causality states that there can be no response before the stimulus occurs, that is,

$$G(t-t') = 0, \quad t < t'. \tag{1.147}$$

It is now convenient to transform to the frequency domain, as we may take advantage of the convolution theorem. The Fourier transforms may be written as

$$f(\omega) = \int f(t)\exp(i\omega t)dt. \tag{1.148}$$

$$X(\omega) = \int X(t)\exp(i\omega t)dt. \tag{1.149}$$

$$G(\omega) = \int G(t-t')\exp[i\omega(t-t')]dt. \tag{1.150}$$

Plugging (1.146) into (1.149) we find

$$\begin{aligned} X(\omega) &= \int dt \exp(i\omega t)\left[\int G(t-t')f(t')dt'\right] \\ &= \int dt' f(t')\left[\int G(t-t')\exp(i\omega t)dt\right] \\ &= \int dt' f(t')(\exp i\omega t)\left[\int G(t-t')\exp i\omega t(t-t')dt\right]. \end{aligned} \tag{1.151}$$

Substituting in (1.142) and (1.144) we find

$$X(\omega) = G(\omega)f(\omega). \tag{1.152}$$

Thus in the frequency domain we see that the response $X(\omega)$ is simply equal to multiplication of a monochromatic stimulus $f(\omega)$ by a number $G(\omega)$, termed the generalized susceptibility.

Generally, $G(\omega)$ is complex, where the real part describes the amplitude of the response and the imaginary part describes the phase difference between the stimulus and the response.

If we consider a complex frequency $\tilde{\omega} = \omega_1 + i\omega_2$, then (1.150) becomes

$$G(\tilde{\omega}) = \int G(t-t')\exp[i\omega_1(t-t')]\exp[-\omega_2(t-t')]dt. \tag{1.153}$$

This integral can be evaluated using Cauchy's integral theorem (Brown and Churchill 1996). However, first we must determine the half space in which the integral is to be determined. The factor $\exp[i\omega_1(t-t')]$ is bounded everywhere, and the factor $\exp[-\omega_2(t-t')]$ is bounded in the upper half plane for $(t-t') > 0$, or bounded in the lower half plane for $(t-t') < 0$. Our requirement of causality (1.147) thus requires that (1.153) be evaluated in the upper half plane.

If we let the contour be a semicircle, as shown in Figure 1.7, with a small bump around the frequency ω, then from Cauchy's theorem we may write

$$\tilde{G}(\omega) = \frac{1}{i\pi}\mathbf{P}\int_{-\infty}^{\infty}\frac{\tilde{G}(\omega')}{\omega'-\omega_0}d\omega', \tag{1.154}$$

where **P** stands for the principal value. We integrate over the contour such that contributions of the semicircle to $G(\omega)$ go to zero as the radius of the semicircle approaches infinity. If we now express our response function as real and imaginary parts, that is, $\tilde{G}(\omega) = G_1(\omega) + iG_2(\omega)$, then we may, now dropping the ω_0 notation, write

$$G_1(\omega) = \frac{1}{\pi}\mathbf{P}\int_{-\infty}^{\infty}\frac{G_2(\omega')}{\omega'-\omega}d\omega' \tag{1.155}$$

$$G_2(\omega) = -\frac{1}{\pi}\mathbf{P}\int_{-\infty}^{\infty}\frac{G_1(\omega')}{\omega'-\omega}d\omega'. \tag{1.156}$$

Notice that (1.156) tells us that if we have determined $G_1(\omega')$ over all frequencies, then we can find $G_2(\omega)$. Thus we observe that the real and imaginary parts of our response function $\tilde{G}(\omega)$ are not independent, but are related by formulas (1.155) and (1.156), known as dispersion relations.

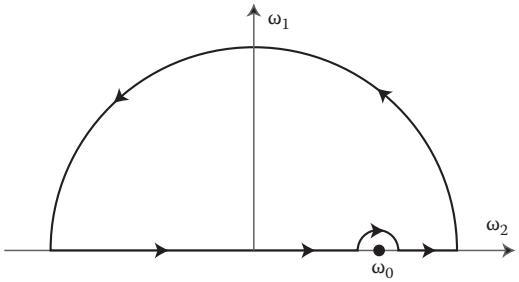


Figure 1.7 Contour for obtaining Equation 1.154.

We now consider a specific example of the response of a system to an external electric field. In (1.126) we found the polarization of an electron due to an external electric field. Upon comparison to (1.152) and (1.6) we see

$$\bar{\mathbf{P}}(\omega) = G(\omega)\bar{\mathbf{E}}(\omega), \tag{1.157}$$

where

$$G(\omega) = \varepsilon_0\chi_e = \frac{ne^2}{m}\frac{1}{\omega_0^2 - \omega^2 - i\gamma\omega}. \tag{1.158}$$

From inspection of (1.155) and (1.156), and considering a complex electric susceptibility $\tilde{\chi}(\omega) = \chi_1(\omega) + i\chi_2(\omega)$ (dropping the subscripted "*e*") we may thus write

$$\chi_1(\omega) = \frac{1}{\pi}\mathbf{P}\int_{-\infty}^{\infty}\frac{\chi_2(\omega')}{\omega' - \omega}d\omega'. \tag{1.159}$$

$$\chi_2(\omega) = -\frac{1}{\pi}\mathbf{P}\int_{-\infty}^{\infty}\frac{\chi_1(\omega')}{\omega' - \omega}d\omega'. \tag{1.160}$$

We are interested in physical systems with real inputs and real outputs. This means we must have

$$\tilde{G}(\omega) = \tilde{G}^*(-\omega), \tag{1.161}$$

or in terms of the real and imaginary parts of the susceptibility

$$\chi_1(\omega) = \chi_1(-\omega). \tag{1.162}$$

$$\chi_2(\omega) = -\chi_2(-\omega). \tag{1.163}$$

From (1.8) we see that $\tilde{\chi}(\omega) = \tilde{\varepsilon}_r(\omega) - 1$. Thus we may write

$$\varepsilon_{1,r}(\omega) - 1 = \frac{2}{\pi}\mathbf{P}\int_0^{\infty}\frac{\omega'\varepsilon_{2,r}(\omega')}{\omega'^2 - \omega^2}d\omega'. \tag{1.164}$$

$$\varepsilon_{2,r}(\omega) = -\frac{2\omega}{\pi}\mathbf{P}\int_0^{\infty}\frac{\varepsilon_{1,r}(\omega') - 1}{\omega'^2 - \omega^2}d\omega' \tag{1.165}$$

Equations 1.164 and 1.165 are known as the Kramers–Kronig dispersion relations.

