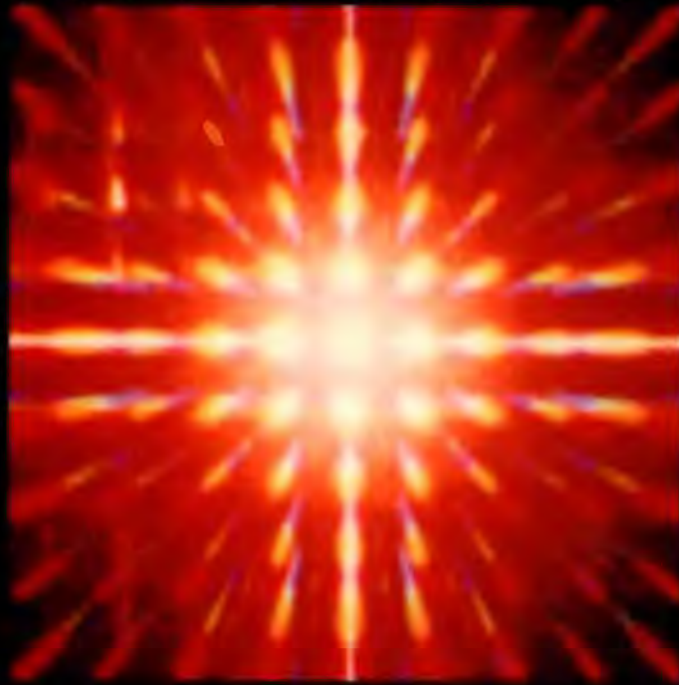


International Edition

OPTICS

FOURTH EDITION

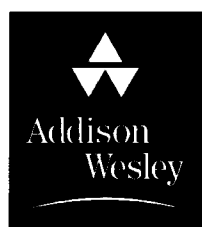


Eugene Hecht

4^{ed} OPTICS

Eugene Hecht

Adelphi University



San Francisco Boston New York
Capetown Hong Kong London Madrid Mexico City
Montreal Munich Paris Singapore Sydney Tokyo Toronto



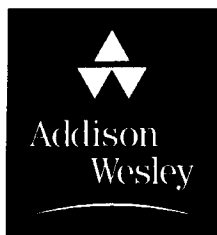
Sponsoring Editor: *Adam Black*
Project Manager: *Nancy Gee*
Manufacturing Supervisor: *Vivian McDougal*
Cover Designer: *Blakeley Kim*
Production Service: **HRS Interactive**
Text Design: **HRS Interactive**
Composition: **HRS Interactive**
Illustration: **HRS Interactive**
Photo Research: *Carolyn Eisen Hecht*

ISBN 0-321-18878-0

If you purchased this book within the United States or Canada you should be aware that it has been wrongfully imported without the approval of the Publisher or the Author.

Copyright ©2002 Pearson Education, Inc., publishing as Addison Wesley, 1301 Sansome St., San Francisco, CA 94111. All rights reserved. Manufactured in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. To obtain permission(s) to use material from this work, please submit a written request to Pearson Education, Inc., Permissions Department, 1900 E. Lake Ave., Glenview, IL 60025. For information regarding permissions, call 847/486/2635.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and the publisher was aware of a trademark claim, the designations have been printed in initial caps or all caps.



1 2 3 4 5 6 7 8 9 10-MA-04 03 02 01
www.aw.com/physics

Preface

To *Ca, b. w. l.*

As with previous revisions, this fourth edition has been guided by three distinct imperatives: to fine-tune the pedagogy, modernize the discourse, and update the content.

During the past several years, using the third edition in the classroom, a number of small but significant pedagogical refinements have evolved, and these have been incorporated throughout this revised text. For instance, all symbols representing vectors now have an arrow directly above them. The intent, as ever, is to be responsive to students' needs.

This edition continues the program of gradually modernizing the treatment. In this, there are several goals, vis-à-vis the student: to impart an appreciation of the central role of atomic scattering in almost every aspect of Optics; to provide an understanding, as early as possible, of the insightful perspective offered by Fourier Theory; and to make clear, from the outset, the underlying quantum-mechanical nature of light.

Technological advances have been made in a wide range of areas from lenses and lasers to telescopes and fiber optics. This fourth edition treats (even if sometimes only in introductory fashion) many of the significant advances that today's students should be aware of.

At the request of users, I have added dozens of new problems throughout the text. Most of these were designed to develop needed analytic skills and are of the "easy-to-intermediate" variety. As in previous editions, the complete solutions to problems without asterisks can be found at the back of the book. It should be noted that the vast majority of the new problems are not provided with such solutions. This was done to increase the selection of potential homework questions.

Much effort was expended in redrawing well over a hundred pieces of existing art to make them clearer and the

physics more readable. To further enhance the pedagogy, this edition also contains many new drawings and photographs.

Chapter 2 (*Wave Motion*), which lays the mathematical groundwork for wave theory, has been somewhat revised in order both to make the existing material clearer and to provide a broader foundation for what's to come. For instance, the discussion of the differential wave equation (p. 13) has been fleshed out a bit, with some of the previously missing steps now provided.

Chapter 3 (*Electromagnetic Theory, Photons, and Light*) has been extensively reworked. Nowadays many students studying Optics come to it with little more than the brief exposure to electromagnetic theory afforded by the traditional introductory course in physics. With them in mind, Section 3.1, *Basic Laws of Electromagnetic Theory* (p. 37), has been expanded, making the material far more accessible to these students. In keeping with the commitment to discuss modern applications of Optics, a new section called *Optical Cooling* examines this important technique (p. 65).

The Propagation of Light (Chapter 4) now contains a discussion of the historical origins of the concept of *index of refraction* (p. 88). This helps to make Snell's Law easier to really understand. The chapter is further enhanced with a more thorough treatment of *Fermat and Mirages* (p. 107). A new section, 4.11.2, called *Photons and the Laws of Reflection and Refraction* (p. 141) completes the chapter.

The treatment of *Geometrical Optics* (Chapter 5) was refined here and there (e.g., pp. 161, 164, 166, 178, and 215) to improve its clarity. The field of telecommunications is so important and so rapidly evolving that each edition of this book has had to treat several major technological advances. Accordingly, the discussion of fiber optics was brought up to date with the consideration of such topics as erbium-doped

fiber amplifiers (p. 197), wavelength division multiplexing, and optical switching via MOEMS (p. 200). Liquid mirrors are briefly considered on page 223. The chapter ends with a new section on *Gravitational Lensing* (p. 231).

The discussion of wavefront aberrations in Chapter 6 (*More on Geometrical Optics*) has been enlarged (p. 254). There is a new piece concerning the upgraded Arecibo Observatory (p. 258) because it beautifully illustrates an important contemporary approach to dealing with spherical aberration.

Chapter 7 (*The Superposition of Waves*) was reworked to make the material generally more accessible (e.g., p. 303). The phasor representation was used to illuminate the creation of both standing waves (p. 288) and partial standing waves (p. 289). Because of a very significant series of experiments published in the last several years, the discussion of *Group Velocity* (p. 296), Section 7.2.2, was enriched and new subsections on *Superluminal* and *Subluminal Light* were added.

In Chapter 8 (*Polarization*), as elsewhere, the prose was tightened and the analysis clarified, here and there. A few new photographs and several fresh diagrams were included. The section (8.7) on *Retarders* was extended, and the concepts of zero-ordered, multiple-ordered, and compound zero-ordered wave plates were introduced. A section on *Liquid Crystals* (8.12) was added, and the operation of both the liquid crystal variable retarder and liquid crystal display (LCD) were explained.

In addition to a few new photos and the occasional clarifying remark, Chapter 9 (*Interference*) now contains a section, 9.8.4, called *Radar Interferometry*.

Over the last two decades there's been some interesting work done on so-called nondiffracting beams. Accordingly, Chapter 10 (*Diffraction*) contains a new section (10.2.7) entitled *The Zeroth-Order Bessel Beam*, that deals with the phenomenon.

Chapters 11 (*Fourier Optics*), and 12 (*Basics of Coherence Theory*) have undergone a line-by-line fine tuning, but little or no overhaul.

Chapter 13, *Modern Optics: Lasers and Other Topics*, has been revised with the addition of a subsection on *Gaussian Laserbeams* and some updating as required (e.g., it now includes material on the Omega laser).

This fourth edition continues the agenda of unifying the discourse, as much as possible, within the framework of a few

grand ideas. Thus the concept of interference, which is one of the premier notions in Optics (and not surprisingly in Quantum Mechanics, as well), is used qualitatively to understand propagation phenomena long before it's studied formally in Chapter 9. Among other benefits, this approach of presenting advanced concepts in simplified form early in the exposition allows the student to develop a cohesive perspective.

Over the years, I have received comments, articles, and photographs, from hundreds of colleagues, and I most sincerely thank them all. I am especially grateful to Professors P. J. Dolan of Northeastern Illinois University, W. A. Mendoza of Jacksonville University, M. W. Coffey of the University of Colorado and H. Fearn of California State University for their contributions and suggestions. Prof. J. R. Peverley of The George Washington University kindly donated several very nice problems on Jones matrices and I thank him for helping to freshen up this edition. Anyone else wishing to contribute their favorite problems, please feel free to do so. Indeed, if you are interested in the discipline and wish to exchange ideas you can contact me by mail at Adelphi University, Physics Department, Garden City, N.Y. 11530 or at genehecht@aol.com.

I'd like to thank the entire team at Addison Wesley, for all their help, without which this edition would never have seen the light of day, as it were. I am especially grateful to Adam Black whose enthusiasm for the project was sustaining, to Joan Marsh whose wise decisions made the whole thing manageable, and to Nancy Gee who handled the day-to-day operation with efficiency and good humor.

The book was produced by HRS Interactive, which did a brilliant job of getting it all together. Lorraine Burke watched over every aspect of the process with incredible patience and skill; Alan Wiener and Jennifer Burke cheerfully brought their production acumen to bear; Ed Burke designed a beautiful book and struggled mightily to maintain the highest standards; Hilda Espreo was the tireless compositor; and as ever, Pat Hannagan, with the able assistance of Chris Burke, produced incomparable art. All have my deepest respect and profound appreciation. Finally, I thank my dear friend, my wife, Carolyn Eisen Hecht for coping with one more edition of one more book.

Freeport, New York

E.H.

Content

1 A Brief History 1

- 1.1 Prolegomenon 1
- 1.2 In the Beginning 1
- 1.3 From the Seventeenth Century 2
- 1.4 The Nineteenth Century 4
- 1.5 Twentieth-Century Optics 7

2 Wave Motion 10

- 2.1 One-Dimensional Waves 10
- 2.2 Harmonic Waves 14
- 2.3 Phase and Phase Velocity 17
- 2.4 The Superposition Principle 20
- 2.5 The Complex Representation 21
- 2.6 Phasors and the Addition of Waves 23
- 2.7 Plane Waves 24
- 2.8 The Three-Dimensional Differential Wave Equation 27
- 2.9 Spherical Waves 28
- 2.10 Cylindrical Waves 31
- Problems 32

3 Electromagnetic Theory, Photons, and Light 36

- 3.1 Basic Laws of Electromagnetic Theory 37
- 3.2 Electromagnetic Waves 44
- 3.3 Energy and Momentum 47
- 3.4 Radiation 58
- 3.5 Light in Bulk Matter 66
- 3.6 The Electromagnetic-Photon Spectrum 73
- 3.7 Quantum Field Theory 80
- Problems 82

4 The Propagation of Light 86

- 4.1 Introduction 86

- 4.2 Rayleigh Scattering 86

- 4.3 Reflection 95

- 4.4 Refraction 100

- 4.5 Fermat's Principle 106

- 4.6 The Electromagnetic Approach 111

- 4.7 Total Internal Reflection 122

- 4.8 Optical Properties of Metals 127

- 4.9 Familiar Aspects of the Interaction of Light and Matter 131

- 4.10 The Stokes Treatment of Reflection and Refraction 136

- 4.11 Photons, Waves, and Probability 137
- Problems 141

5 Geometrical Optics 149

- 5.1 Introductory Remarks 149

- 5.2 Lenses 150

- 5.3 Stops 171

- 5.4 Mirrors 175

- 5.5 Prisms 186

- 5.6 Fiberoptics 193

- 5.7 Optical Systems 201

- 5.8 Wavefront Shaping 226

- 5.9 Gravitational Lensing 231
- Problems 234

6 More on Geometrical Optics 243

- 6.1 Thick Lenses and Lens Systems 243

- 6.2 Analytical Ray Tracing 246

- 6.3 Aberrations 253

- 6.4 GRIN Systems 273

- 6.5 Concluding Remarks 276
- Problems 277

7 The Superposition of Waves 281

- 7.1** The Addition of Waves of the Same Frequency 282
- 7.2** The Addition of Waves of Different Frequency 294
- 7.3** Anharmonic Periodic Waves 302
- 7.4** Nonperiodic Waves 308
Problems 320

8 Polarization 325

- 8.1** The Nature of Polarized Light 325
- 8.2** Polarizers 331
- 8.3** Dichroism 333
- 8.4** Birefringence 336
- 8.5** Scattering and Polarization 344
- 8.6** Polarization by Reflection 348
- 8.7** Retardors 352
- 8.8** Circular Polarizers 357
- 8.9** Polarization of Polychromatic Light 358
- 8.10** Optical Activity 360
- 8.11** Induced Optical Effects—Optical Modulators 365
- 8.12** Liquid Crystals 370
- 8.13** A Mathematical Description of Polarization 372
Problems 379

9 Interference 385

- 9.1** General Considerations 386
- 9.2** Conditions for Interference 390
- 9.3** Wavefront-splitting Interferometers 393
- 9.4** Amplitude-splitting Interferometers 400
- 9.5** Types and Localization of Interference Fringes 414
- 9.6** Multiple-Beam Interference 416
- 9.7** Applications of Single and Multilayer Films 425
- 9.8** Applications of Interferometry 431
Problems 438

10 Diffraction 443

- 10.1** Preliminary Considerations 443
- 10.2** Fraunhofer Diffraction 452
- 10.3** Fresnel Diffraction 485
- 10.4** Kirchhoff's Scalar Diffraction Theory 510
- 10.5** Boundary Diffraction Waves 512
Problems 514

11 Fourier Optics 519

- 11.1** Introduction 519
- 11.2** Fourier Transforms 519
- 11.3** Optical Applications 529
Problems 556

12 Basics of Coherence Theory 560

- 12.1** Introduction 560
- 12.2** Visibility 562
- 12.3** The Mutual Coherence Function and the Degree of Coherence 566
- 12.4** Coherence and Stellar Interferometry 573
Problems 578

13 Modern Optics: Lasers and Other Topics 581

- 13.1** Lasers and Laserlight 581
- 13.2** Imagery — The Spatial Distribution of Optical Information 606
- 13.3** Holography 623
- 13.4** Nonlinear Optics 639
Problems 644

Appendix 1 649

Appendix 2 652

Table 1 653

Solutions to Selected Problems 658

Bibliography 685

Index 689



Electromagnetic Theory, Photons, and Light

The work of J. Clerk Maxwell and subsequent developments since the late 1800s have made it evident that light is most certainly electromagnetic in nature. Classical electrodynamics, as we shall see, unalterably leads to the picture of a continuous transfer of energy by way of electromagnetic waves. In contrast, the more modern view of Quantum Electrodynamics (p. 80) describes electromagnetic interactions and the transport of energy in terms of massless elementary “particles” known as *photons*. The quantum nature of radiant energy is not always apparent, nor is it always of practical concern in Optics. There is a range of situations in which the detecting equipment is such that it is impossible, and desirably so, to distinguish individual quanta.

If the wavelength of light is small in comparison to the size of the apparatus (lenses, mirrors, etc.), one may use, as a first approximation, the techniques of *Geometrical Optics*. A somewhat more precise treatment, which is applicable as well when the dimensions of the apparatus are small, is that of *Physical Optics*. In Physical Optics the dominant property of light is its wave nature. It is even possible to develop most of the treatment without ever specifying the kind of wave one is dealing with. Certainly, as far as the classical study of Physical Optics is concerned, it will suffice admirably to treat light as an electromagnetic wave.

We can think of light as the most tenuous form of matter. Indeed, one of the basic tenets of Quantum Mechanics is that both light and material particles display similar wave-particle properties. As Erwin C. Schrödinger (1887–1961), one of the founders of quantum theory, put it:

In the new setting of ideas the distinction [between particles and waves] has vanished, because it was discovered that all particles have also wave properties, and *vice versa*. Neither of the two concepts must be discarded, they must be amalgamated. Which aspect obtrudes itself depends not on

the physical object, but on the experimental device set up to examine it.*

The quantum-mechanical treatment associates a wave equation with a particle, be it a photon, electron, proton, or whatever. In the case of material particles, the wave aspects are introduced by way of the field equation known as Schrödinger’s Equation. For light we have a representation of the wave nature in the form of the classical electromagnetic field equations of Maxwell. With these as a starting point one can construct a quantum-mechanical theory of photons and their interaction with charges. The dual nature of light is evidenced by the fact that it propagates through space in a wave-like fashion and yet displays particlelike behavior during the processes of emission and absorption. Electromagnetic radiant energy is created and destroyed in quanta or photons and not continuously as a classical wave. Nonetheless, its motion through a lens, a hole, or a set of slits is governed by wave characteristics. If we’re unfamiliar with this kind of behavior in the macroscopic world, it’s because the wavelength of a material object varies inversely with its momentum (p. 56), and even a grain of sand (which is barely moving) has a wavelength so small as to be indiscernible in any conceivable experiment.

The photon has zero mass, and therefore exceedingly large numbers of low-energy photons can be envisioned as present in a beam of light. Within that model, dense streams of photons act on the average to produce well-defined classical fields (p. 52). We can draw a rough analogy with the flow of commuters through a train station during rush hour. Each commuter presumably behaves individually as a quantum of humanity, but all have the same intent and follow fairly simi-

*Erwin C. Schrodinger, *Science Theory and Man*.

lar trajectories. To a distant, myopic observer there is a seemingly smooth and continuous flow. The behavior of the stream *en masse* is predictable from day to day, and so the precise motion of each commuter is unimportant, at least to the observer. The energy transported by a large number of photons is, *on the average*, equivalent to the energy transferred by a classical electromagnetic wave. For these reasons the classical field representation of electromagnetic phenomena has been, and will continue to be, so useful. Nonetheless, it should be understood that the apparent continuous nature of electromagnetic waves is a fiction of the macroscopic world, just as the apparent continuous nature of ordinary matter is a fiction—it just isn't that simple.

Quite pragmatically, then, we can consider light to be a classical electromagnetic wave, keeping in mind that there are situations for which this description is woefully inadequate.

3.1 Basic Laws of Electromagnetic Theory

Our intent in this section is to review and develop some of the ideas needed to appreciate the concept of electromagnetic waves.

We know from experiments that charges, even though separated in vacuum, experience a mutual interaction. Recall the familiar electrostatics demonstration in which a pith ball somehow senses the presence of a charged rod without actually touching it. As a possible explanation we might speculate that each charge emits (and absorbs) a stream of undetected particles (*virtual photons*). The exchange of these particles among the charges may be regarded as the mode of interac-

tion. Alternatively, we can take the classical approach and imagine instead that every charge is surrounded by something called an electric field. We then need only suppose that each charge interacts directly with the electric field in which it is immersed. Thus if a point charge q , experiences a force $\vec{F}_E$, the *electric field* $\vec{E}$ at the position of the charge is defined by $\vec{F}_E = q\vec{E}$. In addition, we observe that a moving charge may experience another force $\vec{F}_M$, which is proportional to its velocity $\vec{v}$. We are thus led to define yet another field, namely, the *magnetic induction* or just the *magnetic field* $\vec{B}$, such that $\vec{F}_M = q\vec{v} \times \vec{B}$. If forces $\vec{F}_E$ and $\vec{F}_M$ occur concurrently, the charge is moving through a region pervaded by both electric and magnetic fields, whereupon $\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$.

As we'll see, electric fields are generated by both electric charges and *time-varying magnetic fields*. Similarly, magnetic fields are generated by electric currents and by *time-varying electric fields*. This interdependence of $\vec{E}$ and $\vec{B}$ is a key point in the description of light.

3.1.1 Faraday's Induction Law

"Convert magnetism into electricity" was the brief remark Michael Faraday jotted in his notebook in 1822, a challenge he set himself with an easy confidence that made it seem so attainable. After several years doing other research, Faraday returned to the problem of electromagnetic induction in 1831. His first apparatus made use of two coils mounted on a wooden spool (Fig. 3.1a). One, called the primary, was attached to a battery and a switch; the other, the secondary, was attached to a galvanometer. He found that the galvanometer deflected in one direction just for a moment whenever the switch was closed, returning to zero almost immediately, despite the con-

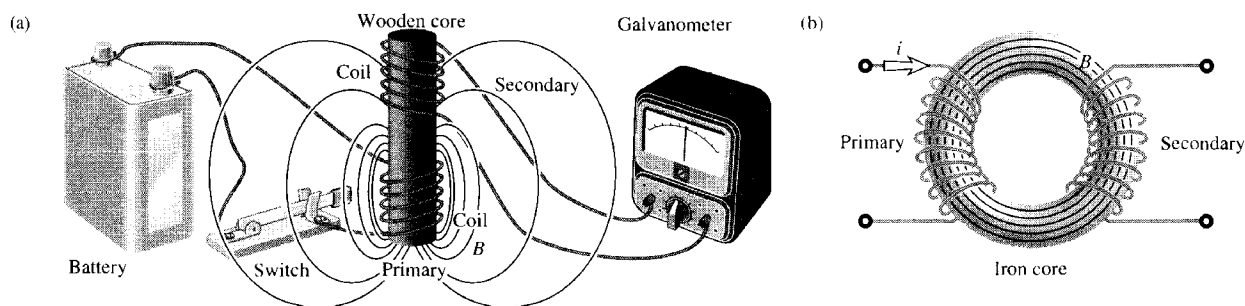


Figure 3.1 (a) The start of a current in one coil produces a time varying magnetic field that induces a current in the other coil. (b) An iron core couples the primary coil to the secondary.

stant current still in the primary. Whenever the switch was opened, interrupting the primary current, the galvanometer in the secondary circuit momentarily swung in the opposite direction and then promptly returned to zero.

Using a ferromagnetic core to concentrate the “magnetic force,” Faraday wound two coils around opposing sections of a soft iron ring (Fig. 3.1*b*). Now the effect was unmistakable—a **changing magnetic field generated a current**. Indeed, as he would continue to discover, *change* was the essential aspect of electromagnetic induction.

By thrusting a magnet into a coil, Faraday showed that there is a voltage—otherwise known as the **induced electromotive force** or **emf**—across the coil’s terminals. (Electromotive force is a dreadful, outmoded term—it’s not a force, but a voltage—so we’ll avoid it and just use emf.) Furthermore, the amplitude of the emf depends on how rapidly the magnet is moved. *The induced emf depends on the rate-of-change of B through the coil and not on B itself.* A weak magnet moved rapidly can induce a greater emf than a strong magnet moved slowly.

When the same changing B -field passes through two different wire loops, as in Fig. 3.2, the induced emf is larger across the terminals of the larger loop. In other words, here where the B -field is changing, **the induced emf is proportional to the area A of the loop penetrated perpendicularly by the field**. If the loop is successively tilted over, as is shown in Fig. 3.3, the area presented perpendicularly to the field ($A_{\perp}$) varies as $A \cos \theta$, and, when $\theta = 90^\circ$, the induced emf is zero because no amount of B -field then penetrates the loop: when $\Delta B / \Delta t \neq 0$, $\text{emf} \propto A_{\perp}$. The converse also holds: **when the field is constant, the induced emf is proportional to the rate-of-change of the perpendicular area penetrated**. If a coil is twisted or rotated or even squashed while in a constant B -field so that the perpendicular area initially penetrated is altered, there will be an induced $\text{emf} \propto \Delta A_{\perp} / \Delta t$ and it will be propor-

tional to B . In summary, when $A_{\perp} = \text{constant}$, $\text{emf} \propto A_{\perp} \Delta B / \Delta t$ and, when $B = \text{constant}$, $\text{emf} \propto B \Delta A_{\perp} / \Delta t$.

All of this suggests that the emf depends on the rate-of-change of both $A_{\perp}$ and B , that is, on the rate of change of their product. This should bring to mind the notion of the flux of the field—the product of field and area where the penetration is perpendicular. Accordingly, the **flux of the magnetic field** through the wire loop is

$$\Phi_M = B_{\perp} A = BA_{\perp} = BA \cos \theta$$

More generally, if B varies in space as it’s likely to, the flux of the magnetic field through any open area A bounded by the conducting loop (Fig. 3.4) is given by

$$\Phi_M = \iint_A \vec{B} \cdot d\vec{S} \quad (3.1)$$

The induced emf, developed around the loop, is then

$$\text{emf} = - \frac{d\Phi_M}{dt} \quad (3.2)$$

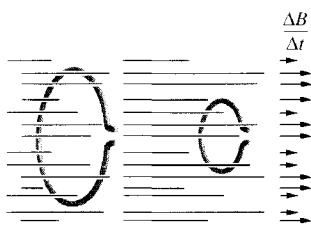


Figure 3.2 A larger time-varying magnetic flux passes through the larger loop and induces a greater emf across its terminals.

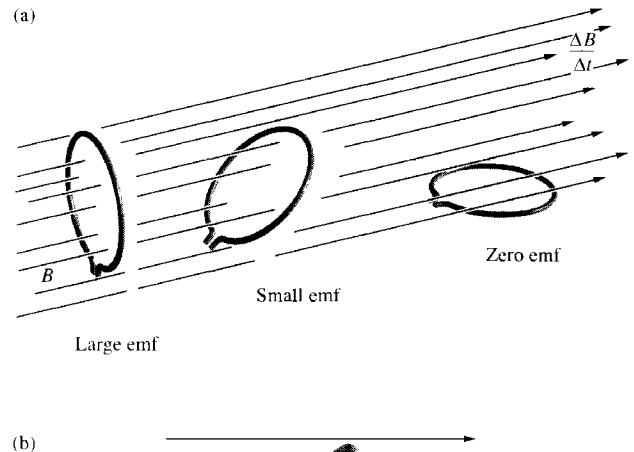


Figure 3.3 (a) The induced emf is proportional to the perpendicular area intercepted by the magnetic field. (b) That perpendicular area varies as $\cos \theta$.

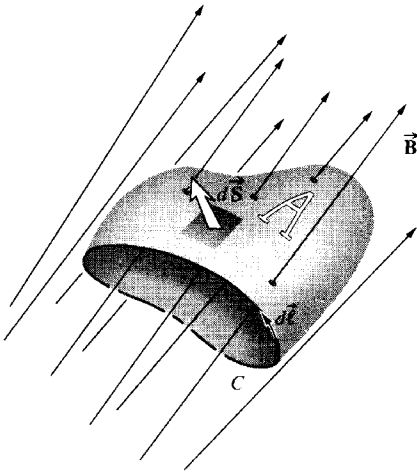


Figure 3.4 $\vec{B}$ -field through an open area A .

We should not, however, get too involved with the image of wires and current and emf. Our present concern is with the electric and magnetic fields themselves.

In very general terms, an emf is a potential difference, and that's a potential-energy difference per unit charge. A potential-energy difference per unit charge corresponds to work done per unit charge, which is force per unit charge times distance, and that's electric field times distance. The emf exists only as a result of the presence of an electric field:

$$\text{emf} = \oint_C \vec{E} \cdot d\vec{\ell} \quad (3.3)$$

taken around the closed curve C , corresponding to the loop. Equating Eqs. (3.2) and (3.3), and making use of Eq. (3.1), we get

$$\oint_C \vec{E} \cdot d\vec{\ell} = -\frac{d}{dt} \iint_A \vec{B} \cdot d\vec{S} \quad (3.4)$$

We began this discussion by examining a conducting loop, and have arrived at Eq. (3.4); this expression, except for the path C , contains no reference to the physical loop. In fact, the path can be chosen arbitrarily and need not be within, or anywhere near, a conductor. The electric field in Eq. (3.4) arises not from the presence of electric charges but rather from the time-varying magnetic field. With no charges to act as sources or sinks, the field lines close on themselves, forming loops (Fig. 3.5).

We are interested in electromagnetic waves traveling in space where there are no wire loops, and the magnetic flux changes because $\vec{B}$ changes. The *Induction Law* (Eq. 3.4) can

then be rewritten as

$$\oint_C \vec{E} \cdot d\vec{\ell} = - \iint_A \frac{\partial \vec{B}}{\partial t} \cdot d\vec{S} \quad (3.5)$$

A partial derivative with respect to t is taken because $\vec{B}$ is usually also a function of the space variables. This expression in itself is rather fascinating, since it indicates that *a time-varying magnetic field will have an electric field associated with it*.

3.1.2 Gauss's Law—Electric

Another fundamental law of electromagnetism is named after the German mathematician Karl Friedrich Gauss (1777–1855). Gauss's Law is about the relationship between the flux of the electric field and the sources of that flux, charge. The ideas derive from fluid dynamics, where both the concepts of field and flux were introduced. The flow of a fluid, as represented by its velocity field, is depicted via streamlines, much as the electric field is pictured via field lines. Figure 3.6 shows a portion of a moving fluid within which there is a region isolated by an imaginary closed surface. The discharge rate, or *volume flux* (Av), is the volume of fluid flowing past a point in the tube per unit time. The volume flux through both end surfaces

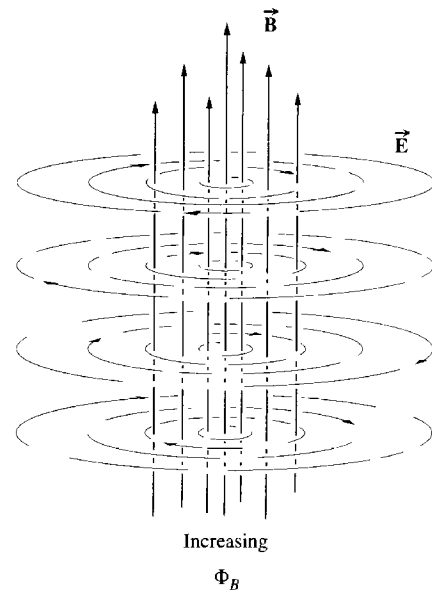
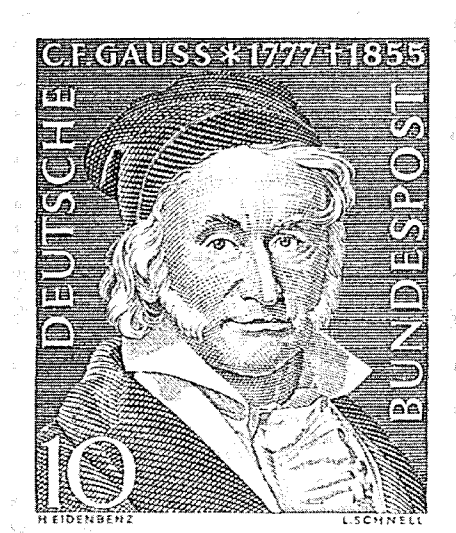


Figure 3.5 A time-varying $\vec{B}$ -field. Surrounding each point where Φ_M is changing, the $\vec{E}$ -field forms closed loops.



Johann Karl
Friedrich Gauss

is equal in magnitude—what flows in per second flows out per second. The net fluid flux (into and out of the closed area) summed over all the surfaces equals zero. If, however, a small pipe is inserted into the region either sucking out fluid (a sink) or delivering fluid (a source), the net flux will then be non-zero.

To apply these ideas to the electric field, consider an imaginary closed area A placed in some arbitrary electric field, as depicted in Fig. 3.7. The flux of electric field through A is taken to be

$$\Phi_E = \oint_A \vec{E} \cdot d\vec{S} \quad (3.6)$$

The circled double integral serves as a reminder that the surface is closed. The vector $d\vec{S}$ is in the direction of an *outward normal*. **When there are no sources or sinks of the electric field within the region encompassed by the closed surface, the net flux through the surface equals zero**—that much is a general rule for all such fields.

In order to find out what would happen in the presence of internal sources and sinks, consider a spherical surface of

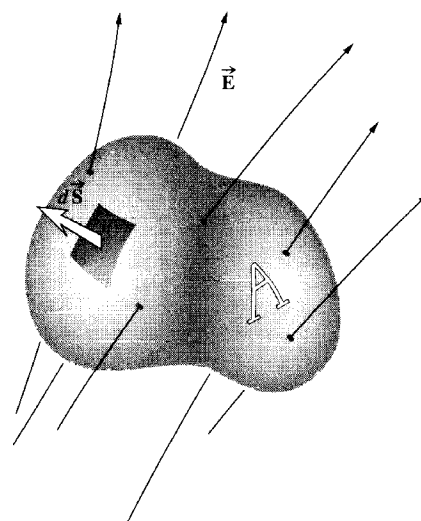


Figure 3.7 $\vec{E}$ -field
through a closed area A

radius r centered on and surrounding a positive point-charge (q) in vacuum. The E -field is everywhere outwardly radial, and at any distance r it is entirely perpendicular to the surface: $E = E_{\perp}$ and so

$$\Phi_E = \oint_A E_{\perp} dS = \oint_A E dS$$

Moreover, since E is constant over the surface of the sphere, it can be taken out of the integral:

$$\Phi_E = E \oint_A dS = E 4\pi r^2$$

But we know from Coulomb's Law that the point-charge has an electric field given by

$$E = \frac{1}{4\pi\epsilon_0} \frac{q}{r^2}$$

and so

$$\Phi_E = \frac{q}{\epsilon_0}$$

This is the electric flux associated with a single point-charge q , within the closed surface. Since all charge distributions are made up of point-charges, it's reasonable that the net flux due to a number of charges contained within any closed area is

$$\Phi_E = \frac{1}{\epsilon_0} \sum q.$$

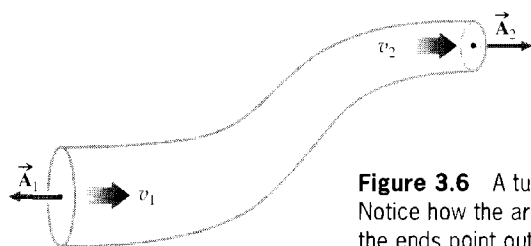


Figure 3.6 A tube of fluid flow.
Notice how the area vectors on
the ends point outward.

Combining the two equations for Φ_E , we get **Gauss's Law**,

$$\oiint_A \vec{E} \cdot d\vec{S} = \frac{1}{\epsilon_0} \sum q.$$

In order to apply the calculus, it's useful to approximate the charge distribution as being continuous. Then if the volume enclosed by A is V and the charge distribution has a density ρ , Gauss's Law becomes

$$\oiint_A \vec{E} \cdot d\vec{S} = \frac{1}{\epsilon_0} \iiint_V \rho dV \quad (3.7)$$

Electric Permittivity

For the special case of vacuum, the electric permittivity *of free space* is given by $\epsilon_0 = 8.8542 \times 10^{-12} \text{ C}^2 \cdot \text{N} \cdot \text{m}^{-2}$. If the charge is embedded in some material medium its permittivity (ϵ) will appear in Eq. (3.7) instead of ϵ_0 . One function of the permittivity in Eq. (3.7) is, of course, to balance out the units, but the concept is basic to the description of the parallel plate capacitor (see Section 3.1.4). There it's the medium-dependent proportionality constant between the device's capacitance and its geometric characteristics. Indeed ϵ is often measured by a procedure in which the material under study is placed within a capacitor. Conceptually, the permittivity embodies the electrical behavior of the medium: in a sense, it is a measure of the degree to which the material is permeated by the electric field in which it is immersed.

In the early days of the development of the subject, people in various areas worked in different systems of units, a state of affairs leading to some obvious difficulties. This necessitated the tabulation of numerical values for ϵ in each of the different systems, which was, at best, a waste of time. The same problem regarding densities was neatly avoided by using specific gravity (i.e., density ratios). Thus it was advantageous to tabulate values not of ϵ but of a new related quantity that is independent of the system of units being used. Accordingly, we define K_E as ϵ/ϵ_0 . This is the **dielectric constant** (or *relative permittivity*), and it is appropriately unitless. The permittivity of a material can then be expressed in terms of ϵ_0 as

$$\epsilon = K_E \epsilon_0 \quad (3.8)$$

Our interest in K_E anticipates the fact that the permittivity is related to the speed of light in dielectric materials, such as glass, air, and quartz.

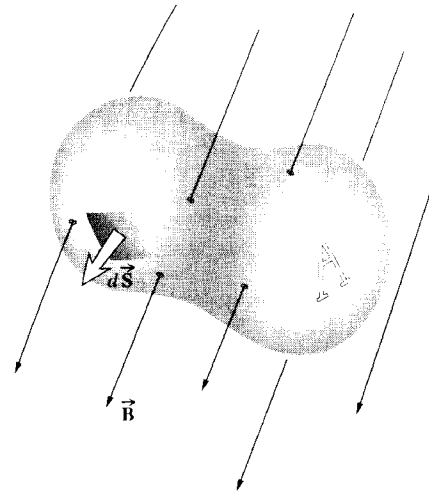


Figure 3.8 $\vec{B}$ -field through a closed area A .

3.1.3 Gauss's Law—Magnetic

There is no known magnetic counterpart to the electric charge; that is, no isolated magnetic poles have ever been found, despite extensive searching, even in lunar soil samples. Unlike the electric field, the magnetic field $\vec{B}$ does not diverge from or converge toward some kind of magnetic charge (a monopole source or sink). Magnetic fields can be described in terms of current distributions. Indeed, we might envision an elementary magnet as a small current loop in which the lines of $\vec{B}$ are continuous and closed. Any closed surface in a region of magnetic field would accordingly have an equal number of lines of $\vec{B}$ entering and emerging from it (Fig. 3.8). This situation arises from the absence of any monopoles within the enclosed volume. The flux of magnetic field Φ_M through such a surface is zero, and we have the magnetic equivalent of Gauss's Law:

$$\Phi_M = \oiint_A \vec{B} \cdot d\vec{S} = 0 \quad (3.9)$$

3.1.4 Ampère's Circuital Law

Another equation that will be of great interest is associated with André Marie Ampère (1775–1836). Known as the *Circuital Law*, its physical origins are a little obscure—it will take a bit of doing to justify it, but it's worth it. Accordingly, imagine a straight current-carrying wire in vacuum and the circular B -field surrounding it (Fig. 3.9). We know from experiments that the magnetic field of a straight wire carrying a current i is $B = \mu_0 i / 2\pi r$. Now, suppose we put ourselves back in time to the nineteenth century when it was common to think of mag-

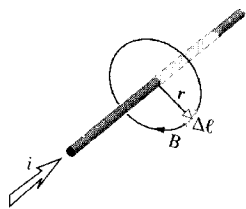


Figure 3.9 The $\vec{B}$ -field surrounding a current-carrying wire.

netic charge (q_m). Let's define this monopole charge so that it experiences a force when placed in a magnetic field B equal to $q_m B$ in the direction of B , just as an electric charge q_e experiences a force $q_e E$. Suppose we carry this north-seeking monopole around a closed circular path perpendicular to and centered on a current-carrying wire and determine the work done in the process. Since the direction of the force changes, because $\vec{B}$ changes direction, we will have to divide the circular path into tiny segments ($\Delta\ell$) and sum up the work done over each. Work is the component of the force parallel to the displacement times the displacement: $\Delta W = q_m B_{\parallel} \Delta\ell$, and the total work done by the field is $\sum q_m B_{\parallel} \Delta\ell$. In this case, $\vec{B}$ is everywhere tangent to the path, so that $B_{\parallel} = B = \mu_0 i / 2\pi r$, which is constant around the circle. With both q_m and B constant, the summation becomes.

$$q_m \sum B_{\parallel} \Delta\ell = q_m B \sum \Delta\ell = q_m B 2\pi r$$

where $\sum \Delta\ell = 2\pi r$ is the circumference of the circular path.

If we substitute for B the equivalent current expression, which varies inversely with r , the radius cancels—the work is independent of the circular path taken. Since no work is done in traveling perpendicular to $\vec{B}$, the work must be the same if we move q_m (out away from the wire or in toward it) along a radius, carrying it from one circular segment to another as we go around. Indeed, W is independent of path altogether—the work will be the same for *any closed path* encompassing the current. Putting in the current expression for B and canceling the “charge” q_m , we get the rather remarkable expression

$$\sum B_{\parallel} \Delta\ell = \mu_0 i$$

which is to be summed over any closed path surrounding i . The magnetic charge has disappeared, which is nice, since we no longer expect to be able to perform this little thought experiment with a monopole. Still, the physics was consistent, and the equation should hold, monopoles or no. Moreover, if there are several current-carrying wires encompassed by the closed path, their fields will superimpose and add, yielding a net

field. The equation is true for the separate fields and must be true as well for the net field. Hence

$$\sum B_{\parallel} \Delta\ell = \mu_0 \sum i$$

As $\Delta\ell \rightarrow 0$, the sum becomes an integral *around a closed path*:

$$\oint_C \vec{B} \cdot d\vec{\ell} = \mu_0 \sum i$$

Today this equation is known as **Ampère's Law**, though at one time it was commonly referred to as the “work rule.” It relates a line integral of $\vec{B}$ tangent to a closed curve C , with the total current i passing within the confines of C .

When the current has a nonuniform cross section, Ampère's Law is written in terms of the *current density* or current per unit area J , integrated over the area:

$$\oint_C \vec{B} \cdot d\vec{\ell} = \mu_0 \iint_A \vec{J} \cdot d\vec{S} \quad (3.10)$$

The open surface A is bounded by C (Fig. 3.10). The quantity μ_0 is called the **permeability of free space** and it's defined as $4\pi \times 10^{-7} \text{N} \cdot \text{s}^2 / \text{C}^2$. When the current is imbedded in a material medium its permeability (μ) will appear in Eq. (3.10). As in Eq. (3.8),

$$\mu = K_M \mu_0 \quad (3.11)$$

with K_M being the dimensionless *relative permeability*.

Equation (3.10), though often adequate, is not the whole

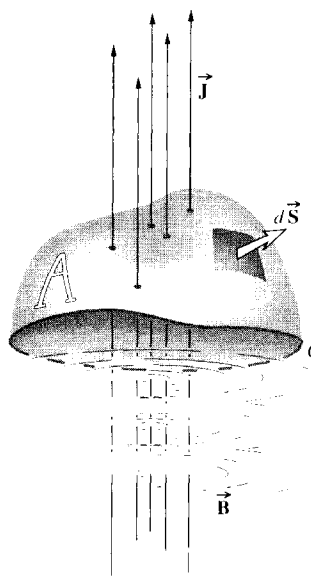


Figure 3.10 Current density through an open area A .

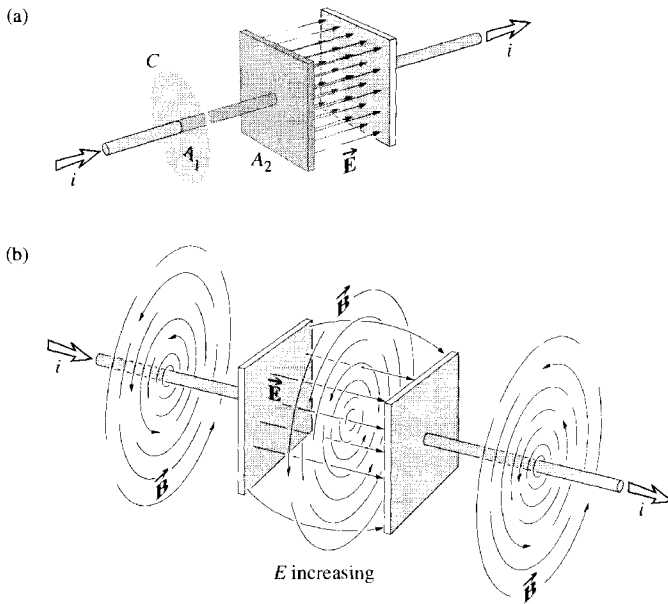


Figure 3.11 (a) Ampère's Law is indifferent to which area A_1 or A_2 is bounded by the path C . Yet a current passes through A_1 and not through A_2 , and that means something is very wrong. (b) $\vec{B}$ -field concomitant with a time-varying $\vec{E}$ -field in the gap of a capacitor.

truth. Ampère's Law is not particular about the area used, provided it's bounded by the curve C , which makes for an obvious problem when charging a capacitor, as shown in Fig. 3.11a. If flat area A_1 is used, a net current of i flows through it and there is a $\vec{B}$ -field along curve C . The right side of Eq. (3.10) is nonzero, so the left side is nonzero. But if area A_2 is used instead to encompass C , no net current passes through it and the field must now be zero, even though nothing physical has actually changed. Something is obviously wrong!

Moving charges are not the only source of a magnetic field. While charging or discharging a capacitor, one can measure a $\vec{B}$ -field in the region between its plates (Fig. 3.11b), which is indistinguishable from the field surrounding the leads, even though no electric current actually traverses the capacitor. Notice, however, that if A is the area of each plate and Q the charge on it,

$$E = \frac{Q}{\epsilon A}$$

As the charge varies, the electric field changes, and taking the derivative of both sides yields

$$\epsilon \frac{\partial E}{\partial t} = \frac{i}{A}$$

which is effectively a $\vec{E}$ current density. James Clerk Maxwell hypothesized the existence of just such a mechanism, which he called the *displacement current density*,* defined by

$$\vec{J}_D \equiv \epsilon \frac{\partial \vec{E}}{\partial t} \quad (3.12)$$

The restatement of Ampère's Law as

$$\oint_C \vec{B} \cdot d\vec{\ell} = \mu \iint_A \left(\vec{J} + \epsilon \frac{\partial \vec{E}}{\partial t} \right) \cdot d\vec{S} \quad (3.13)$$

was one of Maxwell's greatest contributions. It points out that even when $\vec{J} = 0$, a time-varying $\vec{E}$ -field will be accompanied by a $\vec{B}$ -field (Fig. 3.7).

3.1.5 Maxwell's Equations

The set of integral expressions given by Eqs. (3.5), (3.7), (3.9), and (3.13) have come to be known as Maxwell's Equations. Remember that these are generalizations of experimental results. The simplest statement of Maxwell's Equations applies to the behavior of the electric and magnetic fields in free space, where $\epsilon = \epsilon_0$, $\mu = \mu_0$, and both ρ and $\vec{J}$ are zero. In that instance,

$$\oint_C \vec{E} \cdot d\vec{\ell} = - \iint_A \frac{\partial \vec{B}}{\partial t} \cdot d\vec{S} \quad (3.14)$$

$$\oint_C \vec{B} \cdot d\vec{\ell} = \mu_0 \epsilon_0 \iint_A \frac{\partial \vec{E}}{\partial t} \cdot d\vec{S} \quad (3.15)$$

$$\oiint_A \vec{B} \cdot d\vec{S} = 0 \quad (3.16)$$

$$\oiint_A \vec{E} \cdot d\vec{S} = 0 \quad (3.17)$$

Observe that except for a multiplicative scalar, the electric and magnetic fields appear in the equations with a remarkable symmetry. However $\vec{E}$ affects $\vec{B}$. $\vec{B}$ will in turn affect $\vec{E}$. The mathematical symmetry implies a good deal of physical symmetry.

Maxwell's Equations can be written in a differential form, which will be somewhat more useful for our purposes. The

*Maxwell's own words and ideas concerning this mechanism are examined in an article by A. M. Bork, *Am. J. Phys.* **31**, 854 (1963). Incidentally, Clerk is pronounced *clark*.

appropriate calculation is carried out in Appendix 1, and the consequent equations for *free space*, in Cartesian coordinates, are as follows:

$$\begin{aligned} \frac{\partial \vec{E}_z}{\partial y} - \frac{\partial \vec{E}_y}{\partial z} &= -\frac{\partial \vec{B}_x}{\partial t} \quad (i) \\ \frac{\partial \vec{E}_x}{\partial z} - \frac{\partial \vec{E}_z}{\partial x} &= -\frac{\partial \vec{B}_y}{\partial t} \quad (ii) \\ \frac{\partial \vec{E}_x}{\partial x} - \frac{\partial \vec{E}_y}{\partial y} &= -\frac{\partial \vec{B}_z}{\partial t} \quad (iii) \end{aligned} \quad (3.18)$$

$$\begin{aligned} \frac{\partial \vec{B}_z}{\partial y} - \frac{\partial \vec{B}_y}{\partial z} &= \mu_0 \epsilon_0 \frac{\partial \vec{E}_x}{\partial t} \quad (i) \\ \frac{\partial \vec{B}_x}{\partial z} - \frac{\partial \vec{B}_z}{\partial x} &= \mu_0 \epsilon_0 \frac{\partial \vec{E}_y}{\partial t} \quad (ii) \\ \frac{\partial \vec{B}_x}{\partial x} - \frac{\partial \vec{B}_y}{\partial y} &= \mu_0 \epsilon_0 \frac{\partial \vec{E}_z}{\partial t} \quad (iii) \end{aligned} \quad (3.19)$$

$$\frac{\partial \vec{B}_x}{\partial x} + \frac{\partial \vec{B}_y}{\partial y} + \frac{\partial \vec{B}_z}{\partial z} = 0 \quad (3.20)$$

$$\frac{\partial \vec{E}_x}{\partial x} + \frac{\partial \vec{E}_y}{\partial y} + \frac{\partial \vec{E}_z}{\partial z} = 0 \quad (3.21)$$

The transition has thus been made from the formulation of Maxwell's Equations in terms of integrals over *finite regions* to a restatement in terms of derivatives at *points* in space.

We now have all that is needed to comprehend the magnificent process whereby electric and magnetic fields, inseparably coupled and mutually sustaining, propagate out into space as a single entity, free of charges and currents, sans matter, sans aether.

3.2 Electromagnetic Waves

We have relegated to Appendix 1 a complete and mathematically elegant derivation of the electromagnetic wave equation. Here the focus is on the equally important task of developing

a more intuitive appreciation of the physical processes involved. Three observations, from which we might build a qualitative picture, are readily available to us: the general perpendicularity of the fields, the symmetry of Maxwell's Equations, and the interdependence of $\vec{E}$ and $\vec{B}$ in those equations.

In studying electricity and magnetism, one soon becomes aware that a number of relationships are described by vector cross-products or, if you like, right-hand rules. In other words, an occurrence of one sort produces a related, perpendicularly directed response. Of immediate interest is the fact that a time-varying $\vec{E}$ -field generates a $\vec{B}$ -field, which is everywhere perpendicular to the direction in which $\vec{E}$ changes (Fig. 3.12). In the same way, a time-varying $\vec{B}$ -field generates an $\vec{E}$ -field, which is everywhere perpendicular to the direction in which $\vec{B}$ changes (Fig. 3.5). Consequently, we might anticipate the general transverse nature of the $\vec{E}$ - and $\vec{B}$ -fields in an electromagnetic disturbance.