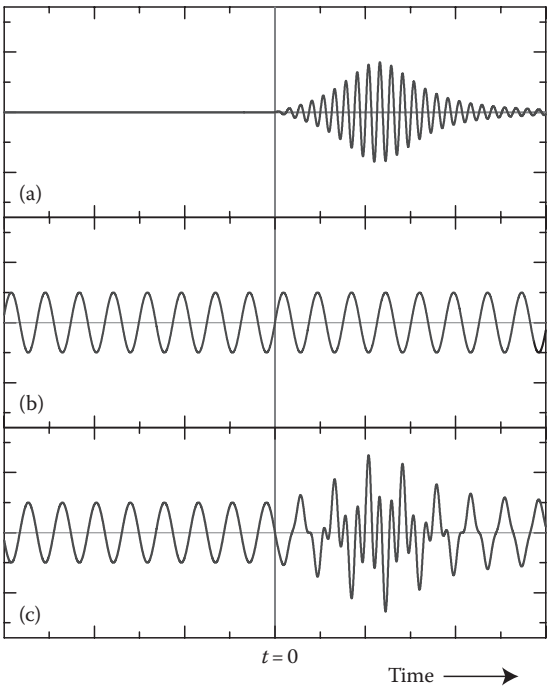


Figure 1.8 Diagram demonstrating the connection between causality and dispersion. (a) Time pulse incident upon a system at $t = 0$. (b) Single frequency component absorbed by the system. (c) Remaining electromagnetic pulse after subtraction of (b) from (a).

Although we have used a purely mathematical treatment to derive the relation between the real and imaginary portions of a causal response function, a simple intuitive argument presents the same picture (Toll 1956). Consider a pulse of electromagnetic radiation which impinges on a system at $t = 0$ as shown in Figure 1.8a. Let us assume this system can only absorb a single frequency, as shown in Figure 1.8b. Then upon absorption of this one component, what remains of our electromagnetic pulse is shown in Figure 1.8c, that is, the subtraction of B from A. No real physical (causal) system can have a response before the arrival of the signal, and thus what is shown in Figure 1.8c is nonsense. Therefore, absorption of the single frequency shown in Figure 1.8b must be accompanied by a phase shift of the other frequency components such that, for $t < 0$, the time signal shown in Figure 1.8c vanishes. In terms of physical systems, this means that a response function cannot describe the absorption alone but must also describe the phase shift and therefore must consist of real and imaginary parts.

1.4.5 Sum Rules

Equation 1.164 may be written as

$$\varepsilon_{1,\mathrm{r}}(\omega)-1=\frac{2}{\pi}\mathbf{P}\int_0^{\omega_c}\frac{\omega'\varepsilon_{2,\mathrm{r}}(\omega')}{\omega^2-\omega^2}d\omega'+\frac{2}{\pi}\mathbf{P}\int_{\omega_c}^{\infty}\frac{\omega'\varepsilon_{2,r}(\omega')}{\omega'^2-\omega^2}d\omega', \tag{1.166}$$

where ω_c denotes a cutoff frequency above which there is no absorption. Thus according to (1.165), $\varepsilon_{2,\mathrm{r}}(\omega) = 0$ for $\omega > \omega_c$. At frequencies much greater than the cutoff frequency, that is,

$\omega \gg \omega_c$, we can neglect ω' in the denominator of the first integral in (1.166), and the second integral is zero since $\varepsilon_{2,r}(\omega) = 0$ for $\omega \gg \omega_c$. Under these conditions, (1.166) becomes

$$\varepsilon_{1,r}(\omega) = 1 - \frac{2}{\pi\omega^2}\int_0^{\omega_c} \omega'\varepsilon_{2,r}(\omega')d\omega', \quad \omega \gg \omega_c. \tag{1.167}$$

Notice also that if we are at sufficiently high frequencies, then (1.132) can be written as

$$\tilde{\varepsilon}_r(\omega) = 1 - \frac{\omega_p^2}{\omega^2}. \tag{1.168}$$

Thus equating (1.167) and (1.168) we see

$$\int_0^{\infty} \omega\varepsilon_{2,r}(\omega)d\omega = \frac{\pi}{2}\omega_p^2, \tag{1.169}$$

where we have extended the upper limit of integration since $\varepsilon_{2,r}(\omega) = 0$ for $\omega > \omega_c$. It is more common to express (1.169) in terms of the conductivity. Using (1.30) and (1.128) we may write

$$\int_0^{\infty} \sigma_1(\omega)d\omega = \frac{\pi}{2}\omega_p^2 = \frac{ne^2\pi}{2\varepsilon_0 m}. \tag{1.170}$$

1.5 SUMMARY

We have provided an abbreviated overview of classical electromagnetic wave theory. Although we were only able to touch upon a limited number of topics, we hope this chapter will provide useful knowledge, not only for the rest of this book but also as a general reference on electricity and magnetism.

ACKNOWLEDGMENTS

I would like to acknowledge the Office of Naval Research for support under contract N00014-07-1-0819.

REFERENCES

Ashcroft, N.W. and Mermin, N.D. 1976, *Solid State Physics*, Brooks Cole, Pacific Grove, CA.

Brillouin, L. 1960, *Wave Propagation and Group Velocity*, Academic Press, New York.

Brown, J.W. and Churchill, R.V. 1996, *Complex Variables and Applications*, 6th edn., McGraw Hill, London, U.K.

Burns, G. 1990, *Solid State Physics*, Academic Press, Boston, MA.

Cheng, D.K. and Kong, J.A. 1968a, Covariant descriptions of bianisotropic media, *Proc. IEEE*, 56, 248.

Cheng, D.K. and Kong, J.A. 1968b, Time-harmonic fields in source-free bianisotropic media, *J. Appl. Phys*., 39, 5792.

Choy, T.C. 1999, *Effective Medium Theory: Principles and Applications*, Oxford University Press, New York.