Consider a charge that is somehow caused to *accelerate* from rest. When the charge is motionless, it has associated with it a constant radial $\vec{E}$ -field extending in all directions presumably to infinity. At the instant the charge begins to move, the $\vec{E}$ -field is altered in the vicinity of the charge, and this alteration propagates out into space at some finite speed. The time-varying electric field induces a magnetic field by means of Eq. (3.15) or (3.19). If the charge's velocity is constant, the rate-of-change of the $\vec{E}$ -field is steady, and the resulting $\vec{B}$ -field is constant. But here the charge is accelerating. $\partial \vec{E} / \partial t$

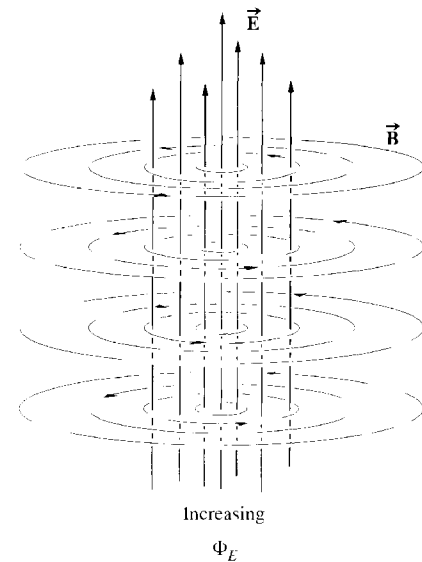


Figure 3.12 A time-varying $\vec{E}$ -field. Surrounding each point where Φ_E is changing, the $\vec{B}$ -field forms closed loops.

is itself not constant, so the induced $\vec{B}$ -field is time-dependent. The time-varying $\vec{B}$ -field generates an $\vec{E}$ -field, (3.14) or (3.18), and the process continues, with $\vec{E}$ and $\vec{B}$ coupled in the form of a pulse. As one field changes, it generates a new field that extends a bit farther, and the pulse moves out from one point to the next through space.

We can draw an overly mechanistic but rather picturesque analogy, if we imagine the electric field lines as a dense radial distribution of strings (p. 59). When somehow plucked, each string is distorted, forming a kink that travels outward from the source. All these kinks combine at any instant to yield a three-dimensional expanding pulse in the continuum of the electric field.

The $\vec{E}$ - and $\vec{B}$ -fields can more appropriately be considered as two aspects of a single physical phenomenon, the *electromagnetic field*, whose source is a moving charge. The disturbance, once it has been generated in the electromagnetic field, is an untethered wave that moves beyond its source and independently of it. Bound together as a single entity, the time-varying electric and magnetic fields regenerate each other in an endless cycle. The electromagnetic waves reaching us from the relatively nearby Andromeda galaxy (which can be seen with the naked eye) have been on the wing for 2 200 000 years.

We have not yet considered the direction of wave propagation with respect to the constituent fields. Notice, however, that the high degree of symmetry in Maxwell's Equations for free space suggests that the disturbance will propagate in a direction that is symmetrical to both $\vec{E}$ and $\vec{B}$. That implies that an electromagnetic wave cannot be purely longitudinal (i.e., as long as $\vec{E}$ and $\vec{B}$ are not parallel). Let's now replace conjecture with a bit of calculation.

Appendix I shows that Maxwell's Equations for free space can be manipulated into the form of two extremely concise vector expressions:

$$\nabla^2 \vec{E} = \epsilon_0 \mu_0 \frac{\partial^2 \vec{E}}{\partial t^2} \quad [\text{A1.26}]$$

$$\text{and} \quad \nabla^2 \vec{B} = \epsilon_0 \mu_0 \frac{\partial^2 \vec{B}}{\partial t^2} \quad [\text{A1.27}]$$

The Laplacian,* ∇^2 , operates on each component of $\vec{E}$ and $\vec{B}$, so that the two vector equations actually represent a total of six

scalar equations. In Cartesian coordinates,

$$\begin{aligned} \frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial^2 E_x}{\partial z^2} &= \epsilon_0 \mu_0 \frac{\partial^2 E_x}{\partial t^2} \\ \frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2} + \frac{\partial^2 E_y}{\partial z^2} &= \epsilon_0 \mu_0 \frac{\partial^2 E_y}{\partial t^2} \end{aligned} \quad (3.22)$$

$$\frac{\partial^2 E_z}{\partial x^2} + \frac{\partial^2 E_z}{\partial y^2} + \frac{\partial^2 E_z}{\partial z^2} = \epsilon_0 \mu_0 \frac{\partial^2 E_z}{\partial t^2}$$

$$\begin{aligned} \frac{\partial^2 B_x}{\partial x^2} + \frac{\partial^2 B_x}{\partial y^2} + \frac{\partial^2 B_x}{\partial z^2} &= \epsilon_0 \mu_0 \frac{\partial^2 B_x}{\partial t^2} \\ \frac{\partial^2 B_y}{\partial x^2} + \frac{\partial^2 B_y}{\partial y^2} + \frac{\partial^2 B_y}{\partial z^2} &= \epsilon_0 \mu_0 \frac{\partial^2 B_y}{\partial t^2} \end{aligned} \quad (3.23)$$

$$\frac{\partial^2 B_z}{\partial x^2} + \frac{\partial^2 B_z}{\partial y^2} + \frac{\partial^2 B_z}{\partial z^2} = \epsilon_0 \mu_0 \frac{\partial^2 B_z}{\partial t^2}$$

Expressions of this sort, which relate the space and time variations of some physical quantity, had been studied long before Maxwell's work and were known to describe wave phenomena (p. 28). Each and every component of the electromagnetic field (E_x , E_y , E_z , B_x , B_y , B_z) obeys the scalar differential wave equation

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 \psi}{\partial t^2} \quad [2.60]$$

provided that

$$v = 1/\sqrt{\epsilon_0 \mu_0} \quad (3.24)$$

To evaluate v , Maxwell made use of the results of electrical experiments performed in 1856 in Leipzig by Wilhelm Weber (1804–1891) and Rudolph Kohlrausch (1809–1858). Equivalently, nowadays μ_0 is assigned a value of $4\pi \times 10^{-7} \text{ m}\cdot\text{kg}/\text{C}^2$ in SI units, and until recently one might determine ϵ_0 directly from simple capacitor measurements. In any event, in modern units

$$\epsilon_0 \mu_0 \approx (8.85 \times 10^{-12} \text{ s}^2 \cdot \text{C}^2/\text{m}^3 \cdot \text{kg}) (4\pi \times 10^{-7} \text{ m}\cdot\text{kg}/\text{C}^2)$$

$$\text{or} \quad \epsilon_0 \mu_0 \approx 11.12 \times 10^{-18} \text{ s}^2/\text{m}^2$$

And now the moment of truth—in free space, the predicted

*In Cartesian coordinates,

$$\nabla^2 \vec{E} = \hat{i} \nabla^2 E_x + \hat{j} \nabla^2 E_y + \hat{k} \nabla^2 E_z$$

speed of all electromagnetic waves would then be

$$v = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx 3 \times 10^8 \text{ m/s}$$

This theoretical value was in remarkable agreement with the previously measured speed of light (315 300 km/s) determined by Fizeau. The results of Fizeau's experiments, performed in 1849 with a rotating toothed wheel, were available to Maxwell and led him to comment:

This velocity [i.e., his theoretical prediction] is so nearly that of light, that it seems we have strong reason to conclude that light itself (including radiant heat, and other radiations if any) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.

This brilliant analysis was one of the great intellectual triumphs of all time. It has become customary to designate the speed of light in vacuum by the symbol c , which comes from the Latin word *celer*, meaning fast. In 1983 the 17th Conférence Générale des Poids et Mesures in Paris adopted a new definition of the meter and thereby fixed the speed of light in vacuum as exactly

$$c = 2.997\,924\,58 \times 10^8 \text{ m/s}$$

3.2.1 Transverse Waves

The experimentally verified transverse character of light must now be explained within the context of electromagnetic theory. To that end, consider the fairly simple case of a plane wave propagating in vacuum in the positive x -direction. The electric field intensity is a solution of Eq. (A1.26), where $\vec{E}$ is constant over each of an infinite set of planes perpendicular to the x -axis. It is therefore a function only of x and t ; that is, $\vec{E} = \vec{E}(x, t)$. We now refer back to Maxwell's Equations, and in particular to Eq. (3.21), which is generally read as *the divergence of $\vec{E}$ equals zero*. Since $\vec{E}$ is not a function of either y or z , the equation can be reduced to

$$\frac{\partial E_x}{\partial x} = 0 \quad (3.25)$$

If E_x is not zero—that is, if there is some component of the field in the direction of propagation—this expression tells us that it does not vary with x . At any given time, E_x is constant

for all values of x , but of course, this possibility cannot therefore correspond to a traveling wave advancing in the positive x -direction. Alternatively, it follows from Eq. (3.25) that for a wave, $E_x = 0$; the electromagnetic wave has no electric field component in the direction of propagation. The $\vec{E}$ -field associated with the plane wave is then exclusively *transverse*.

The fact that the $\vec{E}$ -field is transverse means that in order to completely specify the wave we will have to specify the moment-by-moment direction of $\vec{E}$. Such a description corresponds to the **polarization** of the light, and it will be treated in Chapter 8. Without any loss of generality, we deal here with *plane or linearly polarized waves*, for which the direction of the vibrating $\vec{E}$ -vector is fixed. Thus we orient our coordinate axes so that the electric field is parallel to the y -axis, whereupon

$$\vec{E} = \hat{j} E_y(x, t) \quad (3.26)$$

Returning to Eq. (3.18), it follows that

$$\frac{\partial E_y}{\partial x} = - \frac{\partial B_z}{\partial t} \quad (3.27)$$

Therefore B_x and B_y are constant and of no interest at present. The time-dependent $\vec{B}$ -field can only have a component in the z -direction. Clearly then, *in free space, the plane electromagnetic wave is transverse* (Fig. 3.13). Except in the case of normal incidence, such waves propagating in real material media are generally not transverse—a complication arising from the fact that the medium may be dissipative or contain free charge.

We have not specified the form of the disturbance other than to say that it is a plane wave. Our conclusions are therefore quite general, applying equally well to both pulses and

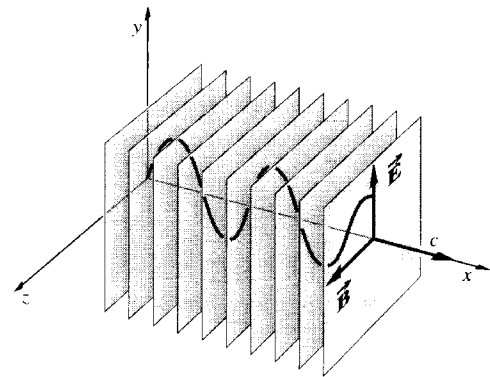


Figure 3.13 The field configuration in a plane harmonic electromagnetic wave.

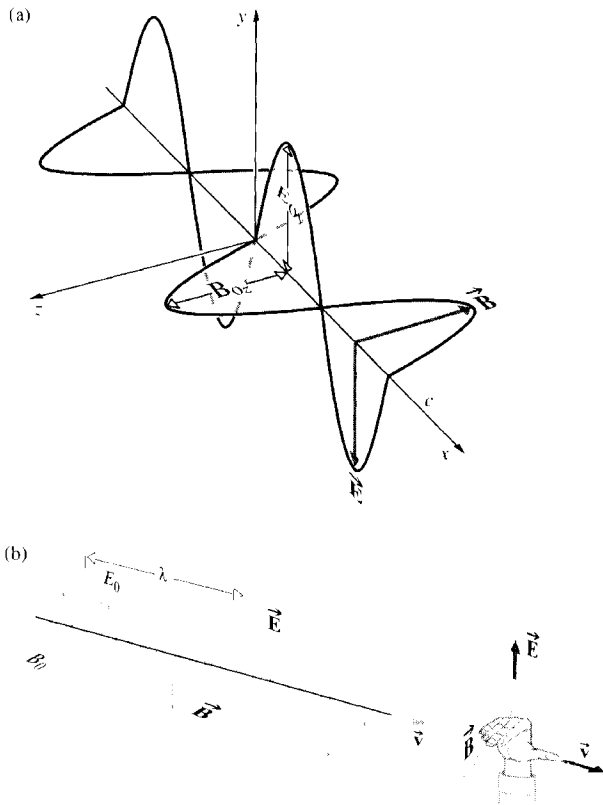


Figure 3.14 (a) Orthogonal harmonic $\vec{E}$ - and $\vec{B}$ -fields for a plane polarized wave. (b) The wave propagates in the direction of $\vec{E} \times \vec{B}$.

continuous waves. We have already pointed out that harmonic functions are of particular interest because any waveform can be expressed in terms of sinusoidal waves using Fourier techniques (p. 302). We therefore limit the discussion to harmonic waves and write $E_y(x, t)$ as

$$E_y(x, t) = E_{0y} \cos [\omega(t - x/c) + \epsilon] \quad (3.28)$$

the speed of propagation being c . The associated magnetic flux density can be found by directly integrating Eq. (3.27), that is,

$$B_z = - \int \frac{\partial E_y}{\partial x} dt$$

Using Eq. (3.28), we obtain

$$B_z = - \frac{E_{0y}\omega}{c} \int \sin [\omega(t - x/c) + \epsilon] dt$$

or
$$B_z(x, t) = \frac{1}{c} E_{0y} \cos [\omega(t - x/c) + \epsilon] \quad (3.29)$$

The constant of integration, which represents a time-independent field, has been disregarded. Comparison of this result with Eq. (3.28) makes it evident that

$$E_y = cB_z \quad (3.30)$$

Since E_y and B_z differ only by a scalar, and so have the same time-dependence, $\vec{E}$ and $\vec{B}$ are *in-phase* at all points in space. Moreover, $\vec{E} = \hat{j}E_y(x, t)$ and $\vec{B} = \hat{k}B_z(x, t)$ are *mutually perpendicular*, and their cross-product, $\vec{E} \times \vec{B}$, points in the propagation direction, $\hat{i}$ (Fig. 3.14).

Plane waves, though important, are not the only solutions to Maxwell's Equations. As we saw in the previous chapter, the differential wave equation allows many solutions, among which are cylindrical and spherical waves (Fig. 3.15).

3.3 Energy and Momentum

One of the most significant properties of the electromagnetic wave is that it transports energy and momentum. The light from even the nearest star beyond the Sun travels 25 million million miles to reach the Earth, yet it still carries enough energy to do work on the electrons within your eye.

3.3.1 The Poynting Vector

Any electromagnetic wave exists within some region of space, and it is therefore natural to consider the *radiant energy per unit volume*, or **energy density**, u . We suppose that the electric field itself can somehow store energy. This is a major logical step since it imparts to the field the attribute of physical reality—if the field has energy it is a *thing-in-itself*. Moreover,

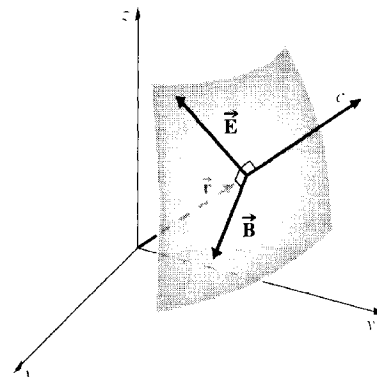


Figure 3.15 Portion of a spherical wavefront far from the source.

inasmuch as the classical field is continuous, its energy is continuous. Let's assume as much and see where it leads. We can compute (Problem 3.6) the energy density of the $\vec{E}$ -field (e.g., between the plates of a charged capacitor) to be

$$u_E = \frac{\epsilon_0}{2} E^2 \quad (3.31)$$

Similarly, the energy density of the B -field alone (as it might be computed within a current-carrying toroid) is

$$u_B = \frac{1}{2\mu_0} B^2 \quad (3.32)$$

The relationship $E = cB$ was derived specifically for plane waves; nonetheless, it's quite general in its applicability. Using $c = 1/\sqrt{\epsilon_0\mu_0}$, it follows that

$$u_E = u_B \quad (3.33)$$

The energy streaming through space in the form of an electromagnetic wave is shared equally between the constituent electric and magnetic fields. Inasmuch as

$$u = u_E + u_B \quad (3.34)$$

$$u = \epsilon_0 E^2 \quad (3.35)$$

or equivalently,

$$u = \frac{1}{\mu_0} B^2 \quad (3.36)$$

To represent the flow of electromagnetic energy associated with a traveling wave, let S symbolize the transport of energy per unit time (the power) across a unit area. In the SI system it has units of W/m^2 . Figure 3.16 depicts an electromagnetic wave traveling with a speed c through an area A . During a very small interval of time Δt , only the energy contained in the cylindrical volume, $u(c \Delta t A)$, will cross A . Thus

$$S = \frac{uc \Delta t A}{\Delta t A} = uc \quad (3.37)$$

or, using Eq. (3.35),

$$S = \frac{1}{\mu_0} EB \quad (3.38)$$

We now make the reasonable assumption (for isotropic media) that the energy flows in the direction of the propagation of the wave. The corresponding *vector* $\vec{S}$ is then

$$\vec{S} = \frac{1}{\mu_0} \vec{E} \times \vec{B} \quad (3.39)$$

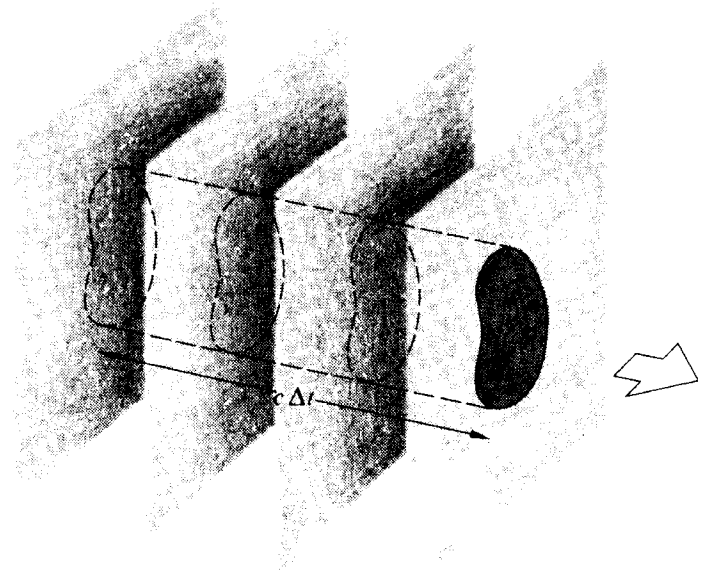


Figure 3.16 The flow of electromagnetic energy.

$$\text{or} \quad \vec{S} = c^2 \epsilon_0 \vec{E} \times \vec{B} \quad (3.40)$$

The magnitude of $\vec{S}$ is the power per unit area crossing a surface whose normal is parallel to $\vec{S}$. Named after John Henry Poynting (1852–1914), it has come to be known as the **Poynting vector**. Let's now apply these considerations to the case of a harmonic, linearly polarized (the directions of the $\vec{E}$ - and $\vec{B}$ -fields are fixed) plane wave traveling through free space in the direction of $\vec{k}$:

$$\vec{E} = \vec{E}_0 \cos(\vec{k} \cdot \vec{r} - \omega t) \quad (3.41)$$

$$\vec{B} = \vec{B}_0 \cos(\vec{k} \cdot \vec{r} - \omega t) \quad (3.42)$$

Using Eq. (3.40), we find

$$\vec{S} = c^2 \epsilon_0 \vec{E}_0 \times \vec{B}_0 \cos^2(\vec{k} \cdot \vec{r} - \omega t) \quad (3.43)$$

This is the instantaneous flow of energy per unit area per unit time.

Averaging Harmonic Functions

It should be evident that $\vec{E} \times \vec{B}$ cycles from maxima to minima. At optical frequencies ($\approx 10^{15}$ Hz), $\vec{S}$ is an extremely

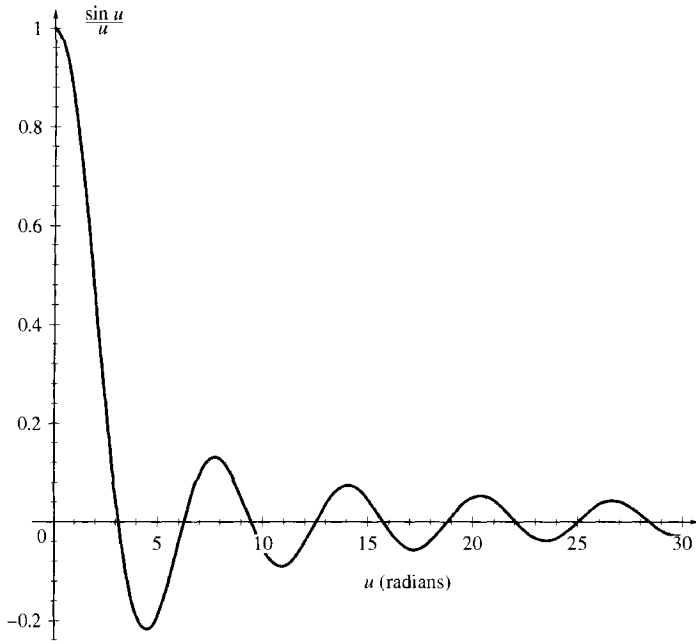


Figure 3.17 $\text{sinc } u$. Notice how the sinc function has a value of zero at $u = \pi, 2\pi, 3\pi$, and so forth.

rapidly varying function of time (indeed, twice as rapid as the fields, since cosine-squared has double the frequency of cosine). Therefore, its instantaneous value would be an impractical quantity to measure directly. This suggests that we employ an averaging procedure. That is, we absorb the radiant energy during some finite interval of time using, for example, a photocell, a film plate, or the retina of a human eye.

The specific form of Eq. (3.43), and the central role played by harmonic functions, suggest that we take a moment to study the average values of such functions. The time-averaged value of some function $f(t)$ over an interval T is written as $\langle f(t) \rangle_T$ and given by

$$\langle f(t) \rangle_T = \frac{1}{T} \int_{t-T/2}^{t+T/2} f(t) dt$$

The resulting value of $\langle f(t) \rangle_T$ very much depends on T . To find the average of a harmonic function, evaluate

$$\langle e^{i\omega t} \rangle_T = \frac{1}{T} \int_{t-T/2}^{t+T/2} e^{i\omega t} dt = \frac{1}{i\omega T} e^{i\omega t} \Big|_{t-T/2}^{t+T/2}$$

$$\langle e^{i\omega t} \rangle_T = \frac{1}{i\omega T} (e^{i\omega(t+T/2)} - e^{i\omega(t-T/2)})$$

$$\text{and} \quad \langle e^{i\omega t} \rangle_T = \frac{1}{i\omega T} e^{i\omega t} (e^{i\omega T/2} - e^{-i\omega T/2})$$

The parenthetical term should remind us (p. 22) of $\sin \omega T/2$. Hence

$$\langle e^{i\omega t} \rangle_T = \left(\frac{\sin \omega T/2}{\omega T/2} \right) e^{i\omega t}$$

The ratio in brackets is so common and important in Optics that it's given its own name; $(\sin u)/u$ is called ($\text{sinc } u$). Taking the real and imaginary parts of the above expression yields

$$\langle \cos \omega t \rangle_T = (\text{sinc } u) \cos \omega t$$

$$\text{and} \quad \langle \sin \omega t \rangle_T = (\text{sinc } u) \sin \omega t$$

The average of the cosine is itself a cosine, oscillating with the same frequency but having a sinc-function amplitude that drops off from its initial value of 1.0 very rapidly (Fig. 3.17 and Table 1 in the Appendix). Since $\text{sinc } u = 0$ at $u = \omega T/2 = \pi$, which happens when $T = \tau$, it follows that $\cos \omega t$ averaged over an interval T equal to one period equals zero. Similarly, $\cos \omega t$ averages to zero over any whole number of periods, as does $\sin \omega t$. That's reasonable in that each of these functions encompasses as much positive area above the axis as negative area below the axis, and that's what the defining integral corresponds to. After an interval of several periods, the sinc term will be so small that the fluctuations around zero will be negli-

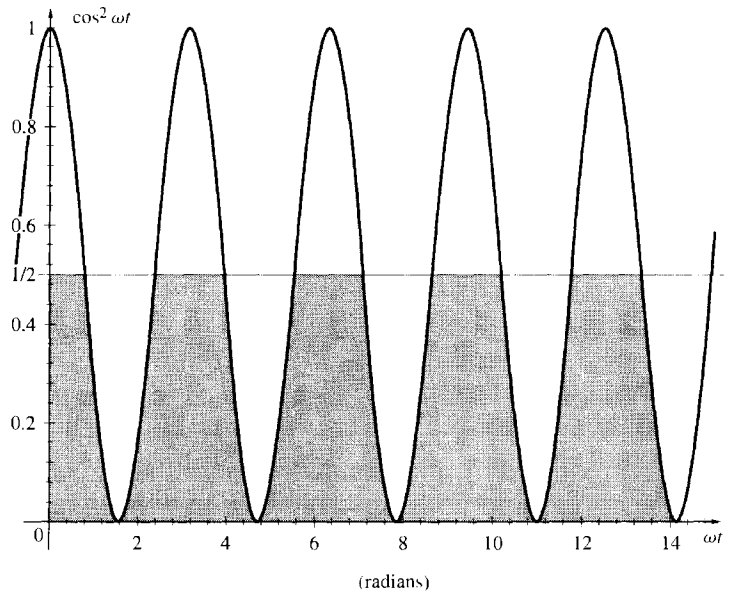


Figure 3.18 Using the peaks above the $\frac{1}{2}$ line to fill the troughs beneath it suggests that the average is $\frac{1}{2}$.

gible: $\langle \cos \omega t \rangle_T$ and $\langle \sin \omega t \rangle_T$ are then essentially zero.

It's left for Problem 3.8 to show that $\langle \cos^2 \omega t \rangle_T = \frac{1}{2}[1 + \text{sinc } \omega T \cos 2\omega t]$, which oscillates about a value of $1/2$ at a frequency of 2ω and rapidly approaches $1/2$ as T increases beyond a few dozen periods. In the case of light $\tau \approx 10^{-15}$ s and so averaging over even a microsecond corresponds to $T \approx 10^9 \tau$, far more than enough to drive the sinc function to some totally negligible value, whereupon $\langle \cos^2 \omega t \rangle_T = 1/2$. Figure 3.18 suggests the same result; we chop off the humps above the $1/2$ line and use them to fill in the missing areas beneath the line. After enough cycles, the area under the $f(t)$ curve divided by T , which is $\langle f(t) \rangle_T$, approaches $1/2$.

3.3.2 Irradiance

When we talk about the “amount” of light illuminating a surface, we are referring to something called the **irradiance**,* denoted by I —the **average energy per unit area per unit time**. Any kind of light-level detector has an entrance window that admits radiant energy through some fixed area A . The dependence on the size of that particular window is removed by dividing the total energy received by A . Furthermore, since the power arriving cannot be measured instantaneously, the detector must integrate the energy flux over some finite time, T . If the quantity to be measured is the *net* energy per unit area received, it depends on T and is therefore of limited utility. Someone else making a similar measurement under the same conditions can get a different result using a different T . If, however, the T is now divided out, a highly practical quantity results, one that corresponds to the average energy per unit area per unit time, namely, I .

The time-averaged value ($T \gg \tau$) of the magnitude of the Poynting vector, symbolized by $\langle S \rangle_T$, is a measure of I . In the specific case of harmonic fields and Eq. (3.43),

$$\langle S \rangle_T = c^2 \epsilon_0 |\vec{E}_0 \times \vec{B}_0| \langle \cos^2(\vec{k} \cdot \vec{r} - \omega t) \rangle$$

Because $\langle \cos^2(\vec{k} \cdot \vec{r} - \omega t) \rangle_T = \frac{1}{2}$ for $T \gg \tau$ (see Problem 3.7)

$$\langle S \rangle_T = \frac{c^2 \epsilon_0}{2} |\vec{E}_0 \times \vec{B}_0|$$

$$\text{or} \quad I \equiv \langle S \rangle_T = \frac{c \epsilon_0}{2} E_0^2 \quad (3.44)$$

*In the past physicists generally used the word *intensity* to mean the flow of energy per unit area per unit time. By international, if not universal, agreement, that term is slowly being replaced in Optics by the word *irradiance*.

The irradiance is proportional to the square of the amplitude of the electric field. Two alternative ways of saying the same thing are simply

$$I = \frac{c}{\mu_0} \langle B^2 \rangle_T \quad (3.45)$$

$$\text{and} \quad I = \epsilon_0 c \langle E^2 \rangle_T \quad (3.46)$$

Within a linear, homogeneous, isotropic dielectric, the expression for the irradiance becomes

$$I = \epsilon v \langle E^2 \rangle_T \quad (3.47)$$

Since, as we have learned, $\vec{E}$ is considerably more effective at exerting forces and doing work on charges than is $\vec{B}$, we shall refer to $\vec{E}$ as the **optical field** and use Eq. (3.46) almost exclusively.

The time rate of flow of radiant energy is the **optical power** P or **radiant flux**, generally expressed in watts. If we divide the radiant flux incident on or exiting from a surface by the area of the surface, we have the **radiant flux density** (W m^{-2}). In the former case, we speak of the *irradiance*, in the latter the *exitance*, and in either instance the **flux density**. The irradiance is a measure of the *concentration* of power.

The Inverse Square Law

We saw earlier that the spherical-wave solution of the differential wave equation has an amplitude that varies inversely with r . Let's now examine this same feature within the context of energy conservation. Consider an isotropic point source in free space, emitting energy equally in all directions (i.e., emitting spherical waves). Surround the source with two concentric imaginary spherical surfaces of radii r_1 and r_2 , as shown in Fig. 3.19. Let $E_0(r_1)$ and $E_0(r_2)$ represent the amplitudes of the waves over the first and second surfaces, respectively. If energy is to be conserved, the total amount of energy flowing through each surface per second must be equal, since there are no other sources or sinks present. Multiplying I by the surface area and taking the square root, we get

$$r_1 E_0(r_1) = r_2 E_0(r_2)$$

Inasmuch as r_1 and r_2 are arbitrary, it follows that

$$r E_0(r) = \text{constant},$$

and the amplitude must drop off inversely with r . The irradiance from a point source is proportional to $1/r^2$. This is the

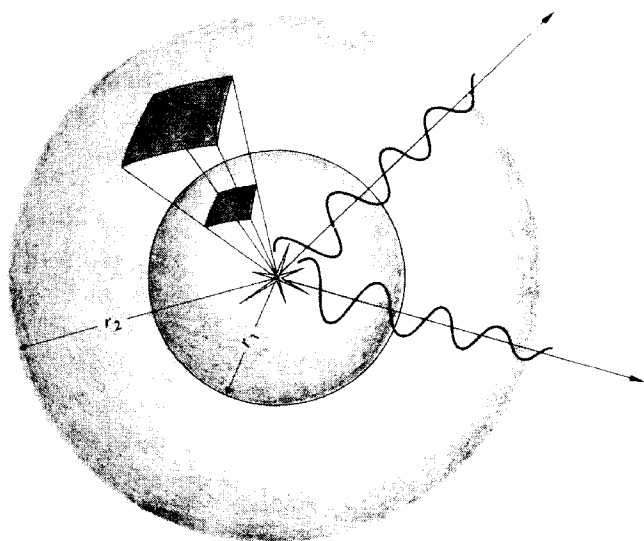


Figure 3.19 The geometry of the Inverse Square Law.

well-known **Inverse Square Law**, which is easily verified with a point source and a photographic exposure meter.

3.3.3 Photons

Light is absorbed and emitted in tiny discrete bursts, in particles of electromagnetic “stuff,” known as photons. That much has been confirmed and is well established, but the question of whether or not light is “really” a stream of photons is far from settled, and that issue will be revisited several times to come.*

Ordinarily, a light beam delivers so many minute energy quanta that its inherent granular nature is totally hidden and a continuous phenomenon is observed macroscopically. That sort of thing is commonplace in Nature; the forces exerted by the individual gas molecules in a wind blend into what seems a continuous pressure, but it obviously isn’t. Indeed, that analogy between a gas and a flow of photons is one we will come back to presently.

*Thus whenever, for brevity’s sake, an expression such as “a beam of photons” is used, the reader is advised to keep in mind that although the light-is-corpuscular model has wide acceptance (especially in high-energy physics) and is a crucial part of the contemporary discourse, like almost everything else in physics, it is not yet established beyond all doubt. See, for example, the summary article by R. Kidd, J. Ardin, and A. Anton, “Evolution of the modern photon,” *Am. J. Phys.* **57** (1), 27 (1989).

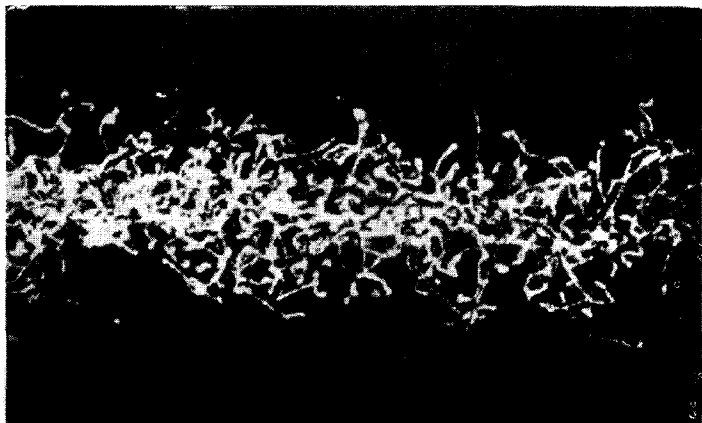
As the great French physicist Louis de Broglie put it, “Light is, in short, the most refined form of matter,” and all matter, including light, is quantized. At base it comes in minute elementary units—quarks, leptons, and photons. That overarching unity is among the most appealing reasons to embrace the photon as particle (whatever a “particle” is).

The Failure of Classical Theory

In 1900 Max Planck produced a rather tentative, and somewhat erroneous, analysis of a process known as *blackbody radiation* (p. 581). Nonetheless, the expression he came up with beautifully fit all the existing experimental data, a feat no other formulation had even come close to. Basically, he considered electromagnetic (EM) waves in equilibrium within an isothermal chamber (or cavity). All the EM-radiation within the cavity is emitted and absorbed by the walls of the enclosure—none enters from outside. This ensures that its spectral composition will match that emanating from an ideal black surface. The goal was to predict the spectrum of the radiation that would emerge from a small opening in the cavity. Totally stymied by the problem, as a last resort, Planck turned to the classical statistical analysis of Maxwell and Boltzmann, which was developed as the basis of the kinetic theory of gases. Philosophically, this is a completely deterministic treatment that assumes one can follow, at least in principle, every atom as it moves around in the system. Consequently, each atom is taken to be recognizable, independent, and enumerable. For purely computational reasons, Planck hypothesized that each one of the oscillators lining the walls of the chamber could absorb and emit only discrete amounts of energy proportional to its oscillatory frequency, ν . These energy jolts were equal to whole number multiples of $h\nu$, where h , now called **Planck’s Constant**, was found to be 6.626×10^{-34} J·s. Being a rather traditional man, Planck otherwise held fast to the classical wave picture of light, insisting that only the oscillators were quantized.

Prophetically, J. J. Thomson (1903)—the discoverer of the electron—extended the idea, suggesting that electromagnetic waves might actually be radically different from other waves; perhaps local concentrations of radiant energy truly existed. Thomson had observed that when a beam of high-frequency EM-radiation (X-rays) was shone onto a gas, only certain of the atoms, here and there, were ionized. It was as if the beam had “hot spots” rather than having its energy distributed continuously over the wavefront (see photo at the top of page 52).

The concept of the photon in its modern incarnation came



A beam of X-rays enters a cloud chamber on the left. The tracks are made by electrons emitted via either the Photoelectric Effect (these tend to leave long tracks at large angles to the beam) or the Compton Effect (short tracks more in the forward direction). Although classically the X-ray beam has its energy uniformly distributed along transverse wavefronts, the scattering seems discrete and random. (From the Smithsonian Report, 1915.)

into being in 1905 by way of Einstein's brilliant theoretical work on the Photoelectric Effect. When a metal is bathed in EM-radiation, it emits electrons. The details of that process had been studied experimentally for decades, but it defied analysis via classical Electromagnetic Theory. Einstein's startling treatment established that *the electromagnetic field itself is quantized*. Each constituent photon has an energy given by the product of Planck's Constant and the frequency of the radiation field:

$$\mathcal{E} = h\nu \quad (3.48)$$

Photons are stable, chargeless, massless elementary particles that exist only at the speed c .

In 1924 Satyendra N. Bose formulated a new and rigorous proof of Planck's blackbody equation using statistical methods applied to light quanta. The cavity was envisioned to be filled with a "gas" of photons, *which were taken to be totally indistinguishable, one from the other*. That was a crucial feature of this quantum-mechanical treatment. It meant that the microparticles were completely interchangeable, and this had a profound effect on the statistical formulation. In a mathematical sense, each particle of this quantum "gas" is related to every other particle, and no one of them can be taken as statistically independent of the system as a whole. That's very different from the independent way classical microparticles

behave in an ordinary gas. The quantum-mechanical probability function that describes the statistical behavior of thermal light is now known as the Bose-Einstein distribution. The photon, whatever it is, became an indispensable tool of theoretical physics.

Unlike ordinary objects, photons cannot be seen directly; what is known of them comes from observing the results of their being either created or annihilated. Light is never *seen* just sailing along through space. A photon is observed by detecting the effect it has on its surroundings, and it only has an observable effect when it either comes into, or goes out of, existence. Photons begin and end on charged particles; most often they are emitted from and absorbed by electrons. And these are usually the electrons circulating in the clouds around atoms. A number of experiments have directly confirmed the quantal nature of the emission process. For example, imagine a very dim source surrounded, at equal distances, by identical photodetectors each capable of measuring a minute amount of light. If the emission, *no matter how faint*, is a continuous wave, as is maintained classically, all the detectors should register each emitted pulse in coincidence. That does not happen; instead, counts are registered by detectors independently, one at a time, in clear agreement with the idea that atoms emit localized light quanta in random directions.

Furthermore, it has been confirmed that when an atom emits light (i.e., a photon), it recoils in the opposite direction, just as a pistol recoils when it fires a bullet. In Fig. 3.20 atoms pumped up with excess energy (i.e., excited, p. 63) are formed into a narrow beam. They soon spontaneously radiate photons in random directions and are themselves kicked backward, often laterally away from the beam. The resulting spread of

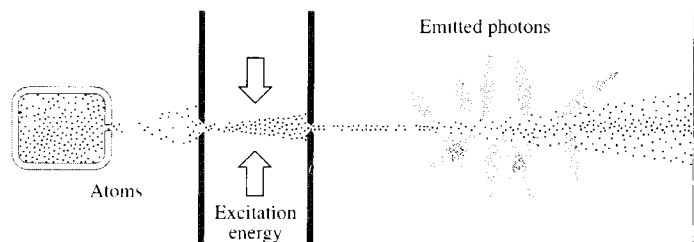


Figure 3.20 When so-called excited atoms forming a narrow beam radiate photons, they recoil laterally and the beam spreads out. Alternatively if the beam is formed of atoms that are not excited (i.e., they are in their ground states), it remains narrow all the way to the screen.

the beam is a quantum-mechanical effect inconsistent with the classical picture of the emission of a symmetrical wave.

A Barrage of Photons

When we analyze phenomena involving the activity of immense numbers of participants, the use of statistical techniques is often the only practical way to proceed. In addition to the classical Maxwell-Boltzmann statistics (for distinguishable particles), there are two kinds of quantum-mechanical statistics (for indistinguishable particles): Bose-Einstein and Fermi-Dirac. The first treats particles that *are not* subject to the Pauli Exclusion Principle (i.e., particles that have zero or integer spins). Fermi-Dirac statistics treats particles that *are* subject to the Pauli Exclusion Principle (i.e., those that have spins that are odd integer multiples of $\frac{1}{2}$). Photons are called **bosons**, they are spin-1 particles, and the manner in which they group together obeys Bose-Einstein statistics. Similarly, electrons are **fermions**; they are spin- $\frac{1}{2}$ particles that obey Fermi-Dirac statistics.

Microparticles have defining physical characteristics such as mass, charge, and spin—characteristics that do not change. When these are given, we have completely specified the kind of particle being considered. Alternatively, there are alterable properties of any given microparticle that describe its momentary condition such as energy, momentum, and spin orientation. When all of these alterable quantities are given, we have specified the particular **state** the particle happens to be in at the moment.

Fermions are committed loners; *only one fermion can occupy any given state*. By comparison, bosons are gregarious; *any number of them can occupy the same state, and moreover, they actually tend to cluster close together*. *When a very large number of photons occupy the same state, the inherent granularity of the light beam essentially vanishes and the electromagnetic field appears as the continuous medium of an electromagnetic wave*. Thus we can associate a monochromatic (monoenergetic) plane wave with a stream of photons having a high population density, all progressing in the same state (with the same energy, same frequency, same momentum, same direction). ***Different monochromatic plane waves represent different photon states.***

Unlike the photon, because electrons are fermions, large numbers of them cannot cluster tightly in the same state, and a monoenergetic beam of electrons does not manifest itself on a macroscopic scale as a continuous classical wave. In that regard, EM-radiation is quite distinctive.

For a uniform monochromatic light beam of frequency ν , the quantity $I/h\nu$ is the average number of photons impinging on a unit area (normal to the beam) per unit time, namely, the photon flux density. More realistically, if the beam is quasi-monochromatic (p. 17) with an average frequency ν_0 , its **mean photon flux density** is $I/h\nu_0$. Given that an incident quasi-monochromatic beam has a cross-sectional area A , its **mean photon flux** is

$$\Phi = AI/h\nu_0 = P/h\nu_0 \quad (3.49)$$

where P is the **optical power** of the beam in watts. The mean photon flux is the average number of photons arriving per unit of time (Table 3.1). For example, a small 1.0 mW He-Ne laserbeam with a mean wavelength of 632.8 nm delivers a mean photon flux of $P/h\nu_0 = (1.0 \times 10^{-3} \text{ W})/[(6.626 \times 10^{-34} \text{ J}\cdot\text{s})(2.998 \times 10^8 \text{ m/s})/(632.8 \times 10^{-9} \text{ m})] = 3.2 \times 10^{15}$ photons per second.

Imagine a uniform beam of light having a constant irradiance (and therefore a constant mean photon flux) incident on a screen. The energy of the beam is deposited on the screen in a random flurry of minute bursts. Individually, the incoming photons register at locations on the plane that are totally unpredictable, and arrive at moments in time that are equally unpredictable. It looks as if the beam is composed of a random stream of photons, but that conclusion, however tempting, goes beyond the observation. What *can* be said is that the light delivers its energy in a staccato of impacts that are random in space and time across the beam.

Suppose that we project a light pattern onto the screen; it might be a set of interference fringes or the image of a woman's face. The barrage of photons forming the image is a

TABLE 3.1 The Mean Photon Flux Density for a Sampling of Common Sources

Light Source	Mean Photon Flux Density Φ/A in units of (photons/s·m ²)
Laserbeam (10 mW, He-Ne, focused to 20 μm)	10^{26}
Laserbeam (1 mW, He-Ne)	10^{21}
Bright sunlight	10^{18}
Indoor light level	10^{16}
Twilight	10^{14}
Moonlight	10^{12}
Starlight	10^{10}



Figure 3.21 These photographs (which were electronically enhanced) are a compelling illustration of the granularity displayed by light in its interaction with matter. Under exceedingly faint illumination, the pattern (each spot corresponding to one photon) seems almost random, but as the light level increases the quantal character of the process gradually becomes obscured. (See *Advances in Biological and Medical Physics* V, 1957, 211–242.) (Photos courtesy Radio Corporation of America.)

statistical tumult; we cannot predict when a photon will arrive at any given location. But we can determine the likelihood of one or more photons hitting any particular point during a substantial time interval. *At any location on the screen, the measured (or classically computed) value of the irradiance is proportional to the probability of detecting a photon at that location* (p. 137).

Figure 1.8, which is a pictorial record of the arrival of individual photons, was produced using a special kind of photomultiplier tube. To underscore the inherent photonic nature of radiant energy, let's now use an entirely different and more straightforward photographic approach to record the incidence of light. A photographic emulsion contains a distribution of microscopic ($\approx 10^{-6}$ m) silver halide crystals, each comprising approximately 10^{10} Ag atoms. A single photon can interact with such a crystal, disrupting a silver-halogen bond and freeing up an Ag atom. One or more silver atoms then serve as a development center on the exposed crystal. The film is developed using a chemical reducing agent. It dissolves each exposed crystal, depositing at that site all of its Ag atoms as a single clump of the metal.

Figure 3.21 shows a series of photographs taken with increasing amounts of illumination. Using extremely dim light, a few thousand photons, the first picture is composed of roughly as many silver clumps, making a pattern that only

begins to suggest an overall image. As the number of participating photons goes up (roughly by a factor of 10 for each successive picture), the image becomes increasingly smooth and recognizable. When there are tens of millions of photons forming the image, the statistical nature of the process is lost and the picture assumes a familiar continuous appearance.

Photon Counting

What, if anything, can be said about the statistical nature of the barrage of photons delivered as a beam of light? To answer that question, researchers within the last 35 years or so have conducted experiments in which they literally counted individual photons. What they found was that the pattern of arrival of photons was characteristic of the type of source.* We cannot go into the theoretical details here, but it is informative at least to look at the results for the two extreme cases of what is often called *coherent* and *chaotic* light.

* See P. Koczyk, P. Wiewior, and C. Radzewicz, "Photon counting statistics—Undergraduate experiment," *Am. J. Phys.* **64** (3), 240 (1996) and A. C. Funk and M. Beck, "Sub-Poissonian photocurrent statistics: Theory and undergraduate experiment," *Am. J. Phys.* **65** (6), 492 (1997).

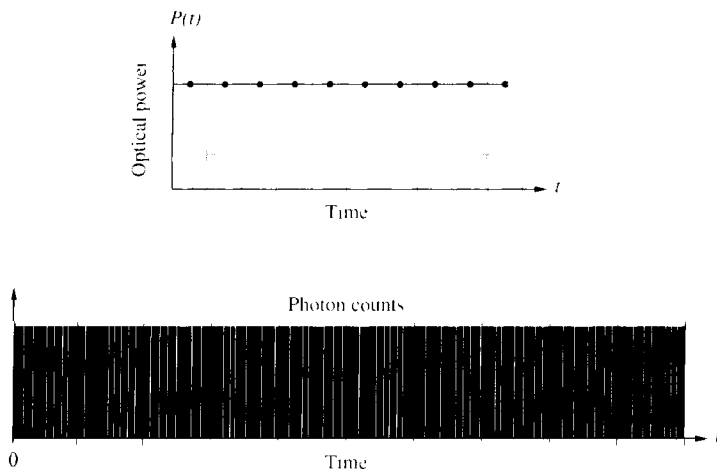


Figure 3.22 A constant optical power and the corresponding random set of photon counts. The arrival of each photon is an independent event.

Consider an ideal continuous laserbeam of *constant irradiance*; remember that irradiance is a time-averaged quantity via Eq. (3.46). The beam has a constant optical power P —which is also a time-averaged quantity—and, from Eq. (3.49), a corresponding mean photon flux Φ . Figure 3.22 depicts the random arrival of photons on a time scale that is short compared to the interval over which the irradiance is averaged. Thus it is possible for the macroscopic quantity P to be measured to be constant, even though there is an underlying discontinuous transfer of energy.

Now pass the beam through a shutter that stays open for a short sampling time T (which might be in the range from about $10\ \mu\text{s}$ to perhaps $10\ \text{ms}$), and count the number of photons arriving at a photodetector during that interval. After a brief pause repeat the procedure, and do it again and again, tens of thousands of times. The results are presented in a histogram (Fig. 3.23), where the number of trials in which N photons

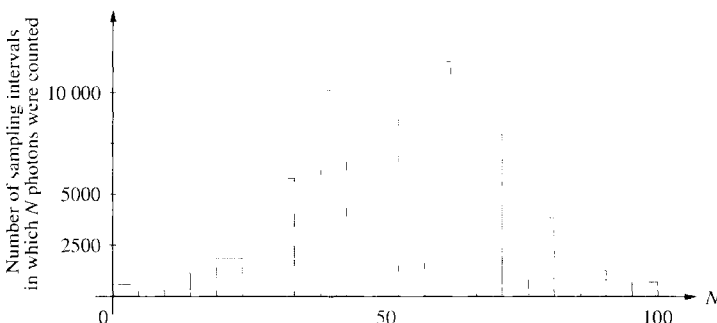


Figure 3.23 A typical histogram showing the probability or photon-count distribution for a beam of constant irradiance.

were counted is plotted against N . Few trials register either very few photons or very many photons. On average, the number of photons per trial is $N_{\text{av}} = \Phi T = PT/h\nu_0$. The shape of the data plot, which can be derived using probability theory, closely approximates the well-known *Poisson distribution*. It represents a graph of the probability that the detector (during a trial interval lasting a time T) will record zero photons, one photon, two photons, and so forth.

The Poisson distribution is the same symmetrical curve one gets when counting either the number of particles randomly emitted by a long-lived radioactive sample, or the number of raindrops randomly descending on an area in a steady shower. It's also the curve of the probability of getting a head, plotted against the number of heads occurring (N), for a coin tossed more than about 20 times. Thus with $N_{\text{max}} = 20$ the highest probability occurs near the average value N_{av} , namely, at $\frac{1}{2}N_{\text{max}}$ or 10 and the lowest at $N = 0$ and $N = 20$. The most probable value will be 10 heads out of 20 tosses, and the likelihood of getting either no heads or all heads is vanishingly small. It would seem that however an ideal laser produces light, it generates a stream of photons whose individual arrival is random and statistically independent. For reasons that will be explored later, an ideal monochromatic beam—a monochromatic plane wave—is the epitome of what is known as **coherent light**.

Not surprisingly, *the statistical distribution of the number of photons arriving at a detector depends on the nature of the source*; it is fundamentally different for an ideal *coherent source* at one extreme, as compared to an equally idealized completely incoherent or *chaotic source* at the other extreme. A stabilized laser resembles a coherent source, and an ordinary thermal source such as a lightbulb or a star or a gas discharge lamp more closely resembles a chaotic source. In the case of ordinary light, there are inherent fluctuations in the irradiance and therefore in the optical power (p. 48). These fluctuations are correlated, and the associated number of emitted photons, though random in time, is correspondingly also correlated (Fig. 3.24). The greater the optical power, the greater the number density of photons. Because the arrival of photons at the detector is not a succession of independent events, Bose-Einstein statistics apply (Fig. 3.25). Here the most likely number of counts per interval is zero, whereas, ideally, for laser light the most likely number of photons to be measured during a sampling interval equals the average number recorded. Thus even if a beam of laser light and a beam of ordinary light have the same average irradiance and the same frequency spectrum, they are still inherently distinguishable—a result that extends beyond classical theory.