Cohen, D.L. 2001, *Demystifying Electromagnetic Equations*, SPIE–The International Society for Optical Engineering, Bellingham, Washington.
Collin, R.E. 1991, *Field Theory of Guided Waves*, 2nd edn., John Wiley & Sons, Inc., New York.
Cottam, M.G. and Tilley, D.R. 1989, *Introduction to Surface and Superlattice Excitations*, Cambridge University Press, Cambridge, U.K.
Eerenstein, W., Mathur, N.D., and Scott, J.F. 2006, Multiferroic and magnetoelectric materials, *Nature*, 442, 759.
Gibbs, W. and Wilson, E.B. 1929, *Vector Analysis. A Text-Book for the Use of Students of Mathematics and Physics*, Yale University Press, New Haven, CT.
Heaviside, O. 1892, On the forces, stresses, and fluxes of energy in the electromagnetic field, *Phil. Trans. R. Soc. A*, 183, 423–480.
Huang, K. 1987, *Statistical Mechanics*, 2nd edn., Wiley, Indianapolis, IN.
Jackson, J.D. 1998, *Classical Electrodynamics*, 3rd edn., John Wiley & Sons, Inc., Indianapolis, IN.
Kittel, C. 1996, *Introduction to Solid State Physics*, 7th edn., John Wiley & Sons, Inc., Indianapolis, IN.
Kong, J.A. 1990, *Electromagnetic Wave Theory*, 2nd edn., John Wiley & Sons, Inc., Indianapolis, IN.
Kramers, H.A. 1927, La diffusion de la lumiere par les atomes, *Atti Cong. Intern. Fisica*, 2, 545–557.
Kronig, R. de L. 1926, On the theory of the dispersion of x-rays, *J. Opt. Soc. Am*., 12, 547–557.
Landau, L.D., Pitaevskii, L.P., and Lifshitz, E.M. 1984, *Electrodynamics of Continuous Media*, 2nd edn., Butterworth-Heinemann, London, U.K.
Markos, P. and Soukoulis, C.M. 2008, *Wave Propagation*, Princeton University Press, Princeton, NJ.
Maxwell, J.C. 1865, A dynamical theory of the electromagnetic field, *R. Soc. Trans*., 155, 459–512.
Pomeroy, S.B. 1999, *Ancient Greece: A Political, Social, and Cultural History*, Oxford University Press, Oxford, U.K.
Ramo, S., Whinnery, J.R., and Van Duzer, T. 1994, *Fields and Waves in Communication Electronics*, 3rd edn., John Wiley & Sons, Inc., Indianapolis, IN.
Schelkunoff, S.A. 1938, The impedance concept and its application to problems of reflection, refraction, shielding and power absorption, *Bell System Tech. J*., 17, p. 17.
Sihvola, A.H. 1999, *Electromagnetic Mixing Formulas and Applications*, The Institution of Engineering and Technology, Herts, U.K.
Sommerfeld, A. 1950, *Lectures on Theoretical Physics*, Academic Press, New York.
Stratton, J.A. 1989, *Electromagnetic Theory*, McGraw-Hill Book Co., New York.
Timusk, T. and Tanner, D.B. 1941, Optical properties of high-temperature superconductors, in *The Physical Properties of High Temperature Superconductors*, ed. D.M. Ginsberg, World Scientific, Singapore, pp. 363–459.
Toll, J.S. 1956, Causality and the dispersion relation: Logical foundations, *Phys. Rev*., 104, 1760–1770.

Chapter 2

SEMICONDUCTORS AND THEIR NANOSTRUCTURES

Jeffrey Davis and Chennupati Jagadish

CONTENTS

2.1 BASIC SEMICONDUCTOR PHYSICS

A semiconducting material is defined as one that is electrically insulating at a temperature of absolute zero but that becomes conducting for temperatures below its melting point. The property that first attracted the attention of physicists in the early nineteenth century was the decreasing resistivity with increasing temperature, in stark contrast to metals where the opposite is true (Faraday 1839, Smith 1964, Ashcroft and Mermin 1976, p. 563). Both of these effects occur in semiconductors because it is possible to thermally excite charge carriers from a full energy band that prevents conduction to an unoccupied band that then allows electrons to move freely. The structure of these energy bands is used to explain most of the properties of semiconductors. In this chapter, we therefore begin with a brief introduction to semiconductor band structures and then continue with our discussion on some of the important and interesting properties observed in semiconductors and their nanostructures.

2.1.1 Band Structure

Consider a single silicon atom; the solution to the Schrödinger equation reveals discrete electronic orbitals, with four electrons in the valence shell. This silicon atom is able to covalently bond with four other silicon atoms, which in turn are able to bond with four silicon atoms, and so on, forming a crystal lattice. In a crystal lattice, the electronic energy structure changes. To a first approximation, the atomic orbitals are replaced by filled bonding and empty antibonding orbitals, with pairs of electrons localized between the two nuclei they are binding. In the absence of interactions between different pairs of electrons, every electron in the lattice will have an identical set of discrete states. In practice, however, this is a strongly interacting system and the Hamiltonian needs to include the electromagnetic interactions between all the valence electrons. The effect of this is that the Pauli exclusion principle prevents more than two electrons from existing in the same quantum state, lifting the degeneracy and generating an additional state for each additional Si–Si bond, with all states split by an amount that is determined by the strength of the interaction. In other words, a single Si–Si bond will have a single molecular orbital that is doubly degenerate (i.e., it can contain two electrons of opposite spin); a pair of Si–Si bonds will have two molecular orbitals (each doubly degenerate) that are split in energy due to the electromagnetic interaction; three Si–Si bonds will have three molecular orbitals that are split in energy; and so on, until, in a semi-infinite crystal lattice, there are a semi-infinite number of Si–Si bonds, and a semi-infinite number of states, split by a semi-infinitesimal amount, as depicted in Figure 2.1a. An alternative way of looking at the formation of bands is depicted in Figure 2.1b, where the energy structure is shown as a function of atom separation. As the nuclei become closer, the discrete atomic states shift and form bands as the interaction between electrons increases (Allison 1971). In a semi-infinite lattice, rather than a series of discrete states, there is, therefore, a quasi-continuum of states that is referred to as an energy band. In a tetravalent system, such as Si, there are actually four valence bands, corresponding to the four molecular bonding orbitals for each Si atom, and another four "conduction" bands for the corresponding antibonding orbitals. At low temperature, the valence bands are all fully occupied, and the conduction bands are empty. The properties of these bands and splitting between them are determined largely by the type of atoms, the crystal structure, and relevant momentum and angular momentum considerations.

The vast majority of semiconductors are tetravalent; however, trivalent, heptavalent, and hexavalent semiconductors also occur, with corresponding crystal structures. Regardless, the periodic nature of semiconductor crystals allows one to introduce the concepts of the unit cell and reciprocal lattice. The unit cell is defined as a volume of the lattice that is repeated within the crystal, completely filling the space with no overlap, and usually represents the full symmetry of the lattice. In this way, the unit cell, together with the associated lattice vectors