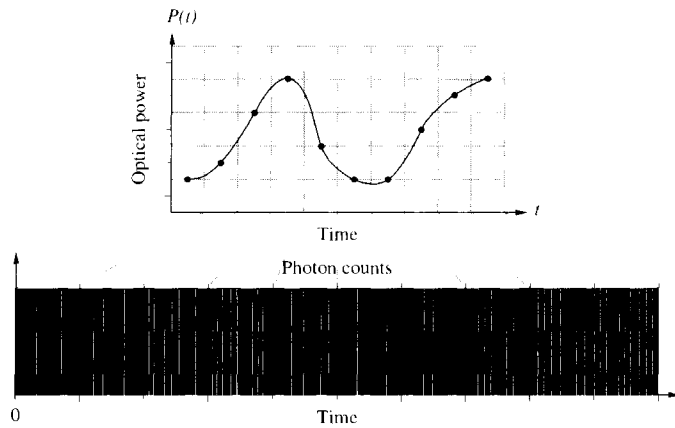


Figure 3.24 A time-varying optical power and the corresponding random set of photon counts. Now there are fluctuations that are correlated, and the photon arrivals are no longer independent.

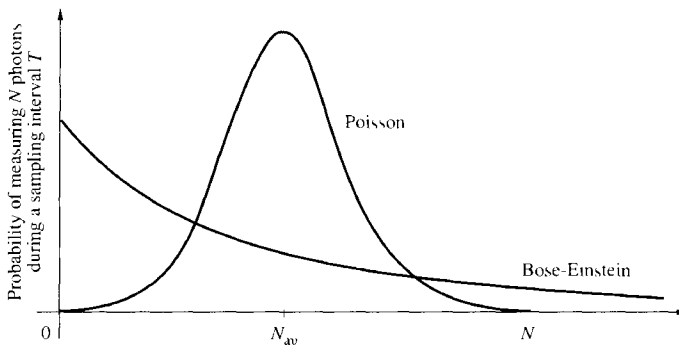


Figure 3.25 Poisson and Bose-Einstein photon counting distributions.

3.3.4 Radiation Pressure and Momentum

As long ago as 1619, Johannes Kepler proposed that it was the pressure of sunlight that blew back a comet's tail so that it always pointed away from the Sun. That argument particularly appealed to the later proponents of the corpuscular theory of light. After all, they envisioned a beam of light as a stream of particles, and such a stream would obviously exert a force as it bombarded matter. For a while it seemed as though this effect might at last establish the superiority of the corpuscular over the wave theory, but all the experimental efforts to that end failed to detect the force of radiation, and interest slowly waned.

Ironically, it was Maxwell in 1873 who revived the subject by establishing theoretically that waves do indeed exert

pressure. “In a medium in which waves are propagated,” wrote Maxwell, “there is a pressure in the direction normal to the waves, and numerically equal to the energy in a unit of volume.”

When an electromagnetic wave impinges on some material surface, it interacts with the charges that constitute bulk matter. Regardless of whether the wave is partially absorbed or reflected, it exerts a force on those charges and hence on the surface itself. For example, in the case of a good conductor, the wave's electric field generates currents, and its magnetic field generates forces on those currents.

It's possible to compute the resulting force via Electromagnetic Theory, whereupon Newton's Second Law (which maintains that force equals the time rate-of-change of momentum) suggests that the *wave itself carries momentum*. Indeed, whenever we have a flow of energy, it's reasonable to expect that there will be an associated momentum—the two are the related time and space aspects of motion.

As Maxwell showed, the **radiation pressure**, $\mathcal{P}$, equals the energy density of the electromagnetic wave. From Eqs. (3.31) and (3.32), for a vacuum, we know that

$$u_E = \frac{\epsilon_0}{2} E^2 \quad \text{and} \quad u_B = \frac{1}{2\mu_0} B^2$$

Since $\mathcal{P} = u = u_E + u_B$,

$$\mathcal{P} = \frac{\epsilon_0}{2} E^2 + \frac{1}{2\mu_0} B^2$$

Alternatively, using Eq. (3.37), we can express the pressure in terms of the magnitude of the Poynting vector, namely,

$$\mathcal{P}(t) = \frac{S(t)}{c} \quad (3.50)$$

Notice that this equation has the units of power divided by area, divided by speed—or equivalently, force times speed divided by area and speed, or just force over area. **This is the instantaneous pressure that would be exerted on a perfectly absorbing surface by a normally incident beam.**

Inasmuch as the $\vec{E}$ - and $\vec{B}$ -fields are rapidly varying, $S(t)$ is rapidly varying, so it is eminently practical to deal with the average radiation pressure, namely,

$$\langle \mathcal{P}(t) \rangle_T = \frac{\langle S(t) \rangle_T}{c} = \frac{I}{c} \quad (3.51)$$

expressed in newtons per square meter. This same pressure is exerted on a source that itself is radiating energy.

Referring back to Fig. 3.16, if p is momentum, the force

exerted by the beam on an absorbing surface is

$$A\mathcal{P} = \frac{\Delta p}{\Delta t} \quad (3.52)$$

If p_v is the momentum per unit volume of the radiation, then an amount of momentum $\Delta p = p_v(c \Delta t A)$ is transported to A during each time interval Δt , and

$$A\mathcal{P} = \frac{p_v(c \Delta t A)}{\Delta t} = A \frac{S}{c}$$

Hence the volume density of electromagnetic momentum is

$$p_v = \frac{S}{c^2} \quad (3.53)$$

When the surface under illumination is **perfectly reflecting**, the beam that entered with a velocity $+c$ will emerge with a velocity $-c$. This corresponds to twice the change in momentum that occurs on absorption, and hence

$$\langle \mathcal{P}(t) \rangle_T = 2 \frac{\langle S(t) \rangle_T}{c}$$

Notice, from Eqs. (3.50) and (3.52), that if some amount of energy $\mathcal{E}$ is transported per square meter per second, then there will be a corresponding momentum $\mathcal{E}/c$ transported per square meter per second.

In the photon picture, each quantum, has an energy $\mathcal{E} = h\nu$. We can then expect a photon to carry a momentum

$$p = \frac{\mathcal{E}}{c} = \frac{h}{\lambda} \quad (3.54)$$

Its vector momentum would be

$$\vec{p} = \hbar \vec{k}$$

where $\vec{k}$ is the propagation vector and $\hbar \equiv h/2\pi$. This all fits in rather nicely with Special Relativity, which relates the mass m , energy, and momentum of a particle by

$$\mathcal{E} = [(cp)^2 + (mc^2)^2]^{1/2}$$

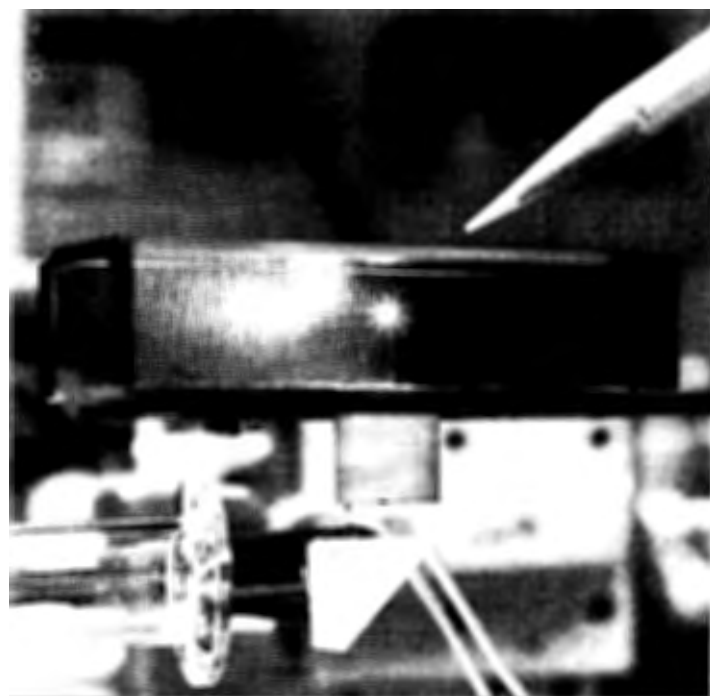
For a photon $m = 0$ and $\mathcal{E} = cp$.

These quantum-mechanical ideas have been confirmed experimentally utilizing the Compton Effect, which detects the energy and momentum transferred to an electron upon interaction with an individual X-ray photon (see photo on page 52).

The average flux density of electromagnetic energy from the Sun impinging normally on a surface just outside the

Earth's atmosphere is about 1400 W/m^2 . Assuming complete absorption, the resulting pressure would be $4.7 \times 10^{-6} \text{ N/m}^2$, or 1.8×10^{-9} ounce/cm², as compared with, say, atmospheric pressure of about 10^5 N/m^2 . The pressure of solar radiation at the Earth is tiny, but it is still responsible for a substantial planetwide force of roughly 10 tons. Even at the very surface of the Sun, radiation pressure is relatively small (see Problem 3.31). As one might expect, it becomes appreciable within the blazing body of a large bright star, where it plays a significant part in supporting the star against gravity. Despite the Sun's modest flux density, it nonetheless can produce appreciable effects over long-acting times. For example, had the pressure of sunlight exerted on the Viking spacecraft during its journey been neglected, it would have missed Mars by about 15 000 km. Calculations show that it is even feasible to use the pressure of sunlight to propel a space vehicle among the inner planets.* Ships with immense reflecting sails driven by solar radiation pressure may some day ply the dark sea of local space.

*The charged-particle flux called the "solar wind" is 1000 to 100 000 times less effective in providing a propulsive force than is sunlight.



The tiny starlike speck is a minute (one-thousandth of an inch diameter) transparent glass sphere suspended in midair on an upward 250-mW laserbeam. (Photo courtesy Bell Laboratories.)

The pressure exerted by light was actually measured as long ago as 1901 by the Russian experimenter Pyotr Nikolaevich Lebedev (1866–1912) and independently by the Americans Ernest Fox Nichols (1869–1924) and Gordon Ferrie Hull (1870–1956). Their accomplishments were formidable, considering the light sources available at the time. Nowadays, with the advent of the laser, light can be focused down to a spot size approaching the theoretical limit of about one wavelength in radius. The resulting irradiance, and therefore the pressure, is appreciable, even with a laser rated at just a few watts. It has thus become practical to consider radiation pressure for all sorts of applications, such as separating isotopes, accelerating particles, cooling and trapping atoms (p. 65), and even optically levitating small objects.

Light can also transport angular momentum, but that raises a number of issues that will be treated later (p. 330).

3.4 Radiation

Electromagnetic radiation comes in a broad range of wavelengths and frequencies, although in vacuum it all travels at the same speed. Despite the fact that we distinguish different regions of the spectrum with names like radiowaves, microwaves, infrared, and so forth, there is only one entity, one essence of electromagnetic wave. Maxwell's Equations are independent of wavelength and so suggest no fundamental differences in kind. Accordingly, it is reasonable to look for a common source-mechanism for all EM-radiation. What we find is that the various types of radiant energy seem to have a common origin in that they are all associated with *nonuniformly moving charges*. We are, of course, dealing with waves in the electromagnetic field, and charge *is* that which gives rise to field, so this is not altogether surprising.

A stationary charge has a constant $\vec{E}$ -field, no $\vec{B}$ -field, and hence produces no radiation—where would the energy come from if it did? A uniformly moving charge has both an $\vec{E}$ - and a $\vec{B}$ -field, but it does not radiate. If you traveled along with the charge, the current would thereupon vanish, hence $\vec{B}$ would vanish, and we would be back at the previous case, uniform motion being relative. That's reasonable, since it would make no sense at all if the charge stopped radiating just because you started walking along next to it. That leaves *nonuniformly moving charges*, which assuredly do radiate. In the photon picture this is underscored by the conviction that the fundamental

interactions between substantial matter and radiant energy are between photons and charges.

We know in general that free charges (those not bound within an atom) emit electromagnetic radiation when accelerated. That much is true for charges changing speed along a straight line within a linear accelerator, sailing around in circles inside a cyclotron, or simply oscillating back and forth in a radio antenna—if a charge moves *nonuniformly, it radiates*. A free charged particle can spontaneously absorb or emit a photon, and an increasing number of important devices, ranging from the free-electron laser (1977) to the synchrotron radiation generator, utilize this mechanism on a practical level.

3.4.1 Linearly Accelerating Charges

Consider a charge moving at a constant speed. It essentially has attached to it an unchanging radial electric field and a surrounding circular magnetic field. Although at any point in space the $\vec{E}$ -field changes from moment to moment, at any instant its value can be determined by supposing that the field lines move along, fixed to the charge. Thus the field does not disengage from the charge, and there is no radiation.

The electric field of a charge at rest can be represented, as in Fig. 3.26, by a uniform, radial distribution of straight field lines. For a charge moving at a constant velocity $\vec{v}$, the field lines are still radial and straight, but they are no longer uniformly distributed. The nonuniformity becomes evident at high speeds and is usually negligible when $v \ll c$.

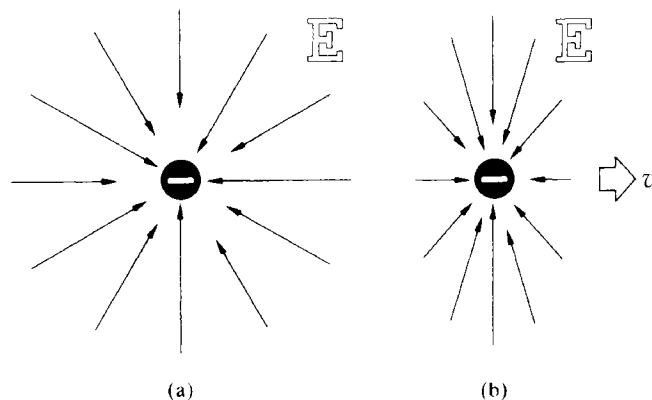


Figure 3.26 (a) Electric field of a stationary electron. (b) Electric field of a moving electron.

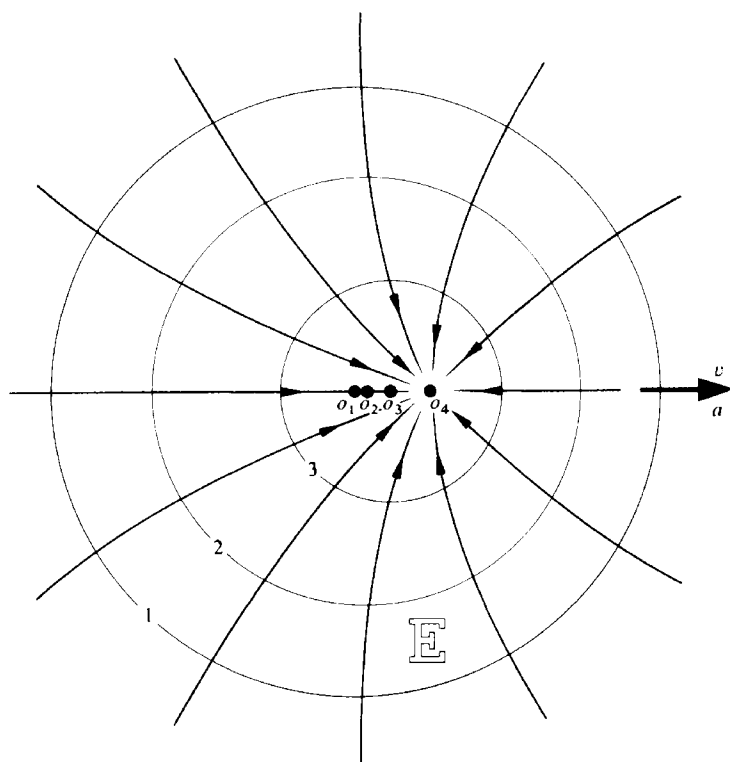


Figure 3.27 Electric field of a uniformly accelerating electron.

In contrast, Fig. 3.27 shows the field lines associated with an electron accelerating uniformly to the right. The points O_1 , O_2 , O_3 , and O_4 are the positions of the electron after equal time intervals. The field lines are now curved, and this is a significant difference. As a further contrast, Fig. 3.28 depicts the field of an electron at some arbitrary time t_2 . Before $t = 0$ the particle was always at rest at the point O . The charge was then uniformly accelerated until time t_1 , reaching a speed v , which was maintained constant thereafter. We can anticipate that the surrounding field lines will somehow carry the information that the electron has accelerated. We have ample reason to assume that this “information” will propagate at the speed c . If, for example, $t_2 = 10^{-8}$ s, no point beyond 3 m from O would be aware of the fact that the charge had even moved. All the lines in that region would be uniform, straight, and centered on O , as if the charge were still there. At time t_2 the electron is at point O_2 moving with a constant speed v . In the vicinity of O_2 the field lines must then resemble those in Fig. 3.26b. Gauss’s Law requires that the lines outside the sphere of radius ct_2 connect to those within the sphere of radius $c(t_2 - t_1)$, since there are no charges between them. It is now appar-

ent that during the interval when the particle accelerated, the field lines became distorted and a kink appeared. The exact shape of the lines within the region of the kink is of little interest here. What is significant is that there now exists a *transverse component* of the electric field $\vec{E}_T$, which propagates outward as a pulse. At some point in space the transverse electric field will be a function of time, and it will therefore be accompanied by a magnetic field.

The radial component of the electric field drops off as $1/r^2$, while the transverse component goes as $1/r$. At large distances from the charge, the only significant field will be the $\vec{E}_T$ -component of the pulse, which is known as the *radiation field*.* For a positive charge moving slowly ($v \ll c$), the electric and magnetic radiation fields can be shown to be proportional to $\vec{r} \times (\vec{r} \times \vec{a})$ and $(\vec{a} \times \vec{r})$, respectively, where $\vec{a}$ is the acceleration. For a negative charge the reverse occurs, as shown in

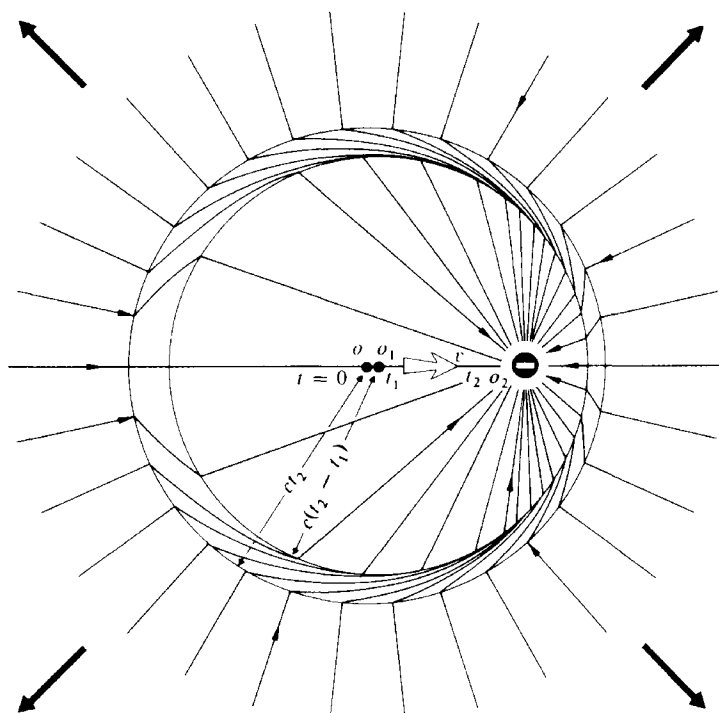


Figure 3.28 A kink in the $\vec{E}$ -field lines.

*The details of this calculation using J. J. Thomson’s method of analyzing the kink can be found in J. R. Tessman and J. T. Finnell, Jr., “Electric Field of an Accelerating Charge,” *Am. J. Phys.* **35**, 523 (1967). As a general reference for radiation, see, for example, Marion and Heald, *Classical Electromagnetic Radiation*, Chapter 7.

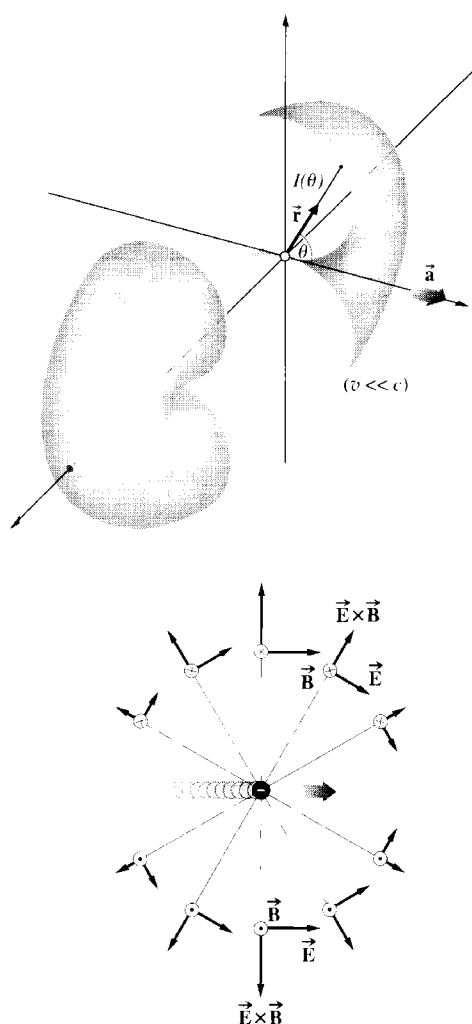


Figure 3.29 The toroidal radiation pattern of a linearly accelerating charge (split to show cross section).

Fig. 3.29. Observe that the irradiance is a function of θ and that $I(0) = I(180^\circ) = 0$ while $I(90^\circ) = I(270^\circ)$ is a maximum. **Energy is most strongly radiated perpendicular to the acceleration causing it.**

The energy that is radiated out into the surrounding space is supplied to the charge by some external agent. That agent is responsible for the accelerating force, which in turn does work on the charge.

3.4.2 Synchrotron Radiation

A free charged particle traveling on any sort of curved path is accelerating and will radiate. This provides a powerful mechanism for producing radiant energy, both naturally and in the

laboratory. The synchrotron radiation generator, a research tool developed in the 1970s, does just that. Clumps of charged particles, usually electrons or positrons, interacting with an applied magnetic field are made to revolve around a large, essentially circular track at a precisely controlled speed. The frequency of the orbit determines the fundamental frequency of the emission (which also contains higher harmonics), and that's continuously variable, more or less, as desired. Incidentally, it's necessary to use clumps of charge; *a uniform loop of current does not radiate*.

A charged particle slowly revolving in a circular orbit radiates a doughnut-shaped pattern similar to the one depicted in Fig. 3.29. Again the distribution of radiation is symmetrical around $\vec{a}$, which is now the centripetal acceleration acting inward along the radius drawn from the center of the circular orbit to the charge. And once more, **energy is most strongly radiated perpendicular to the acceleration causing it**. The higher the speed, the more an observer at rest in the laboratory will "see" the backward lobe of the radiation pattern shrink while the forward lobe elongates in the direction of motion. At speeds approaching c , the particle beam (usually with a diameter comparable to that of a straight pin) radiates essentially along a narrow cone pointing tangent to the orbit in the instantaneous direction of $\vec{v}$ (Fig. 3.30). Moreover, for $v \approx c$ the radiation will be strongly polarized in the plane of the motion.

This "searchlight," often less than a few millimeters in diameter, sweeps around as the particle clumps circle the machine, much like the headlight on a train rounding a turn. With each revolution the beam momentarily ($< \frac{1}{2}$ ns) flashes through one of many windows in the device. As we will learn (p. 314), when a signal has a short duration it must comprise a broad range of frequencies. The result is a tremendously intense source of rapidly pulsating radiation, tunable over a wide range of frequencies, from infrared to light to X-rays.

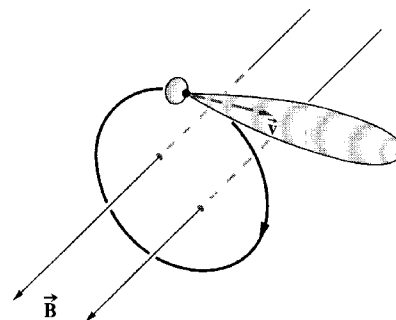


Figure 3.30 Radiation pattern for an orbiting charge.

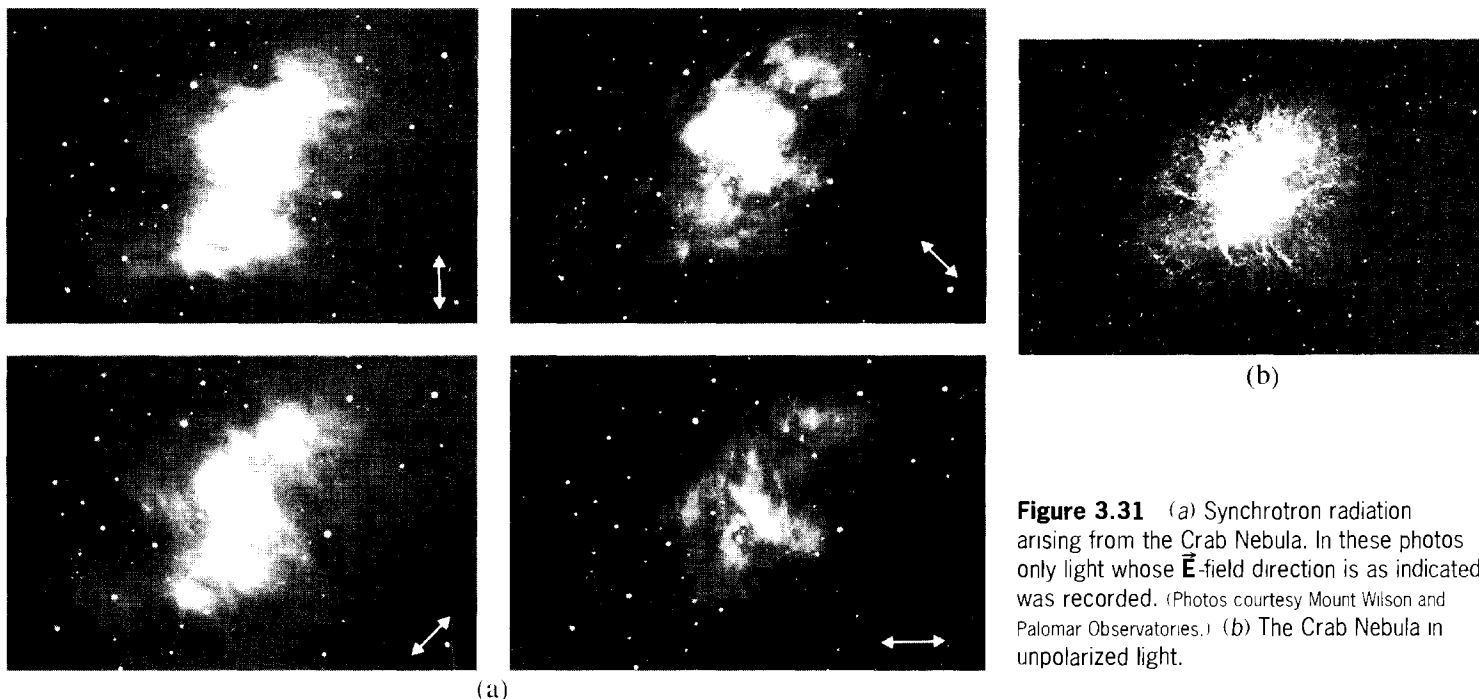


Figure 3.31 (a) Synchrotron radiation arising from the Crab Nebula. In these photos only light whose $\vec{E}$ -field direction is as indicated was recorded. (Photos courtesy Mount Wilson and Palomar Observatories.) (b) The Crab Nebula in unpolarized light.

When magnets are used to make the circulating electrons wiggle in and out of their circular orbits, bursts of high-frequency X-rays of unparalleled intensity can be created. These beams are hundreds of thousands of times more powerful than a dental X-ray (which is roughly a fraction of a watt) and can easily

burn a finger-sized hole through a 3-mm-thick lead plate.

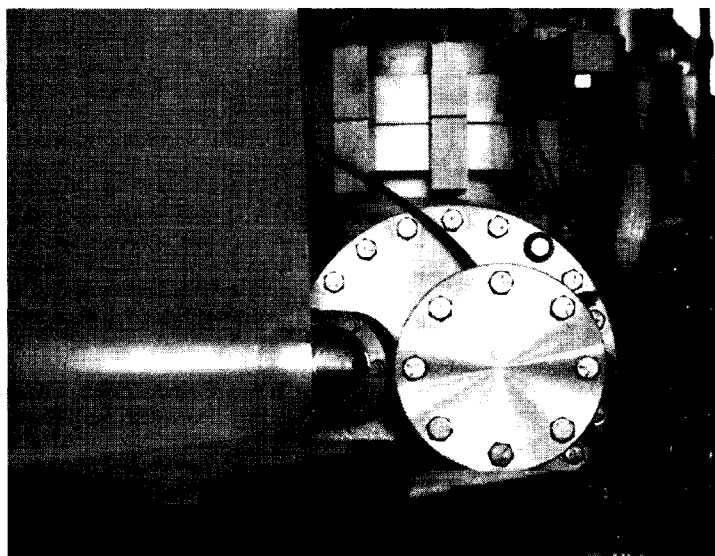
Although this technique was first used to produce light in an electron synchrotron as long ago as 1947, it took several decades to recognize that what was an energy-robbing nuisance to the accelerator people might be a major research tool in itself (see photo).

In the astronomical realm, we can expect that some regions exist that are pervaded by magnetic fields. Charged particles trapped in these fields will move in circular or helical orbits, and if their speeds are high enough, they will emit synchrotron radiation. Figure 3.31 shows five photographs of the extragalactic Crab Nebula.* Radiation emanating from the nebula extends over the range from radio frequencies to the

*The Crab Nebula is believed to be expanding debris left over after the cataclysmic death of a star. From its rate of expansion, astronomers calculated that the explosion took place in A.D. 1050. This was subsequently corroborated when a study of old Chinese records (the chronicles of the Beijing Observatory) revealed the appearance of an extremely bright star, in the same region of the sky, in A.D. 1054.

In the first year of the period Chihha, the fifth moon, the day Chichou [i.e., July 4, 1054], a great star appeared.... After more than a year, it gradually became invisible.

There is little doubt that the Crab Nebula is the remnant of that supernova.



The first beam of "light" from the National Synchrotron Light Source (1982) emanating from its ultraviolet electron storage ring.

extreme ultraviolet. Assuming the source to be trapped circulating charges, we can anticipate strong polarization effects. These are evident in the first four photographs, which were taken through a polarizing filter. The direction of the electric field vector is indicated in each picture. Since in synchrotron radiation, the emitted $\vec{E}$ -field is polarized in the orbital plane, we can conclude that each photograph corresponds to a particular uniform magnetic field orientation normal to the orbits and to $\vec{E}$.

It is believed that a majority of the low-frequency radiowaves reaching the Earth from outer space have their origin in synchrotron radiation. In 1960 radio astronomers used these long-wavelength emissions to identify a class of objects known as quasars. In 1955 bursts of polarized radiowaves were discovered emanating from Jupiter. Their origin is now attributed to spiraling electrons trapped in radiation belts surrounding the planet.

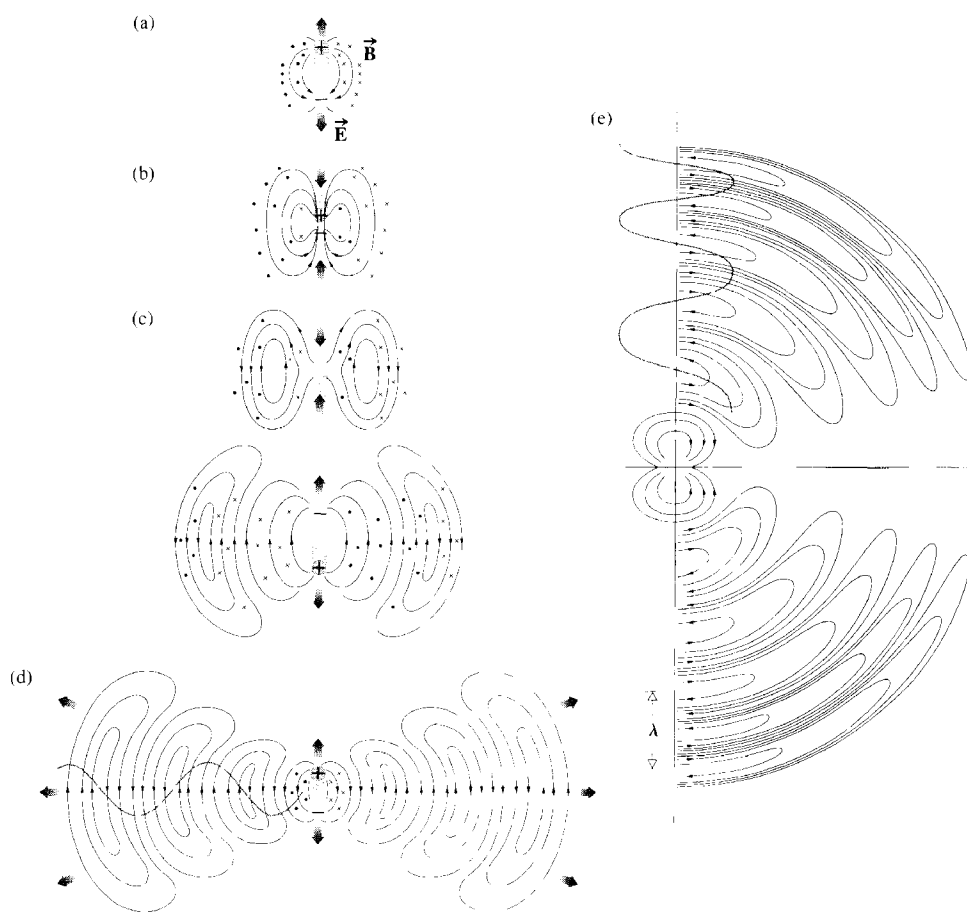


Figure 3.32 The $\vec{E}$ -field of an oscillating electric dipole.

3.4.3 Electric Dipole Radiation

Perhaps the simplest electromagnetic wave-producing mechanism to visualize is the oscillating dipole—two charges, one plus and one minus, vibrating to and fro along a straight line. And yet this arrangement is surely the most important of all.

Both light and ultraviolet radiation arise primarily from the rearrangement of the outermost, or weakly bound, electrons in atoms and molecules. It follows from the quantum-mechanical analysis that the electric dipole moment of the atom is the major source of this radiation. The rate of energy emission from a material system, although a quantum-mechanical process, can be envisioned in terms of the classical oscillating electric dipole. This mechanism is therefore central to understanding the way atoms, molecules, and even nuclei emit and absorb electromagnetic waves. Figure 3.32 schematically depicts the electric field distribution in the region of an electric

dipole. In this configuration, a negative charge oscillates linearly in simple harmonic motion about an equal stationary positive charge. If the angular frequency of the oscillation is ω , the time-dependent dipole moment $\mu(t)$ has the scalar form

$$\mu = \mu_0 \cos \omega t \quad (3.55)$$

Note that $\mu(t)$ could represent the collective moment of the oscillating charge distribution on the atomic scale or even an oscillating current in a linear television antenna.

At $t = 0$, $\mu = \mu_0 = qd$, where d is the initial maximum separation between the centers of the two charges (Fig. 3.32a). The dipole moment is actually a vector in the direction from $-q$ to $+q$. The figure shows a sequence of field line patterns as the displacement, and therefore the dipole moment decreases, then goes to zero, and finally reverses direction. When the charges effectively overlap, $\mu = 0$, and the field lines must close on themselves.

Very near the atom, the $\vec{E}$ -field has the form of a static electric dipole. A bit farther out, in the region where the closed loops form, there is no specific wavelength. The detailed treatment shows that the electric field is composed of five different terms, and things are fairly complicated. Far from the dipole, in what is called the *wave* or *radiation zone*, the field configuration is much simpler. In this zone, a fixed wavelength has been established; $\vec{E}$ and $\vec{B}$ are transverse, mutually perpendicular, and in phase. Specifically,

$$E = \frac{\mu_0 k^2 \sin \theta}{4\pi\epsilon_0} \frac{\cos(kr - \omega t)}{r} \quad (3.56)$$

and $B = E/c$, where the fields are oriented as in Fig. 3.33. The Poynting vector $\vec{S} = \vec{E} \times \vec{B}/\mu_0$ always points radially outward in the wave zone. There, the $\vec{B}$ -field lines are circles concentric with, and in a plane perpendicular to, the dipole axis. This is understandable, since $\vec{B}$ can be considered to arise from the time-varying oscillator current.

The irradiance (radiated radially outward from the source) follows from Eq. (3.44) and is given by

$$I(\theta) = \frac{\mu_0^2 \omega^4}{32\pi^2 c^3 \epsilon_0} \frac{\sin^2 \theta}{r^2} \quad (3.57)$$

again an inverse-square-law dependence on distance. The angular flux density distribution is toroidal, as in Fig. 3.29. The axis along which the acceleration takes place is the symmetry axis of the radiation pattern. Notice the dependence of the irradiance on ω^4 —*the higher the frequency, the stronger the radiation*. That feature will be important when we consider scattering.

It's not difficult to attach an AC generator between two conducting rods and send currents of free electrons oscillating up and down that “transmitting antenna.” Figure 3.34a shows

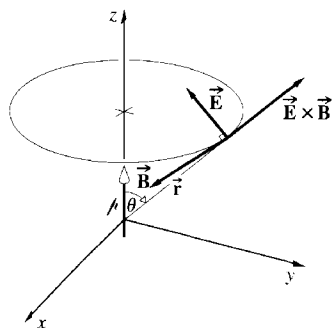


Figure 3.33 Field orientations for an oscillating electric dipole.

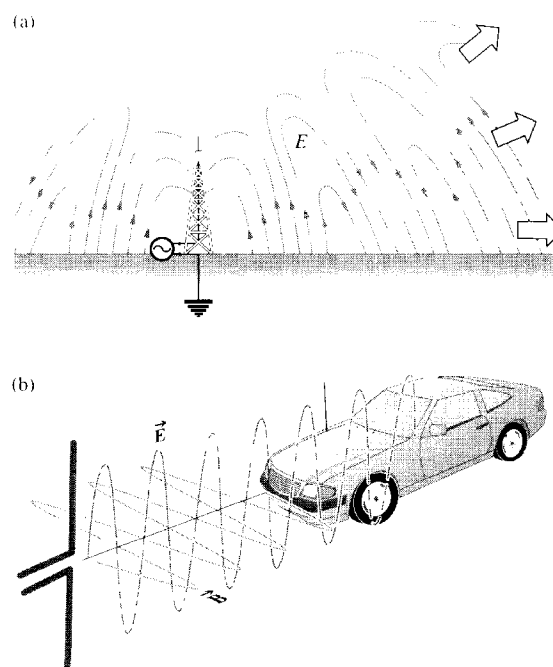


Figure 3.34 (a) Electromagnetic waves from a transmitting tower. (b) Automobiles often have radio antennas that stick straight up about a meter. The vertically oscillating electric field of a passing radiowave induces a voltage along the length of the antenna, and that becomes the input signal to the receiver.

the arrangement carried to its logical conclusion—a fairly standard AM radio tower. An antenna of this sort will function most efficiently if its length corresponds to the wavelength being transmitted or, more conveniently, to $\frac{1}{2}\lambda$. The radiated wave is then formed at the dipole in synchronization with the oscillating current producing it. AM radiowaves are unfortunately several hundred meters long. Consequently, the antenna shown in the figure has half the $\frac{1}{2}\lambda$ -dipole essentially buried in the Earth. That at least saves some height, allowing the device to be built only $\frac{1}{4}\lambda$ tall. Moreover, this use of the Earth also generates a so-called *ground wave* that hugs the planet's surface, where most people with radios are likely to be located. A commercial station usually has a range somewhere between 25 and 100 miles.

3.4.4 The Emission of Light from Atoms

Surely the most significant mechanism responsible for the natural emission and absorption of radiant energy—especially of light—is the *bound charge*, electrons confined within atoms.

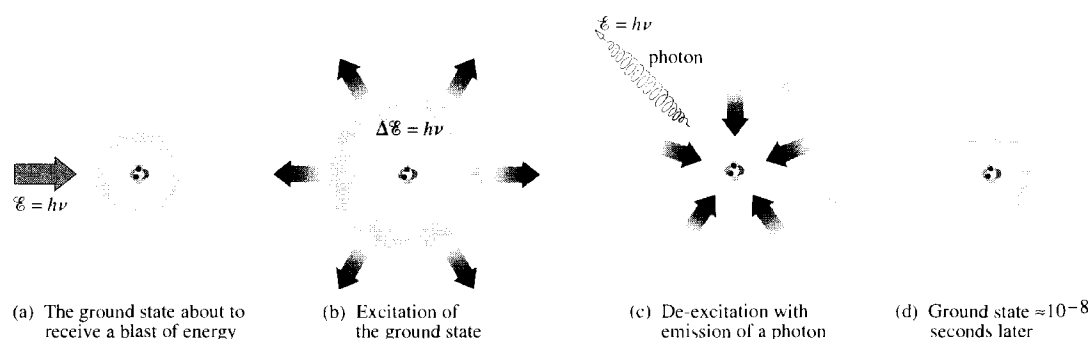


Figure 3.35 The excitation of an atom. (a) Energy in the amount $h\nu$ is delivered to the atom. (b) Since this matches the energy needed to reach an excited state, the atom absorbs the energy and attains a higher energy level. (c) With the emission of a photon, it drops back (d) and returns to the ground state in about 10^{-8} s.

These minute negative particles, which surround the massive positive nucleus of each atom, constitute a kind of distant, tenuous charged cloud. Much of the chemical and optical behavior of ordinary matter is determined by its outer or valence electrons. The remainder of the cloud is ordinarily formed into “closed,” essentially unresponsive, shells around and tightly bound to the nucleus. These closed or filled shells are made up of specific numbers of electron pairs. Even though it is not completely clear what occurs internally when an atom radiates, we do know with some certainty that light is emitted during readjustments in the outer charge distribution of the electron cloud. This mechanism is ultimately the predominant source of light in the world.

Usually, an atom exists with its clutch of electrons arranged in some stable configuration that corresponds to their lowest energy distribution or *level*. Every electron is in the lowest possible energy state available to it, and the atom as a whole is in its so-called **ground-state** configuration. There it will likely remain indefinitely, if left undisturbed. Any mechanism that pumps energy into the atom will alter the ground state. For instance, a collision with another atom, an electron, or a photon can affect the atom’s energy state profoundly. An atom can exist with its electron cloud in only certain specific configurations corresponding to only certain values of energy. In addition to the ground state, there are higher energy levels, the **excited states**, each associated with a specific cloud configuration and a specific well-defined energy. When one or more electrons occupies a level higher than its ground-state level, the atom is said to be **excited**—a condition that is inherently unstable and temporary.

At low temperatures, atoms tend to be in their ground state; at progressively higher temperatures, more and more of them will become excited through atomic collisions. This sort of

mechanism is indicative of a class of relatively gentle excitations—glow discharge, flame, spark, and so forth—which energize only the outermost unpaired valence electrons. We will initially concentrate on these outer electron transitions, which give rise to the emission of light, and the nearby infrared and ultraviolet.

When enough energy is imparted to an atom (typically to the valence electron), whatever the cause, the atom can react by suddenly ascending from a lower to a higher energy level (Fig. 3.35). The electron will make a very rapid transition, a **quantum jump**, from its ground-state orbital configuration to one of the well-delineated excited states, one of the quantized rungs on its energy ladder. As a rule, *the amount of energy taken up in the process equals the energy difference between the initial and final states, and since that is specific and well defined, the amount of energy that can be absorbed by an atom is quantized* (i.e., limited to specific amounts). This state of atomic excitation is a short-lived resonance phenomenon. Usually, after about 10^{-8} s or 10^{-9} s, the excited atom spontaneously relaxes back to a lower state, most often the ground state, losing the excitation energy along the way. This energy readjustment can occur by way of the emission of light or (especially in dense materials) by conversion to thermal energy through interatomic collisions within the medium. (As we’ll soon see, this latter mechanism results in the absorption of light at the resonant frequency and the transmission or reflection of the remaining frequencies—it’s responsible for most of the coloration in the world around us.)

If the atomic transition is accompanied by the emission of light (as it is in a rarefied gas), *the energy of the photon exactly matches the quantized energy decrease of the atom*. That corresponds to a specific frequency, by way of $\Delta\mathcal{E} = h\nu$, a frequency associated with both the photon and the atomic transi-

tion between the two particular states. This is said to be a **resonance frequency**, one of several (each with its own likelihood of occurring) at which the atom very efficiently absorbs and emits energy. The atom radiates a quantum of energy that presumably is created spontaneously, on the spot, by the shifting electron.

Even though what occurs during the atom-transition interval of 10^{-8} s is far from clear, it can be helpful to imagine the orbital electron somehow making its downward energy transition via a gradually damped oscillatory motion at the specific resonance frequency. The radiated light can then be envisioned in a semiclassical way as emitted in a short oscillatory directional pulse, or **wavetrain**, lasting less than roughly 10^{-8} s—a picture that is in agreement with certain experimental observations (see Section 7.4.2, and Fig. 7.36). It's useful to think of this electromagnetic pulse as associated in some inextricable fashion with the photon. In a way, the pulse is a semiclassical representation of the manifest wave nature of the photon. But the two are *not* equivalent in all respects: the electromagnetic wavetrain is a classical creation that describes the propagation and spatial distribution of light extremely well; yet its energy is not quantized, and that is an essential characteristic of the photon. So when we consider photon wavetrains, keep in mind that there is more to the idea than just a classical oscillatory pulse of electromagnetic wave. Of course, the reason even to introduce the notion of the emission of wavetrains is to have a basis for talking about the frequency of the light. This is perhaps the central problem in any naïve photon model: What agency manifests the frequency?

The emission spectra of single atoms or low-pressure gases, whose atoms do not interact appreciably, consist of sharp “lines,” that is, fairly well-defined frequencies characteristic of the atoms. There is always some frequency broadening of that radiation due to atomic motion, collisions, and so forth; hence it's never precisely monochromatic. Generally, however, the atomic transition from one level to another is characterized by the emission of a well-defined, narrow range of frequencies. On the other hand, the spectra of solids and liquids, in which the atoms are interacting with one another, are broadened into wide frequency bands. When two atoms are brought close together, the result is a slight shift in their respective energy levels because they act on each other. The many interacting atoms in a solid create a tremendous number of such shifted levels, in effect spreading out each of their original levels, blurring them into essentially continuous bands. Materials of this nature emit and absorb over broad ranges of frequencies.

Optical Cooling

The linear momentum carried by photons can be transferred to moving atoms or ions, thereby drastically changing their motion. After about ten thousand absorption and subsequent emission cycles, an atom, which was originally moving at perhaps 700 m/s, can be slowed to near zero speed. Since, in general, temperature is proportional to the average kinetic energy (KE) of the particles constituting a system, this process is called **optical** or **laser cooling**. With it, KE temperatures in the microkelvin range are attainable. Laser cooling has become the basis for a variety of applications including the atomic clock, the atom interferometer, and the focusing of atomic beams. For us it brings together the ideas of Sections 3.3.4 and 3.4.4 in a compelling, practical way.

Figure 3.36 depicts a beam of atoms, each of mass m traveling with a velocity $\vec{v}$, colliding with a counterdirected beam of laser photons having a propagation vector $\vec{k}_L$. The laser frequency ν_L is selected to be just beneath the resonant frequency (ν_0) of the atoms. Because of its motion, any particular atom “sees” an oncoming photon with a frequency that is Doppler-shifted* upward by an amount $|\vec{k}_L \cdot \vec{v}|/2\pi = \nu_L v/c$. When the laser frequency is tuned so that $\nu_0 = \nu_L(1 + v/c)$, collisions with the photons will resonate the atoms. In the process, each photon transfers its momentum of $\hbar\vec{k}_L$ to the absorbing atom (see p. 57) whose speed is thereupon reduced by an amount Δv where $m\Delta v = \hbar k_L$.

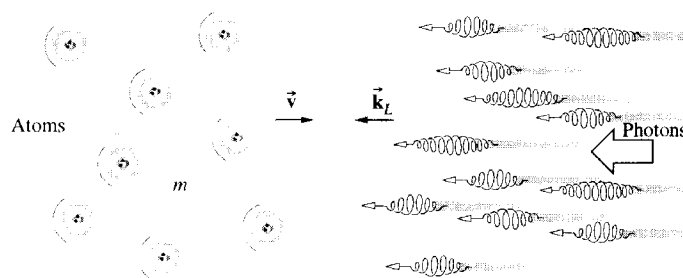


Figure 3.36 A stream of atoms colliding with a laserbeam in a process called laser cooling.

*Imagine an observer moving at v_0 , toward a source that is sending out waves having a speed v at a frequency ν_s . As a result of the Doppler Effect, he will experience a frequency $\nu_o = \nu_s(v + v_0)/v$. For more of the details, see almost any introductory physics text, for example, E. Hecht, *Physics: Calculus*, Sect. 11.11.

The cloud of atoms is not very dense, and each excited atom can drop back to its ground state with the spontaneous emission of a photon of energy $h\nu_0$. This emission is randomly directed, and so although the atom recoils, the average amount of momentum regained by it over thousands of cycles tends to zero. The change in momentum of the atom per photon absorption-emission cycle is therefore effectively $\hbar\mathbf{k}_L$, and it slows down. In each cycle (as seen by someone at rest in the lab), the atom absorbs a photon of energy $h\nu_L$, emits a photon of energy $h\nu_0$, and in the process loses an amount of KE corresponding to $h\nu_L v/c$, which is proportional to the Doppler Shift.

By contrast, an atom moving in the opposite direction, away from the light source, sees photons to have a frequency $\nu_L(1 - v/c)$, far enough away from ν_0 that there can be little or no absorption, and therefore no momentum gain.

Notice that the radiation pressure force is frequency dependent and the atoms experience a speed-dependent force via the Doppler Effect. That means that ν_0 and ν_L have to be kept in the proper relationship as v decreases. There are a number of very clever ways this is accomplished.

3.5 Light in Bulk Matter

The response of dielectric or nonconducting materials to electromagnetic fields is of special concern in Optics. We will, of course, be dealing with transparent dielectrics in the form of lenses, prisms, plates, films, and so forth, not to mention the surrounding sea of air.

The net effect of introducing a homogeneous, isotropic dielectric into a region of free space is to change ϵ_0 to ϵ and μ_0 to μ in Maxwell's Equations. The phase speed in the medium now becomes

$$v = 1/\sqrt{\epsilon\mu} \quad (3.58)$$

The ratio of the speed of an electromagnetic wave in vacuum to that in matter is known as the **absolute index of refraction** n :

$$n \equiv \frac{c}{v} = \sqrt{\frac{\epsilon\mu}{\epsilon_0\mu_0}} \quad (3.59)$$

In terms of the relative permittivity and relative permeability of the medium, n becomes

$$n = \sqrt{K_E K_M} \quad (3.60)$$

There are magnetic substances that are transparent in the infrared and microwave regions of the spectrum. But we are

primarily interested in materials that are transparent in the visible, and these are all essentially “nonmagnetic.” Indeed, K_M generally doesn't deviate from 1 by any more than a few parts in 10^4 (e.g., for diamond $K_M = 1 - 2.2 \times 10^{-5}$). Setting $K_M = 1$ in the formula for n results in an expression known as *Maxwell's Relation*, namely,

$$n \approx \sqrt{K_E} \quad (3.61)$$

wherein K_E is presumed to be the *static dielectric constant*. As indicated in Table 3.2, this relationship seems to work well only for some simple gases. The difficulty arises because K_E and therefore n are actually *frequency-dependent*. The dependence of n on the wavelength (or color) of light is a well-known effect called **dispersion**. It arises on a microscopic level, and so Maxwell's Equations are quite oblivious to it. Sir Isaac Newton used prisms to disperse white light into its constituent colors over three hundred years ago, and the phenomenon was well known, if not well understood, even then.

TABLE 3.2 Maxwell's Relation

Gases at 0°C and 1 atm		
Substance	$\sqrt{K_E}$	n
Air	1.000294	1.000293
Helium	1.000034	1.000036
Hydrogen	1.000131	1.000132
Carbon dioxide	1.00049	1.00045

Liquids at 20°C		
Substance	$\sqrt{K_E}$	n
Benzene	1.51	1.501
Water	8.96	1.333
Ethyl alcohol (ethanol)	5.08	1.361
Carbon tetrachloride	4.63	1.461
Carbon disulfide	5.04	1.628

Solids at room temperature		
Substance	$\sqrt{K_E}$	n
Diamond	4.06	2.419
Amber	1.6	1.55
Fused silica	1.94	1.458
Sodium chloride	2.37	1.50

Values of K_E correspond to the lowest possible frequencies, in some cases as low as 60 Hz, whereas n is measured at about 0.5×10^{15} Hz. Sodium D light was used ($\lambda = 589.29$ nm).

Scattering and Absorption

What is the physical basis for the frequency dependence of n ? The answer to that question can be found by examining the interaction of an incident electromagnetic wave with the array of atoms constituting a dielectric material. An atom can react to incoming light in two different ways, depending on the incident frequency or equivalently on the incoming photon energy ($\mathcal{E} = h\nu$). Generally, the atom will “scatter” the light, redirecting it without otherwise altering it. On the other hand, if the photon’s energy matches that of one of the excited states, the atom will absorb the light, making a quantum jump to that higher energy level. In the dense atomic landscape of ordinary gases (at pressures of about 10^2 Pa and up), solids, and liquids, it’s very likely that this excitation energy will rapidly be transferred, via collisions, to random atomic motion, thermal energy, before a photon can be emitted. This commonplace process (the taking up of a photon and its conversion into thermal energy) was at one time widely known as “absorption,” but nowadays that word is more often used to refer just to the “taking up” aspect, regardless of what then happens to the energy. Consequently, it’s now better referred to as **dissipative absorption**. All material media partake in dissipative absorption to some extent, at one frequency or another.

In contrast to this excitation process, *ground-state* or **non-resonant scattering** occurs with incoming radiant energy of other frequencies—that is, lower than the resonance frequencies. Imagine an atom in its lowest state and suppose that it interacts with a photon whose energy is too small to cause a transition to any of the higher, excited states. Despite that, the electromagnetic field of the light can be supposed to drive the electron cloud into oscillation. There is no resulting atomic transition; the atom remains in its ground state while the cloud vibrates ever so slightly at the frequency of the incident light. Once the electron cloud starts to vibrate with respect to the positive nucleus, the system constitutes an oscillating dipole and so presumably will *immediately* begin to radiate at that same frequency. The resulting scattered light consists of a photon that sails off in some direction carrying the same amount of energy as did the incident photon—the *scattering is elastic*. In effect, the atom resembles a little dipole oscillator, a model employed by Hendrik Antoon Lorentz (1878) in order to extend Maxwell’s Theory, in a classical way, to the atomic domain. If the incident light is unpolarized, the atomic oscillators scatter in random directions.

When an atom is irradiated with light, the process of excitation and spontaneous emission is rapidly repeated. In fact, with an emission lifetime of $\approx 10^{-8}$ s, an atom could emit

upward of 10^8 photons per second in a situation in which there was enough energy to keep reexciting it. Atoms have a very strong tendency to interact with resonant light (they have a large *absorption cross section*). This means that the saturation condition, in which the atoms of a low-pressure gas are constantly emitting and being reexcited, occurs at a modest value of irradiance ($\approx 10^2$ W/m²). So it’s not very difficult to get atoms firing out photons at a rate of 100 million per second.

Generally, we can imagine that in a medium illuminated by an ordinary beam of light, each atom behaves as though it was a “source” of a tremendous number of photons (scattered either elastically or resonantly) that fly off in all directions. A stream of energy like this resembles a classical spherical wave. *Thus we imagine an atom (even though it is simplistic to do so) as a point source of spherical electromagnetic waves*—provided we keep in mind Einstein’s admonition that “outgoing radiation in the form of spherical waves does not exist.”

When a material with no resonances in the visible is bathed in light, nonresonant scattering occurs, and it gives each participating atom the appearance of being a tiny source of spherical wavelets. As a rule, the closer the frequency of the incident beam is to an atomic resonance, the more strongly will the interaction occur and, in dense materials, the more energy will be dissipatively absorbed. It is precisely this mechanism of *selective absorption* (see Section 4.9) that creates much of the visual appearance of things. It is primarily responsible for the color of your hair, skin, and clothing, the color of leaves and apples and paint.

3.5.1 Dispersion

Dispersion corresponds to the phenomenon whereby the index of refraction of a medium is frequency dependent. All material media are dispersive; only vacuum is nondispersive.

Maxwell’s Theory treats substantial matter as continuous, representing its electric and magnetic responses to applied $\vec{E}$ - and $\vec{B}$ -fields in terms of constants, ϵ and μ . Consequently, K_e and K_m are also constant, and n is therefore unrealistically independent of frequency. To deal theoretically with dispersion, it’s necessary to incorporate the atomic nature of matter and to exploit some frequency-dependent aspect of that nature. Following H. A. Lorentz, the contributions of large numbers of atoms can be averaged to represent the behavior of an isotropic dielectric medium.

When a dielectric is subjected to an applied electric field,

the internal charge distribution is distorted. This corresponds to the generation of electric dipole moments, which in turn contribute to the total internal field. More simply stated, the external field separates positive and negative charges in the medium (each pair of which is a dipole), and these charges then contribute an additional field component. The resultant dipole moment per unit volume is called the **electric polarization**, $\vec{P}$. For most materials $\vec{P}$ and $\vec{E}$ are proportional and can satisfactorily be related by

$$(\epsilon - \epsilon_0)\vec{E} = \vec{P} \quad (3.62)$$

The redistribution of charge and the consequent polarization can occur via the following mechanisms. There are molecules that have a permanent dipole moment as a result of unequal sharing of valence electrons. These are known as *polar molecules*; the nonlinear water molecule is a fairly typical example (Fig. 3.37). Each hydrogen–oxygen bond is polar covalent, with the H-end positive with respect to the O-end. Thermal agitation keeps the molecular dipoles randomly oriented. With the introduction of an electric field, the dipoles align themselves, and the dielectric takes on an **orientational polarization**. In the case of *nonpolar molecules and atoms*, the applied field distorts the electron cloud, shifting it relative to the nucleus, thereby producing a dipole moment. In addition to this **electronic polarization**, there is another process that's applicable specifically to molecules, for example, the ionic crystal NaCl. In the presence of an electric field, the positive and negative ions undergo a shift with respect to each other. Dipole moments are therefore induced, resulting in what is called **ionic or atomic polarization**.

If the dielectric is subjected to an incident harmonic electromagnetic wave, its internal charge structure will experience time-varying forces and/or torques. These will be proportional to the electric field component of the wave.* For fluids that are polar dielectrics, the molecules actually undergo rapid rotations, aligning themselves with the $\vec{E}(t)$ -field. But these molecules are relatively large and have appreciable moments of inertia. At high driving frequencies ω , polar molecules will be unable to follow the field alternations. Their contributions to $\vec{P}$ will decrease, and K_E will drop markedly. The relative permittivity of water is fairly constant at approximately 80, up to about 10^{10} Hz, after which it falls off quite rapidly.

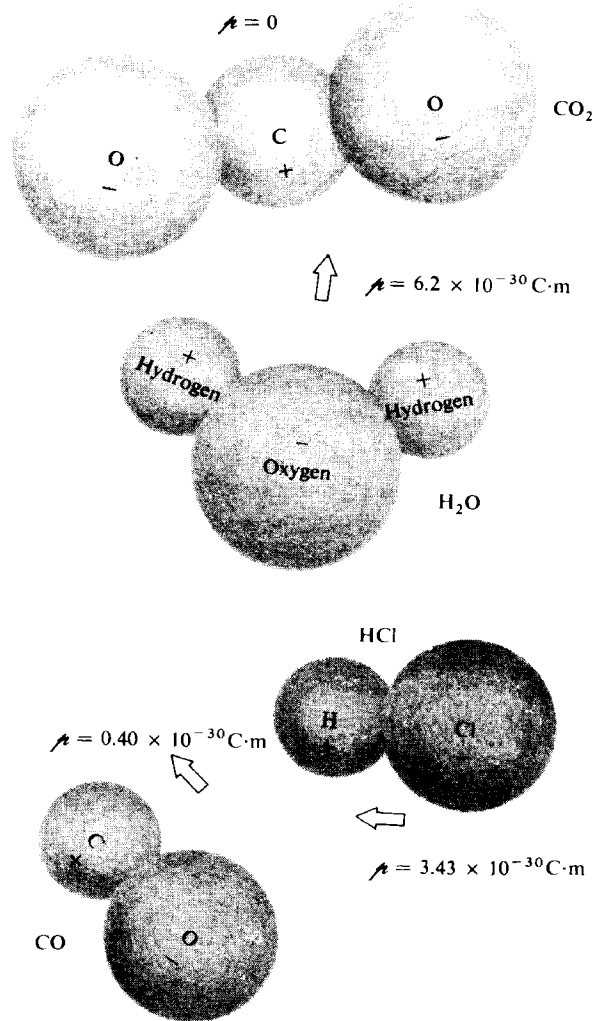


Figure 3.37 Assorted molecules and their dipole moments.

In contrast, electrons have little inertia and can continue to follow the field contributing to $K_E(\omega)$ even at optical frequencies (of about 5×10^{14} Hz). Thus the dependence of n on ω is governed by the interplay of the various electric polarization mechanisms contributing at the particular frequency. With this in mind, it is possible to derive an analytical expression for $n(\omega)$ in terms of what's happening within the medium on an atomic level.

The electron cloud of the atom is bound to the positive nucleus by an attractive electric force that sustains it in some sort of equilibrium configuration. Without knowing much more about the details of all the internal atomic interactions, we can anticipate that, like other stable mechanical systems

*Forces arising from the magnetic component of the field have the form $\vec{F}_M = q\vec{v} \times \vec{B}$ in comparison to $\vec{F}_E = q\vec{E}$ for the electric component; but $v \ll c$, so it follows from Eq. (3.30) that $\vec{F}_M$ is generally negligible.

which are not totally disrupted by small perturbations, a net force, F , must exist that returns the system to equilibrium. Moreover, we can reasonably expect that for very small displacements, x , from equilibrium (where $F = 0$), the force will be linear in x . In other words, a plot of $F(x)$ versus x will cross the x -axis at the equilibrium point ($x = 0$) and will be a straight line very close on either side. Thus for small displacements it can be supposed that the restoring force has the form $F = -k_E x$, where k_E is a kind of elastic constant much like a spring constant. Once somehow momentarily disturbed, an electron bound in this way will oscillate about its equilibrium position with a natural or resonant frequency given by $\omega_0 = \sqrt{k_E/m_e}$, where m_e is its mass. This is the oscillatory frequency of the *undriven* system.

A material medium is envisioned as an assemblage, in vacuum, of a very great many polarizable atoms, each of which is small (by comparison to the wavelength of light) and close to its neighbors. When a lightwave impinges on such a medium, each atom can be thought of as a classical **forced oscillator** being driven by the time-varying electric field $E(t)$ of the wave, which is assumed here to be applied in the x -direction. Figure 3.38b is a mechanical representation of just such an oscillator in an *isotropic medium* where the negatively charged shell is fastened to a stationary positive nucleus by identical springs. Even under the illumination of bright sunlight, the amplitude of the oscillations will be no greater than about 10^{-17} m. The force (F_E) exerted on an electron of charge q_e by the $E(t)$ field of a harmonic wave of frequency ω is of the form

$$F_E = q_e E(t) = q_e E_0 \cos \omega t \quad (3.63)$$

Notice that if the driving force is in one direction the restoring force is in the opposite direction, which is why it has a minus sign. Newton's Second Law provides the equation of motion; that is, the sum of the forces equals the mass times the acceleration:

$$q_e E_0 \cos \omega t - m_e \omega_0^2 x = m_e \frac{d^2 x}{dt^2} \quad (3.64)$$

The first term on the left is the driving force, and the second is the opposing restoring force. To satisfy this expression, x will have to be a function whose second derivative isn't very much different from x itself. Furthermore, we can anticipate that the electron will oscillate at the same frequency as $E(t)$, so we "guess" at the solution

$$x(t) = x_0 \cos \omega t$$

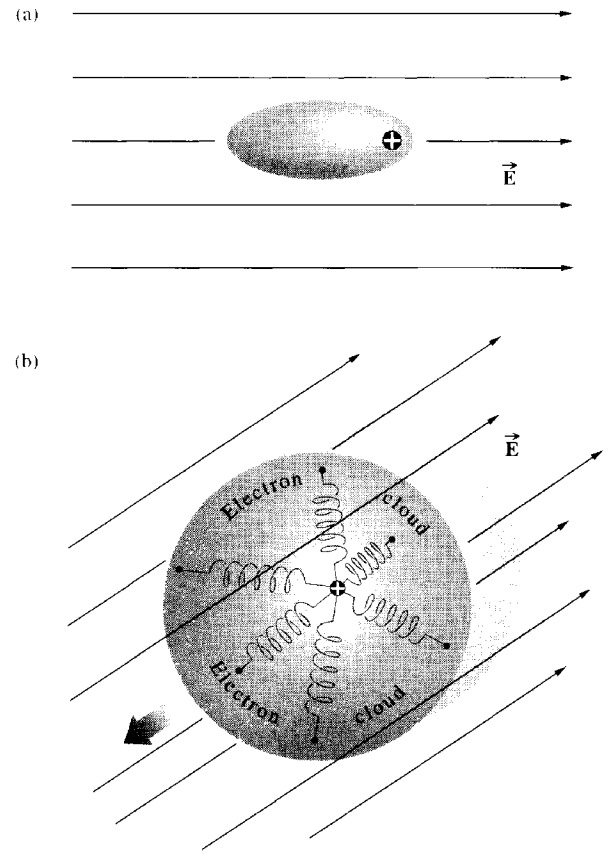


Figure 3.38 (a) Distortion of the electron cloud in response to an applied $\vec{E}$ -field. (b) The mechanical oscillator model for an isotropic medium—all the springs are the same, and the oscillator can vibrate equally in all directions.

and substitute it in the equation to evaluate the amplitude x_0 . In this way we find that

$$x(t) = \frac{q_e/m_e}{(\omega_0^2 - \omega^2)} E_0 \cos \omega t \quad (3.65)$$

or
$$x(t) = \frac{q_e/m_e}{(\omega_0^2 - \omega^2)} E(t) \quad (3.66)$$

This is the relative displacement between the negative cloud and the positive nucleus. It's traditional to leave q_e positive and speak about the displacement of the oscillator. Without a driving force (no incident wave), the oscillator will vibrate at its *resonance frequency* ω_0 . In the presence of a field whose frequency is less than ω_0 , $E(t)$ and $x(t)$ have the same sign, which means that the oscillator can follow the applied force (i.e., is in-phase with it). However, when $\omega > \omega_0$, the displacement $x(t)$ is in a direction opposite to that of the instanta-

neous force $q_e E(t)$ and therefore 180° out-of-phase with it. Remember that we are talking about oscillating dipoles where for $\omega_0 > \omega$, the relative motion of the *positive* charge is a vibration in the direction of the field. Above resonance the positive charge is 180° out-of-phase with the field, and the dipole is said to lag by π rad.

The dipole moment is equal to the charge q_e times its displacement, and if there are N contributing electrons per unit volume, the electric polarization, or density of dipole moments, is

$$P = q_e x N \quad (3.67)$$

Hence

$$P = \frac{q_e^2 N E / m_e}{(\omega_0^2 - \omega^2)} \quad (3.68)$$

and from Eq. (3.62)

$$\epsilon = \epsilon_0 + \frac{P(t)}{E(t)} = \epsilon_0 + \frac{q_e^2 N / m_e}{(\omega_0^2 - \omega^2)} \quad (3.69)$$

Using the fact that $n^2 = K_E = \epsilon / \epsilon_0$, we can arrive at an expression for n as a function of ω , which is known as a **dispersion equation**:

$$n^2(\omega) = 1 + \frac{N q_e^2}{\epsilon_0 m_e} \left(\frac{1}{\omega_0^2 - \omega^2} \right) \quad (3.70)$$

At frequencies increasingly above resonance, $(\omega_0^2 - \omega^2) < 0$, and the oscillator undergoes displacements that are approximately 180° out-of-phase with the driving force. The resulting electric polarization will therefore be similarly out-of-phase with the applied electric field. Hence the dielectric constant and therefore the index of refraction will both be less than 1. At frequencies increasingly below resonance, $(\omega_0^2 - \omega^2) > 0$, the electric polarization will be nearly in-phase with the applied electric field. The dielectric constant and the corresponding index of refraction will then both be greater than 1. This kind of behavior, which actually represents only part of what happens, is nonetheless generally observed in all sorts of materials.

We can test the utility of the analysis using a dispersive prism (p. 187) made of the sample material under study, but first we rewrite Eq. (3.70), as is done in Problem 3.53:

$$(n^2 - 1)^{-1} = -C\lambda^{-2} + C\lambda_0^{-2}$$

where, since $\omega = 2\pi c/\lambda$, the multiplicative constant is given by $C = 4\pi^2 c^2 \epsilon_0 m_e / N q_e^2$. Figure 3.39 is a plot of $(n^2 - 1)^{-1}$ versus λ^{-2} using data from a student experiment. A crown-glass prism was illuminated with the various wavelengths from a He discharge tube, and the index of refraction was

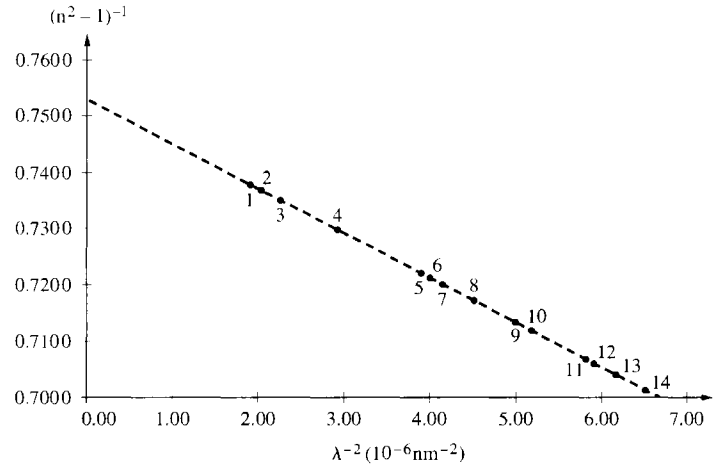


Figure 3.39 Graph of $(n^2 - 1)^{-1}$ versus λ^{-2} for the data shown in Table 3.3. See N. Gauthier, *Phys. Teach.*, **25**, 502 (1987).

measured for each one (Table 3.3). The resulting curve is indeed a straight line; its slope (using $y = mx + b$) equals $-C$, and its y -intercept corresponds to $C\lambda_0^{-2}$. From this it follows that the resonant frequency is 2.95×10^{15} Hz, which is properly in the ultraviolet.

TABLE 3.3 Dispersion of Crown Glass. The wavelengths are those of a He discharge tube. The corresponding indices were measured.

	Wavelength λ (nm)	Index of Refraction n
1.	728.135	1.5346
2.	706.519, 706.570	1.5352
3.	667.815	1.53629
4.	587.562, 587.587	1.53954
5.	504.774	1.54417
6.	501.567	1.54473
7.	492.193	1.54528
8.	471.314	1.54624
9.	447.148	1.54943
10.	438.793	1.55026
11.	414.376	1.55374
12.	412.086	1.55402
13.	402.619	1.55530
14.	388.865	1.55767

As a rule, any given substance will actually undergo several transitions from $n > 1$ to $n < 1$ as the illuminating frequency is made to increase. The implication is that instead of a single frequency ω_0 at which the system resonates, there apparently are several such frequencies. It would seem reasonable to generalize matters by supposing that there are N molecules per unit volume, each with f_j oscillators having natural frequencies ω_{0j} , where $j = 1, 2, 3, \dots$. In that case,

$$n^2(\omega) = 1 + \frac{Nq_e^2}{\epsilon_0 m_e} \sum_j \left(\frac{f_j}{\omega_{0j}^2 - \omega^2} \right) \quad (3.71)$$

This is essentially the same result as that arising from the quantum-mechanical treatment, with the exception that some of the terms must be reinterpreted. Accordingly, the quantities ω_{0j} would then be the characteristic frequencies at which an atom may absorb or emit radiant energy. The f_j terms, which satisfy the requirement that $\sum_j f_j = 1$, are weighting factors known as *oscillator strengths*. They reflect the emphasis that should be placed on each one of the modes. Since they measure the likelihood that a given atomic transition will occur, the f_j terms are also known as *transition probabilities*.

A similar reinterpretation of the f_j terms is even required classically, since agreement with the experimental data demands that they be less than unity. This is obviously contrary to the definition of the f_j that led to Eq. (3.71). One then supposes that a molecule has many oscillatory modes but that each of these has a distinct natural frequency and strength.

Notice that when ω equals any of the characteristic frequencies, n is discontinuous, contrary to actual observation. This is simply the result of having neglected the damping term, which should have appeared in the denominator of the sum. Incidentally, the damping, in part, is attributable to energy lost when the forced oscillators reradiate. In solids, liquids, and gases at high pressure ($\approx 10^3$ atm), the interatomic distances are roughly 10 times less than those of a gas at standard temperature and pressure. Atoms and molecules in this relatively close proximity experience strong interactions and a resulting “frictional” force. The effect is a damping of the oscillators and a dissipation of their energy within the substance in the form of “heat” (random molecular motion).

Had we included a damping force proportional to the speed (of the form $m_e \gamma dx/dt$) in the equation of motion, the dispersion equation (3.71) would have been

$$n^2(\omega) = 1 + \frac{Nq_e^2}{\epsilon_0 m_e} \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 + i\gamma_j\omega} \quad (3.72)$$

Although this expression is fine for rarefied media such as

gases, there is another complication that must be overcome if the equation is to be applied to dense substances. Each atom interacts with the local electric field in which it is immersed. Yet unlike the isolated atoms considered above, those in a dense material will also experience the induced field set up by their brethren. Consequently, an atom “sees” in addition to the applied field $E(t)$ another field,* namely, $P(t)/3\epsilon_0$. Without going into the details here, it can be shown that

$$\frac{n^2 - 1}{n^2 + 2} = \frac{Nq_e^2}{3\epsilon_0 m_e} \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 + i\gamma_j\omega} \quad (3.73)$$

Thus far we have been considering electron-oscillators almost exclusively, but the same results would have been applicable to ions bound to fixed atomic sites as well. In that instance m_e would be replaced by the considerably larger ion mass. Thus, although electronic polarization is important over the entire optical spectrum, the contributions from ionic polarization significantly affect n only in regions of resonance ($\omega_{0j} = \omega$).

The implications of a complex index of refraction will be considered later, in Section 4.8. At the moment we limit the discussion, for the most part, to situations in which absorption is negligible (i.e., $\omega_{0j}^2 - \omega^2 \gg \gamma_j\omega$) and n is real, so that

$$\frac{n^2 - 1}{n^2 + 2} = \frac{Nq_e^2}{3\epsilon_0 m_e} \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2} \quad (3.74)$$

Colorless, transparent materials have their characteristic frequencies outside the visible region of the spectrum (which is why they are, in fact, colorless and transparent). In particular, glasses have effective natural frequencies above the visible in the ultraviolet, where they become opaque. In cases for which $\omega_{0j}^2 \gg \omega^2$, by comparison, ω^2 may be neglected in Eq. (3.74), yielding an essentially constant index of refraction over that frequency region. For example, the important characteristic frequencies for glasses occur at wavelengths of about 100 nm. The middle of the visible range is roughly five times that value, and there, $\omega_{0j}^2 \gg \omega^2$. Notice that as ω increases toward ω_{0j} , $(\omega_{0j}^2 - \omega^2)$ decreases and n gradually increases with frequency, as is clearly evident in Fig. 3.40. This is called **normal dispersion**. In the ultraviolet region, as ω approaches a natural frequency, the oscillators will begin to resonate. Their amplitudes will increase markedly, and this will be accompanied by damping and a strong absorption of energy from the

*This result, which applies to isotropic media, is derived in almost any text on Electromagnetic Theory.

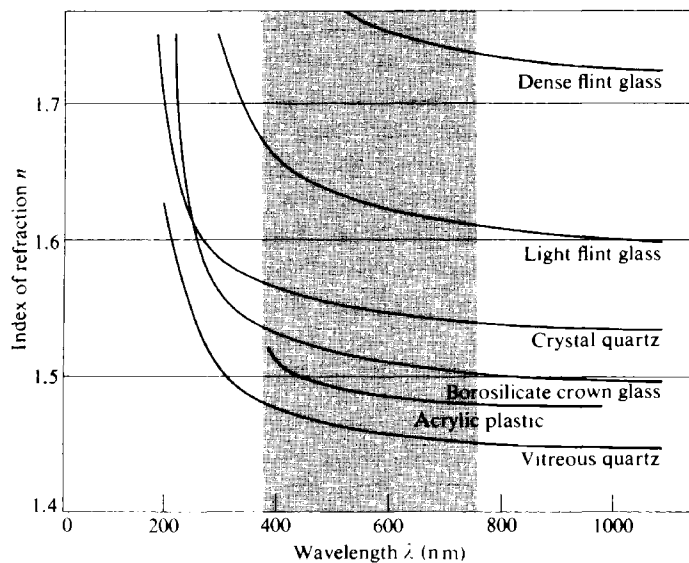


Figure 3.40 The wavelength dependence of the index of refraction for various materials.

incident wave. When $\omega_{0j} = \omega$ in Eq. (3.73), the damping term obviously becomes dominant. The regions immediately surrounding the various ω_{0j} in Fig. 3.41 are called **absorption bands**. There $dn/d\omega$ is negative, and the process is spoken of as **anomalous** (i.e., abnormal) **dispersion**. When white light passes through a glass prism, the blue constituent has a higher index than the red and is therefore deviated through a larger angle (see Section 5.5.1). In contrast, when we use a liquid-cell prism containing a dye solution with an absorption band in the visible, the spectrum is altered markedly (see Problem 3.50). All substances possess absorption bands somewhere within the electromagnetic frequency spectrum, so that the term *anomalous dispersion*, being a carryover from the late 1800s, is certainly a misnomer.

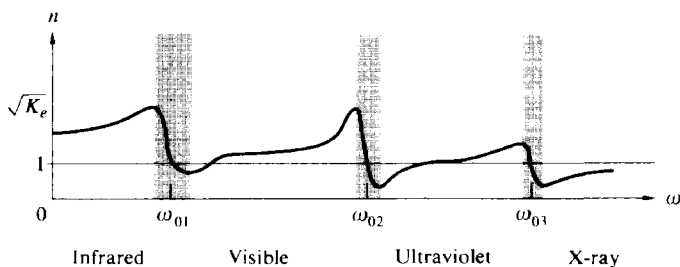


Figure 3.41 Refractive index versus frequency.

As we have seen, atoms within a molecule can also vibrate about their equilibrium positions. But the nuclei are massive, and so the natural oscillatory frequencies are low, in the infrared. Molecules such as H_2O and CO_2 have resonances in both the infrared and ultraviolet. When water is trapped within a piece of glass during its manufacture, these molecular oscillators are available, and an infrared absorption band exists. The presence of oxides also results in infrared absorption. Figure 3.42 shows the $n(\omega)$ curves (ranging from the ultraviolet to the infrared) for a number of important optical crystals. Note how they rise in the ultraviolet and fall in the infrared. At the even lower frequencies of radiowaves, glass is again transparent. In comparison, a piece of stained glass evidently has a resonance in the visible where it absorbs out a particular range of frequencies, transmitting the complementary color.

As a final point, notice that if the driving frequency is greater than any of the ω_{0j} terms, then $n^2 < 1$ and $n < 1$. Such a situation can occur, for example, if we beam X-rays onto a glass plate. This is an intriguing result, since it leads to $v > c$, in seeming contradiction to Special Relativity. We will consider this behavior again later on, when we discuss the group velocity (Section 7.6).

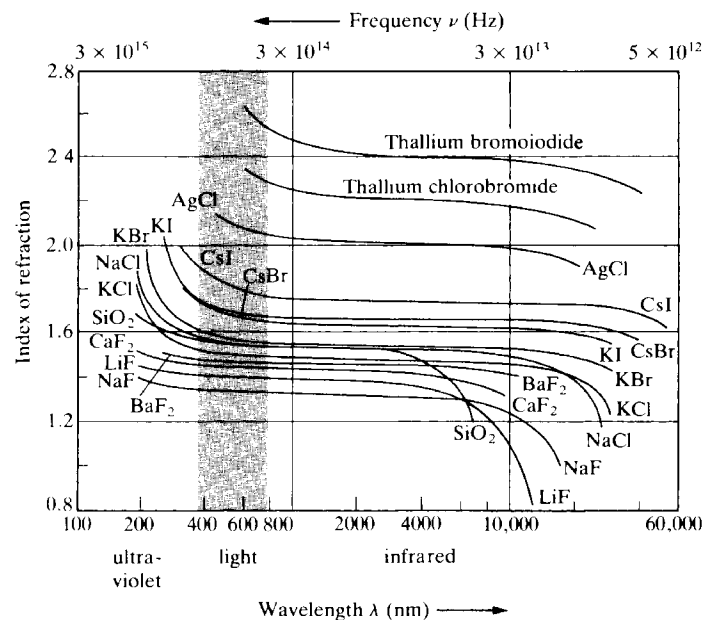
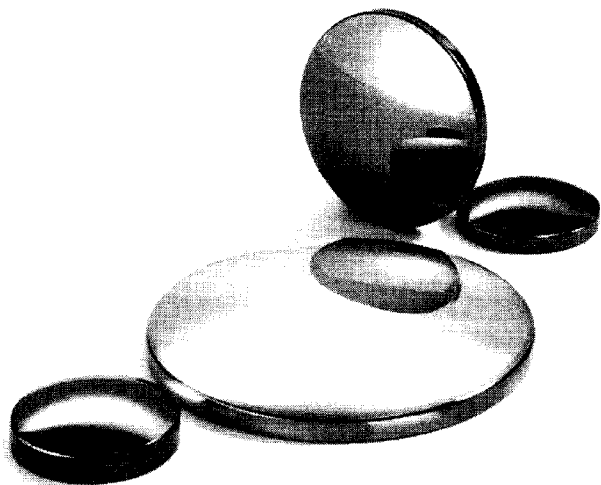


Figure 3.42 Index of refraction versus wavelength and frequency for several important optical crystals. (Adapted from data published by The Harshaw Chemical Co.)

In partial summary then, over the visible region of the spectrum, electronic polarization is the operative mechanism determining $n(\omega)$. Classically, one imagines electron-oscillators vibrating at the frequency of the incident wave. When the wave's frequency is appreciably different from a characteristic or natural frequency, the oscillations are small, and there is little dissipative absorption. At resonance, however, the oscillator amplitudes are increased, and the field does an increased amount of work on the charges. Electromagnetic energy removed from the wave and converted into mechanical energy is dissipated thermally within the substance, and one speaks of an absorption peak or band. The material, although essentially transparent at other frequencies, is fairly opaque to incident radiation at its characteristic frequencies (see photo).

3.6 The Electromagnetic-Photon Spectrum

In 1867, when Maxwell published the first extensive account of his Electromagnetic Theory, the frequency band was only known to extend from the infrared, across the visible, to the ultraviolet. Although this region is of major concern in Optics,



A group of semiconductor lenses made from ZnSe, CdTe, GaAs, and Ge. These materials are particularly useful in the infrared ($2\text{ }\mu\text{m}$ to $30\text{ }\mu\text{m}$), where they are highly transparent despite the fact that they are quite opaque in the visible region of the spectrum. (Photo courtesy Two-Six Incorporated.)

it is a small segment of the vast electromagnetic spectrum (see Fig. 3.43). This section enumerates the main categories (there is actually some overlapping) into which the spectrum is usually divided.

3.6.1 Radiofrequency Waves

In 1887, eight years after Maxwell's death, Heinrich Hertz, then professor of physics at the Technische Hochschule in Karlsruhe, Germany, succeeded in generating and detecting electromagnetic waves.* His transmitter was essentially an oscillatory discharge across a spark gap (a form of oscillating electric dipole). For a receiving antenna, he used an open loop of wire with a brass knob on one end and a fine copper point on the other. A small spark visible between the two ends marked the detection of an incident electromagnetic wave. Hertz focused the radiation, determined its polarization, reflected and refracted it, caused it to interfere setting up standing waves, and then even measured its wavelength (on the order of a meter). As he put it:

I have succeeded in producing distinct rays of electric force, and in carrying out with them the elementary experiments which are commonly performed with light and radiant heat. ... We may perhaps further designate them as rays of light of very great wavelength. The experiments described appear to me, at any rate, eminently adapted to remove any doubt as to the identity of light, radiant heat, and electromagnetic wave motion.

The waves used by Hertz are now classified in the *radiofrequency* range, which extends from a few hertz to about 10^9 Hz (λ , from many kilometers to 0.3 m or so). These are generally emitted by an assortment of electric circuits. For example, the 60-Hz alternating current circulating in power lines radiates with a wavelength of $5 \times 10^6\text{ m}$, or about 3×10^3 miles. There is no upper limit to the theoretical wavelength; one could leisurely swing the proverbial charged pith ball and, in so doing, produce a rather long, if not very strong, wave. Indeed, waves more than 18 million miles long have been detected streaming down toward Earth from outer space. The

.....
*David Hughes may well have been the first person who actually performed this feat, but his experiments in 1879 went unpublished and unnoticed for many years.

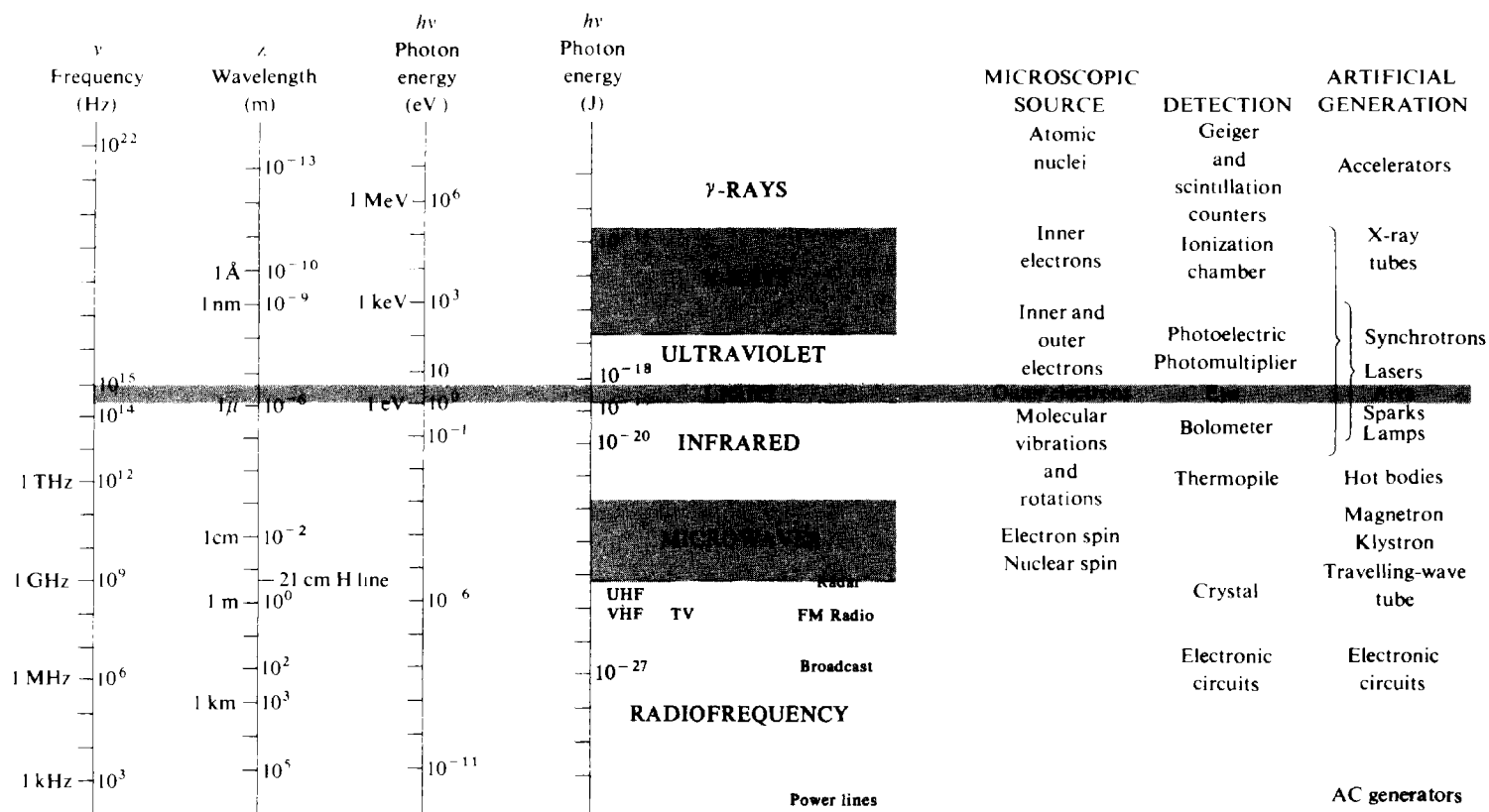


Figure 3.43 The electromagnetic-photon spectrum.

higher frequency end of the band is used for television and radio broadcasting.

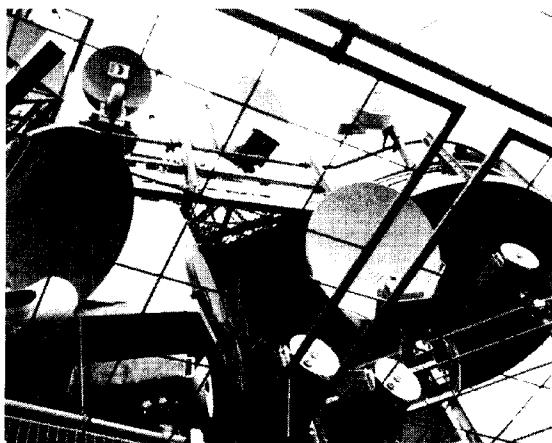
At 1 MHz (10^6 Hz), a radiofrequency photon has an energy of 6.62×10^{-28} J or 4×10^{-9} eV, a very small quantity by any measure. The granular nature of the radiation is generally obscured, and only a smooth transfer of radiofrequency energy is apparent.

3.6.2 Microwaves

The microwave region extends from about 10^9 Hz up to about 3×10^{11} Hz. The corresponding wavelengths go from roughly 30 cm to 1.0 mm. Radiation capable of penetrating the Earth's atmosphere ranges from less than 1 cm to about 30 m. Microwaves are therefore of interest in space-vehicle communications, as well as radio astronomy. In particular, neutral hydrogen atoms, distributed over vast regions of space, emit 21-cm (1420-MHz) microwaves. A good deal of information about the structure of our own and other galaxies has been gleaned from this particular emission.

Molecules can absorb and emit energy by altering the state

of motion of their constituent atoms—they can be made to vibrate and rotate. Again, the energy associated with either motion is quantized, and molecules possess rotational and vibrational energy levels in addition to those due to their electrons. Only polar molecules will experience forces via the $\vec{E}$ -field of an incident electromagnetic wave that will cause them to rotate into alignment, and only they can absorb a photon and make a rotational transition to an excited state. Since massive molecules are not able to swing around easily, we can anticipate that they will have low-frequency rotational resonances (far IR, 0.1 mm, to microwave, 1 cm). For instance, water molecules are polar (see Fig. 3.37), and if exposed to an electromagnetic wave, they will swing around, trying to stay lined up with the alternating $\vec{E}$ -field. This will occur with particular vigor at any one of its rotational resonances. Consequently, water molecules efficiently and dissipatively absorb microwave radiation at or near such a frequency. The microwave oven (12.2 cm, 2.45 GHz) is an obvious application. On the other hand, nonpolar molecules, such as carbon dioxide, hydrogen, nitrogen, oxygen, and methane, cannot make rotational transitions by way of the absorption of photons.



Microwave antennae on the top of the Eiffel Tower in Paris. (Photo by E.H.)

Nowadays microwaves are used for everything from transmitting telephone conversations and interstation television to cooking hamburgers, from guiding planes and catching speeders (by radar) to studying the origins of the Universe, opening garage doors, and viewing the surface of the planet (see photo). They are also quite useful for studying Physical Optics with experimental arrangements that are scaled up to convenient dimensions.

Photons in the low-frequency end of the microwave spectrum have little energy, and one might expect their sources to be electric circuits exclusively. Emissions of this sort can, however, arise from atomic transitions, if the energy levels involved are quite near each other. The apparent ground state



A picture of a candy bar made using T-rays. The nuts, which were hidden beneath the chocolate, are visible as a result of refraction. (Photo courtesy V. Rudd, Picometrix, Inc.)

of the cesium atom is a good example. It is actually a pair of closely spaced energy levels, and transitions between them involve an energy of only 4.14×10^{-5} eV. The resulting microwave emission has a frequency of $9.192\,631\,77 \times 10^9$ Hz. This is the basis for the well-known cesium clock, the standard of frequency and time.

The range of radiation that straddles both microwaves and infrared (roughly 50 GHz to 10 THz) is often called terahertz radiation or T-rays. They're not absorbed by most dry, non-polar materials such as plastic, paper, or fat. Water will absorb T-rays, and they're reflected by metals because of the free electrons. As a result, they can be used to image internal structure that would otherwise be hidden from view (see photo).



A photograph of an 18 by 75 mile area northeast of Alaska. It was taken by the Seasat satellite 800 kilometers (500 miles) above the Earth. The overall appearance is somewhat strange because this is actually a radar or microwave picture. The wrinkled gray region on the right is Canada. The small, bright shell shape is Banks Island, embedded in a black band of shore-fast, first-year sea ice. Adjacent to that is open water, which appears smooth and gray. The dark gray blotchy area at the far left is the main polar ice pack. There are no clouds because the radar "sees" right through them.

3.6.3 Infrared

The infrared region, which extends roughly from 3×10^{11} Hz to about 4×10^{14} Hz, was first detected by the renowned astronomer Sir William Herschel (1738–1822) in 1800. As the name implies, this band of EM-radiation lies just beneath red light. The infrared, or IR, is often subdivided into four regions: the *near IR*, that is, near the visible (780–3000 nm); the *intermediate IR* (3000–6000 nm); the *far IR* (6000–15 000 nm); and the *extreme IR* (15 000 nm–1.0 mm). This is again a rather loose division, and there is no universality in the nomenclature. Radiant energy at the long-wavelength extreme can be generated by either microwave oscillators or incandescent sources (i.e., molecular oscillators). Indeed, any material will radiate and absorb IR via thermal agitation of its constituent molecules.

The molecules of any object at a temperature above absolute zero (-273°C) will radiate IR, even if only weakly (see Section 13.1.1). On the other hand, infrared is copiously emitted in a continuous spectrum from hot bodies, such as electric heaters, glowing coals, and ordinary house radiators. Roughly half the electromagnetic energy from the Sun is IR, and the common lightbulb actually radiates far more IR than light. Like all warm-blooded creatures, we too are infrared emitters. The human body radiates IR quite weakly, starting at around 3000 nm, peaking in the vicinity of 10 000 nm, and trailing off from there into the extreme IR and, negligibly, beyond. This emission is exploited by see-in-the-dark sniper-scopes, as well as by some rather nasty “heat”-sensitive snakes (Crotalidae, pit vipers, and Boidae, constrictors) that tend to be active at night.

Besides rotating, a molecule can vibrate in several different ways, with its atoms moving in various directions with respect to one another. The molecule need not be polar, and even a linear system such as CO_2 has three basic vibrational modes and

a number of energy levels, each of which can be excited by photons. The associated vibrational emission and absorption spectra are, as a rule, in the IR (1000 nm to 0.1 mm). Many molecules have both vibrational and rotational resonances in the IR and are good absorbers, which is one reason IR is often misleadingly called “heat waves”—just put your face in the sunshine and feel the resulting buildup of thermal energy.

Infrared radiant energy is generally measured with a device that responds to the heat generated on absorption of IR by a blackened surface. There are, for example, thermocouple, pneumatic (e.g., Golay cells), pyroelectric, and bolometer detectors. These in turn depend on temperature-dependent variations in induced voltage, gas volume, permanent electric polarization and resistance, respectively. The detector can be coupled by way of a scanning system to a cathode ray tube to produce an instantaneous televisionlike IR picture (see photo) known as a thermograph (which is quite useful for diagnosing all sorts of problems, from faulty transformers to faulty people). Photographic films sensitive to near IR (< 1300 nm) are also available. There are IR spy satellites that look out for rocket launchings, IR resource satellites that look out for crop diseases, and IR astronomical satellites that look out into space. There are “heat-seeking” missiles guided by IR, and IR lasers and telescopes peering into the heavens.

Small differences in the temperatures of objects and their surroundings result in characteristic IR emission that can be used in many ways, from detecting brain tumors and breast cancers to spotting a lurking burglar. The CO_2 laser, because it is a convenient source of continuous power at appreciable levels of 100 W and more, is widely used in industry, especially in precision cutting and heat treating. Its extreme-IR emissions ($18.3\text{ }\mu\text{m}$ – $23.0\text{ }\mu\text{m}$) are readily absorbed by human tissue, making the laserbeam an effective bloodless scalpel that cauterizes as it cuts.

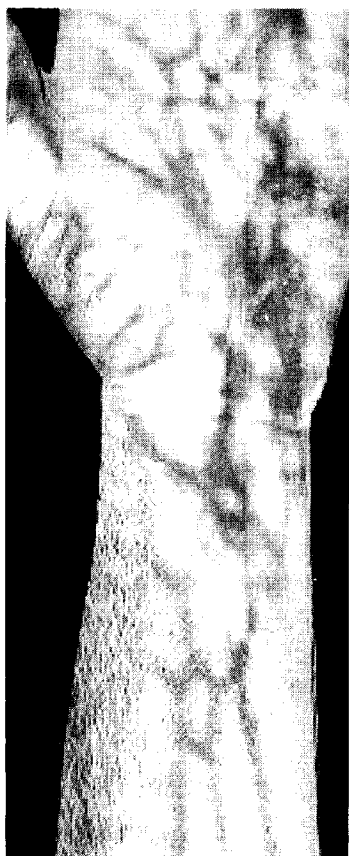
3.6.4 Light

Light corresponds to the electromagnetic radiation in the narrow band of frequencies from about 3.84×10^{14} Hz to roughly 7.69×10^{14} Hz (see Table 3.4). It is generally produced by a rearrangement of the outer electrons in atoms and molecules. (Don’t forget synchrotron radiation, which is a different mechanism.)*



Thermograph of the author. Note the cool beard and how far the hairline has receded since the first edition of this book. (Photo by E.H.)

*There is no need here to define light in terms of human physiology. On the contrary, there is plenty of evidence to indicate that this would not be a very good idea. For example, see T. J. Wang, “Visual Response of the Human Eye to X Radiation.” *Am. J. Phys.* **35**, 779 (1967).



An arm viewed in a broad band of radiant energy extending from 468.5 nm (which is light) to 827.3 nm (which is near-infrared). The technique has many biomedical applications, among which is the early detection of skin cancer.

In an incandescent material, a hot, glowing metal filament, or the solar fireball, electrons are randomly accelerated and undergo frequent collisions. The resulting broad emission spectrum is called **thermal radiation**, and it is a major source of light. In contrast, if we fill a tube with some gas and pass an electric discharge through it, the atoms therein will become excited and radiate. The emitted light is characteristic of the particular energy levels of those atoms, and it is made up of a series of well-defined frequency bands or lines. Such a device is known as a gas discharge tube. When the gas is the krypton 86 isotope, the lines are particularly narrow (zero nuclear spin, therefore no hyperfine structure). The orange-red line of Kr86, whose vacuum wavelength is 605.780 210 5 nm, has a width (at half height) of only 0.000 47 nm, or about 400 MHz. Accordingly, until 1983 it was the international standard of length (with 1 650 763.73 wavelengths equaling a meter).

Newton was the first to recognize that **white light** is actually a mixture of all the colors of the visible spectrum, that the prism does not create color by altering white light to different degrees, as had been thought for centuries, but simply fans out the light, separating it into its constituent colors. Not surpris-

TABLE 3.4 Approximate Frequency and Vacuum Wavelength Ranges for the Various Colors

Color	λ_0 (nm)	ν (THz)*
Red	780–622	384–482
Orange	622–597	482–503
Yellow	597–577	503–520
Green	577–492	520–610
Blue	492–455	610–659
Violet	455–390	659–769

*1 terahertz (THz) = 10^{12} Hz, 1 nanometer (nm) = 10^{-9} m.

ingly, the very concept of *whiteness* seems dependent on our perception of the Earth's daylight spectrum—a broad frequency distribution that falls off more rapidly in the violet than in the red (Fig. 3.44). The human eye-brain detector perceives as white a wide mix of frequencies, usually with about the same amount of energy in each portion. That is what we mean when we speak about “white light”—much of the color of the spectrum, with no region predominating. Nonetheless, many different distributions will appear more or less white. We recognize a piece of paper to be white whether it's seen indoors under incandescent light or outside under skylight, even though those whites are quite different. In fact, there are many pairs of colored light beams (e.g., 656-nm red and 492-nm cyan) that will produce the sensation of whiteness, and the eye cannot always distinguish one white from another; it can-

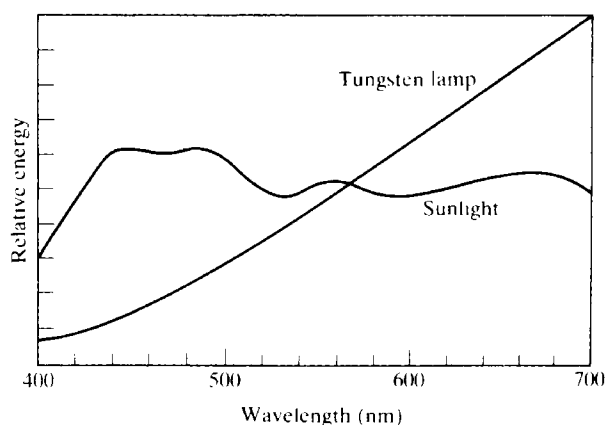


Figure 3.44 A graph of sunlight compared with the light from a tungsten lamp.

not frequency analyze light into its harmonic components the way the ear can analyze sound (see Section 7.3).