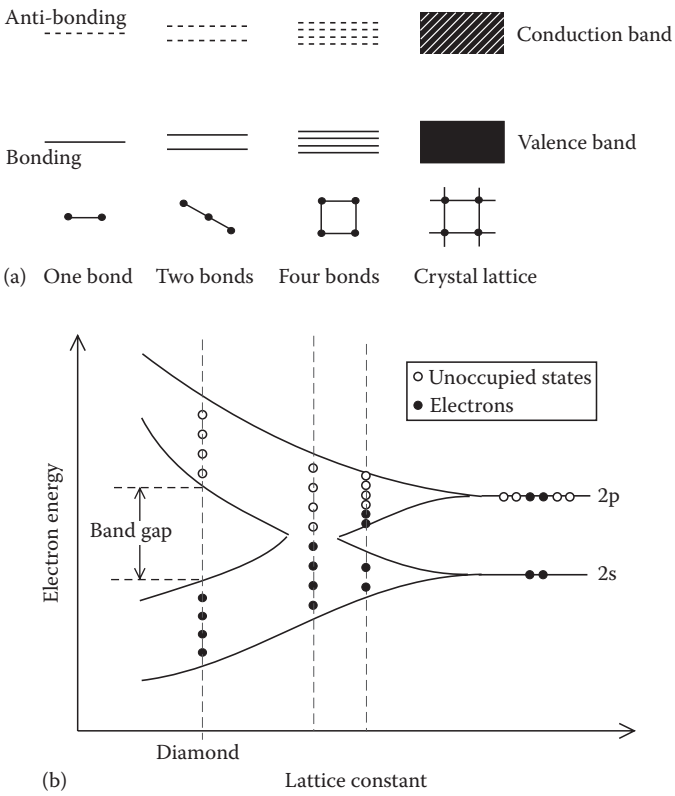


Figure 2.1 (a) As the number of bonds increases, so does the number of distinct bonding and antibonding states. In a crystal lattice, these distinct states form a quasi-continuum of states, leading to the formation of energy bands. Alternatively in (b), as the separation between carbon atoms decreases, and interactions between their electrons increase, the electron energies transition from discrete atomic states to bands separated by a band gap. (Adapted from Allison, J., Electronic Engineering Materials and Devices, McGraw Hill, London, 1971. With permission.)

(i.e., the vectors that define the unit cell) can be used to describe the entire crystal lattice. The choice of unit cell is not unique, and different unit cells may be defined for different purposes. In addition to the conventional unit cell described earlier, one particularly useful and uniquely defined unit cell is the Wigner–Seitz unit cell (Ashcroft and Mermin 1976). The advantage of this representation is that it defines a primitive unit cell (i.e., it contains exactly one repeating unit per cell) and represents the full symmetry of the crystal. The Wigner–Seitz unit cell is shown together with the conventional unit cell for a face-centered cubic lattice in Figure 2.2.

The reciprocal lattice is defined by the set of wave vectors, $\mathbf{k}$, which satisfy $e^{-i\mathbf{k}\cdot\mathbf{R}} = 1$ (defining a set of plane waves that represent the symmetry of the direct lattice), where $\mathbf{R}$ are the lattice vectors of the real space Bravais lattice.* Just as we defined a primitive unit cell in the real lattice, an equivalent primitive cell of the reciprocal lattice can be defined. One particular formation of such a primitive cell is referred to as the first Brillouin zone and is determined in the same way as the Wigner–Seitz primitive cell in the real lattice. More details on these fundamental concepts can be found in any solid state textbook, for example, Ashcroft and Mermin (1976) and Kittel (2005).

* A Bravais lattice is defined as an infinite array of discrete points with an arrangement and orientation that appear exactly the same regardless of which point an array is viewed from. Where this is not the case, a basis that contains multiple atoms can usually be defined, which can then be represented as a Bravais lattice.

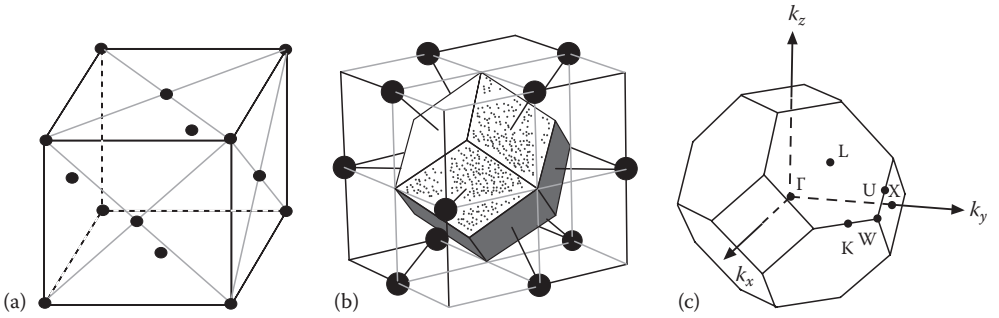


Figure 2.2 (a) The conventional unit cell for a face-centered cubic lattice (f.c.c.). (b) The corresponding Wigner–Seitz unit cell (note the cubic lattice depicted here is different from that in (a)). (c) The first Brillouin zone of an f.c.c. crystal, with the points of high symmetry labeled.

The unique definition of the first Brillouin zone allows several points of high symmetry to be defined, as depicted in Figure 2.2c for a face-centered cubic lattice. The most obvious is at the zone center and is given the label Γ; others are found at the edges of the Brillouin zone at different vertices and plane centers. To explain the purpose of introducing the reciprocal lattice and these high symmetry points, consider again an electron in the crystal lattice. The energy bands described earlier assume stationary electrons; however, one of the key properties of semiconductors (and metals) is that the electrons are not always stationary, and so the kinetic energy (KE) needs to be included. In a crystal, however, this is not simply KE = $\frac{1}{2}mv^2$, with m the mass and v the velocity, as the electrons interact with the nuclei (leading to a change in their effective mass), and the extent of this interaction is dependent on the electron's speed and direction of motion. In order to fully describe the energy of an electron, it then becomes necessary to define the kinetic energy for motion in different directions. This is where the reciprocal lattice and key symmetry points become useful.

The defined points of high symmetry in the reciprocal lattice correspond to wave vectors that define the momentum of electrons in the real crystal lattice. Plotting the energy of the electrons for these points in the reciprocal lattice, and along the lines joining these points, thereby provides a good description of the total energy landscape.

This type of representation is known as a dispersion relation, examples of which are shown in Figure 2.3. The horizontal axis in these figures is labeled by the high symmetry points in the reciprocal lattice. Between these points the position is varied linearly along the line connecting them in the reciprocal lattice. In real space, this can be thought of as changing either the direction or magnitude of the electron wave vector/momentum. The Γ-point is located at the center of the Brillouin zone and corresponds to zero momentum. The regions between the Γ-point and any other point, or between two points on the surface of the Brillouin zone, therefore correspond to increasing the magnitude of the wave vector/momentum in the direction defined by the other high symmetry point, or changing the direction of the wave vector/momentum, respectively.

The complete band structure of semiconductors is usually defined in this way, and many of their properties are described with respect to this representation. For example, the effective mass of electrons in the crystal lattice is determined by the curvature of the relevant band, where a steeper dispersion curve corresponds to a smaller effective mass. Conversely, if the band structure is not known, the effective mass can be determined experimentally, as described in Section 2.2.6, and this helps to define the band structure.