Colors are the subjective human physiological and psychological responses, primarily, to the various frequency regions extending from about 384 THz for red, through orange, yellow, green, and blue, to violet at about 769 THz (Table 3.4). Color is not a property of the light itself but a manifestation of the electrochemical sensing system—eye, nerves, brain. To be more precise, we should not say “yellow light” but rather “light that is seen as yellow.” Remarkably, a variety of different frequency mixtures can evoke the same color response from the eye-brain sensor. A beam of red light (peaking at, say, 430 THz) overlapping a beam of green light (peaking at, say, 540 THz) will result, believe it or not, in the perception of *yellow* light, even though no frequencies are actually present in the so-called yellow band. Apparently, the eye-brain averages the input and “sees” yellow (Section 4.9). That’s why a color television screen can manage with only three phosphors: red, green, and blue.

In a flood of bright sunlight where the photon flux density might be 10^{21} photons/m²·s, we can generally expect the quantum nature of the energy transport to be thoroughly masked. However, in very weak beams, since photons in the visible range ($h\nu \approx 1.6$ eV to 3.2 eV) are energetic enough to produce effects on a distinctly individual basis, the granularity will become evident. Research on human vision indicates that as few as 10 light photons, and possibly even 1, may be detectable by the eye.

3.6.5 Ultraviolet

Adjacent to and just beyond light in the spectrum is the ultraviolet region (approximately 8×10^{14} Hz to about 3.4×10^{16} Hz), discovered by Johann Wilhelm Ritter (1776–1810). Photon energies therein range from roughly 3.2 eV to 100 eV. Ultraviolet, or UV, rays from the Sun will thus have more than enough energy to ionize atoms in the upper atmosphere and in so doing create the ionosphere. These photon energies are also of the order of the magnitude of many chemical reactions, and ultraviolet rays become important in triggering those reactions. Fortunately, ozone (O₃) in the atmosphere absorbs what would otherwise be a lethal stream of solar UV. At wavelengths less than around 290 nm, UV is germicidal (i.e., it kills microorganisms). The particlelike aspects of radiant energy become increasingly evident as the frequency rises.

Humans cannot see UV very well because the cornea absorbs it, particularly at the shorter wavelengths, while the eye lens absorbs most strongly beyond 300 nm. A person who has had a lens removed because of cataracts can see UV ($\lambda > 300$ nm). In addition to insects, such as honeybees, a fair number of other creatures can visually respond to UV. Pigeons, for example, are capable of recognizing patterns illuminated by UV and probably employ that ability to navigate by the Sun even on overcast days.

An atom emits a UV photon when an electron makes a long jump down from a highly excited state. For example, the outermost electron of a sodium atom can be raised to higher and higher energy levels until it is ultimately torn loose altogether at 5.1 eV, and the atom is ionized. If the ion subsequently recombines with a free electron, the latter will rapidly descend to the ground state, most likely in a series of jumps, each resulting in the emission of a photon. It is possible, however, for the electron to make one long plunge to the ground state, radiating a single 5.1-eV UV photon. Even more energetic UV can be generated when the inner, tightly bound electrons of an atom are excited.

The unpaired valence electrons of isolated atoms can be an important source of colored light. But when these same atoms combine to form molecules or solids, the valence electrons are ordinarily paired in the process of creating the chemical bonds that hold the thing together. Consequently, the electrons are often more tightly bound, and their molecular-excited states are higher up in the UV. Molecules in the atmosphere, such as



An ultraviolet photograph of Venus taken by *Mariner 10*.

N_2 , O_2 , CO_2 , and H_2O , have just this sort of electronic resonance in the UV.

Nowadays there are ultraviolet photographic films and microscopes, UV orbiting celestial telescopes, synchrotron sources, and ultraviolet lasers.

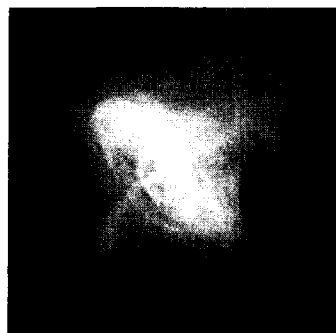
3.6.6 X-rays

X-rays were rather fortuitously discovered in 1895 by Wilhelm Conrad Röntgen (1845–1923). Extending in frequency from roughly 2.4×10^{16} Hz to 5×10^{19} Hz, they have extremely short wavelengths; most are smaller than an atom. Their photon energies (100 eV to 0.2 MeV) are large enough so that X-ray quanta can interact with matter one at a time in a clearly granular fashion, almost like bullets of energy. One of the most practical mechanisms for producing X-rays is the rapid deceleration of high-speed charged particles. The resulting broad-frequency *bremsstrahlung* (German for “braking radiation”) arises when a beam of energetic electrons is fired at a material target, such as a copper plate. Collisions with the Cu nuclei produce deflections of the beam electrons, which in turn radiate X-ray photons.

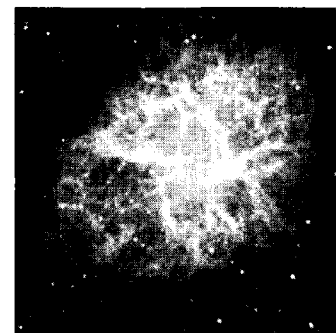
In addition, the atoms of the target may become ionized during the bombardment. Should that occur through removal of an inner electron strongly bound to the nucleus, the atom will emit X-rays as the electron cloud returns to the ground state. The resulting quantized emissions are specific to the tar-



An early X-ray photograph of the Sun taken March 1970. The limb of the Moon is visible in the southeast corner. (Photo Courtesy Dr. G. Vaiana and NASA.)



This amazingly detailed X-ray image of the Crab Nebula (see p. 61) was recently taken by the orbiting Chandra X-Ray Observatory. The picture reveals the locations of the most energetic particles in the pulsar.



An optical image of the Crab Nebula. The light forming this picture comes from particles of intermediate energy. The filaments are due to hot gases at temperatures of tens of thousands of degrees.

get atom, revealing its energy-level structure, and accordingly are called *characteristic radiation*.

Traditional medical film-radiography generally produces little more than simple shadow castings, rather than photographic images in the usual sense; it has not been possible to fabricate useful X-ray lenses. But modern focusing methods using mirrors (see Section 5.4) have begun an era of X-ray imagery, creating detailed pictures of all sorts of things, from imploding fusion pellets to celestial sources, such as the Sun (see photo), distant pulsars, and black holes—objects at temperatures of millions of degrees that emit predominantly in the X-ray region. Orbiting X-ray telescopes have given us an exciting new eye on the Universe (see photo). There are X-ray microscopes, picosecond X-ray streak cameras, X-ray diffraction gratings, and interferometers, and work continues on X-ray holography. In 1984 a group at the Lawrence Livermore National Laboratory succeeded in producing laser radiation at a wavelength of 20.6 nm. Although this is more accurately in the extreme ultraviolet (XUV), it's close enough to the X-ray region to qualify as the first soft X-ray laser.

3.6.7 Gamma Rays

Gamma rays are the highest energy (10^4 eV to about 10^{19} eV), shortest wavelength electromagnetic radiations. They are emitted by particles undergoing transitions within the atomic nucleus. A single gamma-ray photon carries so much energy

that it can be detected with little difficulty. At the same time, its wavelength is so small that it is now extremely difficult to observe any wavelike properties.

We have gone full cycle from the radiofrequency wavelike response to gamma-ray particlelike behavior. Somewhere, not far from the (logarithmic) center of the spectrum, there is light. As with all electromagnetic radiation, its energy is quantized, but here in particular what we “see” will depend on how we “look.”

3.7 Quantum Field Theory

A charged particle exerts forces on other charged particles. It creates a web of electromagnetic interaction around itself that extends out into space. That imagery leads to the concept of the electric field, which is a representation of the way the electromagnetic interaction reveals itself on a macroscopic level. The static electric field is, in effect, a spatial conception summarizing the interaction among charges. Through Faraday’s vision, the idea of the field was extended, and it became appropriate to imagine that one charge sets up an $\vec{E}$ -field in space and another charge, immersed in that field, interacts directly with it, and vice versa. What began as a mapping of the force distribution (whatever its cause) became a thing, a field, capable itself of exerting force. Still the picture seems straightforward, even if many questions come to mind. Does the static $\vec{E}$ -field have a physical reality in-and-of itself? If it does, does it fill space with energy and how exactly does that happen? Is anything actually flowing? How does the field produce a force on a charge? Does it take time to exert its influence?

Once the electromagnetic field became a reality, physicists could imagine disturbances of that tenuous medium which so conveniently spans the void of space; **light was an electromagnetic wave in the electromagnetic field**. Although it’s easy enough to envision a wave sweeping through an existing field (p. 59), it’s not so obvious how a localized pulse launched into space, like the one shown in Fig. 3.45, might be conceptualized. There is no static field filling space, extending out in front of the pulse; if the pulse advances through the medium of the electromagnetic field, it must first create that medium itself as it progresses. That’s not impossible to imagine on some level, but it’s hardly what one would call a classical wave. For any traditional wave, a medium in equilibrium is the fundamental starting point; it exists at any location before and after the wave passes. So this idea of an electromagnetic

wave, which is so beautiful mathematically, is not quite so transparent conceptually.

As early as 1905, Einstein already considered the classical equations of Electromagnetic Theory to be descriptions of the average values of the quantities being considered. “To me it seems absurd,” he wrote to Planck, “to have energy continuously distributed in space without assuming an aether. . . . While Faraday’s representation was useful in the development of electrodynamics, it does not follow in my opinion that this view must be maintained in all its detail.” Classical theory wonderfully accounted for everything being measured, but it was oblivious to the exceedingly fine granular structure of the phenomenon. Using thermodynamic arguments, Einstein proposed that electric and magnetic fields were quantized, that they are particulate rather than continuous. After all, classical theory evolved decades before the electron was even discovered. If charge (the fundamental source of electromagnetism) is quantized, shouldn’t the theory reflect that in some basic way?

Today, we are guided by Quantum Mechanics, a highly mathematical theory that provides tremendous computational and predictive power but is nonetheless disconcertingly abstract. In particular, the subdiscipline that treats microparticles and their interactions, Quantum Field Theory (QFT), in its various forms, is the most fundamental and arguably the most successful of all physical theories. Light quanta come out of the theory in a completely natural way by quantizing the electromagnetic field. The apparent implication of this is

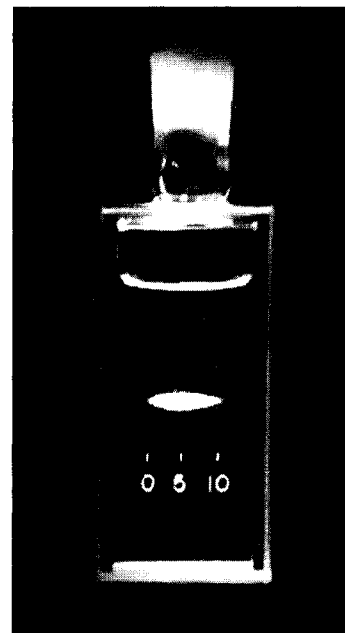


Figure 3.45 An ultrashort pulse of green light from a neodymium-doped glass laser. The pulse passed through a water cell whose wall is marked in millimeters. During the 10-picosecond exposure, the pulse moved about 2.2 mm. (Photo courtesy Bell Laboratories.)

that all microparticles originate in the same way from their own individual fields: **the field's the thing**, as it were. Thus the electron is the quantum of the electron field, the proton is the quantum of the proton field, and so forth. Filling in the details has been the business of field theorists for much of this century.

There are two distinct philosophical currents in contemporary QFT: the field-centered and the particle-centered. In the field-centered view, *fields are the fundamental entities, particles are just the quanta of the fields*. In the particle-centered view, *particles are the fundamental entities, fields are just the macroscopic coherent states of particles*. The field tradition goes back to L. de Broglie (1923), E. Schrödinger, P. Jordan, and W. Pauli, whose research laid the foundations of the quantum-mechanical variant sometimes called Wave Mechanics. The particle tradition began with the early work of W. Heisenberg (1925), although its spiritual mentor was P.A.M. Dirac who set the particle agenda with his theory of the electron-positron pair. The particular offshoot of QFT that strives to provide a relativistic quantum-mechanical treatment of the electromagnetic interaction is called Quantum Electrodynamics (QED), and it too has its particle-centered and field-centered proponents. Some of the basic ideas of QED have been made accessible on this level by R. P. Feynman, and insofar as they illuminate Optics we'll explore them later in this text (p. 137).

Contemporary physics by way of QFT holds that all fields are quantized; that each of the fundamental Four Forces (Gravitational, Electromagnetic, Strong, and Weak) is mediated by a special kind of field particle. These *messenger bosons* are continuously absorbed and emitted by the interacting material particles (electrons, protons, etc.). This ongoing exchange *is* the interaction. The mediating particle of the electric field is the **virtual photon**. This massless messenger trav-

els at the speed of light and transports momentum and energy. When two electrons repel one another, or an electron and proton attract, it is by emitting and absorbing virtual photons and thereby transferring momentum from one to the other, that transfer being a measure of the action of force. The messenger particles of the electromagnetic force are called *virtual* photons because they are bound to the interaction. Virtual photons can never escape to be detected directly by some instrument, however unsettling that is philosophically and however hard that makes it to establish their existence. Indeed, virtual photons (as distinct from real photons) exist only as the means of interaction. They are creatures of theory whose metaphysical status is yet to be determined.*

On a macroscopic level, messenger particles can manifest themselves as a continuous field provided they can group in very large numbers. Fundamental particles have an intrinsic angular momentum, or *spin*, that determines their grouping characteristics. Quantum Theory tells us that the desired field behavior can occur only if forces are mediated by messenger particles having angular momenta equal to integer multiples of $\hbar/2\pi$ (i.e., $0, \hbar/2\pi, 2\hbar/2\pi, 3\hbar/2\pi, \dots$). The angular momentum of the virtual photon is $\hbar/2\pi$; it's a spin-1 particle. The exceedingly important class of interactions that have spin-1 messengers are known as **gauge forces**, and the electromagnetic force is the model for all the gauge forces. Today, the magic of action-at-a-distance is understood via the no less mysterious exchange of virtual particles, but at least now a highly predictive mathematical theory is in place that describes the phenomenon.

*For a discussion of the issues being struggled with, see H. R. Brown and R. Harré, *Philosophical Foundations of Quantum Field Theory*.

PROBLEMS

Complete solutions to all problems—except those with an asterisk—can be found in the back of the book.

3.1 Consider the plane electromagnetic wave (in SI units) given by the expressions $E_x = 0$, $E_y = 2 \cos[2\pi \times 10^{14}(t - x/c) + \pi/2]$, and $E_z = 0$.

(a) What are the frequency, wavelength, direction of motion, amplitude, initial phase angle, and polarization of the wave?

(b) Write an expression for the magnetic flux density.

3.2 Write an expression for the $\vec{E}$ - and $\vec{B}$ -fields that constitute a plane harmonic wave traveling in the $+z$ -direction. The wave is linearly polarized with its plane of vibration at 45° to the yz -plane.

3.3* Considering Eq. (3.30), show that the expression

$$\vec{k} \times \vec{E} = \omega \vec{B}$$

is correct as it applies to a plane wave for which the direction of the electric field is constant.

3.4* Imagine an electromagnetic wave with its $\vec{E}$ -field in the y -direction. Show that Eq. (3.27)

$$\frac{\partial E}{\partial x} = -\frac{\partial B}{\partial t}$$

applied to the harmonic wave $\vec{B}$

$$\vec{E} = \vec{E}_0 \cos(kx - \omega t) \quad \vec{B} = \vec{B}_0 \cos(kx - \omega t)$$

yields the fact that

$$E_0 = cB_0$$

in agreement with Eq. (3.30).

3.5* An electromagnetic wave is specified (in SI units) by the following function:

$$\vec{E} = (-6\hat{i} + 3\sqrt{5}\hat{j})(10^4 \text{ V/m})e^{i[(1.5 \times 10^6 + 2.5)\pi \times 10^7 - 9.42 \times 10^7 t]}$$

Find (a) the direction along which the electric field oscillates, (b) the scalar value of amplitude of the electric field, (c) the direction of propagation of the wave, (d) the propagation number and wavelength, (e) the frequency and angular frequency, and (f) the speed.

3.6 The electric field of an electromagnetic wave traveling in the positive x -direction is given by

$$\vec{E} = E_0 \hat{j} \sin \frac{\pi z}{z_0} \cos(kx - \omega t)$$

(a) Describe the field verbally. (b) Determine an expression for k . (c) Find the phase speed of the wave.

3.7* A 550-nm harmonic EM-wave whose electric field is in the z -direction is traveling in the y -direction in vacuum. (a) What is the frequency of the wave? (b) Determine both ω and k for this wave. (c) If the electric field amplitude is 600 V/m, what is the amplitude of the magnetic field? (d) Write an expression for both $E(t)$ and $B(t)$ given that each is zero at $x = 0$ and $t = 0$. Put in all the appropriate units.

3.8* Calculate the energy input necessary to charge a parallel plate capacitor by carrying charge from one plate to the other. Assume the energy is stored in the field between the plates and compute the energy per unit volume, u_E , of that region, that is, Eq. (3.31). *Hint:* since the electric field increases throughout the process, either integrate or use its average value $E/2$.

3.9* Starting with Eq. (3.32), prove that the energy densities of the electric and magnet fields are equal ($u_E = u_B$) for an electromagnetic wave.

3.10 The time average of some function $f(t)$ taken over an interval T is given by

$$\langle f(t) \rangle_T = \frac{1}{T} \int_t^{t+T} f(t') dt'$$

where t' is just a dummy variable. If $\tau = 2\pi/\omega$ is the period of a harmonic function, show that $\vec{r}$

$$\langle \sin^2(\vec{k} \cdot \vec{r} - \omega t) \rangle = \frac{1}{2}$$

$$\langle \cos^2(\vec{k} \cdot \vec{r} - \omega t) \rangle = \frac{1}{2}$$

and

$$\langle \sin(\vec{k} \cdot \vec{r} - \omega t) \cos(\vec{k} \cdot \vec{r} - \omega t) \rangle = 0$$

when $T = \tau$ and when $T \gg \tau$.

3.11* Show that a more general formulation of the previous problem yields

$$\langle \cos^2 \omega t \rangle_T = \frac{1}{2} [1 + \text{sinc } \omega T \cos 2\omega t]$$

for any interval T .

3.12* With the previous problem in mind, prove that

$$\langle \sin^2 \omega t \rangle_T = \frac{1}{2} [1 - \text{sinc } \omega T \cos 2\omega t]$$

for any interval T .

3.13* Prove that the irradiance of a harmonic EM-wave is given by

$$I = \frac{1}{2c\mu_0} E_0^2$$

and then determine the average rate at which energy is transported per unit area by a plane wave having an amplitude of 15.0 V/m.

3.14* A light bulb puts out 20 W of radiant energy (most of it IR). Assume it to be a point source and calculate the irradiance 1.00 m away.

3.15* Consider a linearly polarized plane electromagnetic wave traveling in the $+x$ -direction in free space having as its plane of vibration the xy -plane. Given that its frequency is 10 MHz and its amplitude is $E_0 = 0.08$ V/m.

(a) Find the period and wavelength of the wave.

(b) Write an expression for $E(t)$ and $B(t)$.

(c) Find the flux density, $\langle S \rangle$, of the wave.

3.16* On average, the net electromagnetic power radiated by the Sun, its so-called *luminosity* (L), is 3.9×10^{26} W. Determine the mean amplitude of the electric field due to all the radiant energy arriving at the top of Earth's atmosphere (1.5×10^{11} m from the Sun).

3.17 A linearly polarized harmonic plane wave with a scalar amplitude of 10 V/m is propagating along a line in the xy -plane at 45° to the x -axis with the xy -plane as its plane of vibration. Please write a vector expression describing the wave assuming both k_x and k_y are positive. Calculate the flux density taking the wave to be in vacuum.

3.18 Pulses of UV lasting 2.00 ns each are emitted from a laser that has a beam of diameter 2.5 mm. Given that each burst carries an energy of 6.0 J, (a) determine the length in space of each wavetrain, and (b) find the average energy per unit volume for such a pulse.

3.19* A laser provides pulses of EM-radiation in vacuum lasting 10^{-12} s. If the radiant flux density is 10^{20} W/m², determine the amplitude of the electric field of the beam.

3.20 A 1.0-mW laser has a beam diameter of 2 mm. Assuming the divergence of the beam to be negligible, compute its energy density in the vicinity of the laser.

3.21* A cloud of locusts having a density of 100 insects per cubic meter is flying north at a rate of 6 m/min. What is the flux density of locusts? That is, how many cross an area of 1 m² perpendicular to their flight path per second?

3.22 Imagine that you are standing in the path of an antenna that is radiating plane waves of frequency 100 MHz and flux density 19.88×10^{-2} W/m². Compute the photon flux density, that is, the number of photons per unit time per unit area. How many photons, on the average, will be found in a cubic meter of this region?

3.23* How many photons per second are emitted from a 100-W yellow lightbulb if we assume negligible thermal losses and a quasi-monochromatic wavelength of 550 nm? In actuality only about 2.5% of the total dissipated power emerges as visible radiation in an ordinary 100-W lamp.

3.24 A 3.0-V flashlight bulb draws 0.25 A, converting about 1.0% of the dissipated power into light ($\lambda \approx 550$ nm). If the beam has a cross-sectional area of 10 cm² and is approximately cylindrical,

(a) How many photons are emitted per second?

(b) How many photons occupy each meter of the beam?

(c) What is the flux density of the beam as it leaves the flashlight?

3.25* An isotropic quasimonochromatic point source radiates at a rate of 100 W. What is the flux density at a distance of 1 m? What are the amplitudes of the $\vec{E}$ - and $\vec{B}$ -fields at that point?

3.26 Using energy arguments, show that the amplitude of a cylindrical wave must vary inversely with $\sqrt{r}$. Draw a diagram indicating what's happening.

3.27* What is the momentum of a 10^{19} -Hz X-ray photon?

3.28 Consider an electromagnetic wave impinging on an electron. It is easy to show kinematically that the average value of the time rate-of-change of the electron's momentum $\vec{p}$ is proportional to the average value of the time rate-of-change of the work, W , done on it by the wave. In particular,

$$\left\langle \frac{d\vec{p}}{dt} \right\rangle = \frac{1}{c} \left\langle \frac{dW}{dt} \right\rangle \hat{i}$$

Accordingly, if this momentum change is imparted to some completely absorbing material, show that the pressure is given by Eq. (3.51).

3.29* Derive an expression for the radiation pressure when the normally incident beam of light is totally reflected. Generalize this result to the case of oblique incidence at an angle θ with the normal.

3.30 A completely absorbing screen receives 300 W of light for 100 s. Compute the total linear momentum transferred to the screen.

3.31 The average magnitude of the Poynting vector for sunlight arriving at the top of Earth's atmosphere (1.5×10^{11} m from the Sun) is about 1.4 kW/m².

- (a) Compute the average radiation pressure exerted on a metal reflector facing the Sun.
- (b) Approximate the average radiation pressure at the surface of the Sun whose diameter is 1.4×10^9 m.

3.32* A surface is placed perpendicular to a beam of light of constant irradiance (I). Suppose that the fraction of the irradiance absorbed by the surface is α . Show that the pressure on the surface is given by

$$\mathcal{P} = (2 - \alpha)I/c$$

3.33* A light beam with an irradiance of 2.00×10^6 W/m² impinges normally on a surface that reflects 70.0% and absorbs 30.0%. Compute the resulting radiation pressure on the surface.

3.34 What force on the average will be exerted on the ($40 \text{ m} \times 50 \text{ m}$) flat, highly reflecting side of a space station wall if it's facing the Sun while orbiting Earth?

3.35 A parabolic radar antenna with a 2-m diameter transmits 200-kW pulses of energy. If its repetition rate is 500 pulses per second, each lasting $2 \mu\text{s}$, determine the average reaction force on the antenna.

3.36 Consider the plight of an astronaut floating in free space with only a 10-W lantern (inexhaustibly supplied with power). How long will it take to reach a speed of 10 m/s using the radiation as propulsion? The astronaut's total mass is 100 kg.

3.37 Consider the uniformly moving charge depicted in Fig. 3.26*b*. Draw a sphere surrounding it and show via the Poynting vector that the charge does not radiate.

3.38* A plane, harmonic, linearly polarized light wave has an electric field intensity given by

$$E_z = E_0 \cos \pi 10^{15} \left(t - \frac{x}{0.65c} \right)$$

while traveling in a piece of glass. Find

- (a) The frequency of the light.
- (b) Its wavelength.
- (c) The index of refraction of the glass.

3.39* What is the speed of light in diamond if the index of refraction is 2.42?

3.40* Given that the wavelength of a lightwave in vacuum is 540 nm, what will it be in water, where $n = 1.33$?

3.41* Determine the index of refraction of a medium if it is to reduce the speed of light by 10% as compared to its speed in vacuum?

3.42 If the speed of light (the phase speed) in Fabulite (SrTiO_3) is 1.245×10^8 m/s, what is its index of refraction?

3.43* What is the distance that yellow light travels in water (where $n = 1.33$) in 1.00 s?

3.44* A 500-nm lightwave in vacuum enters a glass plate of index 1.60 and propagates perpendicularly across it. How many waves span the glass if it's 1.00 cm thick?

3.45* Yellow light from a sodium lamp ($\lambda_0 = 589 \text{ nm}$) traverses a tank of glycerin (of index 1.47), which is 20.0 m long, in a time t_1 . If it takes a time t_2 for the light to pass through the same tank when filled with carbon disulfide (of index 1.63), determine the value of $t_2 - t_1$.

3.46* A lightwave travels from point A to point B in vacuum. Suppose we introduce into its path a flat glass plate ($n_g = 1.50$) of thickness $L = 1.00 \text{ mm}$. If the vacuum wavelength is 500 nm, how many waves span the space from A to B with and without the glass in place? What phase shift is introduced with the insertion of the plate?

3.47 The low-frequency relative permittivity of water varies from 88.00 at 0°C to 55.33 at 100°C . Explain this behavior. Over the same range in temperature, the index of refraction ($\lambda = 589.3 \text{ nm}$) goes from roughly 1.33 to 1.32. Why is the change in n so much smaller than the corresponding change in K_E ?

3.48 Show that for substances of low density, such as gases, which have a single resonant frequency ω_0 , the index of refraction is given by

$$n \approx 1 + \frac{Nq_e^2}{2\epsilon_0 m_e(\omega_0^2 - \omega^2)}$$

3.49* In the next chapter, Eq. (4.47), we'll see that a substance reflects radiant energy appreciably when its index differs most from the medium in which it is embedded.

- (a) The dielectric constant of ice measured at microwave frequencies is roughly 1, whereas that for water is about 80 times greater—why?
- (b) How is it that a radar beam easily passes through ice but is considerably reflected when encountering a dense rain?

3.50 Fuchsin is a strong (aniline) dye, which in solution with alcohol has a deep red color. It appears red because it absorbs the green component of the spectrum. (As you might expect, the surfaces of crystals of fuchsin reflect green light rather strongly.) Imagine that you have a thin-walled hollow prism filled with this solution. What will the spectrum look like for incident white light? By the way, anomalous dispersion was first observed in about 1840 by Fox Talbot, and the effect was christened in 1862 by Le Roux. His work was

promptly forgotten, only to be rediscovered eight years later by C. Christiansen.

3.51* Take Eq. (3.71) and check out the units to make sure that they agree on both sides.

3.52 The resonant frequency of lead glass is in the UV fairly near the visible, whereas that for fused silica is far into the UV. Use the dispersion equation to make a rough sketch of n versus ω for the visible region of the spectrum.

3.53* Show that Eq. (3.70) can be rewritten as

$$(n^2 - 1)^{-1} = -C\lambda^{-2} + C\lambda_0^{-2}$$

where $C = 4\pi^2 c^2 \epsilon_0 m_e / N q_e^2$.

3.54 Augustin Louis Cauchy (1789–1857) determined an empirical equation for $n(\lambda)$ for substances that are transparent in the visible. His expression corresponded to the power series relation

$$n = C_1 + C_2/\lambda^2 + C_3/\lambda^4 + \dots$$

where the C s are all constants. In light of Fig. 3.41, what is the physical significance of C_1 ?

3.55 Referring to the previous problem, realize that there is a region between each pair of absorption bands for which the Cauchy Equa-

tion (with a new set of constants) works fairly well. Examine Fig. 3.41: what can you say about the various values of C_1 as ω decreases across the spectrum? Dropping all but the first two terms, use Fig. 3.40 to determine approximate values for C_1 and C_2 for borosilicate crown glass in the visible.

3.56* Crystal quartz has refractive indexes of 1.557 and 1.547 at wavelengths of 410.0 nm and 550.0 nm, respectively. Using only the first two terms in Cauchy's Equation, calculate C_1 and C_2 and determine the index of refraction of quartz at 610.0 nm.

3.57* In 1871 Sellmeier derived the equation

$$n^2 = 1 + \sum_j \frac{A_j \lambda^2}{\lambda^2 - \lambda_{0j}^2}$$

where the A_j terms are constants and each λ_{0j} is the vacuum wavelength associated with a natural frequency ν_{0j} , such that $\lambda_{0j}\nu_{0j} = c$. This formulation is a considerable practical improvement over the Cauchy Equation. Show that where $\lambda \gg \lambda_{0j}$, Cauchy's Equation is an approximation of Sellmeier's. *Hint:* Write the above expression with only the first term in the sum; expand it by the binomial theorem; take the square root of n^2 and expand again.

3.58* If an ultraviolet photon is to dissociate the oxygen and carbon atoms in the carbon monoxide molecule, it must provide 11 eV of energy. What is the minimum frequency of the appropriate radiation?



The Propagation of Light

4.1 Introduction

Our present concern is with the basic phenomena of *transmission* (p. 92), *reflection* (p. 95), and *refraction* (p. 100). These will be described classically in two ways: first, via the general notions of waves and rays (p. 104) and then from the more specific perspective of Electromagnetic Theory (p. 111). After that, we'll turn to a highly simplified treatment of Quantum Electrodynamics (QED) for a modern interpretation of what's happening (p. 139).

Most students have already studied these fundamental propagation phenomena in some introductory way and found ideas like the Laws of Reflection and Refraction to be straightforward and simple. But that's only because such treatments are from a macroscopic perspective that tends to be misleadingly superficial. For instance, reflection, which looks as obvious as light "bouncing off a surface," is a wonderfully subtle affair usually involving the coordinated behavior of countless atoms. The more deeply we explore these processes, the more challenging they become. Beyond that, many fascinating questions need to be addressed: How does light move through a material medium? What happens to it as it does? Why does light appear to travel at a speed other than c when photons can only exist at c ?

Each encounter of light with bulk matter can be viewed as a cooperative event arising when a stream of photons sails through, and interacts with, an array of atoms suspended (via electromagnetic fields) in the void. The details of that journey determine why the sky is blue and blood is red, why your cornea is transparent and your hand opaque, why snow is white and rain is not. At its core, this chapter is about **scattering**, in particular, the absorption and prompt re-emission of EM-radiation by electrons associated with atoms and molecules. *The processes of transmission, reflection, and refraction*

are macroscopic manifestations of scattering occurring on a submicroscopic level.

To begin the analysis, let's first consider the propagation of radiant energy through various homogeneous media.

4.2 Rayleigh Scattering

Imagine a narrow beam of sunlight having a broad range of frequencies advancing through empty space. As it progresses, the beam spreads out very slightly, but apart from that, all the energy continues forward at c . There is no scattering, and the beam cannot be seen from the side. Nor does the light tire or diminish in any way. When a star in a nearby galaxy 1.7×10^5 light-years away was seen to explode in 1987, the flash of light that reached Earth had been sailing through space for 170 000 years before it got here. **Photons are timeless.**

Now, suppose we mix a wisp of air into the void—some molecules of nitrogen, oxygen, and so forth. These molecules have no resonances in the visible, no one of them can be raised into an excited state by absorbing a quantum of light, and the gas is therefore transparent. Instead, each molecule behaves as a little oscillator whose electron cloud can be driven into a ground-state vibration by an incoming photon. Immediately upon being set vibrating, the molecule initiates the re-emission of light. A photon is absorbed, and without delay another photon of the same frequency (and wavelength) is emitted; the light is *elastically scattered*. The molecules are randomly oriented, and photons scatter out every which way (Fig. 4.1). Even when the light is fairly dim, the number of photons is immense, and it looks as if the molecules are scattering little classical spherical wavelets (Fig. 4.2)—energy streams out in every direction. Still, the scattering process is quite weak and

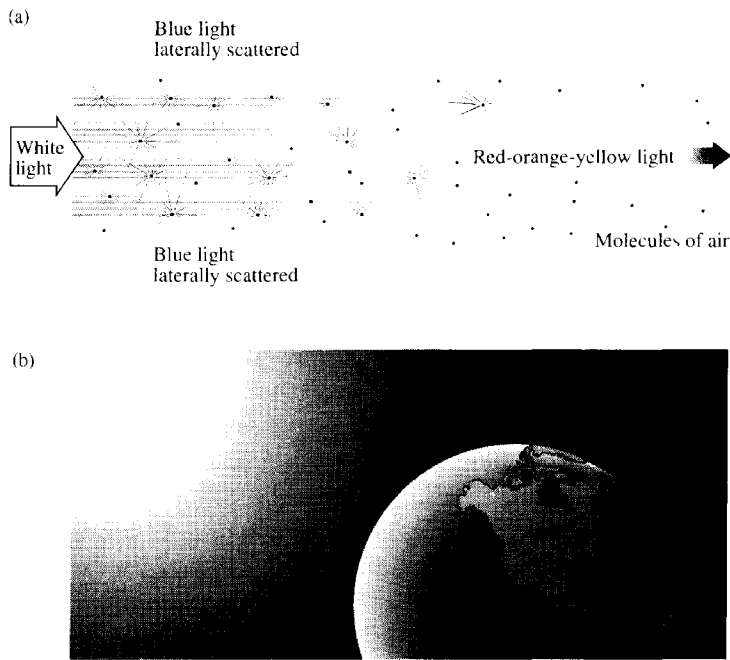


Figure 4.1 (a) Sunlight traversing a region of widely spaced air molecules. The light laterally scattered is mostly blue, and that's why the sky is blue. The unscattered light, which is rich in red, is viewed only when the Sun is low in the sky at sunrise and sunset. (b) Solar rays reach about 18° beyond the daytime terminator because of atmospheric scattering. Over this twilight band the skylight fades to the complete darkness of night.

the gas tenuous, so the beam is very little attenuated unless it passes through a tremendous volume of air.

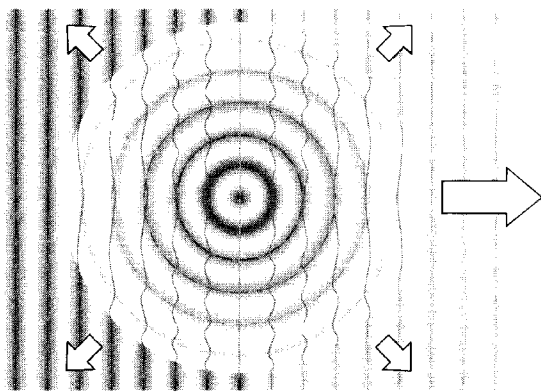


Figure 4.2 A plane wave, incident from the left, sweeps across an atom and spherical wavelets are scattered. The process is continuous, and hundreds of millions of photons per second stream out of the scattering atom in all directions.

The amplitudes of these ground-state vibrations, and therefore the amplitudes of the scattered light, increase with frequency because all the molecules have electronic resonances in the UV. The closer the driving frequency is to a resonance, the more vigorously the oscillator responds. So, violet light is strongly scattered laterally out of the beam, as is blue to a slightly lesser degree, as is green to a considerably lesser degree, as is yellow to a still lesser degree, and so on. The beam that traverses the gas will thus be richer in the red end of the spectrum, while the light scattered out (sunlight not having very much violet in it in the first place) will be richer in blue. That, in part, is why the sky is blue.

Long before Quantum Mechanics, Lord Rayleigh (1871) analyzed scattered sunlight in terms of molecular oscillators. Using a simple argument based on dimensional analysis (see Problem 4.1), he correctly concluded that the intensity of the scattered light was proportional to $1/\lambda^4$ and therefore increases with ν^4 . Before this work, it was widely believed that the sky was blue because of scattering from minute dust particles. Since that time, scattering involving particles smaller than a wavelength (i.e., less than about $\lambda/15$) has been referred to as Rayleigh Scattering. Atoms and ordinary molecules fit the bill since they are a few tenths of a nanometer in diameter, whereas light has a wavelength of around 500 nm. A human's blue eyes, a bluejay's feathers, the blue-tailed skinks's blue tail, and the baboon's blue buttocks are all colored via Rayleigh Scattering. Indeed, in the animal kingdom scattering is the cause of almost all the blue, much of the green, and even some of the purple coloration. Scattering from the tiny alveolar cells in the barbs of the jay's feathers make it blue, whereas a parrot's green is a blend of yellow arising from preferential absorption (p. 134) and blue via scattering.

As we will see in a moment, a dense uniform substance will not appreciably scatter laterally, and that applies to much of the lower atmosphere. After all, if blue light were strongly scattered out at sea level, a far-off mountain would appear red-dish and that's not the case even over distances of tens of kilometers. Even in the middle regions of the atmosphere, the density is great enough to suppress Rayleigh Scattering; something else must be contributing to the blue of the sky. What happens in the mid-atmosphere is that thermal motion of the air results in rapidly changing *density fluctuations* on a local scale. These momentary, fairly random fluctuations cause more molecules to be in one place than another and to radiate more in one direction than another. M. Smoluchowski (1908) and A. Einstein (1910) independently provided the basic ideas for the theory of scattering from these fluctuations, which



Without an atmosphere to scatter sunlight, the Moon's sky is an eerie black.

gives similar results to those of Rayleigh. Scattering from inhomogeneities in density is of interest whenever light travels great distances in a medium, such as the glass fiber of a communications link (p. 197).

Sunlight streaming into the atmosphere from one direction is scattered in all directions. Without an atmosphere, the daytime sky would be as black as the void of space, as black as the Moon sky. When the Sun is low over the horizon, its light passes through a great thickness of air (far more so than it does at noon). With the blue-end appreciably attenuated, the reds and yellows propagate along the line-of-sight from the Sun to produce Earth's familiar fiery sunsets.

4.2.1 Scattering and Interference

In dense media, a tremendous number of close-together atoms or molecules contribute an equally tremendous number of scattered electromagnetic wavelets. These wavelets overlap and interfere in a way that does not occur in a tenuous medium. As a rule, *the denser the substance through which light advances, the less the lateral scattering*, and to understand why that's so, we must examine the interference taking place.

Interference has already been discussed (p. 20) and will be treated in further detail in Chapters 7 and 9; here, the basics suffice. Recall that interference is *the superposition of two or more waves producing a resultant disturbance that is the sum of the overlapping wave contributions*. Figure 2.14 shows two harmonic waves of the same frequency traveling in the same direction. When they are precisely in-phase (Fig. 2.14a), the resultant at every point is the sum of the two wave-height values. This extreme case is called **total constructive interference**. When the phase difference reaches 180° , the waves tend to cancel, and we have the other extreme, called **total destructive interference** (Fig. 2.14d).

The theory of Rayleigh Scattering has independent mole-

cules randomly arrayed in space so that the phases of the secondary wavelets scattered off to the side have no particular relationship to one another and there is no sustained pattern of interference. That situation occurs when the separation between the molecular scatterers is roughly a wavelength or more, as it is in a tenuous gas. In Fig. 4.3a a parallel beam of light is inci-

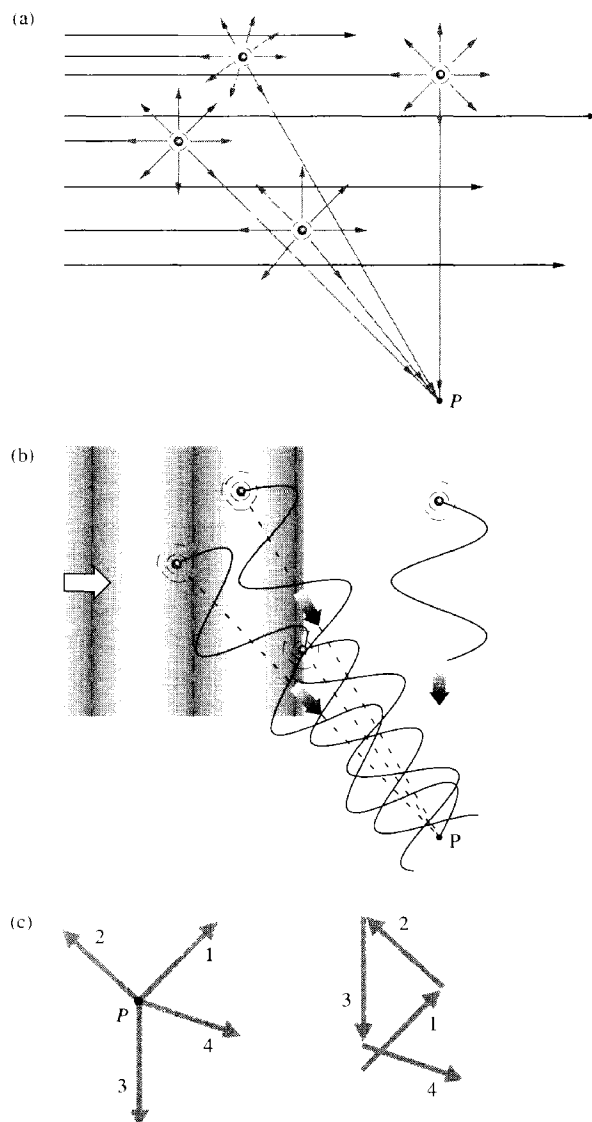


Figure 4.3 (a) The scattering of light from a widely spaced distribution of molecules. (b) The wavelets arriving at a lateral point P have a jumble of different phases and tend not to interfere in a sustained constructive fashion. (c) That can probably be appreciated most easily using phasors. As they arrive at P the phasors have large phase angle differences with respect to each other. When added tip-to-tail they therefore tend to spiral around keeping the resultant phasor quite small. Remember that we are really dealing with millions of tiny phasors rather than four substantial ones.

dent from the left. This so-called *primary light field* (in this instance composed of plane waves) illuminates a group of widely spaced molecules. A continuing progression of primary wavefronts sweep over and successively energize and re-energize each molecule, which, in turn, scatters light in all directions, and in particular out to some lateral point P . Because the lengths of their individual paths to P differ greatly in comparison to λ , some of the wavelets arriving at P are ahead of others while some are behind, and that by substantial fractions of a wavelength (Fig. 4.3b). In other words, the phases of the wavelets at P differ greatly. (Remember that the molecules are also moving around, and that changes the phases as well.) At any moment some wavelets interfere constructively, some destructively, and the shifting random hodgepodge of overlapping wavelets effectively averages away the interference. ***Random, widely spaced scatterers driven by an incident primary wave emit wavelets that are essentially independent of one another in all directions except forward. Laterally scattered light, unimpeded by interference, streams out of the beam.*** This is approximately the situation existing about 100 miles up in the Earth's tenuous high-altitude atmosphere, where a good deal of blue-light scattering takes place.

That the scattered irradiance should depend on $1/\lambda^4$ is easily seen by returning to the concept of dipole radiation (Section 3.4.3). Each molecule is taken as an electron oscillator driven into vibration by the incident field. Being far apart, they are assumed to be independent of one another and each radiates in accord with Eq. (3.56). The scattered electric fields are essentially independent, and there is no interference laterally. Accordingly, the net irradiance at P is the algebraic sum of the scattered irradiances from each molecule (p. 285). For an individual scatterer the irradiance is given by Eq. (3.57), and it varies with ω^4 .

The advent of the laser has made it relatively easy to observe Rayleigh Scattering directly in low-pressure gases, and the results confirm the theory.

Forward Propagation

To see why the forward direction is special, why the wave advances in any medium, consider Fig. 4.4. Notice that for a forward point P all the different paths taken by the light are about the same length; scattering alters the various path lengths by very little. (The scattered wavelets arrive at P more or less in-phase and essentially interfere constructively.) A more detailed description is provided by Fig. 4.5. It depicts a sequence in time showing two molecules A and B , interacting with an incoming primary plane wave—a solid arc represents

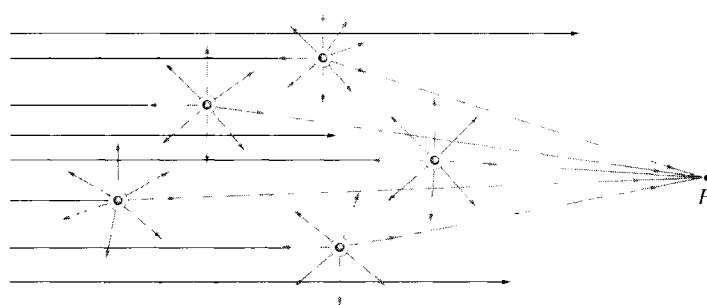


Figure 4.4 Scattering in the forward direction doesn't change the light paths very much, and the waves all arrive at P pretty much in-phase.

a secondary wavelet peak (a positive maximum); a dashed arc corresponds to a trough (a negative maximum). In (a), the primary wavefront impinges on molecule A , which begins to scatter a spherical wavelet. For the moment, suppose the

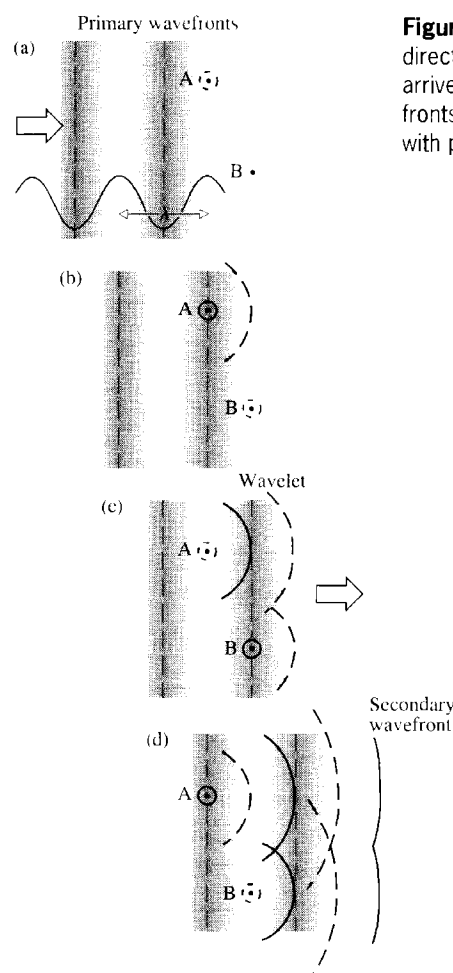


Figure 4.5 In the forward direction the scattered wavelets arrive in-phase on planar wavefronts—trough with trough, peak with peak.

wavelet is 180° out-of-phase with the incident wave. (A driven oscillator is usually out-of-phase with the driver: p. 93.) Thus *A* begins to radiate a trough (a negative *E*-field) in response to being driven by a peak (a positive *E*-field). Part (b) shows the spherical wavelet and the plane wave overlapping, marching out-of-step but marching together. The incident wavefront impinges on *B*, and it, in turn, begins to reradiate a wavelet, which must also be out-of-phase by 180° . In (c) and (d), we see the point of all of this, namely, that both wavelets are moving forward—they are in-phase with each other. That condition would be true for all such wavelets regardless of both how many molecules there were and how they were distributed. Because of the asymmetry introduced by the beam itself, ***all the scattered wavelets add constructively with each other in the forward direction.***

4.2.2 The Transmission of Light Through Dense Media

Now, suppose the amount of air in the region under consideration is increased. In fact, imagine that each little cube of air, one wavelength on a side, contains a great many molecules, whereupon it is said to have an appreciable *optical density*. (This usage probably derives from the fact that early experiments on gases indicated that an increase in density is accompanied by a proportionate increase in the index of refraction.) At the wavelengths of light, the Earth's atmosphere at STP has about 3 million molecules in such a λ^3 -cube. The scattered wavelets ($\lambda \approx 500$ nm) radiated by sources so close together (≈ 3 nm) cannot properly be assumed to arrive at some point *P* with random phases—interference will be important. This is equally true in liquids and solids where the atoms are 10 times closer and arrayed in a far more orderly fashion. In such cases, the light beam effectively encounters a uniform medium with no discontinuities to destroy the symmetry. Again, the scattered wavelets interfere constructively in the forward direction (that much is independent of the arrangement of the molecules), but now destructive interference predominates in all other directions. ***Little or no light ends up scattered laterally or backwards in a dense homogeneous medium.***

To illustrate the phenomenon, Fig. 4.6 shows a beam moving through an ordered array of close-together scatterers. All along wavefronts throughout the beam, sheets of molecules are energized in-phase, radiate, and are re-energized, over and over again as the light sweeps past. Thus some molecule *A* radiates spherically out of the beam, but because of the ordered close arrangement, there will be a molecule *B*, a dis-

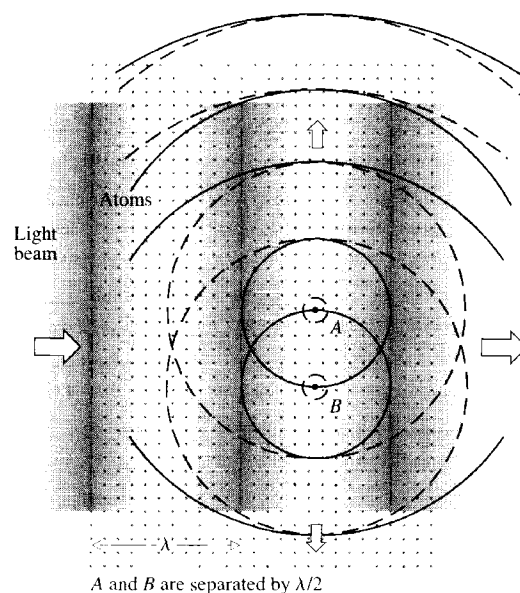


Figure 4.6 A plane wave impinging from the left. The medium is composed of many closely spaced atoms. Among countless others, a wavefront stimulates two atoms, *A* and *B*, that are very nearly one-half wavelength apart. The wavelets they emit interfere destructively. Trough overlaps crest, and they completely cancel each other in the direction perpendicular to the beam. That process happens over and over again, and little or no light is scattered laterally.

tance $\approx \lambda/2$ away, such that both wavelets cancel in the transverse direction. Here, where λ is thousands of times larger than the scatterers and their spacing, there will likely always be pairs of molecules that tend to negate each other's wavelets in any given lateral direction. Even if the medium is not perfectly ordered, the net electric field at a point in any transverse direction will be the sum of a great many *tiny* scattered fields, each somewhat out-of-phase with the next, so that the sum (which will be different from point to point) will always be

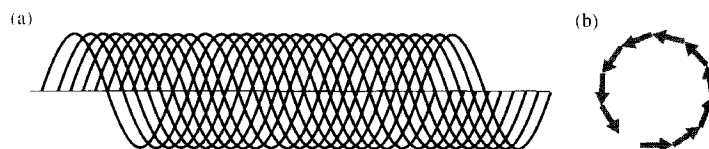


Figure 4.7 (a) When a great many tiny slightly shifted waves arrive at a point in space, there is generally as much positive *E*-field as negative, and the resultant disturbance is nearly zero. (b) The tiny phasors representing those waves form a very small circular figure, and the resultant (which oscillates in a way that depends on the number of waves) is never large.

small (Fig. 4.7). This makes sense from the perspective of conservation of energy—we can't have constructive interference in every direction. *Interference produces a redistribution of energy, out of the regions where it's destructive into the regions where it's constructive.*

The more dense, uniform, and ordered the medium is (the more nearly homogeneous), the more complete will be the lateral destructive interference and the smaller the amount of nonforward scattering. Thus most of the energy will go into the forward direction, and the beam will advance essentially undiminished (Fig. 4.8).