Conduction, thermal, and optical properties are also described within the band structure formalism, and similarly can also be used to experimentally determine aspects of the band

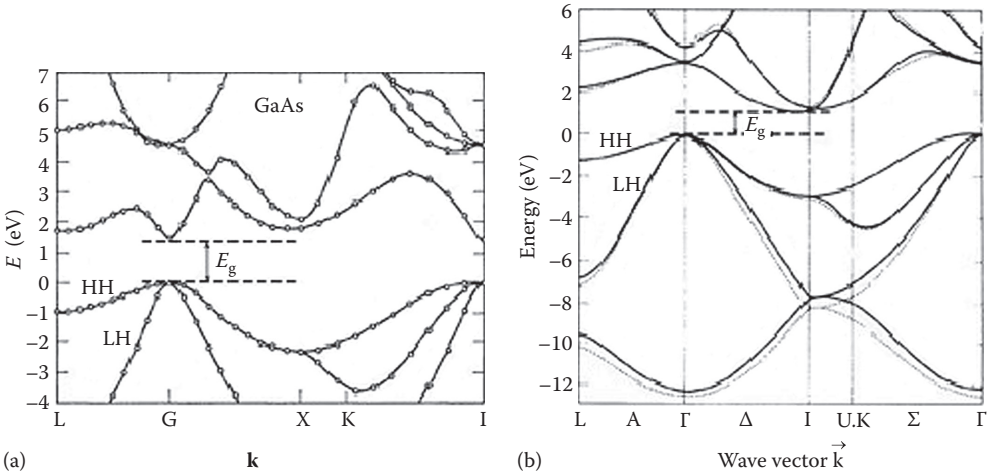


Figure 2.3 (a) The band structure of GaAs (Reprinted figure with permission from Cohen, M.L. and Bergstresser, T.K., *Phys. Rev.*, 141, 789, 1966. Copyright (1966) by the American Physical Society.) and (b) the band structure of Si (Reprinted figure with permission from Chelikowsky, J.R. and Cohen, M.L., *Phys. Rev. B*, 10, 5095, 1974. Copyright (1966) by the American Physical Society.) The horizontal axis represents the magnitude and direction of the wave vector by placing symmetry labels that correspond to specific points of high symmetry in the reciprocal lattice, and in this way identifies the orientation of the wave vector, while between these labels the magnitude is varied.

structure. The band structure can also be calculated by solving the Schrödinger equation for the crystal lattice, which is often achieved practically by making use of the $\mathbf{k}\cdot\mathbf{p}$ formalism. For details of this type of calculation, we refer the reader to Haug and Koch (2004), for example.

Perhaps, the single most important property of semiconductors is the band gap, so it is worth taking the time to detail exactly what is meant by this term. In the dispersion relations in Figure 2.3, it can be seen that the energy gap between the valence band and conduction band varies as a function of wave vector, and so it is impossible to give a single value. "The" band gap of a material is defined as the energy difference between the highest maximum of the highest energy valence band and the lowest minimum of the lowest energy conduction band. In the case of direct gap semiconductors, as shown in Figure 2.3a for GaAs, this is straightforward, as they both occur at the same point in **k**-space (i.e., at the center of the Brillouin zone, $\mathbf{k} = 0$). In some materials, however, the valence band maximum and conduction band minimum can be displaced from each other in momentum space, as shown in Figure 2.3b for silicon. In this case, the band gap is still defined as the energy difference between the valence band maximum and conduction band minimum, and is referred to as an indirect gap.

2.1.2 Conduction Properties

At low temperature and in pure, intrinsic, semiconductors, the valence band is fully occupied and the conduction band is empty. In other words, all electrons are bound to their nuclei, and in order to move to another atomic site the energy of the band gap would need to be overcome. This is very unlikely at low temperatures, meaning there are no free electrons and the material is insulating. If, however, an electron can in some way be excited to the conduction band, there is a continuum of empty states at the same or similar energies, so no excess energy is required to move throughout the crystal lattice, allowing it to conduct electricity.

In addition, the vacancy created in the valence band can be filled by another valence band electron, thereby moving the vacancy to another site, where another electron can move into it, and so on, thereby allowing the free movement of charge. Rather than thinking of this in terms of many electrons moving one at a time in sequence, it is helpful to think of the vacancy moving. In this way, by exciting the electron from the valence band to the conduction band, a quasiparticle in the valence band, referred to as a hole, is generated. The hole then has the opposite properties to an electron, including a positive charge, and it can be treated as a charge carrier that conducts electricity.

In thermal equilibrium, the probability of an electron being excited from the valence band into the conduction band is given by Fermi–Dirac statistics. This introduces the concept of the chemical potential or Fermi level. In semiconductors, the definition of the Fermi level is somewhat different from metals, since it usually lies in the band gap. The Fermi level, or chemical potential, defines the energy at which the probability for occupation is 0.5. At absolute zero, where the valence (conduction) band is completely full (empty), the Fermi level is in the middle of the band gap. At finite temperatures, the Fermi level lies within the order of $k_B T$ (where k_B is the Boltzmann constant and T is the temperature) of the center of the band gap. For intrinsic semiconductors where the band gap is large compared to $k_B T$, the Fermi level remains well within the band gap, and the probability of finding conduction electrons remains low.

In order to promote electrons from the valence band to the conduction band, energy greater than or equal to the band gap needs to be absorbed. For semiconductors, the band gap can range from a few hundred meV (e.g., InSb) up to ~6 eV (e.g., AlN) and often corresponds to optical photon energies (1–4 eV), meaning the absorption of visible light can lead to a transient conductivity. This is the basis for many light-detection and light-harvesting devices. For many applications, however, it is also desirable to have permanent conductivity. This can be achieved by doping the crystal lattice with impurities that have more (or fewer) valence electrons than the host semiconductor. For example, if an atom with five valence electrons replaces an atom with four valence electrons (e.g., P in Si), then four of the electrons will simply take the role of those in the original atom, leaving one extra electron. The energy level occupied by this electron is typically located just below the conduction band, as illustrated in Figure 2.4a. It can then be easily promoted thermally into the conduction band, allowing free movement of the charge. This is known as n-type doping and the dopant atoms are known as donors. In other words, the Fermi level is raised into or close to the conduction band, allowing free conduction. On the other hand, if an atom with three valence electrons is included in the lattice (e.g., B in Si), then an empty state is generated just above the valence band, into which an electron from the valence band can be excited, leaving a hole in the valence band that is

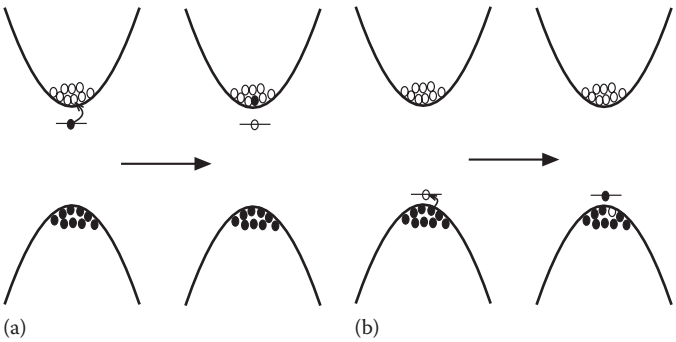


Figure 2.4 Schematic showing the incorporation of (a) n-type and (b) p-type dopants, and the excitation that allows them to conduct electrons and holes, respectively.

free to move and conduct electrical current, as shown in Figure 2.4b. This is known as p-type doping, with the dopant atoms known as acceptors, and corresponds to lowering the Fermi level close to the valence band. Most consumer electronics are made up of combinations of intrinsic (pure) and extrinsic (doped) semiconductor regions, which make use of both p- and n-type doping and the interface between such regions. For more information on p–n junctions and semiconductor devices, we refer the reader to Ashcroft and Mermin (1976), Kittel (2005), Sze and Kwok (2007), and Seeger (2004).

2.1.3 Compound Semiconductors

To date, we have been assuming that semiconductors are pure homogeneous materials made up of one single type of atom (with the exception of impurities). This need not be the case, however, as it is possible to form ionic crystals that are also semiconducting. These types of semiconductors are typically made by combining group III atoms with group V atoms (e.g., GaAs), or group II with group VI atoms (e.g., ZnO). In this way, the number of valence electrons remains the same as if all atoms were from group IV, and the bonding remains tetrahedral. The fact that the nuclei are no longer identical does not affect most properties, as it is the crystal structure and average number of electrons per site that determine such properties, and these are unchanged.

One example of a situation in which the local environment and the heterogeneous nature of the unit cell can play an important role is at boundaries and interfaces. Since the unit cell in compound semiconductors contains at least two different types of atoms, it is likely that in at least one direction, taken in isolation, it will be electrically polarized, as the electron affinity of the two atoms will almost certainly not be the same. This spontaneous polarization is not seen in an infinite lattice, as the fields from neighboring unit cells counter each other. At a surface, however, there is no neighboring unit cell, and an excess of charge can exist, leading to an electric field. This type of charge imbalance can also be found at an interface between two different materials, where the spontaneous polarization of the two unit cells is different. This can be particularly relevant in semiconductor nanostructures.

The major reason compound semiconductors are of particular interest is the flexibility they allow. Take GaAs, for example; by replacing a fraction, x, of the Ga atoms with Al atoms (another group III element) to form $Al_xGa_{1-x}As$, the band gap can be increased by an amount that is dependent on the Al concentration. Alternatively, the band gap can be decreased by replacing Ga with In to form $In_xGa_{1-x}As$. This ability to tune the band gap introduces an extra degree of freedom that is particularly useful when designing semiconductor devices and nanostructures, especially those that emit or absorb light.

2.1.4 Phonons

Much like any other crystalline solid, vibrations within semiconductors are quantized; in classical mechanics, this is similar to the normal modes of vibration. When a bulk system is treated quantum mechanically, however, these "normal modes" have particle-like properties and can be treated as quasiparticles* known as phonons. It then becomes possible to talk about the number or density of phonons, which increases with increasing temperature. For

* A quasiparticle is an elementary excitation that exists only in a solid (e.g., phonons, excitons, holes) with particle-like properties like position, momentum, and mass that are modified by interactions with its environment. This contrasts with "real" particles like electrons, protons, and photons that can exist in vacuum (Prasankumar et al. 2009).

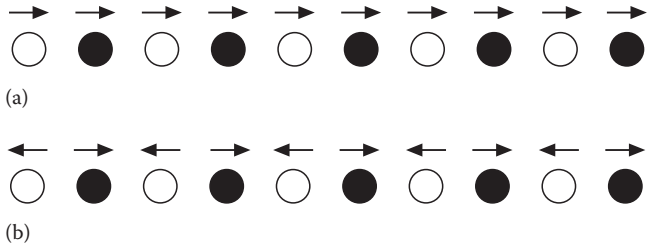


Figure 2.5 (a) Longitudinal acoustic and (b) longitudinal optical phonon modes for a 1D lattice. For acoustic modes the atoms in the unit cell move in phase, while for optical modes they move out of phase.

details regarding the derivation of phonons, see, for example, Ashcroft and Mermin (1976, pp. 437–443) and Ridley (1997).

As an example, consider the simplest case of a one-dimensional (1D) lattice. In this case, there are two general types of vibrations possible, as depicted in Figure 2.5. In the first, (a), the two different atoms that make up the unit cell move in phase, while in the second, (b), they move out of phase. The former case represents acoustic phonons while the latter corresponds to optical phonons, so called because in an ionic crystal this oscillating separation is in fact an oscillating dipole, which makes this type of phonon optically active. Extending this picture to three dimensions is not so straightforward conceptually, but the basic principles remain the same: Acoustic phonons correspond to intercellular interactions (i.e., nuclei within the unit cell move in phase), while optical phonons correspond to intracellular interactions (i.e., nuclei within the unit cell move out of phase) (Ashcroft and Mermin 1976).

One further consideration is the orientation of the vibrational motion relative to the phonon propagation direction. Consistent with the quasiparticle picture, phonons do not exist across the entire lattice simultaneously, but rather are localized, and move through the lattice with a given velocity. This defines the propagation direction and allows the definition of transverse and longitudinal modes. In general, one can define three orthogonal orientations for the vibrational motion: one that is parallel to the propagation direction (longitudinal) (Figure 2.5), and two that are orthogonal to the propagation direction (transverse) (Figure 2.6). This definition of transverse and longitudinal is not always going to be strictly true, as it depends on the direction of propagation relative to the principal axes of the crystal lattice; nonetheless, this terminology remains useful and continues to be used.

A corollary of phonon propagation is that they transfer momentum through the crystal lattice. The momentum of a phonon is manifested in the vibrational motion of the nuclei, and is transferred throughout the lattice as the phonon propagates. The phonon's momentum is one

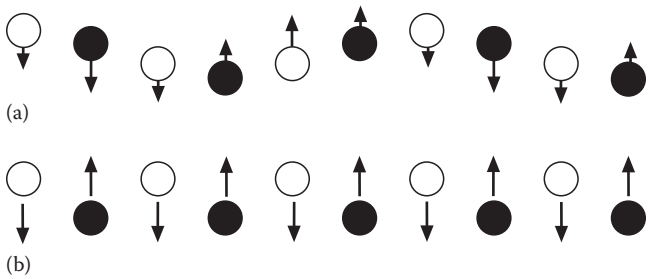


Figure 2.6 (a) Transverse acoustic and (b) transverse optical phonon modes for a 1D lattice. Again, for acoustic modes the atoms in the unit cell move in phase, while for optical modes they move out of phase.