Scattering on a per-molecule basis is extremely weak. In

order to have half its energy scattered, a beam of green light will have to traverse ≈ 150 km of atmosphere. Since about 1000 times more molecules are in a given volume of liquid than in the same volume of vapor (at atmospheric pressure), we can expect to see an increase in scattering. Still, the liquid is a far more ordered state with much less pronounced density fluctuations, and that should suppress the nonforward scattering appreciably. Accordingly, an increased scattering per unit volume is observed in liquids, but it's more like 5 to 50 times as much rather than 1000 times. Molecule for molecule, liquids scatter substantially less than gases. Put a few drops of milk in a tank of water and illuminate it with a bright flash-

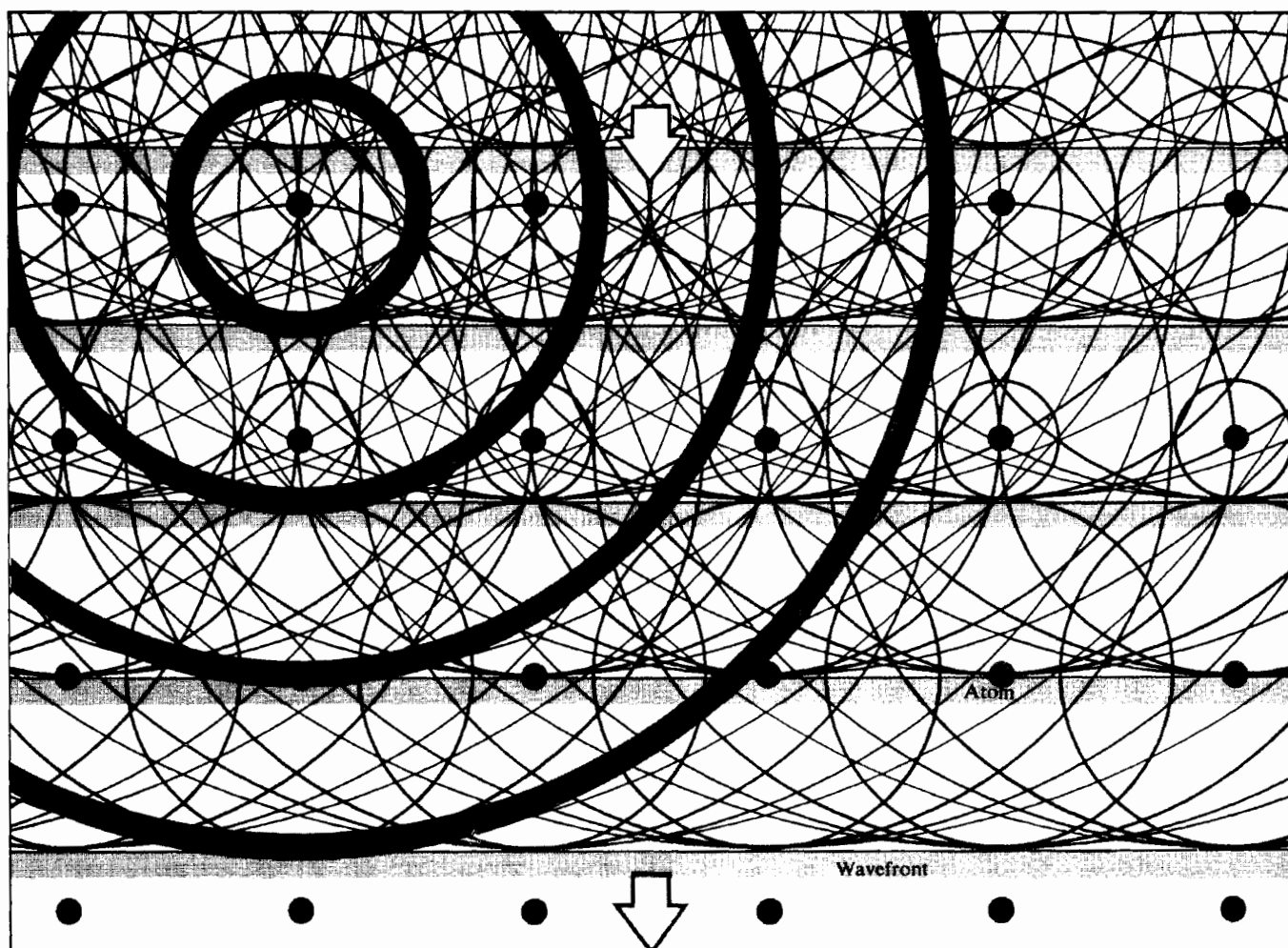


Figure 4.8 A downward plane wave incident on an ordered array of atoms. Wavelets scatter in all directions and overlap to form an ongoing secondary plane wave traveling downward.

light beam. A faint but unmistakable blue haze will scatter out laterally, and the direct beam will emerge decidedly reddened.

Transparent amorphous solids, such as glass and plastic, will also scatter light laterally, but very weakly. Good crystals, like quartz and mica, with their almost perfectly ordered structures, scatter even more faintly. Of course, imperfections of all sorts (dust and bubbles in liquids, flaws and impurities in solids) will serve as scatterers, and when these are small, as in the gem moonstone, the emerging light will be bluish.

In 1869 John Tyndall experimentally studied the scattering produced by small particles. He found that as the size of the particles increased (from a fraction of a wavelength), the amount of scattering of the longer wavelengths increased proportionately. Ordinary clouds in the sky testify to the fact that relatively large droplets of water scatter white light with no appreciable coloration. The same is true of the microscopic globules of fat and protein in milk.

When the number of molecules in a particle is small, they are all close to one another and act in unison; their wavelets interfere constructively, and the scattering is strong. As the size of the particle approaches a wavelength, the atoms at its extremities no longer radiate wavelets that are necessarily in-phase and the scattering begins to diminish. This happens first at the short wavelengths (blue), and so as the particle size increases, it scatters proportionately more of the red end of the spectrum (and it does so increasingly in the forward direction). The theoretical analysis of scattering from spherical particles of any size was first published by Gustav Mie in 1908. Mie Scattering depends only weakly on wavelength and becomes independent of it (white light in, white light out) when the particle size exceeds λ . Reasonably enough, Rayleigh Scattering is the small-size limiting case of Mie Scattering.

On an overcast day, the sky looks hazy gray because of large water droplets. In the same way, some inexpensive plastic food containers and white garbage-bag plastic look pale blue-white in scattered light and are distinctly orange in transmitted light. The garbage bags, in order to be made opaque, contain (2 to 2.5%) clear TiO_2 spheres ($n = 2.76$) about 200 nm in diameter, and these Mie scatter bluish-white.*

4.2.3 Transmission and the Index of Refraction

The transmission of light through a homogeneous medium is an ongoing repetitive process of scattering and rescattering. Each such event introduces a phase shift into the light field, which ultimately shows up as a shift in the apparent phase velocity of the transmitted beam from its nominal value of c . That corresponds to an index of refraction for the medium ($n = c/v$) which is other than one, even though **photons exist only at a speed c** .

To see how this comes about, return to Fig. 4.5. Recall that the scattered wavelets all combine in-phase in the forward direction to form what might best be called the *secondary wave*. For empirical reasons alone we can anticipate that the secondary wave will combine with what is left of the primary wave to yield the only observed disturbance within the medium, namely, the **transmitted wave**. *Both the primary and secondary electromagnetic waves propagate through the interatomic void with the speed c* . Yet the medium can certainly possess an index of refraction other than 1. The refracted wave may appear to have a phase velocity less than, equal to, or even greater than c . The key to this apparent contradiction resides in the phase relationship between the secondary and primary waves.

The classical model predicts that the electron-oscillators will be able to vibrate almost completely in-phase with the driving force (i.e., the primary disturbance) only at relatively low frequencies. As the frequency of the electromagnetic field increases, the oscillators will fall behind, lagging in phase by a proportionately larger amount. A detailed analysis reveals that at resonance the phase lag will reach 90° , increasing thereafter to almost 180° , or half a wavelength, at frequencies well above the particular characteristic value. Problem 4.4 explores this phase lag for a damped driven oscillator, and Fig. 4.9 summarizes the results.

In addition to these lags there is another effect that must be considered. When the scattered wavelets recombine, the resultant secondary wave* itself lags the oscillators by 90° .

The combined effect of both these mechanisms is that at frequencies below resonance, the secondary wave lags the pri-

*It has only recently been observed (and that was by chance) that inhomogeneous opaque materials, such as milk and white paint, can reduce the effective speed of light to as little as one-tenth the value anticipated for the medium. See S. John, "Localization of light," *Phys. Today* **44**, 32 (1991).

*This point will be made more plausible when we consider the predictions of the Huygens-Fresnel theory in the diffraction chapter. Most texts on E & M treat the problem of radiation from a sheet of oscillating charges, in which case the 90° phase lag is a natural result (see Problem 4.5).

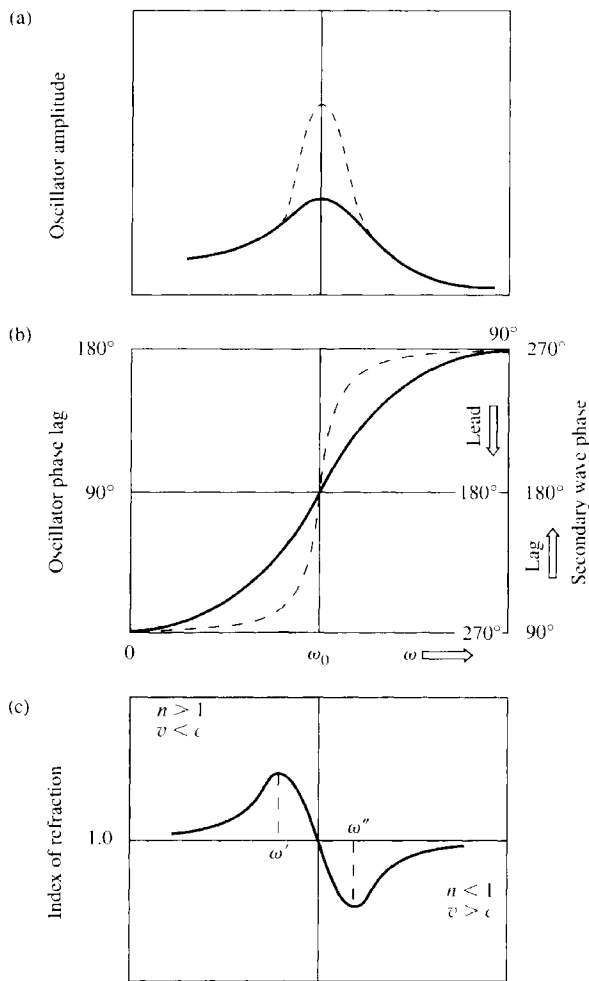


Figure 4.9 A schematic representation of (a) amplitude and (b) phase lag versus driving frequency for a damped oscillator. The dashed curves correspond to decreased damping. The corresponding index of refraction is shown in (c).

mary (Fig. 4.10) by some amount between approximately 90° and 180° , and at frequencies above resonance, the lag ranges from about 180° to 270° . But a phase lag of $\delta \geq 180^\circ$ is equivalent to a phase lead of $360^\circ - \delta$, [e.g., $\cos(\theta - 270^\circ) = \cos(\theta + 90^\circ)$]. This much can be seen on the right side of Fig. 4.9b.

Within the transparent medium, the primary and secondary waves overlap and, depending on their amplitudes and relative phase, generate the net transmitted disturbance. Except for the fact that it is weakened by scattering, the primary wave travels into the material just as if it were traversing free space. By

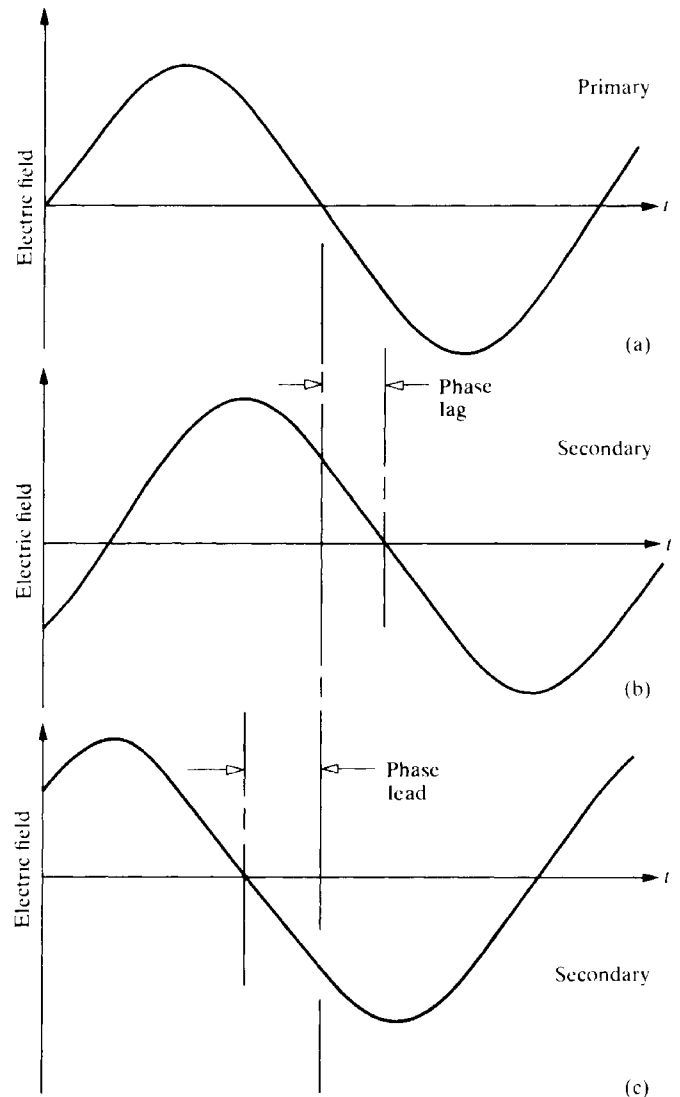


Figure 4.10 A primary wave (a) and two possible secondary waves. In (b) the secondary lags the primary—it takes longer to reach any given value. In (c) the secondary wave reaches any given value before (at an earlier time than) the primary; that is, it leads.

comparison to this free-space wave, which initiated the process, the resultant transmitted wave is phase shifted, and this phase difference is crucial.

When the secondary wave lags (or leads) the primary, the resultant transmitted wave must also lag (or lead) it by some amount (Fig. 4.11). This qualitative relationship will serve our purposes for the moment, although it should be noted that the phase of the resultant also depends on the amplitudes of the

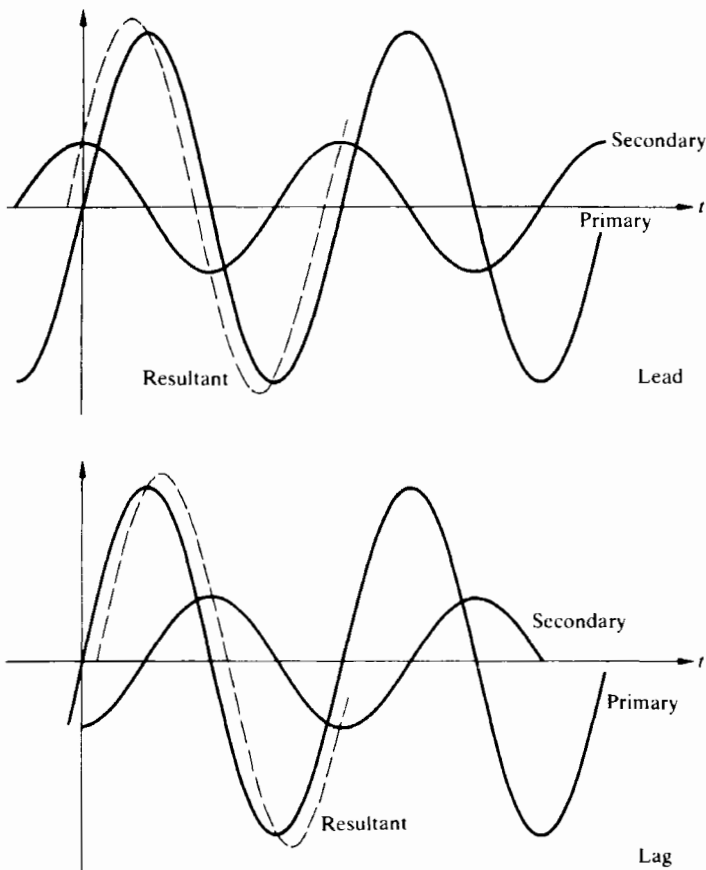


Figure 4.11 If the secondary leads the primary, the resultant will also lead it.

interacting waves [see Eq. (7.10)]. At frequencies below ω_0 the transmitted wave lags the free-space wave, whereas at frequencies above ω_0 it leads the free-space wave. For the special case in which $\omega = \omega_0$ the secondary and primary waves are out-of-phase by 180° . The former works against the latter, so that the refracted wave is appreciably reduced in amplitude although unaffected in phase.

As the transmitted wave advances through the medium, scattering occurs over and over again. Light traversing the substance is progressively retarded (or advanced) in phase. Evidently, since the speed of the wave is the rate of advance of the condition of constant phase, a change in the phase should correspond to a change in the speed.

We now wish to show that a phase shift is indeed tantamount to a difference in phase velocity. In free space, the dis-

turbance at some point P may be written as

$$E_P(t) = E_0 \cos \omega t \quad (4.1)$$

If P is surrounded by a dielectric, there will be a cumulative phase shift ϵ_P , which was built up as the wave moved through the medium to P . At ordinary levels of irradiance the medium will behave linearly, and the frequency in the dielectric will be the same as that in vacuum, even though the wavelength and speed may differ. Once again, but this time in the medium, the disturbance at P is

$$E_P(t) = E_0 \cos (\omega t - \epsilon_P) \quad (4.2)$$

where the subtraction of ϵ_P corresponds to a phase lag. An observer at P will have to wait a longer time for a given crest to arrive when she is in the medium than she would have had to wait in vacuum. That is, if you imagine two parallel waves of the same frequency, one in vacuum and one in the material, the vacuum wave will pass P a time ϵ_P/ω before the other wave. Clearly then, a phase lag of ϵ_P corresponds to a reduction in speed, $v < c$ and $n > 1$. Similarly, a phase lead yields an increase in speed, $v > c$ and $n < 1$. Again, the scattering process is a continuous one, and the cumulative phase shift builds as the light penetrates the medium. That is to say, ϵ is a function of the length of dielectric traversed, as it must be if v is to be constant (see Problem 4.5). In the vast majority of situations encountered in Optics $v < c$ and $n > 1$; see Table 4.1. The important exception is the case of X-ray propagation where $\omega > \omega_0$, $v > c$, and $n < 1$.

The overall form of $n(\omega)$, as depicted in Fig. 4.9c, can now be understood as well. At frequencies far below ω_0 the amplitudes of the oscillators and therefore of the secondary waves are very small, and the phase angles are approximately 90° . Consequently, the refracted wave lags only slightly, and n is only slightly greater than 1. As ω increases, the secondary waves have greater amplitudes and lag by greater amounts. The result is a gradually decreasing wave speed and an increasing value of $n > 1$. Although the amplitudes of the secondary waves continue to increase, their relative phases approach 180° as ω approaches ω_0 . Consequently, their ability to cause a further increase in the resultant phase lag diminishes. A turning point ($\omega = \omega'$) is reached where the refracted wave begins to experience a decreasing phase lag and an increasing speed ($dn/d\omega < 0$). That continues until $\omega = \omega_0$, whereupon the transmitted wave is appreciably reduced in amplitude but unaltered in phase and speed. At that point, $n = 1$, $v = c$, and we are more or less at the center of the absorption band.

TABLE 4.1 Approximate Indices of Refraction of Various Substances*

Air	1.000 29
Ice	1.31
Water	1.333
Ethyl alcohol (C ₂ H ₅ OH)	1.36
Fused quartz (SiO ₂)	1.458 4
Carbon tetrachloride (CCl ₄)	1.46
Turpentine	1.472
Benzene (C ₆ H ₆)	1.501
Plexiglass	1.51
Crown glass	1.52
Sodium chloride (NaCl)	1.544
Light flint glass	1.58
Polystyrene	1.59
Carbon disulfide (CS ₂)	1.628
Dense flint glass	1.66
Lanthanum flint glass	1.80
Zircon (ZrO ₂ ·SiO ₂)	1.923
Fabulite (SrTiO ₃)	2.409
Diamond (C)	2.417
Rutile (TiO ₂)	2.907
Gallium phosphide	3.50

*Values vary with physical conditions—purity, pressure, etc. These correspond to a wavelength of 589 nm

At frequencies just beyond ω_0 the relatively large-amplitude secondary waves lead; the transmitted wave is advanced in phase, and its speed exceeds c ($n < 1$). As ω increases, the whole scenario is played out again in reverse (with some asymmetry due to frequency-dependent asymmetry in oscillator amplitudes and scattering). At even higher frequencies the secondary waves, which now have very small amplitudes, lead by nearly 90° . The resulting transmitted wave is advanced very slightly in phase, and n gradually approaches 1.

The precise shape of a particular $n(\omega)$ curve depends on the specific oscillator damping, as well as on the amount of absorption, which in turn depends on the number of oscillators participating.

A rigorous solution to the propagation problem is known as the *Ewald–Oseen Extinction Theorem*. Although the mathematical formalism, involving integro-differential equations, is far too complicated to treat here, the results are certainly of interest. It is found that the electron-oscillators generate an

electromagnetic wave having essentially two terms. One of these precisely cancels the primary wave within the medium. The other, and only remaining disturbance, moves through the dielectric at a speed $v = c/n$ as the transmitted wave.* *Henceforth we shall simply assume that a lightwave propagating through any substantive medium travels at a speed $v \neq c$.*

Apparently, any quantum-mechanical model we construct will somehow have to associate a wavelength with the photon. That's easily done mathematically via the expression $p = h/\lambda$, even if it's not clear at this point what is doing the waving. Still the wave nature of light seems inescapable; it will have to be infused into the theory one way or another. And once we have the idea of a photon-wavelength, it's natural to bring in the concept of relative phase. Thus *the index of refraction arises when the absorption and emission process advances or retards the phases of the scattered photons, even as they travel at speed c .*

4.3 Reflection

When a beam of light impinges on the surface of a transparent material, such as a sheet of glass, the wave “sees” a vast array of closely spaced atoms that will somehow scatter it. Remember that the wave may be ≈ 500 nm long, whereas the atoms and their separations (≈ 0.2 nm) are thousands of times smaller. In the case of transmission through a dense medium, the scattered wavelets cancel each other in all but the forward direction, and just the ongoing beam is sustained. But that can only happen if there are no discontinuities. This is not the case at an interface between two different transparent media (such as air and glass), which is a jolting discontinuity. When a beam of light strikes such an interface, some light is always scattered backward, and we call this phenomenon **reflection**.

If the transition between two media is gradual—that is, if the dielectric constant (or the index of refraction) changes from that of one medium to that of the other over a distance of a wavelength or more—there will be very little reflection; the interface effectively vanishes. On the other hand, a transition from one medium to the other over a distance of $1/4$ wavelength or less behaves very much like a totally discontinuous change.

*For a discussion of the Ewald–Oseen theorem, see *Principles of Optics* by Born and Wolf, Section 2.4.2; this is heavy reading. Also look at Reali, “Reflection from dielectric materials.” *Am. J. Phys.* **50**, 1133 (1982).

Internal and External Reflection

Imagine that light is traveling across a large homogeneous block of glass (Fig. 4.12). Now, suppose that the block is sheared in half perpendicular to the beam. The two segments are then separated, exposing the smooth flat surfaces depicted in Fig. 4.12*b*. Just before the cut was made, there was no light-wave traveling to the left inside the glass—we know the beam only advances. Now there must be a wave (beam-I) moving to the left, reflected from the surface of the right-hand block. The implication is that a region of scatterers on and beneath the exposed surface of the right-hand block is now “unpaired,” and the backward radiation they emit can no longer be canceled. The region of oscillators that was adjacent to these, prior to the cut, is now on the section of the glass that is to the left. When the two sections were together, these scatterers presumably also emitted wavelets in the backward direction that were 180° out-of-phase with, and canceled, beam-I. Now they produce reflected beam-II. Each molecule scatters light in the backward direction, and, in principle, *each and every molecule contributes to the reflected wave*. Nonetheless, in practice, it is a thin layer ($\approx \lambda/2$ deep) of unpaired atomic oscillators near the surface that is effectively responsible for the reflection. For an air–glass interface, about 4% of the energy of an incident beam falling perpendicularly *in air on glass* will be reflected straight back out by this layer of unpaired scatterers (p. 90). And that’s true whether the glass is 1.0 mm thick or 1.00 m thick.

Beam-I reflects off the right-hand block, and because light was initially traveling from a less to a more optically dense medium, this is called **external reflection**. In other words, the index of the incident medium (n_i) is less than the index of the transmitting medium (n_t). Since the same thing happens to the unpaired layer on the section that was moved to the left, it, too, reflects backwards. With the beam incident perpendicularly *in glass on air*, 4% must again be reflected, this time as beam-II.

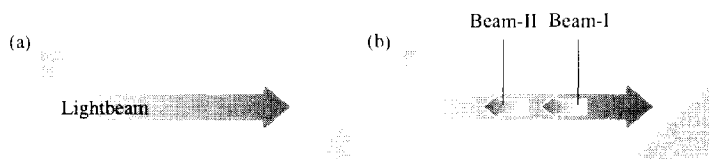


Figure 4.12 (a) A lightbeam propagating through a dense homogeneous medium such as glass. (b) When the block of glass is cut and parted, the light is reflected backward at the two new interfaces. Beam-I is externally reflected, and beam-II is internally reflected. Ideally, when the two pieces are pressed back together, the two reflected beams cancel one another.

This process is referred to as **internal reflection** because $n_i > n_t$. If the two glass regions are made to approach one another increasingly closely (so that we can imagine the gap to be a thin film of, say, air—p. 402), the reflected light will diminish until it ultimately vanishes as the two faces merge and disappear and the block becomes continuous again. Remember this ***180° relative phase shift between internally and externally reflected light*** (see Section 4.10 for a more rigorous treatment)—we will come back to it later on.

Experience with the common mirror makes it obvious that white light is reflected as white—it certainly isn’t blue. To see why, first remember that the layer of scatterers responsible for the reflection is effectively about $\lambda/2$ thick (as per Fig. 4.6). Thus the larger the wavelength, the deeper the region contributing (typically upward of a thousand atom layers), and the more scatterers there are acting together. This tends to balance out the fact that each scatterer is less efficient as λ

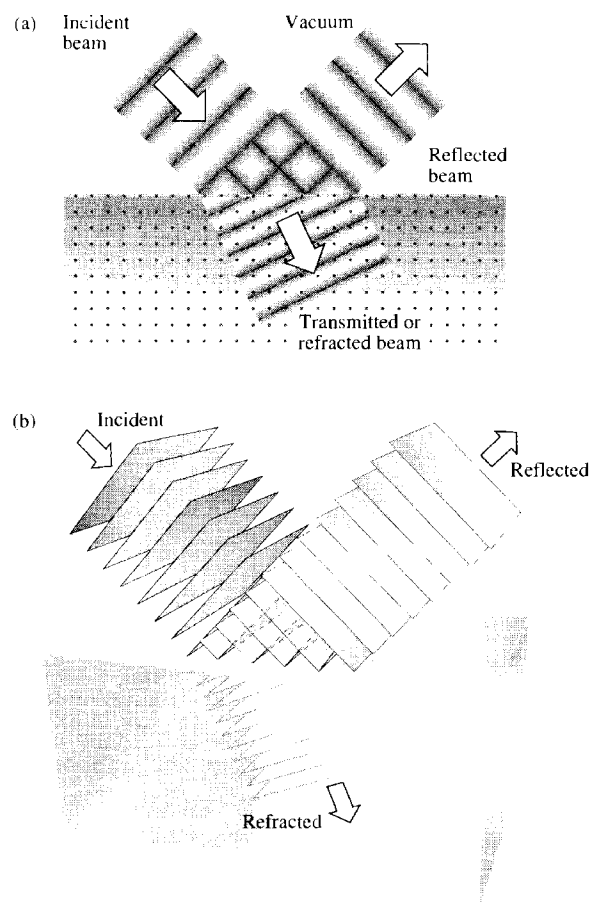


Figure 4.13 A beam of plane waves incident on a distribution of molecules constituting a piece of clear glass or plastic. Part of the incident light is reflected and part refracted.

increases (remember $1/\lambda^4$). The combined result is that *the surface of a transparent medium reflects all wavelengths about equally and doesn't appear colored in any way*. That, as we will see, is why this page looks white under white-light illumination.

4.3.1 The Law of Reflection

Figure 4.13 shows a beam composed of plane wavefronts impinging at some angle on the smooth, flat surface of an optically dense medium (let it be glass). Assume that the surrounding environment is vacuum. Follow one wavefront as it sweeps in and across the molecules on the surface. For the sake of simplicity, in Figs. 4.14 and 4.15 we have omitted everything but one molecular layer at the interface. As the wavefront descends, it energizes and re-energizes one scatterer after another, each of which radiates a stream of photons that can be regarded as a hemispherical wavelet in the incident medium. Because the wavelength is so much greater than the separation between the molecules, the wavelets emitted back into the incident medium advance together and add constructively in only one direction, and there is one well-defined *reflected beam*. That would not be true if the incident radiation was short-wavelength X-rays, in which circumstance there would be several reflected beams. And it would not be true if the scatterers were far apart compared to λ , as they are for a diffraction grating (p. 476), in which case there would also be several reflected beams. The direction of the reflected beam is

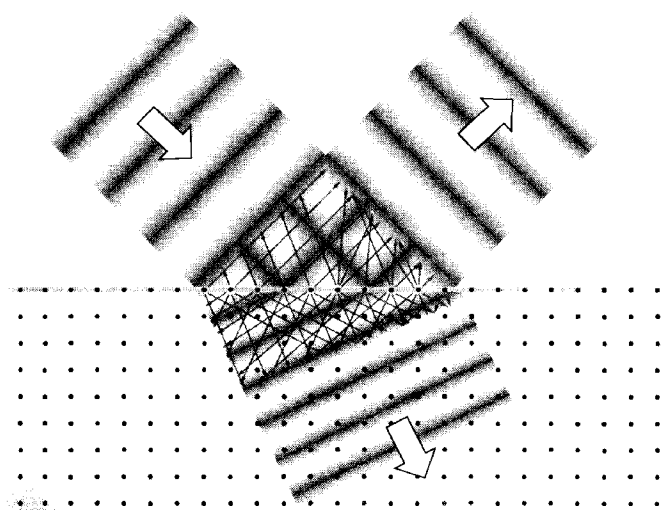


Figure 4.14 A plane wave sweeps in stimulating atoms across the interface. These radiate and reradiate, thereby giving rise to both the reflected and transmitted waves.

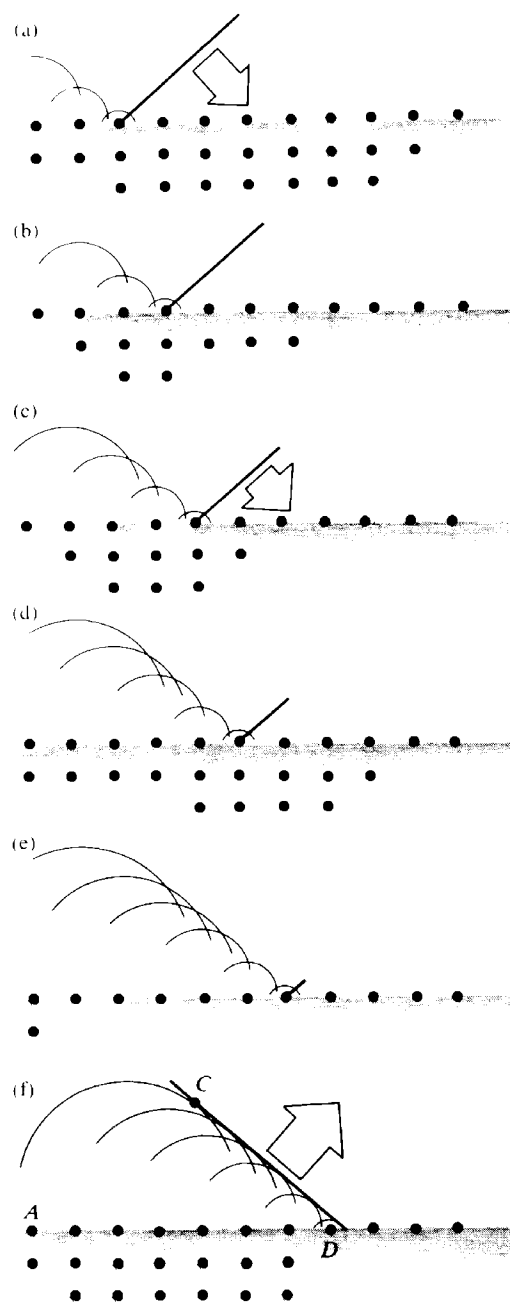


Figure 4.15 The reflection of a wave as the result of scattering.

determined by the constant phase difference between the atomic scatterers. That, in turn, is determined by the angle made by the incident wave and the surface, the so-called **angle-of-incidence**.

In Fig. 4.16, the line $\overline{AB}$ lies along an incoming wavefront, while $\overline{CD}$ lies on an outgoing wavefront—in effect, $\overline{AB}$ transforms on reflection into $\overline{CD}$. With Fig. 4.15 in mind, we see

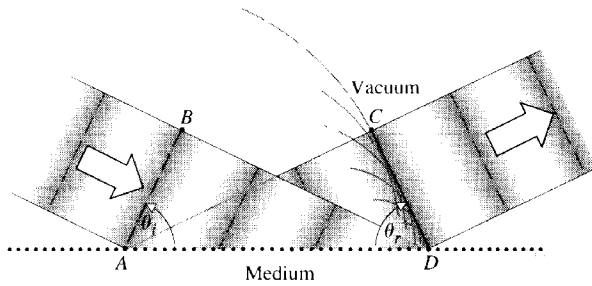


Figure 4.16 Plane waves enter from the left and are reflected off to the right. The reflected wavefront $\overline{CD}$ is formed of waves scattered by the atoms on the surface from A to D . Just as the first wavelet arrives at C from A , the atom at D emits, and the wavefront along $\overline{CD}$ is completed.

that the wavelet emitted from A will arrive at C in-phase with the wavelet just being emitted from D (as it is stimulated by B), as long as the distances $\overline{AC}$ and $\overline{BD}$ are equal. In other words, if all the wavelets emitted from all the surface scatterers are to overlap in-phase and form a single reflected plane wave, it must be that $\overline{AC} = \overline{BD}$. Then, since the two triangles have a common hypotenuse

$$\frac{\sin \theta_i}{\overline{BD}} = \frac{\sin \theta_r}{\overline{AC}}$$

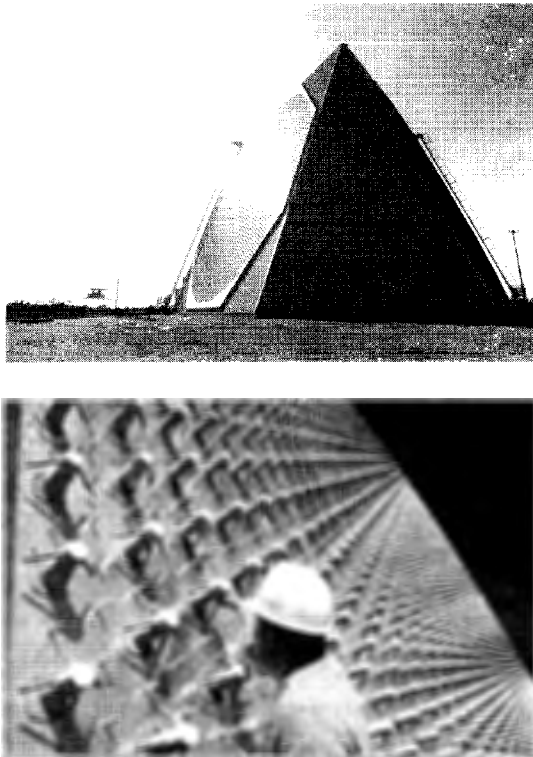
All the waves travel in the incident medium with the same speed v_i . It follows that in the time (Δt) it takes for point B on the wavefront to reach point D on the surface, the wavelet emitted from A reaches point C . In other words, $\overline{BD} = v_i \Delta t = \overline{AC}$, and so from the above equation, $\sin \theta_i = \sin \theta_r$, which means that

$$\theta_i = \theta_r \quad (4.3)$$

The angle-of-incidence equals the angle-of-reflection. This equation is the first part of the **Law of Reflection**. It initially appeared in the book *Catoptrics*, which was purported to have been written by Euclid. We say that a beam is *normally incident* when $\theta_i = 0^\circ$, in which case $\theta_r = 0^\circ$ and for a mirror the beam reflects back on itself. Similarly, *glancing incidence* corresponds to $\theta_i \approx 90^\circ$ and perforce $\theta_r \approx 90^\circ$.

Rays

Drawing wavefronts can get things a bit cluttered, so we introduce another convenient scheme for visualizing the progression of light. The imagery of antiquity was in terms of straight-line streams of light, a notion that got into Latin as “radii” and reached English as “rays.” **A ray is a line drawn in space corresponding to the direction of flow of radiant energy.** It is a mathematical construct and not a physical entity. In a medium that is uniform (homogeneous), rays are straight. If the medium behaves in the same manner in every direction (isotropic), **the rays are perpendicular to the wavefronts.** Thus for a point source emitting spherical waves, the rays, which are perpendicular to them, point radially outward from the source. Similarly, the rays associated with plane waves are all parallel. Rather than sketching bundles of rays, we can simply draw one incident ray and one reflected ray (Fig. 4.17a). *All the angles are now measured from the perpendicular (or*



A modern phased-array radar system. The field of individual small antennas behaves very much like the atoms on a smooth surface. By introducing a proper phase shift between adjacent rows the antenna can “look” in any direction. A reflecting surface has a similar phase shift determined by θ_i as the incident wave sweeps over the array of atoms. (Courtesy Raytheon Corp.)

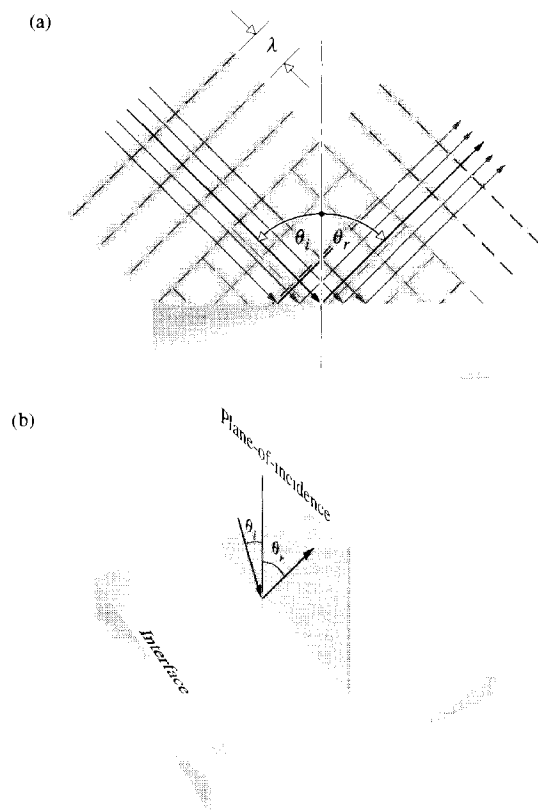


Figure 4.17 (a) Select one ray to represent the beam of plane waves. Both the angle-of-incidence θ_i and the angle-of-reflection θ_r are measured from a perpendicular drawn to the reflecting surface. (b) The incident ray and the reflected ray define the *plane-of-incidence*, perpendicular to the reflecting surface.

normal) to the surface, and θ_i and θ_r have the same numerical values as before (Fig. 4.16).

The ancient Greeks knew the Law of Reflection. It can be deduced by observing the behavior of a flat mirror, and nowadays that observation can be done most simply with a flashlight or, even better, a low-power laser. The second part of the Law of Reflection maintains that *the incident ray, the perpendicular to the surface, and the reflected ray all lie in a plane* called the **plane-of-incidence** (Fig. 4.17b)—this is a three-dimensional business. Try to hit some target in a room with a flashlight beam by reflecting it off a stationary mirror, and the importance of this second part of the law becomes obvious!

Figure 4.18a shows a beam of light incident upon a reflecting surface that is smooth (one for which any irregularities are small compared to a wavelength). In that case, the light re-

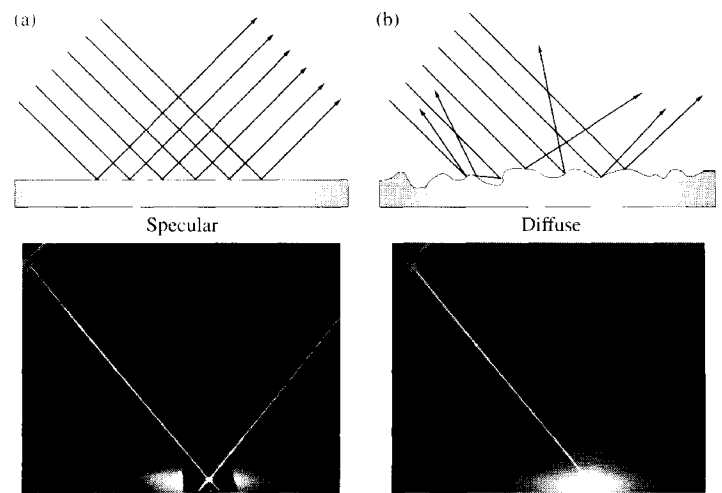


Figure 4.18 (a) Specular reflection. (b) Diffuse reflection. (Photos courtesy Donald Dunitz.)

emitted by millions upon millions of atoms will combine to form a single well-defined beam in a process called **specular reflection** (from the word for a common mirror alloy in ancient times, *speculum*). Provided the ridges and valleys are small compared to λ , the scattered wavelets will still arrive more or less in-phase when $\theta_i = \theta_r$. This is the situation assumed in Figs. 4.13, 4.15, 4.16, and 4.17. On the other hand, when the surface is rough in comparison to λ , although the angle-of-incidence will equal the angle-of-reflection for each ray, the whole lot of rays will emerge every which way, constituting what is called **diffuse reflection** (see photo). Both of these conditions are extremes; the reflecting behavior of most



The cruiser *Aurora*, which played a key role in the Communist Revolution (1917), docked in St. Petersburg. Where the water is still, the reflection is specular. The image blurs where the water is rough and the reflection diffuse.



The F-117A Stealth fighter has an extremely small radar profile, that is, it returns very little of the incoming microwaves back to the station that sent them. That's accomplished mostly by constructing the aircraft with flat tilted-planes that use the Law of Reflection to scatter the radar waves away from their source. One wants to avoid $\theta_i = \theta_r \approx 0$.

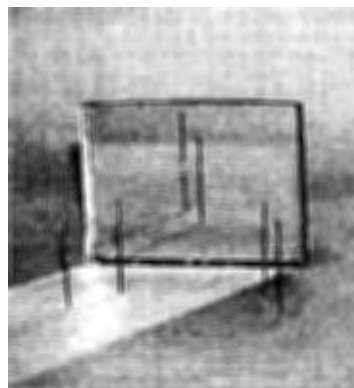
surfaces lies somewhere between them. Thus, although the paper of this page was deliberately manufactured to be a fairly diffuse scatterer, the cover of the book reflects in a manner that is somewhere between diffuse and specular.

4.4 Refraction

Figure 4.13 shows a beam of light impinging on an interface at some angle ($\theta_i \neq 0$). The interface corresponds to a major inhomogeneity, and the atoms that compose it scatter light both backward, as the reflected beam, and forward, as the transmitted beam. The fact that the incident rays are bent or “turned out of their way,” as Newton put it, is called **refraction**.

Examine the transmitted or refracted beam. Speaking classically, each energized molecule on the interface radiates wavelets into the glass that expand out at speed c . These can be imagined as combining into a secondary wave that then recombines with the unscattered remainder of the primary wave, to form the net transmitted wave. The process continues over and over again as the wave advances in the transmitting medium.

However we visualize it, immediately on entering the transmitting medium, there is a single net field, a single net wave. As we have seen, this transmitted wave usually propagates with an effective speed $v_t < c$. It's essentially as if the atoms at the interface scattered “slow wavelets” into the glass that combine to form the “slow transmitted wave.” We'll come back to this imagery when we talk about Huygens's Principle. In any event, because the cooperative phenomenon



By placing a pair of pins in front of a flat mirror and aligning their images with another pair of pins, you can easily verify that $\theta_i = \theta_r$. (Photo by E.H.)

known as the transmitted electromagnetic wave *is* slower than the incident electromagnetic wave, the transmitted wavefronts are refracted, displaced (turned with respect to the incident wavefronts), and the beam bends.

4.4.1 The Law of Refraction

Figure 4.19 picks up where we left off with Figs. 4.13 and 4.16. The diagram depicts several wavefronts, all shown at a single instant in time. Remember that each wavefront is a surface of constant phase, and, to the degree that the phase of the

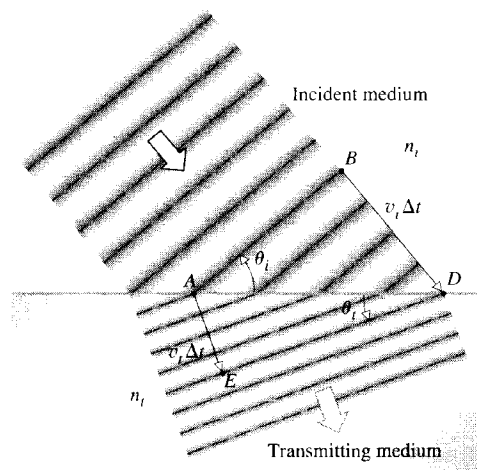


Figure 4.19 The refraction of waves. The atoms in the region of the surface of the transmitting medium reradiate wavelets that combine constructively to form a refracted beam.

net field is retarded by the transmitting medium, each wavefront is held back, as it were. The wavefronts “bend” as they cross the boundary because of the speed change. Alternatively, we can envision Fig. 4.19 as a multiple-exposure picture of a single wavefront showing it after successive equal intervals of time. Notice that in the time Δt , which it takes for point B on a wavefront (traveling at speed v_i) to reach point D , the transmitted portion of that same wavefront (traveling at speed v_t) has reached point E . If the glass ($n_t = 1.5$) is immersed in an incident medium that is vacuum ($n_i = 1$) or air ($n_i = 1.0003$) or anything else where $n_t > n_i$, $v_t < v_i$ and $\overline{AE} < \overline{BD}$, the wavefront bends. The refracted wavefront extends from E to D , making an angle with the interface of θ_t . As before, the two triangles ABD and AED in Fig. 4.19 share a common hypotenuse ($\overline{AD}$), and so

$$\frac{\sin \theta_i}{\overline{BD}} = \frac{\sin \theta_t}{\overline{AE}}$$

where $\overline{BD} = v_i \Delta t$ and $\overline{AE} = v_t \Delta t$. Hence

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

Multiply both sides by c , and since $n_i = c/v_i$ and $n_t = c/v_t$

$$n_i \sin \theta_i = n_t \sin \theta_t \quad (4.4)$$

This equation is the first portion of the **Law of Refraction**, also known as **Snell's Law** after the man who proposed it (1621), Willebrord Snel van Royen (1591–1626). Snel's analysis has been lost, but contemporary accounts follow the treatment shown in Fig. 4.20. What was found through observation was that the bending of the rays could be quantified via

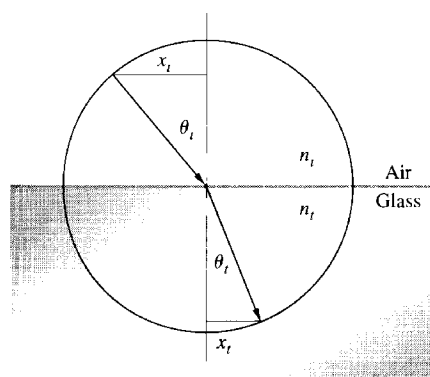


Figure 4.20 Descartes's arrangement for deriving the Law of Refraction.