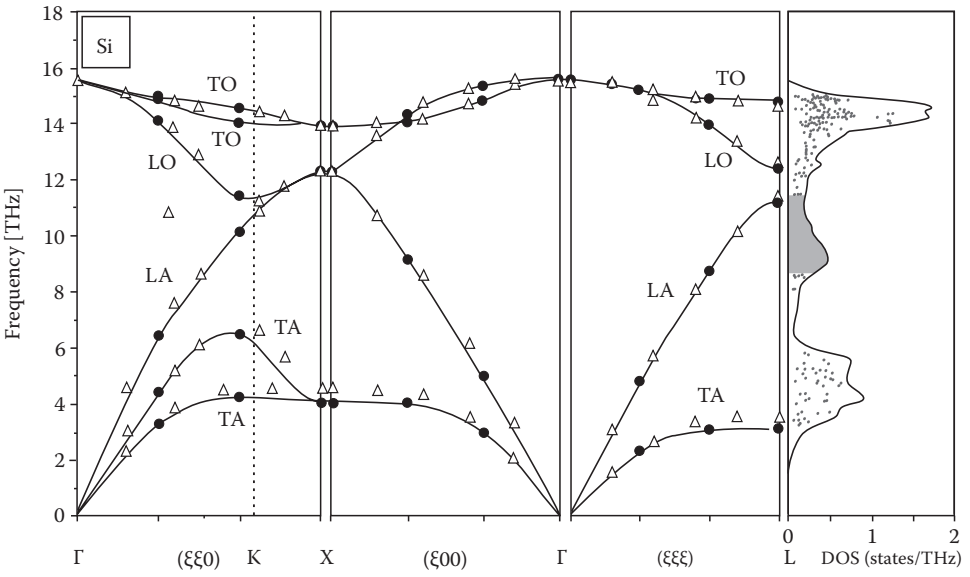


Figure 2.7 Example of phonon dispersion curve and density of states for Si. (Reprinted figure with permission from Savrasov, B., *Phys. Rev. B*, 54, 16470, 1996. Copyright (1966) by the American Physical Society.)

of its key properties, and similar to electrons moving throughout the crystal lattice, phonons propagating in different directions will have different energies depending on the magnitude and direction of their wave vector with respect to the principal axes of the lattice. For this reason, it is useful to introduce phonon dispersion curves that plot phonon energy as a function of momentum (or wave vector), an example of which is shown in Figure 2.7 (Savrasov 1996). It is evident in this figure that the dispersion of optical phonons is rather flat, whereas the energy of the acoustic phonons is strongly dependent on momentum and zero at the zone center (i.e., $\mathbf{k} = 0$). This is typical for most crystals, as is the observation that for a given wave vector, the energy of optical phonons is generally greater than for acoustic phonons, particularly close to the zone center. This means that at low temperatures it is possible to "freeze out" the optical phonons, allowing the effects of acoustic phonons to be isolated. Optical phonon energies are typically in the range of a few tens of meV, which is comparable to the thermal energy at room temperature (25 meV), thus as the temperature is reduced, the prevalence of optical phonons rapidly decreases. At even lower temperatures, it becomes possible to minimize the acoustic phonon density, but because the linear dispersion curve goes to zero at the Γ point, there can always be a finite acoustic phonon population at finite temperatures; under certain conditions, however, zero-phonon effects can still be observed (Borri et al. 2001).

A corollary of treating phonons as quasiparticles is that they are able to interact and scatter with other particles and quasiparticles. For example, phonons can scatter with electrons, holes, and photons in either elastic or inelastic collisions, and can transfer momentum. The effects of these interactions may include alteration of the electrical conductivity, non-radiative recombination of optically excited carriers, or dephasing of a coherently excited system. Changes to these properties as a function of temperature are therefore usually attributed to phonon interactions, as the phonon density (corresponding to the number of lattice vibrations) increases with temperature (Shah 1999).

Interactions of phonons with photons can be divided into two different cases: infrared (IR) absorption and Raman scattering. In order for an IR photon to be absorbed by a crystal as a

phonon, both energy and momentum must be conserved. From the phonon dispersion curve, it is clear that the only place where a photon will have the same energy and momentum as an acoustic phonon is at $\mathbf{k} = 0$, corresponding to a static electric field. The photon dispersion will however overlap with the optical phonon branches at finite values of E and $\mathbf{k}$. In order for the electromagnetic field of the photon to interact with the phonon modes, the crystal must have a permanent dipole moment, that is, the bonds must be at least a little polar, as will be the case in compound semiconductors, but not monotonic semiconductors such as Si. Furthermore, since the electric field of the photons is transverse, it will interact only with the TO phonon modes. Phonons that satisfy these selection rules are said to be IR active (Klingshirn 2007).

Raman scattering and Brillouin scattering (which is equivalent to Raman scattering, but with acoustic phonons) are discussed in greater detail later in this chapter (and in Chapters 6 and 14). At this point, however, we point out that all phonon modes can potentially be Raman active depending on the crystal structure. The precise selection rules to determine whether a phonon mode is Raman active is somewhat more complicated, requiring group theory and a more detailed analysis of the symmetry of the crystal and corresponding phonon modes. In general, however, if the crystal structure has an inversion center, then the phonon modes will be either IR or Raman active; if the crystal lacks such inversion symmetry, then some phonon modes may be both IR and Raman active. Fox (2001) gives a more detailed summary of the optical properties of phonons.

2.1.5 Other Lattice Excitations

In addition to phonons, there are other forms of lattice interactions that can play an important role in semiconductor physics. We briefly mention two of these here: polarons and magnons.

Polarons are formed by the interaction of the electric field of an excited electron (or hole) with that of the nuclei. This interaction causes a deformation of the lattice in the vicinity of the charge carrier, which leads to a local polarization of the lattice. As the free charge carrier moves throughout the lattice, the deformation and polarization move with it. This combination of the charge carrier and its polarization field is treated as a quasiparticle known as a polaron. One important effect of the local induced field is that it hinders the motion of the electron through the lattice, thus reducing its mobility (Klingshirn 2007). This often leads to a "hopping"-type conductivity when polarons are the dominant charge carriers.

Magnons are quasiparticles that represent quantized spin waves and are completely analogous to phonons. Where phonons represent excitations of the nuclear lattice and carry momentum, magnons represent excitations of the spin lattice and carry angular momentum. Similarly, just as phonons can have an important role in energy loss mechanisms, so too can magnons in magnetic semiconductors (Klingshirn 2007).

2.2 OPTICAL PROPERTIES OF SEMICONDUCTORS

The optical response of semiconductors is one of their most useful characteristics and one of the most important tools for determining their band structure. In discussing the optical response, the most important property is the size of the band gap and whether it is direct or indirect. The presence of a band gap between the full valence band and empty conduction band means that a photon with energy greater than the band gap can be readily absorbed, whereas photons with energy less than the band gap will typically be transmitted. For many semiconductors of interest, the band gap is around the optical region of the electromagnetic spectrum, and so it is not surprising that semiconductors now form the basis for many light emitting and detecting devices.

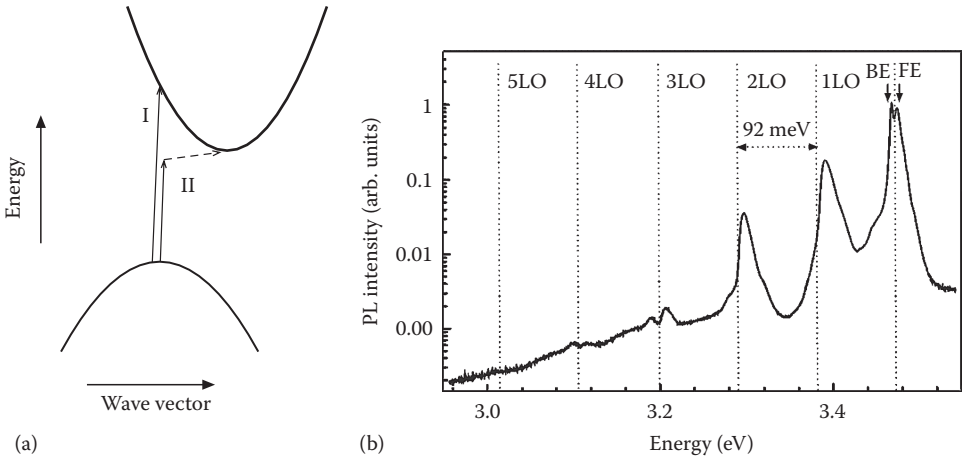


Figure 2.8 (a) The direct transition (I) in an indirect gap material requires far greater photon energy than the phonon-assisted transition (II) between the band edges. In the phonon-assisted transition, the phonon (dashed line) provides the additional momentum to allow the transition. (b) Similar processes can also occur in direct gap materials such as GaN and lead to the formation of phonon replica peaks labeled by the dashed lines and the labels indicating the number of phonons involved in the transition. (Reprinted figure with permission from Kovalev, D. et al., *Phys. Rev. B*, 54, 2518, 1996. Copyright (1966) by the American Physical Society.)

Knowledge only of the band gap, however, is not enough to fully describe the optical properties. In addition to energy conservation, which can be specified with respect to the band gap, momentum and angular momentum must also be conserved.

The momentum carried by a photon ($p = \hbar\mathbf{k} = h/\lambda$) is small compared to the momentum of electrons/holes, and so optical transitions where the change in momentum (or electron/hole wave vector) is close to zero tend to dominate. This is particularly the case for direct gap semiconductors. For indirect gap semiconductors, a valence to conduction band transition with minimum photon energy can occur only if a phonon is absorbed or emitted simultaneously to facilitate conservation of momentum. This type of phonon-assisted transition tends to dominate for indirect gap materials, such as silicon, causing the optical properties to become increasingly temperature dependent. This type of phonon-assisted transition is depicted in Figure 2.8a, where path I represents the direct transition, which requires a photon with energy greater than the band gap, and path II represents the phonon-assisted transition. In this case, a photon and phonon are absorbed simultaneously to allow the indirect transition between the band edges, conserving both energy and momentum.

Phonon-assisted transitions can also occur in direct gap semiconductors and appear as additional "phonon-replica" peaks in the absorption/emission spectrum, separated from the "zero-phonon" peak by multiples of the phonon energy. The presence of phonon replicas is significantly enhanced for strongly polar crystals due to the strong interaction of the TO phonons with incident photons (Fox 2001). An emission spectrum for GaN, a very polar material, is shown in Figure 2.8b, with five phonon replica peaks present.

Another important consideration in determining the optical properties is that photons always carry exactly one unit of angular momentum, and so only transitions that result in such a change of angular momentum are allowed. In order to assess the allowed transitions, we again refer to the band structure. Figure 2.9 shows the three highest valence bands, corresponding to the atomic p-orbitals, with orbital angular momentum $l = 1$. The sixfold

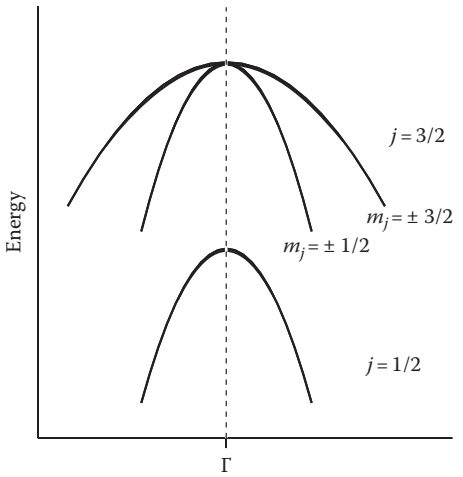


Figure 2.9 The three highest valence bands are shown. The SO band, with $j = 1/2$, is split from the $j = 3/2$ bands due to the spin–orbit interaction. The two $j = 3/2$ bands are also split except at the point of highest symmetry.

degenerate p-like valence band is split by the spin–orbit interaction* into a fourfold degenerate band with total angular momentum $j = 3/2$ and doubly degenerate band with $j = 1/2$. The lowest band, with $j = 1/2$, is referred to as the spin–orbit split-off (SO) band and in most cases does not significantly contribute to the optical properties. At points of high symmetry, the $j = 3/2$ band is fourfold degenerate; however, away from the zone center (i.e., the Γ-point) a splitting occurs, as can be seen in Figure 2.9. The bands are split according to the magnitude of the z-projection of the angular momentum, or magnetic quantum number, m_j, so that the states with $m_j = \pm 1/2$ form one band and those with $m_j = \pm 3/2$ form the other. The dispersion relation of a free particle is parabolic, since $E = p^2/2m$, and the curvature is related to the particle mass. The different curvature of the two $j = 3/2$ bands is interpreted as the effective mass of the holes that are generated when electrons are excited. The band with $m_j = \pm 1/2$, which has greater curvature, is therefore termed the light-hole (LH) band, and that with $m_j = \pm 3/2$, the heavy-hole (HH) band. The physical reason for the different curvatures is that HHs, with a greater magnetic quantum number, and hence greater magnetic moment, interact more strongly with the lattice, giving the appearance of being "heavier" and therefore leading to a shallower dispersion relation.

The angular momentum of the conduction band is determined entirely by the electron spin, as the orbital angular momentum, l, equals zero for the lowest conduction band. The total angular momentum is, therefore, $j = 1/2$ and the band is doubly degenerate with $m_j = \pm 1/2$. If we represent each of these energy bands as discrete states, as shown in Figure 2.10, the allowed transitions can be determined by the requirement for a change of one unit of angular momentum in absorbing a photon. The states involved in these transitions are all pure spin states and require a change in m_j of ±1, which means that these pure transitions are circularly polarized, with σ^+ polarized photons adding 1 to m_j and σ^- subtracting 1 from m_j. Excitation with linearly polarized light would, in contrast, excite a linear superposition of the

* The spin and orbital angular momentum of charged particles generate finite magnetic moments that interact with each other, leading to a splitting of otherwise degenerate states. This interaction is referred to as the spin–orbit interaction.