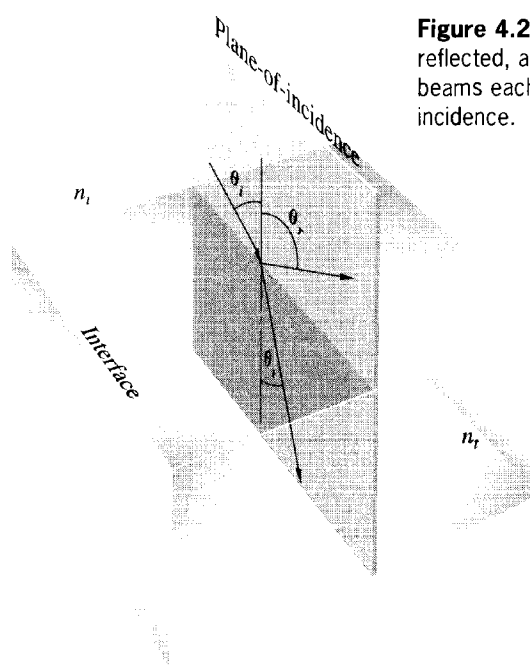


Figure 4.21 The incident, reflected, and transmitted beams each lie in the plane-of-incidence.

the ratio of x_i to x_t , which was constant for all θ_i . That constant was naturally enough called the *index of refraction*. In other words,

$$\frac{x_i}{x_t} \equiv n_t$$

and in air that's equivalent to Eq. (4.4). We now know that the Englishman Thomas Harriot had come to the same conclusion before 1601, but he kept it to himself.

At first, the indices of refraction were simply experimentally determined constants of the physical media. Later, Newton was actually able to derive Snell's Law using his own corpuscular theory. By then, the significance of n as a measure of the speed of light was evident. Still later, Snell's Law was shown to be a natural consequence of Maxwell's Electromagnetic Theory (p. 112).

It is again convenient to transform the diagram into a ray representation (Fig. 4.21) wherein all the angles are measured from the perpendicular. Along with Eq. (4.4), there goes the understanding that *the incident, reflected, and refracted rays all lie in the plane-of-incidence*. In other words, the respective unit propagation vectors $\hat{\mathbf{k}}_i$, $\hat{\mathbf{k}}_r$, and $\hat{\mathbf{k}}_t$ are coplanar (Fig. 4.22).

When $n_i < n_t$ (that is, when the light is initially traveling within the lower-index medium), it follows from Snell's Law that $\sin \theta_i > \sin \theta_t$, and since the same function is everywhere positive between 0° and 90° , then $\theta_i > \theta_t$. Rather than going

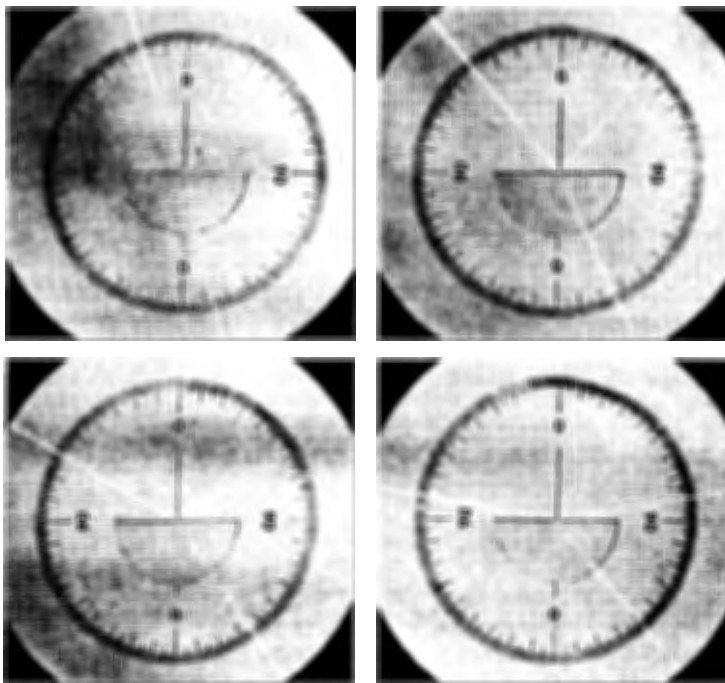
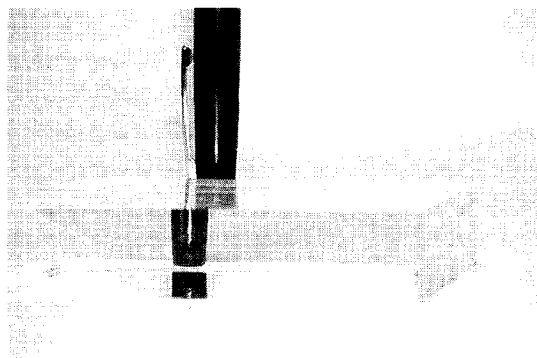


Figure 4.22 Refraction at various angles of incidence. Notice that the bottom surface is cut circular so that the transmitted beam within the glass always lies along a radius and is normal to the lower surface in every case. (Photos courtesy PSSC College Physics, D. C. Heath & Co., 1968.)



The image of a pen seen through a thick block of clear plastic. The displacement of the image arises from the refraction of light toward the normal at the air-plastic interface. If this arrangement is set up with a narrow object (e.g., an illuminated slit) and the angles are carefully measured, you can confirm Snell's Law directly. (Photo by E.H.)

straight through, *the ray entering a higher-index medium bends toward the normal* (Fig. 4.23a). The reverse is also true (Fig. 4.23b); that is, *on entering a medium having a lower index, the ray, rather than going straight through, will bend away from the normal* (see photo). Notice that this implies that the rays will traverse the same path going either way, into or out of either medium. The arrows can be reversed and the resulting picture is still true.

Snell's Law can be rewritten in the form

$$\frac{\sin\theta_i}{\sin\theta_r} = n_{ri} \quad (4.5)$$

where $n_{ri} \equiv n_r/n_i$ is the **relative index of refraction** of the two media.

Let $\hat{\mathbf{u}}_n$ be a unit vector normal to the interface pointing in the direction from the incident to the transmitting medium (Fig. 4.24). As you will have the opportunity to prove in Problem 4.29, the complete statement of the Law of Refraction can be written vectorially as

$$n_i(\hat{\mathbf{k}}_i \times \hat{\mathbf{u}}_n) = n_r(\hat{\mathbf{k}}_r \times \hat{\mathbf{u}}_n) \quad (4.6)$$

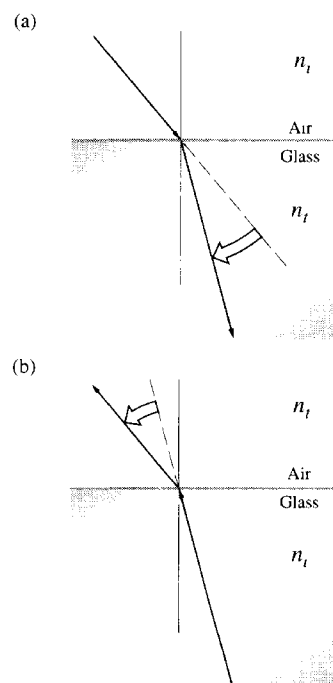
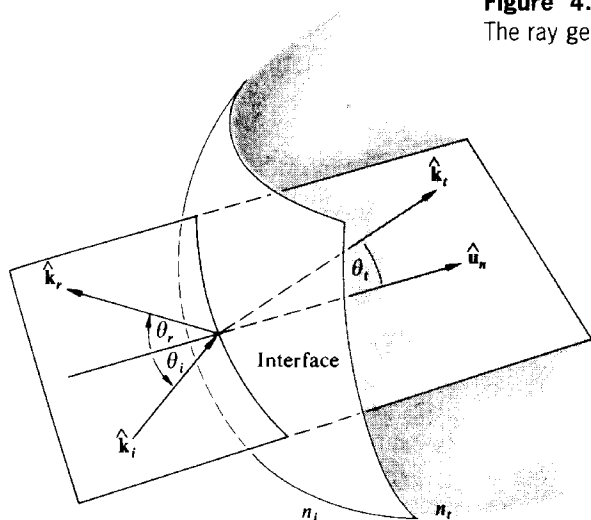


Figure 4.23 The bending of rays at an interface. (a) When a beam of light enters a more optically dense medium, one with a greater index of refraction ($n_i < n_r$), it bends toward the perpendicular. (b) When a beam goes from a more dense to a less dense medium ($n_i > n_r$), it bends away from the perpendicular.

Figure 4.24
The ray geometry.



or alternatively,

$$n_t \hat{\mathbf{k}}_t - n_i \hat{\mathbf{k}}_i = (n_t \cos \theta_t - n_i \cos \theta_i) \hat{\mathbf{u}}_n \quad (4.7)$$

Fig. 4.19 illustrates the three important changes that occur in the beam traversing the interface. (1) It changes direction. Because the leading portion of the wavefront in the glass slows down, the part still in the air advances more rapidly, sweeping past and bending the wave toward the normal. (2) The beam in the glass has a broader cross section than the beam in the air; hence, the transmitted energy is spread thinner. (3) The wavelength decreases because the frequency is unchanged while the speed decreases; $\lambda = v/\nu = c/n\nu$ and

$$\lambda = \frac{\lambda_0}{n} \quad (4.8)$$

This latter notion suggests that the color aspect of light is better thought of as associated with its frequency (or energy, $\mathcal{E} = h\nu$) than its wavelength, since the wavelength changes with the medium through which the light moves. Color is so much a physio-psychological phenomenon (p. 131) that it must be treated rather gingerly. Still, even though it's a bit simplistic, it's useful to remember that blue photons are more energetic than red photons. When we talk about wavelengths and colors, we should always be referring to **vacuum wavelengths** (henceforth to be given as λ_0).

In all the situations treated thus far, it was assumed that the reflected and refracted beams always had the same frequency

as the incident beam, and that's ordinarily a reasonable assumption. Light of frequency ν impinges on a medium and presumably drives the molecules into simple harmonic motion. That's certainly the case when the amplitude of the vibration is fairly small, as it is when the electric field driving the molecules is small. The E -field for bright sunlight is only about 1000 V/m (while the B -field is less than a tenth of the Earth's surface field). This isn't very large compared to the fields keeping a crystal together, which are of the order of 10^{11} V/m—just about the same magnitude as the cohesive field holding the electron in an atom. We can usually expect the oscillators to vibrate in simple harmonic motion, and so the frequency will remain constant—the medium will ordinarily respond linearly. That will not be true, however, if the incident beam has an exceedingly large-amplitude E -field, as can be the case with a high-power laser. So driven, at some frequency ν the medium can behave in a nonlinear fashion, resulting in reflection and refraction of harmonics (2ν , 3ν , etc.) in addition to ν . Nowadays, second-harmonic generators (p. 641) are available commercially. You shine red light (694.3 nm) into an appropriately oriented transparent nonlinear crystal (of, for example, potassium dihydrogen phosphate, KDP, or ammonium dihydrogen phosphate, ADP) and out will come a beam of UV (347.15 nm).

One feature of the above treatment merits some further discussion. It was reasonably assumed that each point on the interface in Fig. 4.13 coincides with a particular point on each of the incident, reflected, and transmitted waves. In other words, there is a fixed phase relationship between each of the



Rays from the submerged portion of the pencil bend on leaving the water as they rise toward the viewer. (Photo by E.H.)

waves at all points along the interface. As the incident front sweeps across the interface, every point on it in contact with the interface is also a point on both a corresponding reflected front and a corresponding transmitted front. This situation is known as **wavefront continuity**, and it will be justified in a more mathematically rigorous treatment in Section 4.6.1. Interestingly, Sommerfeld* has shown that the laws of reflection and refraction (independent of the kind of wave involved) can be derived directly from the requirement of wavefront continuity and the solution to Problem 4.26 demonstrates as much.

4.4.2 Huygens's Principle

Suppose that light passes through a nonuniform sheet of glass, as in Fig. 4.25, so that the wavefront Σ is distorted. How can we determine its new form Σ' ? Or for that matter, what will Σ' look like at some later time, if it is allowed to continue unobstructed?

A preliminary step toward the solution of this problem appeared in print in 1690 in the work entitled *Traité de la Lumière*, which had been written 12 years earlier by the Dutch physicist Christiaan Huygens. It was there that he enunciated what has since become known as **Huygens's Principle**: *every point on a propagating wavefront serves as the source of spherical secondary wavelets, such that the wavefront at some later time is the envelope of these wavelets.*

A further crucial point is that *if the propagating wave has a frequency ν , and is transmitted through the medium at a speed v , then the secondary wavelets have that same frequency and speed.* Huygens was a brilliant scientist, and this is the basis of a remarkably insightful, though quite naive, scattering theory. It's a very early treatment and naturally has several shortcomings, one of which is that it doesn't overtly incorporate the concept of interference and perforce cannot deal with lateral scattering. Moreover, the idea that the secondary wavelets propagate at a speed determined by the medium (a speed that may even be anisotropic, e.g., p. 340) is a happy guess. Nonetheless, Huygens's Principle can be used to arrive at Snell's Law in a way that's similar to the treatment that led to Eq. (4.4). It's probably best not to fuss over the physical details (such as how to rationalize propagation in vac-

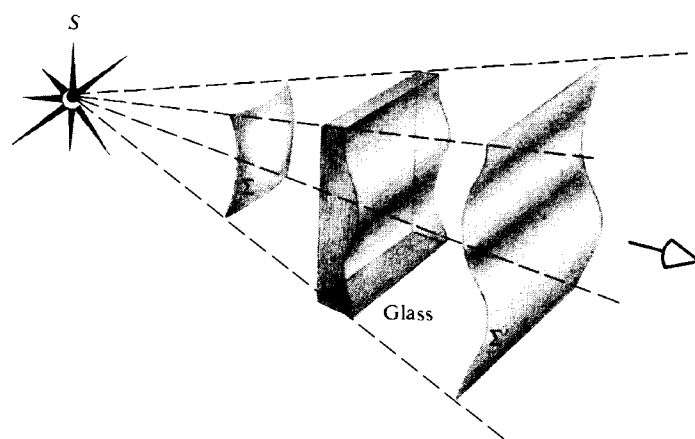


Figure 4.25 Distortion of a portion of a wavefront on passing through a material of nonuniform thickness.

uum) and just use the Principle as a tool—a highly useful fiction that works. After all, if Einstein is right, there are only scattered photons; the wavelets themselves are a theoretical construct.

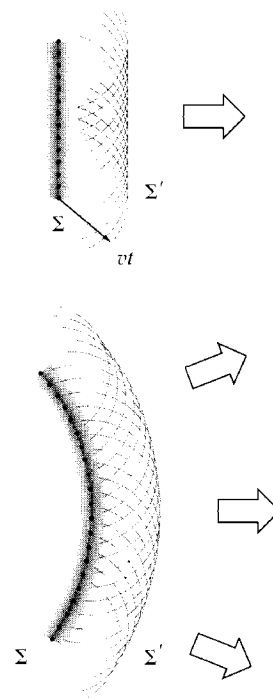


Figure 4.26 According to Huygens's Principle, a wave propagates as if the wavefront were composed of an array of point sources, each emitting a spherical wave.

*A. Sommerfeld, *Optics*, p. 151. See also J. J. Sein, *Am. J. Phys.* **50**, 180 (1982).

If the medium is homogeneous, the wavelets may be constructed with finite radii, whereas if it is inhomogeneous, the wavelets must have infinitesimal radii. Figure 4.26 should make this fairly clear; it shows a view of a wavefront Σ , as well as a number of spherical secondary wavelets, which, after a time t , have propagated out to a radius of vt . The envelope of all these wavelets is then asserted to correspond to the advanced wave Σ' . It is easy to visualize the process in terms of mechanical vibrations of an elastic medium. Indeed, this is the way that Huygens envisioned it within the context of an all-pervading aether, as is evident from his comment:

We have still to consider, in studying the spreading out of these waves, that each particle of matter in which a wave proceeds not only communicates its motion to the next particle to it, which is on the straight line drawn from the luminous point, but that it also necessarily gives a motion to all the others which touch it and which oppose its motion. The result is that around each particle there arises a wave of which this particle is a center.

Fresnel, in the 1800s, successfully modified Huygens's Principle mathematically adding in the concept of interference. A little later on, Kirchhoff showed that the *Huygens–Fresnel Principle* was a direct consequence of the differential wave equation [Eq. (2.60)], thereby putting it on a firm mathematical base. That there was a need for a reformulation of the principle is evident from Fig. 4.26, where we deceptively only drew hemispherical wavelets.* Had we drawn them as spheres, there would have been a *backwave* moving toward the source—something that is not observed. Since this difficulty was taken care of theoretically by Fresnel and Kirchhoff, we need not be disturbed by it.

4.4.3 Light Rays and Normal Congruence

In practice, one can produce very narrow *beams* or *pencils* of light (e.g., a laserbeam), and we might imagine a ray to be the unattainable limit on the narrowness of such a beam. Bear in mind that in an *isotropic medium* (i.e., one whose properties are the same in all directions) **rays are orthogonal trajectories of the wavefronts**. That is to say, *they are lines normal to the wavefronts at every point of intersection*. Evidently, in such a

medium a ray is parallel to the propagation vector $\vec{k}$. As you might suspect, this is not true in *anisotropic* substances, which we will consider later (see Section 8.4.1). Within homogeneous isotropic materials, rays will be straight lines, since by symmetry they cannot bend in any preferred direction, there being none. Moreover, because the speed of propagation is identical in all directions within a given medium, the spatial separation between two wavefronts, measured along rays, must be the same everywhere.* Points where a single ray intersects a set of wavefronts are called *corresponding points*, for example, A , A' , and A'' in Fig. 4.27. Evidently, the separation in time between any two corresponding points on any two sequential wavefronts is identical. If wavefront Σ is transformed into Σ'' after a time t'' , the distance between corresponding points on any and all rays will be traversed in that same time t'' . This will be true even if the wavefronts pass from one homogeneous isotropic medium into another. This just means that each point on Σ can be imagined as following the path of a ray to arrive at Σ'' in the time t'' .

If a group of rays is such that we can find a surface that is orthogonal to each and every one of them, they are said to form a **normal congruence**. For example, the rays emanating from a point source are perpendicular to a sphere centered at the source and consequently form a normal congruence.

We can now briefly consider a scheme that will also allow us to follow the progress of light through various isotropic media. The basis for this approach is the **Theorem of Malus and Dupin** (introduced in 1808 by E. Malus and modified in

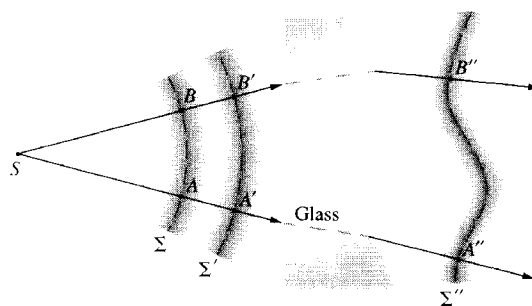


Figure 4.27 Wavefronts and rays.

*When the material is inhomogeneous or when there is more than one medium involved, it will be the *optical path length* (see Section 4.5) between the two wavefronts that is the same.

*See E. Hecht, *Phys. Teach.* **18**, 149 (1980).

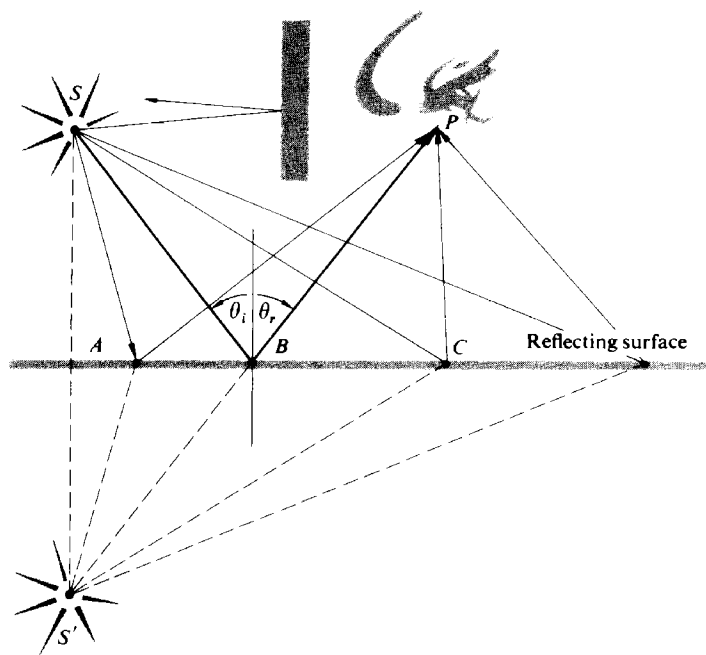


Figure 4.28 Minimum path from the source S to the observer's eye at P .

1816 by C. Dupin), according to which **a group of rays will preserve its normal congruence after any number of reflections and refractions** (as in Fig. 4.27). From our present vantage point of the wave theory, this is equivalent to the statement that rays remain orthogonal to wavefronts throughout all propagation processes in isotropic media. As shown in Problem 4.28, the theorem can be used to derive the Law of Reflection as well as Snell's Law. It is often most convenient to carry out a ray trace through an optical system and then reconstruct the wavefronts using the idea of equal transit times between corresponding points and the orthogonality of the rays and wavefronts.

4.5 Fermat's Principle

The laws of reflection and refraction, and indeed the manner in which light propagates in general, can be viewed from an entirely different and intriguing perspective afforded us by **Fermat's Principle**. The ideas that will unfold presently have had a tremendous influence on the development of physical thought in and beyond the study of Classical Optics.

Hero of Alexandria, who lived some time between 150 B.C.E. and A.D. 250, was the first to propose what has since become known as a *variational principle*. In his treatment of reflection, he asserted that *the path taken by light in going from some point S to a point P via a reflecting surface was the shortest possible one*. This can be seen rather easily in Fig. 4.28, which depicts a point source S emitting a number of rays that are then “reflected” toward P . Presumably, only one of these paths will have any physical reality. If we draw the rays as if they emanated from S' (the image of S), none of the distances to P will have been altered (i.e., $SAP = S'AP$, $SBP = S'BP$, etc.). But obviously the straight-line path $S'BP$, which corresponds to $\theta_i = \theta_r$, is the shortest possible one. The same kind of reasoning (Problem 4.31) makes it evident that points S , B , and P must lie in what has previously been defined as the plane-of-incidence. For over fifteen hundred years Hero's curious observation stood alone, until in 1657 Fermat propounded his celebrated *Principle of Least Time*, which encompassed both reflection and refraction. A beam of light traversing an interface does not take a straight line or *minimum spatial path* between a point in the incident medium and one in the transmitting medium. Fermat consequently reformulated Hero's statement to read: *the actual path between two points taken by a beam of light is the one that is traversed in the least time*. As we shall see, even this form of the statement is incomplete and a bit erroneous at that. For the moment then, let us embrace it but not passionately.

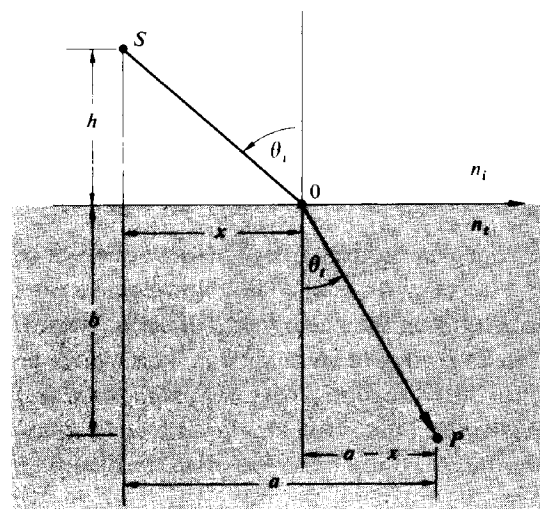


Figure 4.29 Fermat's Principle applied to refraction.

As an example of the application of the principle to the case of refraction, refer to Fig. 4.29, where we minimize t , the transit time from S to P , with respect to the variable x . In other words, changing x shifts point O , changing the ray from S to P . The smallest transit time will then presumably coincide with the actual path. Hence

$$t = \frac{\overline{SO}}{v_i} + \frac{\overline{OP}}{v_t}$$

or

$$t = \frac{(h^2 + x^2)^{1/2}}{v_i} + \frac{[b^2 + (a - x)^2]^{1/2}}{v_t}$$

To minimize $t(x)$ with respect to variations in x , we set $dt/dx = 0$, that is,

$$\frac{dt}{dx} = \frac{x}{v_i(h^2 + x^2)^{1/2}} + \frac{-(a - x)}{v_t[b^2 + (a - x)^2]^{1/2}} = 0$$

Using the diagram, we can rewrite the expression as

$$\frac{\sin \theta_i}{v_i} = \frac{\sin \theta_t}{v_t}$$

which is no less than Snell's Law (Eq. 4.4). If a beam of light is to advance from S to P in the least possible time, it must comply with the Law of Refraction.

Suppose that we have a stratified material composed of m layers, each having a different index of refraction, as in Fig. 4.30. The transit time from S to P will then be

$$t = \frac{s_1}{v_1} + \frac{s_2}{v_2} + \cdots + \frac{s_m}{v_m}$$

or

$$t = \sum_{i=1}^m s_i/v_i$$

where s_i and v_i are the path length and speed, respectively, associated with the i th contribution. Thus

$$t = \frac{1}{c} \sum_{i=1}^m n_i s_i \quad (4.9)$$

in which the summation is known as the **optical path length** (OPL) traversed by the ray, in contrast to the spatial path length $\sum_{i=1}^m s_i$. Clearly, for an inhomogeneous medium where n is a function of position, the summation must be changed to an integral:

$$OPL = \int_S^P n(s) ds \quad (4.10)$$

The optical path length corresponds to the distance in vacuum

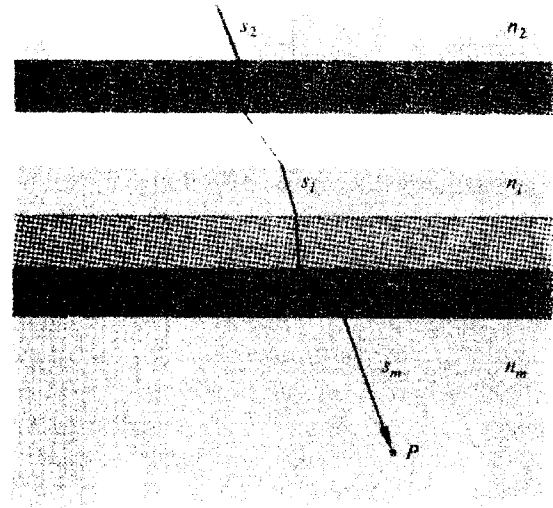


Figure 4.30 A ray propagating through a layered material.

equivalent to the distance traversed (s) in the medium of index n . That is, the two will correspond to the same number of wavelengths, $(OPL)/\lambda_0 = s/\lambda$, and the same phase change as the light advances.

Inasmuch as $t = (OPL)/c$, we can restate Fermat's Principle: *light, in going from point S to P , traverses the route having the smallest optical path length.*

Fermat and Mirages

When light rays from the Sun pass through the inhomogeneous atmosphere of the Earth, as shown in Fig. 4.31, they bend so as to traverse the lower, denser regions as abruptly as

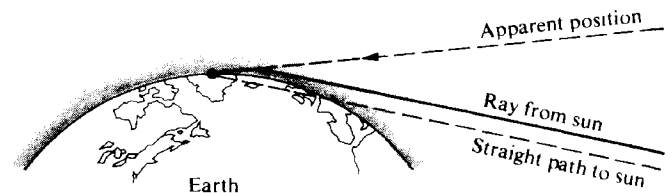


Figure 4.31 The bending of rays through inhomogeneous media. Because the rays bend as they pass through the atmosphere the Sun appears higher in the sky.

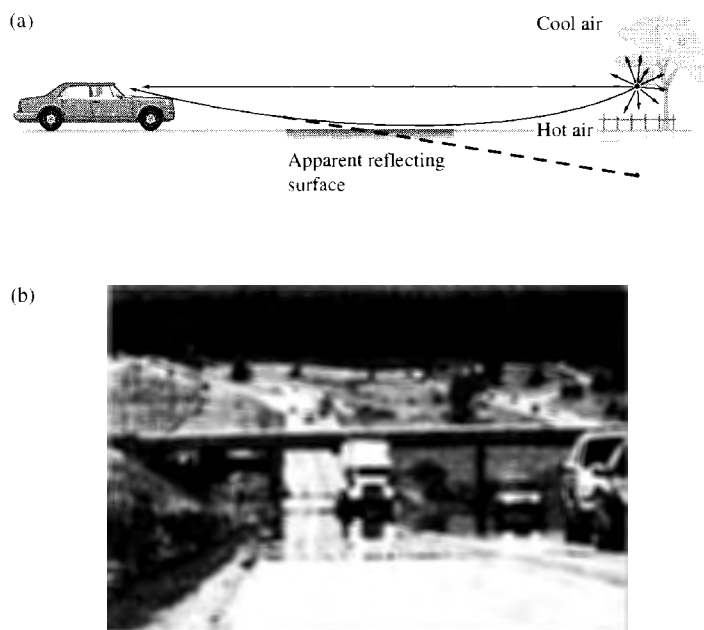


Figure 4.32 (a) At very low angles the rays appear to be coming from beneath the road as if reflected in a puddle. (b) A photo of this puddle effect (Photo courtesy Matt Malloy and Dan MacIsaac, NAU Physics & Astronomy.)

possible, minimizing the *OPL*. Ergo, one can still see the Sun after it has actually passed below the horizon.

In the same way, a road viewed at a glancing angle, as in Fig. 4.32, appears to reflect the environs as if it were covered with a sheet of water. The air near the roadway is warmer and less dense than that farther above it. It was established experimentally by Gladstone and Dale that for a gas of density ρ

$$(n - 1) \propto \rho$$

It follows from the Ideal Gas Law that at a fixed pressure, since $\rho \propto P/T$, $(n - 1) \propto 1/T$; the hotter the road, the lower the index of refraction of the air immediately above it.

According to Fermat's Principle, a ray leaving a branch in Fig. 4.32a heading somewhat downward would take a route that minimized the *OPL*. Such a ray would bend upward, passing through more of the less dense air than if it had traveled straight. To appreciate how that works, imagine the air divided into an infinite number of infinitesimally thin constant- n horizontal layers. A ray passing from layer to layer would bend (via Snell's Law) slightly upward at each interface (much as in Fig. 4.29 held upside down with the ray run backwards). Of course, if the ray comes down nearly vertically it makes a small angle-of-incidence at each interface between

layers, only bends slightly, and soon strikes the ground where no one will "see" it.

On the other hand, it is possible that a light ray that comes in at a shallow enough angle to begin with could ultimately approach an interface at glancing incidence (p. 116). It would then be completely reflected (p. 122), thereby starting its climb back up into the more dense air (much as in Fig. 4.29 held upside down with the ray run forwards).

Any viewer, off on the left in Fig. 4.32, who receives these bent rays, naturally projects them straight backward as if they were reflected from a mirrored surface. Depending on where you stand, you'll see a different mirage puddle, but it will always be far from you and so will always disappear as you approach it. The effect is particularly easy to view on long modern highways. The only requirement is that you look at the road at near glancing incidence, because the rays bend very gradually.*

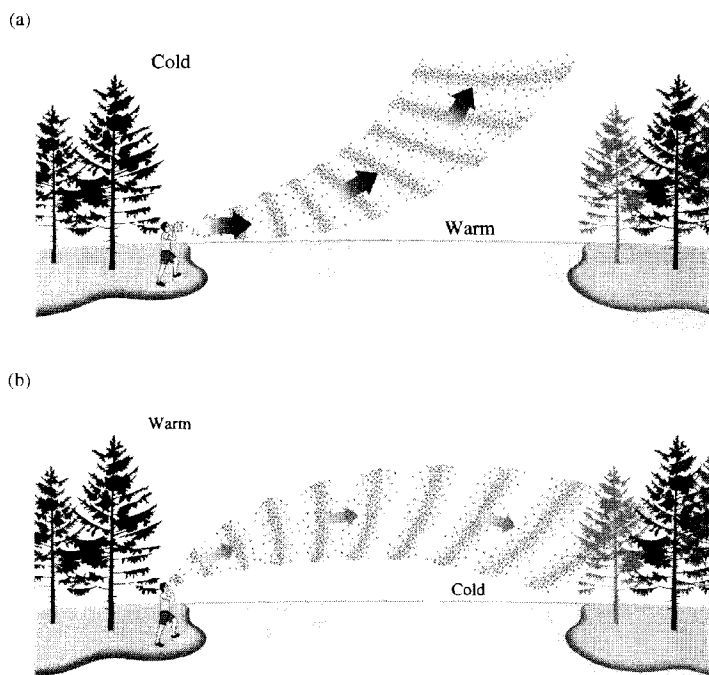


Figure 4.33 The puddle mirage can be understood via waves; the speed, and therefore the wavelength, increase in the less dense medium. That bends the wavefronts and the rays. The same effect is common with sound waves; (a) when the surface air is cold, sounds can be heard much farther than normal. (b) And when it's warm, sounds seem to vanish into the air.

* See for example T. Kosa and P. Palfy-Muhoray, "Mirage mirror on the wall," *Am. J. Phys.* **68** (12), 1120 (2000).

The same effect is well known as it applies to sound. Figure 4.33 depicts the alternative understanding in terms of waves. The wavefronts bend because of temperature-induced changes in speed and therefore in wavelength. (The speed of sound is proportional to the square root of the temperature.) The noises of people on a hot beach climb up and away, and the place can seem strangely quiet. The opposite occurs in the evening when the ground cools before the upper air and distant sounds can clearly be heard.

The Modern Formulation of Fermat's Principle

The original statement of Fermat's *Principle of Least Time* has some serious failings and is in need of alteration. To that end, recall that if we have a function, say $f(x)$, we can determine the specific value of the variable x that causes $f(x)$ to have a *stationary* value by setting $df/dx = 0$ and solving for x . By a stationary value, we mean one for which the slope of $f(x)$ versus x is zero or equivalently where the function has a maximum $\wedge$, minimum $\vee$, or a point of inflection with a horizontal tangent $\nearrow$.

Fermat's Principle in its modern form reads: **a light ray in going from point S to point P must traverse an optical path length that is stationary with respect to variations of that path.** In essence what that means is that the curve of the *OPL* versus x will have a somewhat flattened region in the vicinity of where the slope goes to zero. The zero-slope point corresponds to the actual path taken. In other words, the *OPL* for the true trajectory will equal, to a first approximation, the *OPL* of paths immediately adjacent to it.* For example, in a situation where the *OPL* is a minimum, as with the refraction illustrated in Fig. 4.29, the *OPL* curve will look something like

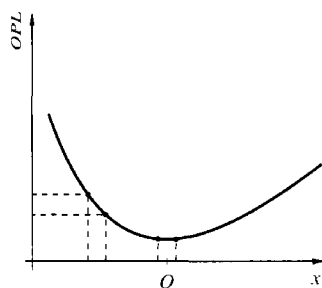


Figure 4.34 In the situation shown in Fig. 4.29 the actual location of point O corresponds to a path of minimum *OPL*.

*The first derivative of the *OPL* vanishes in its Taylor series expansion, since the path is stationary.

Fig. 4.34. A small change in x in the vicinity of O has little effect on the *OPL*, but a similar change in x anywhere well away from O results in a substantial change in *OPL*. Thus there will be many paths neighboring the actual one that would take nearly the same time for the light to traverse. This latter insight makes it possible to begin to understand how light manages to be so clever in its meanderings.

Suppose that a beam of light advances through a homogeneous isotropic medium (Fig. 4.35) so that a ray passes from points S to P . Atoms within the material are driven by the incident disturbance, and they reradiate in all directions. Wavelets progressing along paths in the immediate vicinity of a stationary straight-line path will reach P by routes that differ only slightly in *OPL* (as with group-I in Fig. 4.35b). They will therefore arrive nearly in-phase and reinforce each other. Think of each wavelet represented by a tiny phasor that rotates once around as the wave advances one wavelength (p. 23) along any ray path. Because the *OPL*s are all about the same, the phasors at P all point in more or less the same direction, and even though they're all small they combine to make the dominant contribution.

Wavelets taking other paths far from the stationary one (as with group-II in Fig. 4.35b) will arrive at P appreciably out-of-phase with each other and will therefore tend to cancel. In oth-

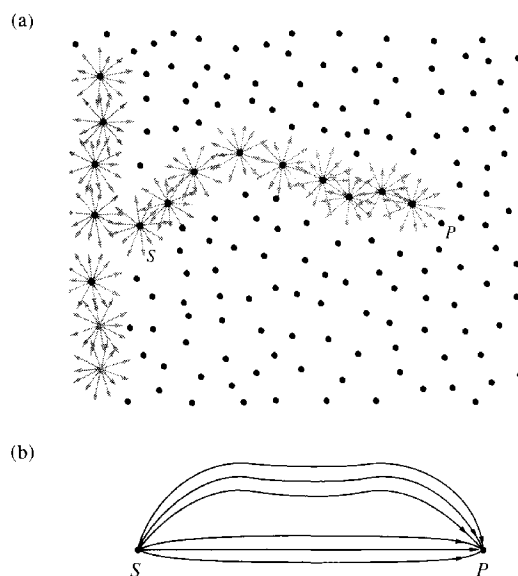


Figure 4.35 (a) Light can presumably take any number of paths from S to P , but it apparently takes only the one that corresponds to a stationary *OPL*. All other routes effectively cancel out. (b) For example, if some light takes each of the three upper paths in the diagram, it arrives at P with three very different phases and interferes more or less destructively.

er words, there will be large angles between the little phasors; placed tip-to-tail they'll spiral around producing only a tiny net contribution. Keep in mind that we've just drawn three ray paths—the argument would be better made with millions of them in each group.

We can conclude that energy will effectively propagate along the ray from S to P that satisfies Fermat's Principle. And this is true whether we're talking about interfering electromagnetic waves or photon probability amplitudes (p. 138).

We can expect that this same logic holds for all propagation processes,* such as, for example, reflection from a plane mirror (Fig. 4.28). There, spherical waves leaving S sweep across the entire mirror, and yet an observer at P sees a well defined point source and not a great blotch of light covering the whole surface. Only rays for which $\theta_i \approx \theta_r$ (as with group-I in Fig. 4.36) have a stationary OPL ; the associated wavelets

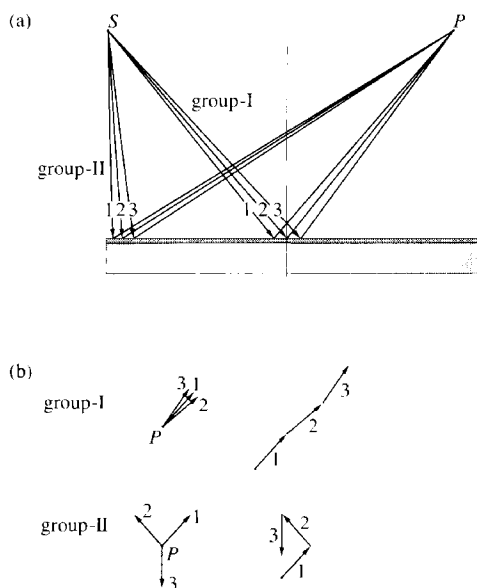


Figure 4.36 Rays reflecting off a plane mirror. Only those in group-I for which the OPL is stationary will correspond to waves that arrive at point- P more or less in-phase. There phasors will add along an almost straight line, producing a substantial resultant wave amplitude (going from the tail of 1 to the tip of 3). The phasors for group-II have large phase angle differences and so when added they essentially spiral around producing a very small resultant wave amplitude (going from the tail of 1 to the tip of 3). Of course, we should really be drawing millions of very tiny phasors in each group and not just three relatively large ones.

*We'll come back to these ideas when we consider QED in this chapter and the Fresnel zone plate in Chapter 10.

will arrive at P nearly in-phase and reinforce each other. All other rays (e.g., group-II in Fig. 4.36) will make negligible contributions to the energy reaching P .

Stationary Paths

To see that the OPL for a ray need not always be a minimum, examine Fig. 4.37, which depicts a segment of a hollow three-dimensional ellipsoidal mirror. If the source S and the observer P are at the foci of the ellipsoid, then by definition the length SQP will be constant, regardless of where on the perimeter Q happens to be. It is also a geometrical property of the ellipse that $\theta_i = \theta_r$ for any location of Q . All optical paths from S to P via a reflection are therefore precisely equal. None is a minimum, and the OPL is clearly stationary with respect to variations. Rays leaving S and striking the mirror will arrive at the focus P . From another viewpoint we can say that radiant energy emitted by S will be scattered by electrons in the mirrored surface such that the wavelets will substantially reinforce each other only at P , where they have traveled the same distance and have the same phase. In any case, if a plane mirror was tangent to the ellipse at Q , the exact same path SQP traversed by a ray would then be a relative minimum. At the other extreme, if the mirrored surface conformed to a curve lying within the ellipse, like the dashed one shown, that same ray along SQP would now negotiate a relative maximum OPL . This is true even though other unused paths (where $\theta_i \neq \theta_r$) would actually be shorter (i.e., apart from inadmissible curved paths). Thus in all cases the rays travel a stationary OPL in accord with the reformulated Fermat's Principle. Note that since the Principle speaks only about the path and not the direction along it, a ray going from P to S will trace the same route as one from S to P . This is the very useful *Principle of Reversibility*.

Fermat's achievement stimulated a great deal of effort to supersede Newton's laws of mechanics with a similar variational formulation. The work of many men, notably Pierre de Maupertuis (1698–1759) and Leonhard Euler, finally led to the mechanics of Joseph Louis Lagrange (1736–1813) and hence to the *Principle of Least Action*, formulated by William Rowan Hamilton (1805–1865). The striking similarity between the principles of Fermat and Hamilton played an important part in Schrödinger's development of Quantum Mechanics. In 1942 Richard Phillips Feynman (1918–1988) showed that Quantum Mechanics can be fashioned in an alternative way using a variational approach. The continuing evo-

lution of variational principles brings us back to Optics via the modern formalism of Quantum Optics.

Fermat's Principle is not so much a computational device as it is a concise way of thinking about the propagation of light. It is a statement about the grand scheme of things without any concern for the contributing mechanisms, and as such it will yield insights under a myriad of circumstances.

4.6 The Electromagnetic Approach

Thus far, we have studied reflection and refraction from the perspectives of Scattering Theory, the Theorem of Malus and Dupin, and Fermat's Principle. Yet another and even more powerful approach is provided by Electromagnetic Theory. Unlike the previous techniques, which say nothing about the incident, reflected, and transmitted radiant flux densities (i.e., I_i , I_r , I_t , respectively), Electromagnetic Theory treats these within the framework of a far more complete description.

4.6.1 Waves at an Interface

Suppose that the incident monochromatic lightwave is planar, so that it has the form

$$\vec{E}_i = \vec{E}_{0i} \exp [i(\vec{k}_i \cdot \vec{r} - \omega_i t)] \quad (4.11)$$

or, more simply,

$$\vec{E}_i = \vec{E}_{0i} \cos (\vec{k}_i \cdot \vec{r} - \omega_i t) \quad (4.12)$$

Assume that $\vec{E}_{0i}$ is constant in time; that is, the wave is linearly or plane polarized. We'll find in Chapter 8 that any form of light can be represented by two orthogonal linearly polarized waves, so that this doesn't actually represent a restriction. Note that just as the origin in time, $t = 0$, is arbitrary, so too is the origin O in space, where $\vec{r} = 0$. Thus, making no assumptions about their directions, frequencies, wavelengths, phases, or amplitudes, we can write the reflected and transmitted waves as

$$\vec{E}_r = \vec{E}_{0r} \cos (\vec{k}_r \cdot \vec{r} - \omega_r t + \varepsilon_r) \quad (4.13)$$

$$\text{and} \quad \vec{E}_t = \vec{E}_{0t} \cos (\vec{k}_t \cdot \vec{r} - \omega_t t + \varepsilon_t) \quad (4.14)$$

Here ε_r and ε_t are *phase constants* relative to $\vec{E}_i$, and are introduced because the position of the origin is not unique. Figure 4.38 depicts the waves in the vicinity of the planar interface between two homogeneous lossless dielectric media of indices n_i and n_r .

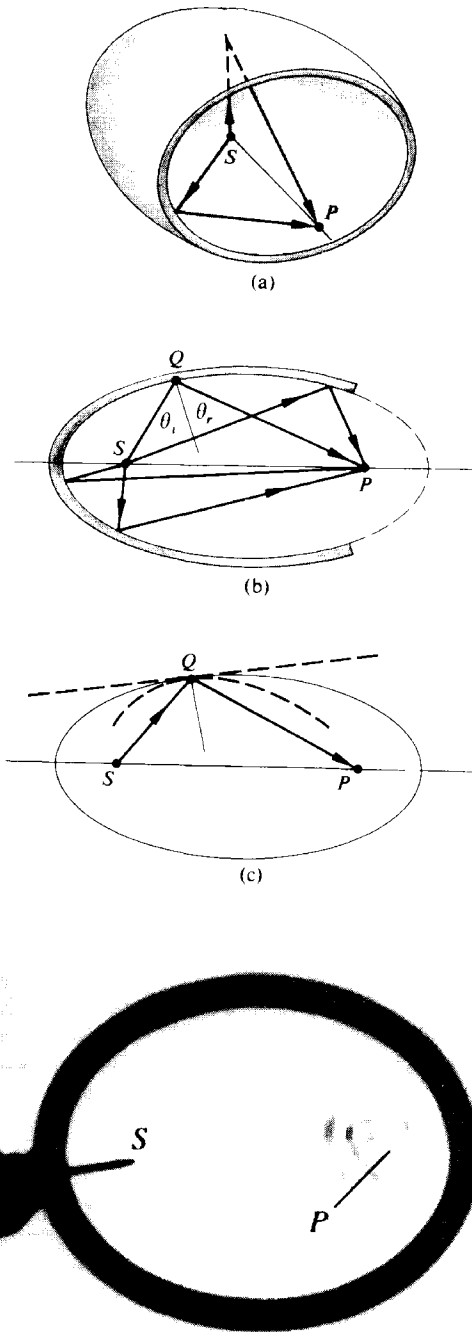


Figure 4.37 Reflection off an ellipsoidal surface. Observe the reflection of waves using a frying pan filled with water. Even though these are usually circular, it is well worth playing with. (Photo courtesy PSSC College Physics, D. C. Heath & Co., 1968.)

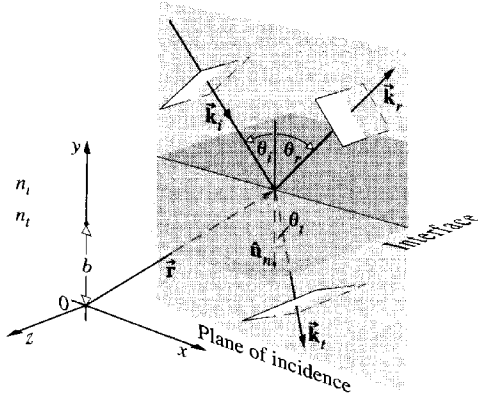


Figure 4.38 Plane waves incident on the boundary between two homogeneous, isotropic, lossless dielectric media.

The laws of Electromagnetic Theory (Section 3.1) lead to certain requirements that must be met by the fields, and they are referred to as the *boundary conditions*. Specifically, one of these is that the component of the electric field $\vec{E}$ that is tangent to the interface must be continuous across it (the same is true for $\vec{H}$). In other words, the total tangential component of $\vec{E}$ on one side of the surface must equal that on the other (Problem 4.37). Thus, since $\hat{u}_n$ is the unit vector normal to the interface, regardless of the direction of the electric field within the wavefront, the cross-product of it with $\hat{u}_n$ will be perpendicular to $\hat{u}_n$ and therefore tangent to the interface. Hence

$$\hat{u}_n \times \vec{E}_i + \hat{u}_n \times \vec{E}_r = \hat{u}_n \times \vec{E}_t \quad (4.15)$$

or

$$\begin{aligned} & \hat{u}_n \times \vec{E}_{0i} \cos(\vec{k}_i \cdot \vec{r} - \omega_i t) \\ & + \hat{u}_n \times \vec{E}_{0r} \cos(\vec{k}_r \cdot \vec{r} - \omega_r t + \varepsilon_r) \\ & = \hat{u}_n \times \vec{E}_{0t} \cos(\vec{k}_t \cdot \vec{r} - \omega_t t + \varepsilon_t) \end{aligned} \quad (4.16)$$

This relationship must obtain at any instant in time and at any point on the interface ($y = b$). Consequently, $\vec{E}_i$, $\vec{E}_r$, and $\vec{E}_t$ must have precisely the same functional dependence on the variables t and r , which means that

$$\begin{aligned} (\vec{k}_i \cdot \vec{r} - \omega_i t)|_{y=b} &= (\vec{k}_r \cdot \vec{r} - \omega_r t + \varepsilon_r)|_{y=b} \\ &= (\vec{k}_t \cdot \vec{r} - \omega_t t + \varepsilon_t)|_{y=b} \end{aligned} \quad (4.17)$$

With this as the case, the cosines in Eq. (4.16) cancel, leaving an expression independent of t and r , as indeed it must be.

Inasmuch as this has to be true for all values of time, the coefficients of t must be equal, to wit

$$\omega_i = \omega_r = \omega_t \quad (4.18)$$

Recall that the electrons within the media are undergoing (linear) forced vibrations at the frequency of the incident wave. Whatever light is scattered has that same frequency. Furthermore,

$$(\vec{k}_i \cdot \vec{r})|_{y=b} = (\vec{k}_r \cdot \vec{r} + \varepsilon_r)|_{y=b} = (\vec{k}_t \cdot \vec{r} + \varepsilon_t)|_{y=b} \quad (4.19)$$

wherein $\vec{r}$ terminates on the interface. The values of ε_r and ε_t correspond to a given position of O , and thus they allow the relation to be valid regardless of that location. (For example, the origin might be chosen such that $\vec{r}$ was perpendicular to $\vec{k}_i$ but not to $\vec{k}_r$ or $\vec{k}_t$.) From the first two terms we obtain

$$[(\vec{k}_i - \vec{k}_r) \cdot \vec{r}]_{y=b} = \varepsilon_r \quad (4.20)$$

Recalling Eq. (2.43), this expression simply says that the endpoint of $\vec{r}$ sweeps out a plane (which is of course the interface) perpendicular to the vector $(\vec{k}_i - \vec{k}_r)$. To phrase it slightly differently, $(\vec{k}_i - \vec{k}_r)$ is parallel to $\hat{u}_n$. Notice, however, that since the incident and reflected waves are in the same medium, $k_i = k_r$. From the fact that $(\vec{k}_i - \vec{k}_r)$ has no component in the plane of the interface, that is, $\hat{u}_n \times (\vec{k}_i - \vec{k}_r) = 0$, we conclude that

$$k_i \sin \theta_i = k_r \sin \theta_r$$

Hence we have the Law of Reflection; that is,

$$\theta_i = \theta_r$$

Furthermore, since $(\vec{k}_i - \vec{k}_r)$ is parallel to $\hat{u}_n$ all three vectors, $\vec{k}_i$, $\vec{k}_r$, and $\hat{u}_n$, are in the same plane, the plane-of-incidence. Again, from Eq. (4.19)

$$[(\vec{k}_i - \vec{k}_r) \cdot \vec{r}]_{y=b} = \varepsilon_t \quad (4.21)$$

and therefore $(\vec{k}_i - \vec{k}_r)$ is also normal to the interface. Thus $\vec{k}_i$, $\vec{k}_r$, $\vec{k}_t$, and $\hat{u}_n$ are all coplanar. As before, the tangential components of $\vec{k}_i$ and $\vec{k}_t$ must be equal, and consequently

$$k_i \sin \theta_i = k_t \sin \theta_t \quad (4.22)$$

But because $\omega_i = \omega_t$, we can multiply both sides by c/ω_i to get

$$n_i \sin \theta_i = n_t \sin \theta_t$$

which is Snell's Law. Finally, if we had chosen the origin O to

be in the interface, it is evident from Eqs. (4.20) and (4.21) that ϵ_r and ϵ_t would both have been zero. That arrangement, though not as instructive, is certainly simpler, and we'll use it from here on.

4.6.2 The Fresnel Equations

We have just found the relationship that exists among the phases of $\vec{E}_i(\vec{r}, t)$, $\vec{E}_r(\vec{r}, t)$, and $\vec{E}_t(\vec{r}, t)$ at the boundary. There is still an interdependence shared by the amplitudes $\vec{E}_{0i}$, $\vec{E}_{0r}$, and $\vec{E}_{0t}$, which can now be evaluated. To that end, suppose that a plane monochromatic wave is incident on the planar surface separating two isotropic media. Whatever the polarization of the wave, we shall resolve its $\vec{E}$ - and $\vec{B}$ -fields into components parallel and perpendicular to the plane-of-incidence and treat these constituents separately.

Case 1: $\vec{E}$ perpendicular to the plane-of-incidence. Assume that $\vec{E}$ is perpendicular to the plane-of-incidence and that $\vec{B}$ is parallel to it (Fig. 4.39). Recall that $E = vB$, so that

$$\hat{k} \times \vec{E} = v\vec{B} \quad (4.23)$$

and
$$\hat{k} \cdot \vec{E} = 0 \quad (4.24)$$

(i.e., $\vec{E}$, $\vec{B}$, and the unit propagation vector $\hat{k}$ form a right-handed system). Again, making use of the continuity of the tangential components of the $\vec{E}$ -field, we have at the boundary at any time and any point

$$\vec{E}_{0i} + \vec{E}_{0r} = \vec{E}_{0t} \quad (4.25)$$

where the cosines cancel. Realize that the field vectors as shown really ought to be envisioned at $y = 0$ (i.e., at the surface), from which they have been displaced for the sake of clarity. Note too that although $\vec{E}_r$ and $\vec{E}_t$ must be normal to the plane-of-incidence by symmetry, *we are guessing that they point outward* at the interface when $\vec{E}_i$ does. The directions of the $\vec{B}$ -fields then follow from Eq. (4.23).

We will need to invoke another of the boundary conditions in order to get one more equation. The presence of material substances that become electrically polarized by the wave has a definite effect on the field configuration. Thus, although the tangential component of $\vec{E}$ is continuous across the boundary, its normal component is not. Instead, the normal component of the product $\epsilon\vec{E}$ is the same on either side of the interface. Sim-

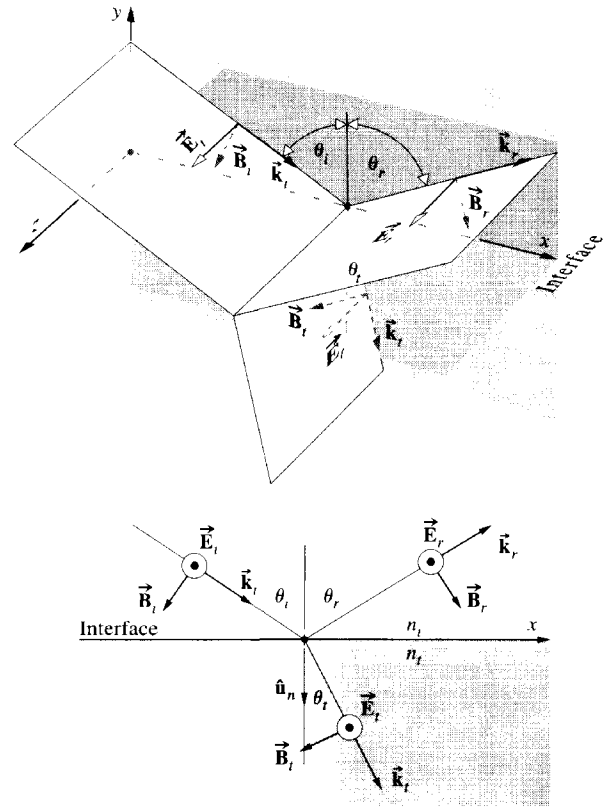


Figure 4.39 An incoming wave whose $\vec{E}$ -field is normal to the plane-of-incidence.

ilarly, the normal component of $\vec{B}$ is continuous, as is the tangential component of $\mu^{-1}\vec{B}$. Here the magnetic effect of the two media appears via their permeabilities μ_i and μ_t . This boundary condition will be the simplest to use, particularly as applied to reflection from the surface of a conductor.* Thus the continuity of the tangential component of $\vec{B}/\mu$ requires that

$$-\frac{B_i}{\mu_i} \cos \theta_i + \frac{B_r}{\mu_i} \cos \theta_r = -\frac{B_t}{\mu_t} \cos \theta_t \quad (4.26)$$

where the left and right sides are the total magnitudes of $\vec{B}/\mu$ parallel to the interface in the incident and transmitting media, respectively. The positive direction is that of increasing x , so

*In keeping with our intent to use only the $\vec{E}$ - and $\vec{B}$ -fields, at least in the early part of this exposition, we have avoided the usual statements in terms of $\vec{H}$, where

$$\vec{H} = \mu^{-1}\vec{B} \quad [\text{A1.14}]$$

that the scalar components of $\vec{B}_i$ and $\vec{B}_r$ appear with minus signs. From Eq. (4.23) we have

$$B_i = E_i/v_i \quad (4.27)$$

$$B_r = E_r/v_r \quad (4.28)$$

and
$$B_t = E_t/v_t \quad (4.29)$$

Since $v_i = v_r$ and $\theta_i = \theta_r$, Eq. (4.26) can be written as

$$\frac{1}{\mu_i v_i} (E_i - E_r) \cos \theta_i = \frac{1}{\mu_t v_t} E_t \cos \theta_t \quad (4.30)$$

Making use of Eqs. (4.12), (4.13), and (4.14) and remembering that the cosines therein equal one another at $y = 0$, we obtain

$$\frac{n_i}{\mu_i} (E_{0i} - E_{0r}) \cos \theta_i = \frac{n_t}{\mu_t} E_{0t} \cos \theta_t \quad (4.31)$$

Combined with Eq. (4.25), this yields

$$\left(\frac{E_{0r}}{E_{0i}} \right)_\perp = \frac{\frac{n_i}{\mu_i} \cos \theta_i - \frac{n_t}{\mu_t} \cos \theta_t}{\frac{n_i}{\mu_i} \cos \theta_i + \frac{n_t}{\mu_t} \cos \theta_t} \quad (4.32)$$

and
$$\left(\frac{E_{0t}}{E_{0i}} \right)_\perp = \frac{2 \frac{n_i}{\mu_i} \cos \theta_i}{\frac{n_i}{\mu_i} \cos \theta_i + \frac{n_t}{\mu_t} \cos \theta_t} \quad (4.33)$$

The $\perp$ subscript serves as a reminder that we are dealing with the case in which $\vec{E}$ is perpendicular to the plane-of-incidence. These two expressions, which are *completely general statements applying to any linear, isotropic, homogeneous media*, are two of the **Fresnel Equations**. Most often one deals with dielectrics for which $\mu_i \approx \mu_t \approx \mu_0$; consequently, the common form of these equations is simply

$$r_\perp \equiv \left(\frac{E_{0r}}{E_{0i}} \right)_\perp = \frac{n_i \cos \theta_i - n_t \cos \theta_t}{n_i \cos \theta_i + n_t \cos \theta_t} \quad (4.34)$$

and

$$t_\perp \equiv \left(\frac{E_{0t}}{E_{0i}} \right)_\perp = \frac{2n_i \cos \theta_i}{n_i \cos \theta_i + n_t \cos \theta_t} \quad (4.35)$$

Here $r_\perp$ denotes the **amplitude reflection coefficient**, and $t_\perp$ is the **amplitude transmission coefficient**.

Case 2: $\vec{E}$ parallel to the plane-of-incidence. A similar pair of equations can be derived when the incoming $\vec{E}$ -field lies in

the plane-of-incidence, as shown in Fig. 4.40. Continuity of the tangential components of $\vec{E}$ on either side of the boundary leads to

$$E_{0i} \cos \theta_i - E_{0r} \cos \theta_r = E_{0t} \cos \theta_t \quad (4.36)$$

In much the same way as before, continuity of the tangential components of $\vec{B}/\mu$ yields

$$\frac{1}{\mu_i v_i} E_{0i} + \frac{1}{\mu_r v_r} E_{0r} = \frac{1}{\mu_t v_t} E_{0t} \quad (4.37)$$

Using the fact that $\mu_i = \mu_r$ and $\theta_i = \theta_r$, we can combine these formulas to obtain two more of the *Fresnel Equations*:

$$r_\parallel \equiv \left(\frac{E_{0r}}{E_{0i}} \right)_\parallel = \frac{\frac{n_t}{\mu_t} \cos \theta_i - \frac{n_i}{\mu_i} \cos \theta_t}{\frac{n_i}{\mu_i} \cos \theta_i + \frac{n_t}{\mu_t} \cos \theta_t} \quad (4.38)$$

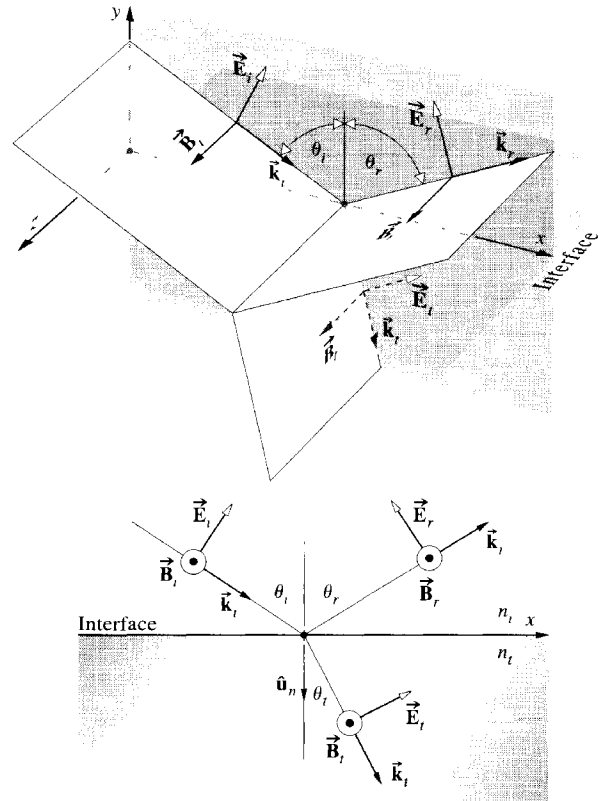


Figure 4.40 An incoming wave whose $\vec{E}$ -field is in the plane-of-incidence.

and

$$t_{\parallel} \equiv \left(\frac{E_{0t}}{E_{0i}} \right)_{\parallel} = \frac{2 \frac{n_i}{\mu_i} \cos \theta_i}{\frac{n_i}{\mu_i} \cos \theta_i + \frac{n_t}{\mu_t} \cos \theta_t} \quad (4.39)$$

When both media forming the interface are dielectrics that are essentially “nonmagnetic” (p. 66), the amplitude coefficients become

$$r_{\parallel} = \frac{n_t \cos \theta_i - n_i \cos \theta_t}{n_i \cos \theta_t + n_t \cos \theta_i} \quad (4.40)$$

and

$$t_{\parallel} = \frac{2n_i \cos \theta_i}{n_i \cos \theta_t + n_t \cos \theta_i} \quad (4.41)$$

One further notational simplification can be made using Snell’s Law, whereupon the Fresnel Equations for dielectric media become (Problem 4.39)

$$r_{\perp} = - \frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \quad (4.42)$$

$$r_{\parallel} = + \frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \quad (4.43)$$

$$t_{\perp} = + \frac{2 \sin \theta_i \cos \theta_t}{\sin(\theta_i + \theta_t)} \quad (4.44)$$

$$t_{\parallel} = + \frac{2 \sin \theta_t \cos \theta_i}{\sin(\theta_i + \theta_t) \cos(\theta_i - \theta_t)} \quad (4.45)$$

A note of caution must be introduced here. Bear in mind that the directions (or more precisely, the phases) of the fields in Figs. 4.39 and 4.40 were selected rather arbitrarily. For example, in Fig. 4.39 we could have assumed that $\vec{E}_r$ pointed inward, whereupon $\vec{B}_r$ would have had to be reversed as well. Had we done that, the sign of $r_{\perp}$ would have turned out to be positive, leaving the other amplitude coefficients unchanged. The signs appearing in Eqs. (4.42) through (4.45), which are positive except for the first, correspond to the particular set of field directions selected. The minus sign in Eq. (4.42), as we will see, just means that we didn’t guess correctly concerning $\vec{E}_r$ in Fig. 4.39. Nonetheless, be aware that the literature is not standardized, and all possible sign variations have been labeled the *Fresnel Equations*. To avoid confusion *they must be related to the specific field directions from which they were derived*.

4.6.3 Interpretation of the Fresnel Equations

This section examines the physical implications of the Fresnel Equations. In particular, we are interested in determining the fractional amplitudes and flux densities that are reflected and refracted. In addition we shall be concerned with any possible phase shifts that might be incurred in the process.

Amplitude Coefficients

Let’s briefly examine the form of the amplitude coefficients over the entire range of θ_i values. At nearly normal incidence ($\theta_i \approx 0$) the tangents in Eq. (4.43) are essentially equal to sines, in which case

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \left[\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \right]_{\theta_i=0}$$

We will come back to the physical significance of the minus sign presently. After expanding the sines and using Snell’s Law, this expression becomes

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \left[\frac{n_t \cos \theta_i - n_i \cos \theta_t}{n_t \cos \theta_i + n_i \cos \theta_t} \right]_{\theta_i=0} \quad (4.46)$$

which follows as well from Eqs. (4.34) and (4.40). In the limit, as θ_i goes to 0, $\cos \theta_i$ and $\cos \theta_t$ both approach one, and consequently

$$[r_{\parallel}]_{\theta_i=0} = [-r_{\perp}]_{\theta_i=0} = \frac{n_t - n_i}{n_t + n_i} \quad (4.47)$$

This equality of the reflection coefficients arises because the plane-of-incidence is no longer specified when $\theta_i = 0$. Thus, for example, at an air ($n_i = 1$) glass ($n_t = 1.5$) interface at nearly normal incidence, the amplitude reflection coefficients equal ± 0.2 . (See Problem 4.45.)

When $n_t > n_i$ it follows from Snell’s Law that $\theta_i > \theta_t$, and $r_{\perp}$ is negative for all values of θ_i (Fig. 4.41). In contrast, Eq. (4.43) tells us that $r_{\parallel}$ starts out positive at $\theta_i = 0$ and decreases gradually until it equals zero when $(\theta_i + \theta_t) = 90^\circ$, since there $\tan \pi/2$ is infinite. The particular value of the incident angle for which this occurs is denoted by θ_p and referred to as the **polarization angle** (see Section 8.6.1). Notice that $r_{\parallel} \rightarrow 0$ at θ_p , just when the phase shifts 180° . That means we won’t see the $\vec{E}$ -field do any flipping when θ_i approaches θ_p from either side. As θ_i increases beyond θ_p , $r_{\parallel}$ becomes progressively more negative, reaching -1.0 at 90° .

If you place a single sheet of glass, a microscope slide, on this page and look straight down into it ($\theta_i = 0$), the region

beneath the glass will seem decidedly grayer than the rest of the paper, because the slide will reflect at both its interfaces, and the light reaching and returning from the paper will be diminished appreciably. Now hold the slide near your eye and again view the page through it as you tilt it, increasing θ_i . The amount of light reflected will increase, and it will become more difficult to see the page through the glass. When $\theta_i \approx 90^\circ$ the slide will look like a perfect mirror as the reflection coefficients (Fig. 4.41) go to -1.0 . Even a rather poor surface (see photo), such as the cover of this book, will be mirrorlike at glancing incidence. Hold the book horizontally at the level of the middle of your eye and face a bright light; you will see the source reflected rather nicely in the cover. This suggests that even X-rays could be mirror-reflected at glancing incidence (p. 242), and modern X-ray telescopes are based on that very fact.

At normal incidence Eqs. (4.35) and (4.41) lead rather straightforwardly to

$$[t_{\parallel}]_{\theta_i=0} = [t_{\perp}]_{\theta_i=0} = \frac{2n_i}{n_i + n_t} \quad (4.48)$$

It will be shown in Problem 4.50 that the expression

$$t_{\perp} + (-r_{\perp}) = 1 \quad (4.49)$$

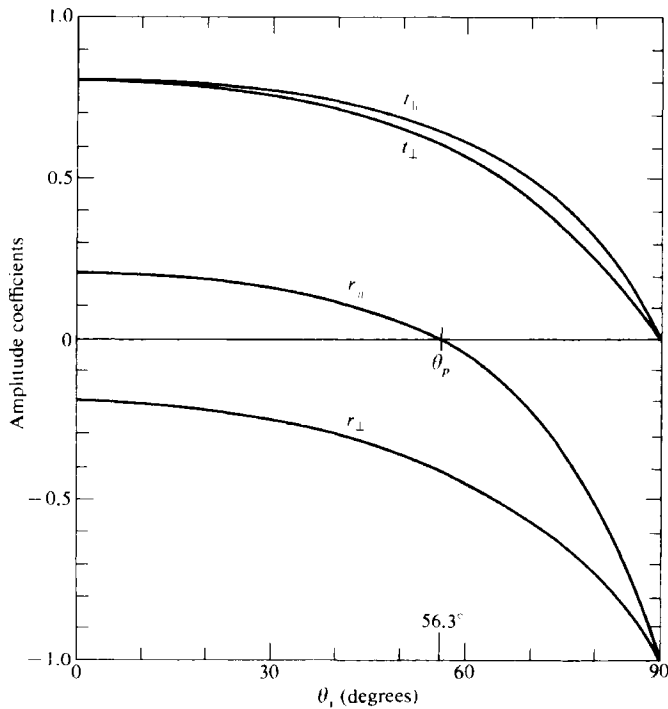


Figure 4.41 The amplitude coefficients of reflection and transmission as a function of incident angle. These correspond to external reflection $n_t > n_i$ at an air-glass interface ($n_t = 1.5$).



At near-glancing incidence the walls and floor are mirrorlike—this despite the fact that the surfaces are rather poor reflectors at $\theta_i = 0^\circ$. (Photo by E.H.)

holds for all θ_i , whereas

$$t_{\parallel} + r_{\parallel} = 1 \quad (4.50)$$

is true only at normal incidence.

The foregoing discussion, for the most part, was restricted to the case of **external reflection** (i.e., $n_t > n_i$). The opposite situation of **internal reflection**, in which the incident medium is the more dense ($n_i > n_t$), is of interest as well. In that instance $\theta_t > \theta_i$, and $r_{\perp}$, as described by Eq. (4.42), will always be positive. Figure 4.42 shows that $r_{\perp}$ increases from its initial value [Eq. (4.47)] at $\theta_i = 0$, reaching $+1$ at what is called the **critical angle**, θ_c . Specifically, θ_c is the special value of the incident angle (p. 122) for which $\theta_t = \pi/2$. Likewise, $r_{\parallel}$ starts off negatively [Eq. (4.47)] at $\theta_i = 0$ and thereafter increases, reaching $+1$ at $\theta_i = \theta_c$, as is evident from the Fresnel Equation (4.40). Again, $r_{\parallel}$ passes through zero at the **polarization angle** θ_p . It is left for Problem 4.66 to show that the polarization angles θ_p and θ_p' for internal and external reflection at the interface between the same media are simply the complements of each other. We will return to internal reflection in Section 4.7, where it will be shown that $r_{\perp}$ and $r_{\parallel}$ are complex quantities for $\theta_i > \theta_c$.

Phase Shifts

It should be evident from Eq. (4.42) that $r_{\perp}$ is negative regardless of θ_i when $n_t > n_i$. Yet we saw earlier that had we chosen $[\vec{E}_r]_{\perp}$ in Fig. 4.37 to be in the opposite direction, the first Fres-

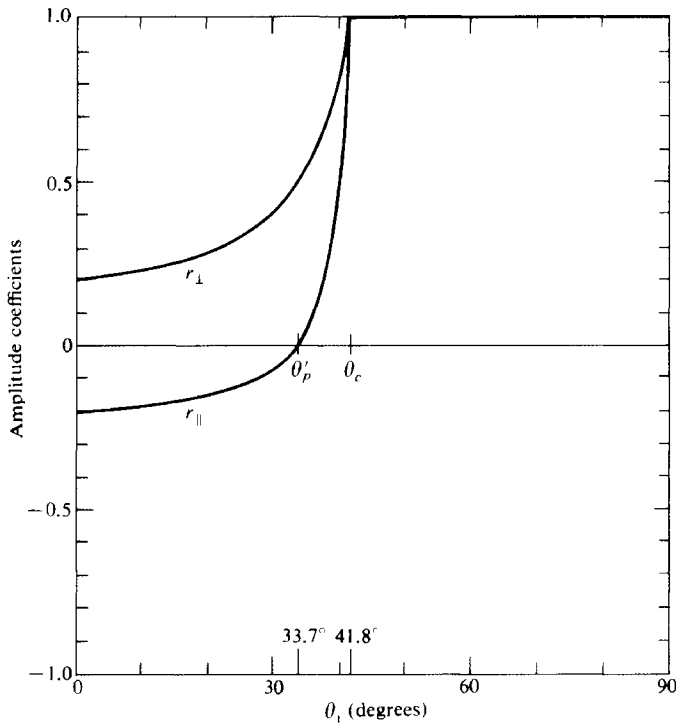


Figure 4.42 The amplitude coefficients of reflection as a function of incident angle. These correspond to internal reflection $n_t < n_i$ at an air-glass interface ($n_t = 1/1.5$).

nel Equation (4.42) would have changed signs, causing $r_{\perp}$ to become a positive quantity. The sign of $r_{\perp}$ is associated with the relative directions of $[\vec{E}_{0i}]_{\perp}$ and $[\vec{E}_{0r}]_{\perp}$. Bear in mind that a reversal of $[\vec{E}_{0r}]_{\perp}$ is tantamount to introducing a phase shift, $\Delta\varphi_{\perp}$, of π radians into $[\vec{E}_r]_{\perp}$. Hence at the boundary $[\vec{E}_i]_{\perp}$ and $[\vec{E}_r]_{\perp}$ will be antiparallel and therefore π out-of-phase with each other, as indicated by the negative value of $r_{\perp}$. When we consider components normal to the plane-of-incidence, there is no confusion as to whether two fields are in-phase or π radians out-of-phase: if parallel, they're in-phase; if antiparallel, they're π out-of-phase. In summary, then, **the component of the electric field normal to the plane-of-incidence undergoes a phase shift of π radians upon reflection when the incident medium has a lower index than the transmitting medium.** Similarly, $t_{\perp}$ and $t_{\parallel}$ are always positive and $\Delta\varphi = 0$. Furthermore, when $n_i > n_t$ no phase shift in the normal component results on reflection, that is, $\Delta\varphi_{\perp} = 0$ so long as $\theta_i < \theta_c$.

Things are a bit less obvious when we deal with $[\vec{E}_i]_{\parallel}$, $[\vec{E}_r]_{\parallel}$, and $[\vec{E}_t]_{\parallel}$. It now becomes necessary to define more explicitly what is meant by *in-phase*, since the field vectors are coplanar

but generally not colinear. The field directions were chosen in Figs. 4.39 and 4.40 such that if you looked down any one of the propagation vectors toward the direction from which the light was coming, $\vec{E}$, $\vec{B}$, and $\vec{k}$ would appear to have the same relative orientation whether the ray was incident, reflected, or transmitted. We can use this as the required condition for two $\vec{E}$ -fields to be in-phase. Equivalently, but more simply, **two fields in the incident plane are in-phase if their y-components are parallel and are out-of-phase if the components are antiparallel.** Notice that when two $\vec{E}$ -fields are out-of-phase so too are their associated $\vec{B}$ -fields and vice versa. With this definition we need only look at the vectors normal to the plane-of-incidence, whether they be $\vec{E}$ or $\vec{B}$, to determine the relative phase of the accompanying fields in the incident plane. Thus in Fig. 4.43a $\vec{E}_i$ and $\vec{E}_r$ are in-phase, as are $\vec{B}_i$ and $\vec{B}_r$, whereas $\vec{E}_i$ and $\vec{E}_t$ are out-of-phase, along with $\vec{B}_i$ and $\vec{B}_t$. Similarly, in Fig. 4.43b $\vec{E}_i$, $\vec{E}_r$, and $\vec{E}_t$ are in-phase, as are $\vec{B}_i$, $\vec{B}_r$, and $\vec{B}_t$.

Now, the amplitude reflection coefficient for the parallel component is given by

$$r_{\parallel} = \frac{n_t \cos \theta_i - n_i \cos \theta_r}{n_i \cos \theta_i + n_t \cos \theta_r}$$

which is positive ($\Delta\varphi_{\parallel} = 0$) as long as

$$n_t \cos \theta_i - n_i \cos \theta_r > 0$$

that is, if

$$\sin \theta_i \cos \theta_i - \cos \theta_i \sin \theta_i > 0$$

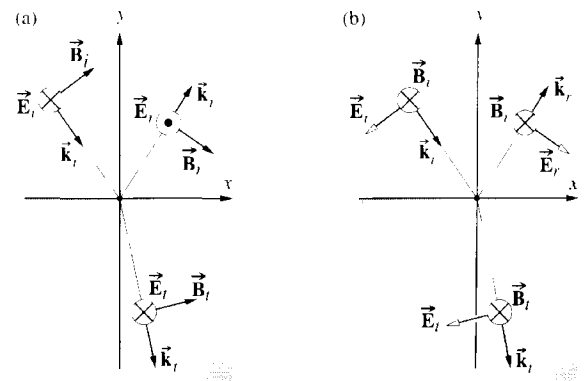


Figure 4.43 Field orientations and phase shifts.

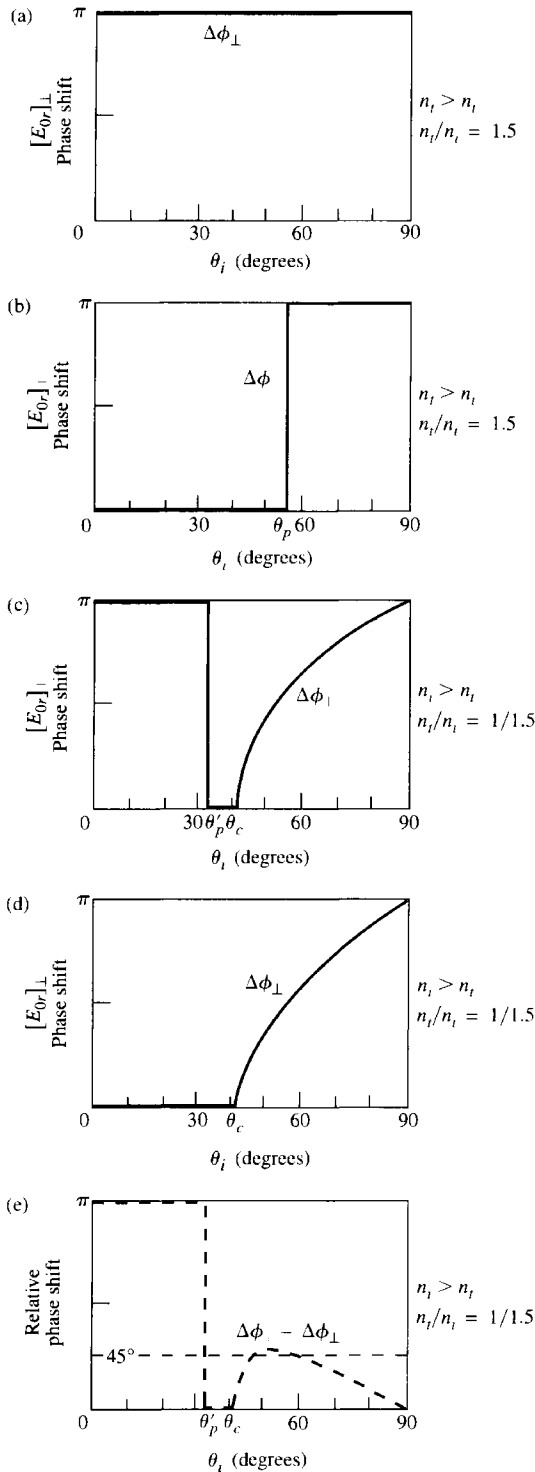


Figure 4.44 Phase shifts for the parallel and perpendicular components of the $\vec{E}$ -field corresponding to internal and external reflection.

or equivalently

$$\sin(\theta_i - \theta_t) \cos(\theta_i + \theta_t) > 0 \quad (4.51)$$

This will be the case for $n_i < n_t$ if

$$(\theta_i + \theta_t) < \pi/2 \quad (4.52)$$

and for $n_i > n_t$ when

$$(\theta_i + \theta_t) > \pi/2 \quad (4.53)$$

Thus when $n_i < n_t$, $[\vec{E}_{0r}]_{\parallel}$ and $[\vec{E}_{0i}]_{\parallel}$ will be in-phase ($\Delta\phi_{\parallel} = 0$) until $\theta_i = \theta_p$ and out-of-phase by π radians thereafter. The transition is not actually discontinuous, since $[\vec{E}_{0r}]_{\parallel}$ goes to zero at θ_p . In contrast, for internal reflection $r_{\parallel}$ is negative until θ'_p , which means that $\Delta\phi_{\parallel} = \pi$. From θ'_p to θ_c , $r_{\parallel}$ is positive and $\Delta\phi_{\parallel} = 0$. Beyond θ_c , $r_{\parallel}$ becomes complex, and $\Delta\phi_{\parallel}$ gradually increases to π at $\theta_i = 90^\circ$.

Figure 4.44, which summarizes these conclusions, will be of continued use to us. The actual functional form of $\Delta\phi_{\parallel}$ and $\Delta\phi_{\perp}$ for internal reflection in the region where $\theta_i > \theta_c$ can be found in the literature,* but the curves depicted here will suffice for our purposes. Figure 4.44e is a plot of the relative phase shift between the parallel and perpendicular components, that is, $\Delta\phi_{\parallel} - \Delta\phi_{\perp}$. It is included here because it will be useful later on (e.g., when we consider polarization effects). Finally, many of the essential features of this discussion are illustrated in Figs. 4.45 and 4.46. The amplitudes of the reflected vectors are in accord with those of Figs. 4.41 and 4.42 (for an air-glass interface), and the phase shifts agree with those of Fig. 4.44.

Many of these conclusions can be verified with the simplest experimental equipment, namely, two linear polarizers, a piece of glass, and a small source, such as a flashlight or high-intensity lamp. By placing one polarizer in front of the source (at 45° to the plane-of-incidence), you can easily duplicate the conditions of Fig. 4.45. For example, when $\theta_i = \theta_p$ (Fig. 4.45b) no light will pass through the second polarizer if its transmission axis is parallel to the plane-of-incidence. In comparison, at near-glancing incidence the reflected beam will vanish when the axes of the two polarizers are almost normal to each other.

*Born and Wolf, *Principles of Optics*, p. 49.

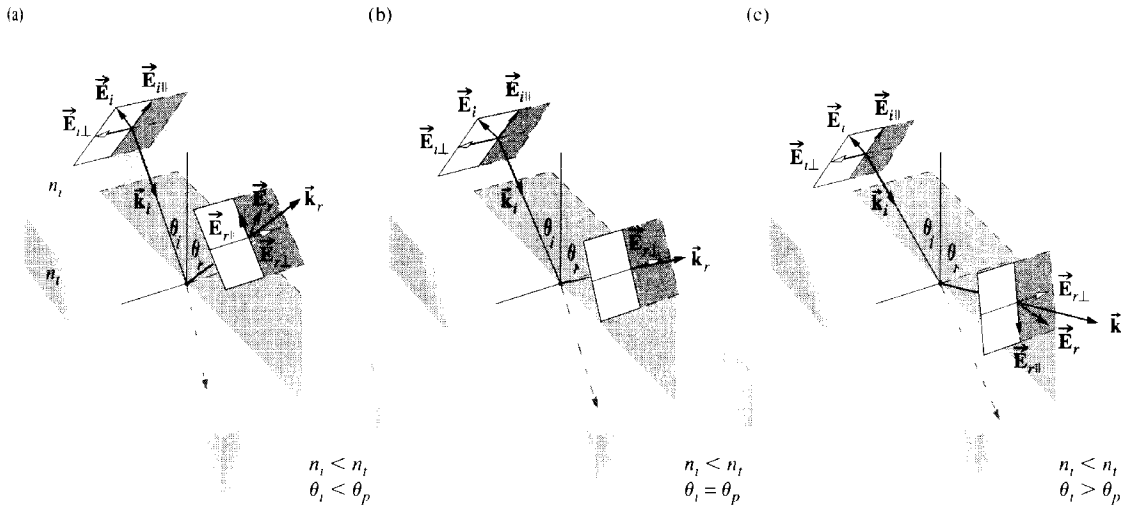


Figure 4.45 The reflected $\vec{E}$ -field at various angles concomitant with external reflection.

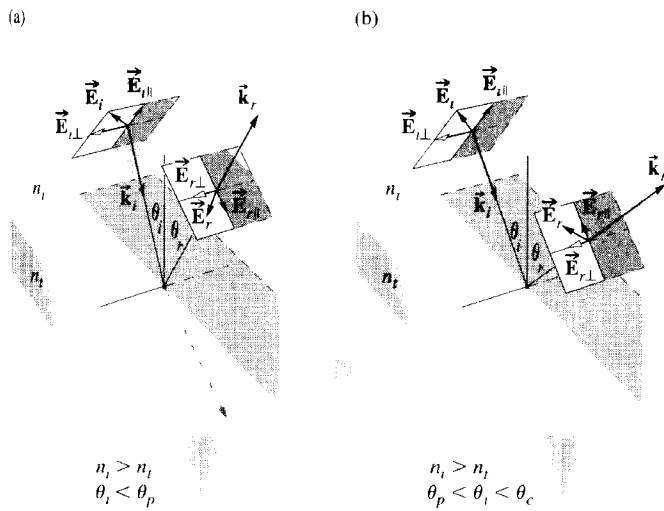


Figure 4.46 The reflected $\vec{E}$ -field at various angles concomitant with internal reflection.

Reflectance and Transmittance

Consider a circular beam of light incident on a surface, as shown in Fig. 4.47, such that there is an illuminated spot of area A . Recall that the power per unit area crossing a surface in vacuum whose normal is parallel to $\vec{S}$, the Poynting vector, is given by

$$\vec{S} = c^2 \epsilon_0 \vec{E} \times \vec{B} \quad [3.40]$$

Furthermore, the radiant flux density (W/m^2) or irradiance is

$$I = \langle S \rangle_T = \frac{c \epsilon_0}{2} E_0^2 \quad [3.44]$$

This is the average energy per unit time crossing a unit area normal to $\vec{S}$ (in isotropic media $\vec{S}$ is parallel to $\vec{k}$). In the case at hand (Fig. 4.47), let I_i , I_r , and I_t be the incident, reflected, and transmitted flux densities, respectively. The cross-sectional areas of the incident, reflected, and transmitted beams are, respectively, $A \cos \theta_i$, $A \cos \theta_r$, and $A \cos \theta_t$. Accordingly, the incident power is $I_i A \cos \theta_i$; this is the energy per unit time flowing in the incident beam, and it's therefore the power

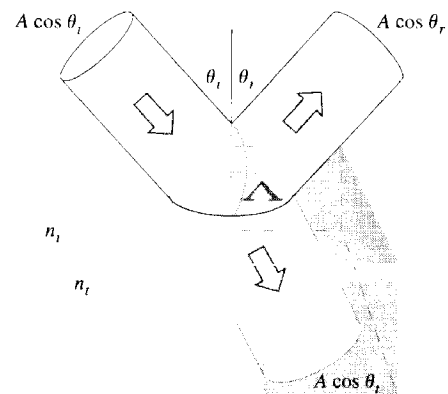


Figure 4.47 Reflection and transmission of an incident beam.

arriving on the surface over A . Similarly, $I_r A \cos \theta_r$ is the power in the reflected beam, and $I_t A \cos \theta_t$ is the power being transmitted through A . We define the **reflectance** R to be the ratio of the reflected power (or flux) to the incident power:

$$R \equiv \frac{I_r A \cos \theta_r}{I_i A \cos \theta_i} = \frac{I_r}{I_i} \quad (4.54)$$

In the same way, the **transmittance** T is defined as the ratio of the transmitted to the incident flux and is given by

$$T \equiv \frac{I_t \cos \theta_t}{I_i \cos \theta_i} \quad (4.55)$$

The quotient I_r/I_i equals $(v_r \epsilon_r E_{0r}^2/2)/(v_i \epsilon_i E_{0i}^2/2)$, and since the incident and reflected waves are in the same medium, $v_r = v_i$, $\epsilon_r = \epsilon_i$, and

$$R = \left(\frac{E_{0r}}{E_{0i}} \right)^2 = r^2 \quad (4.56)$$

In like fashion (assuming $\mu_i = \mu_r = \mu_0$),

$$T = \frac{n_t \cos \theta_t}{n_i \cos \theta_i} \left(\frac{E_{0t}}{E_{0i}} \right)^2 = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i} \right) t^2 \quad (4.57)$$

where use was made of the fact that $\mu_0 \epsilon_i = 1/v_i^2$ and $\mu_0 v_i \epsilon_i = n_i/c$. Notice that at normal incidence, which is a situation of great practical interest, $\theta_r = \theta_i = 0$, and the transmittance [Eq. (4.55)], like the reflectance [Eq. (4.54)], is then simply the ratio of the appropriate irradiances. Since $R = r^2$, we need not worry about the sign of r in any particular formulation, and that makes reflectance a convenient notion. Observe that in Eq. (4.57) T is not simply equal to t^2 , for two reasons. First, the ratio of the indices of refraction must be there, since the speeds at which energy is transported into and out of the interface are different, in other words, $I \propto v$, from Eq. (3.47). Second, the cross-sectional areas of the incident and refracted beams are different. The energy flow per unit area is affected accordingly, and that manifests itself in the presence of the ratio of the cosine terms.

Let's now write an expression representing the conservation of energy for the configuration depicted in Fig. 4.47. In other words, the total energy flowing into area A per unit time must equal the energy flowing outward from it per unit time:

$$I_i A \cos \theta_i = I_r A \cos \theta_r + I_t A \cos \theta_t \quad (4.58)$$

When both sides are multiplied by c , this expression becomes

$$n_i E_{0i}^2 \cos \theta_i = n_i E_{0r}^2 \cos \theta_r + n_i E_{0t}^2 \cos \theta_t$$

$$\text{or} \quad 1 = \left(\frac{E_{0r}}{E_{0i}} \right)^2 + \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i} \right) \left(\frac{E_{0t}}{E_{0i}} \right)^2 \quad (4.59)$$

But this is simply

$$R + T = 1 \quad (4.60)$$

where there was no absorption. It is convenient to use the component forms, that is,

$$R_{\perp} = r_{\perp}^2 \quad (4.61)$$

$$R_{\parallel} = r_{\parallel}^2 \quad (4.62)$$

$$T_{\perp} = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i} \right) t_{\perp}^2 \quad (4.63)$$

and

$$T_{\parallel} = \left(\frac{n_t \cos \theta_t}{n_i \cos \theta_i} \right) t_{\parallel}^2 \quad (4.64)$$

which are illustrated in Fig. 4.48. Furthermore, it can be shown (Problem 4.71) that

$$R_{\parallel} + T_{\parallel} = 1 \quad (4.65)$$

and

$$R_{\perp} + T_{\perp} = 1 \quad (4.66)$$

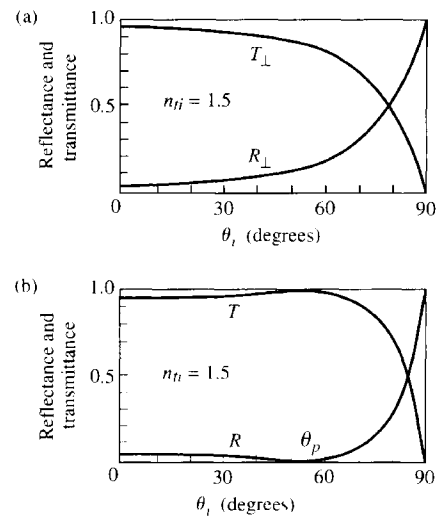


Figure 4.48 Reflectance and transmittance versus incident angle.



Looking down into a puddle (that's melting snow on the right) we see a reflection of the surrounding trees. At normal incidence water reflects about 2% of the light. As the viewing angle increases here its about 40° that percentage increases. (Photo by E.H.)

When $\theta_i = 0$, the incident plane becomes undefined, and any distinction between the parallel and perpendicular components of R and T vanishes. In this case Eqs. (4.61) through (4.64), along with (4.47) and (4.48), lead to

$$R = R_{\parallel} = R_{\perp} = \left(\frac{n_t - n_i}{n_t + n_i} \right)^2 \quad (4.67)$$

and

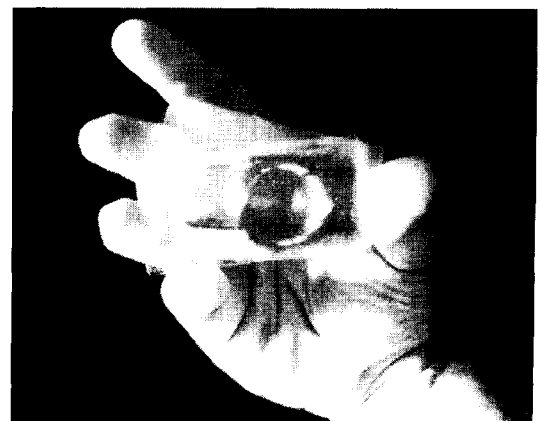
$$T = T_{\parallel} = T_{\perp} = \frac{4n_t n_i}{(n_t + n_i)^2} \quad (4.68)$$



At near normal incidence about 4% of the light is reflected back off each air-glass interface. Here because it's a lot brighter outside than inside the building, you have no trouble seeing the photographer. (Photo by E.H.)

Thus 4% of the light incident normally on an air-glass ($n_g = 1.5$) interface will be reflected back, whether internally, $n_i > n_t$, or externally, $n_i < n_t$ (Problem 4.72). This will be of concern to anyone who is working with a complicated lens system, which might have 10 or 20 such air-glass boundaries. Indeed, if you look perpendicularly into a stack of about 50 microscope slides (cover-glass sliders are much thinner and easier to handle in large quantities), most of the light will be reflected. The stack will look very much like a mirror (see photo). Roll up a thin sheet of clear plastic into a multiturned cylinder and it too will look like shiny metal. The many interfaces produce a large number of closely spaced *specular* reflections that send much of the light back into the incident medium, more or less, as if it had undergone a single frequency-independent reflection. A smooth gray-metal surface does pretty much the same thing—it has a large, frequency-independent specular reflectance—and looks shiny (that's what “shiny” is). If the reflection is diffuse, the surface will appear gray or even white if the reflectance is large enough.

Figure 4.49 is a plot of the reflectance at a single interface, assuming normal incidence for various transmitting media in air. Figure 4.50 depicts the corresponding dependence of the transmittance at normal incidence on the number of interfaces and the index of the medium. Of course, this is why you can't see through a roll of “clear” smooth-surfaced plastic tape, and it's also why the many elements in a periscope must be coated with antireflection films (Section 9.9.2).



Near normal reflection off a stack of microscope slides. You can see the image of the camera that took the picture. (Photo by E.H.)

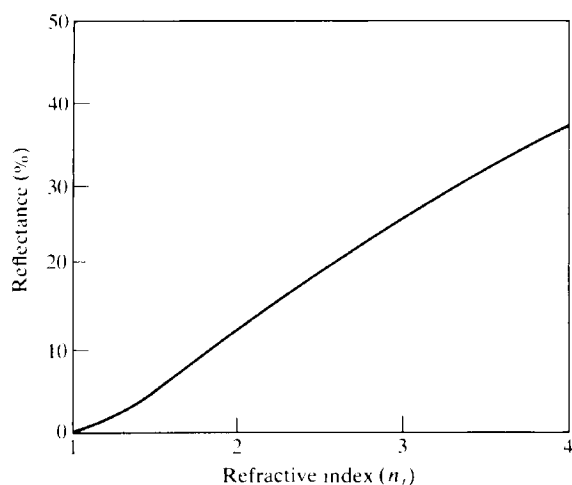


Figure 4.49 Reflectance at normal incidence in air ($n_i = 1.0$) at a single interface.

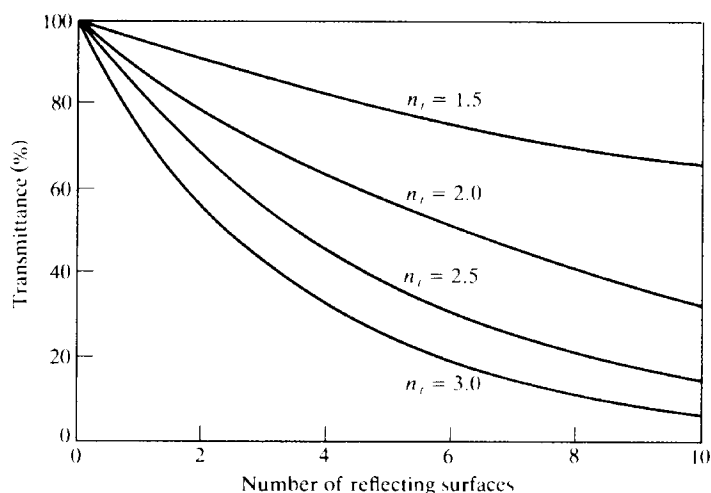


Figure 4.50 Transmittance through a number of surfaces in air ($n_i = 1.0$) at normal incidence.

4.7 Total Internal Reflection

In the previous section it was evident that something rather interesting was happening in the case of internal reflection ($n_i > n_t$) when θ_i was equal to or greater than θ_c , the so-called **critical angle**. Let's now return to that situation for a closer look. Suppose that we have a source embedded in an optically dense medium, and we allow θ_i to increase gradually, as indicated in Fig. 4.51. We know from the preceding section (Fig. 4.42) that $r_{\parallel}$ and $r_{\perp}$ increase with increasing θ_i , and therefore $t_{\parallel}$

and $t_{\perp}$ both decrease. Moreover $\theta_t > \theta_i$, since

$$\sin \theta_t = \frac{n_i}{n_t} \sin \theta_i$$

and $n_i > n_t$, in which case $n_{ti} < 1$. Thus as θ_i becomes larger, the transmitted ray gradually approaches tangency with the boundary, and as it does more and more of the available energy appears in the reflected beam. Finally, when $\theta_i = 90^\circ$, $\sin \theta_t = 1$ and

$$\sin \theta_c = n_{ti} \quad (4.69)$$

As noted earlier, *the critical angle is that special value of θ_i for which $\theta_t = 90^\circ$* . The larger n_i is, the smaller n_{ti} is, and the smaller θ_c is. For incident angles greater than or equal to θ_c , all the incoming energy is reflected back into the incident medium in the process known as **total internal reflection** (see photo).

It should be stressed that the transition from the conditions depicted in Fig. 4.51a to those of 4.51d takes place without any discontinuities. As θ_i becomes larger, the reflected beam grows stronger and stronger while the transmitted beam grows weaker, until the latter vanishes and the former carries off all the energy at $\theta_i = \theta_c$. It's an easy matter to observe the diminution of the transmitted beam as θ_i is made larger. Just place a glass microscope slide on a printed page, this time blocking out any specularly reflected light. At $\theta_i \approx 0$, θ_t is roughly zero, and the page as seen through the glass is fairly bright and clear. But if you move your head, allowing θ_i (the angle at which you view the interface) to increase, the region



Notice that you can't see the two front flames through the water along a bright horizontal band. That's due to total internal reflection. Look at the bottom of a drinking glass through its side. Now add a few inches of water. What happens? (Photo by E.H.)

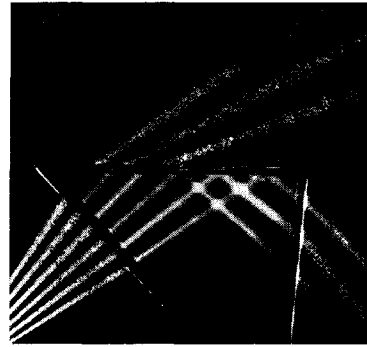
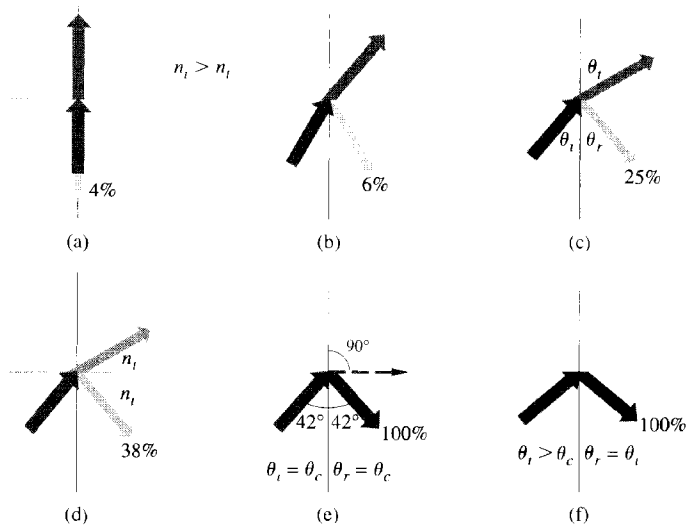


Figure 4.51 Internal reflection and the critical angle. (Photo courtesy of Educational Service, Inc.)

of the printed page covered by the glass will appear darker and darker, indicating that T has indeed been markedly reduced.

The critical angle for our air–glass interface is roughly 42° (see Table 4.2). Consequently, a ray incident normally on the

left face of either of the prisms in Fig. 4.52 will have a $\theta_i > 42^\circ$ and therefore be internally reflected. This is a convenient way to reflect nearly 100% of the incident light without having to worry about the deterioration that can occur with metallic surfaces (see photo).

Another useful way to view the situation is via Fig. 4.53, which shows a simplified representation of scattering off atomic oscillators. We know that the net effect of the presence of the homogeneous isotropic media is to alter the speed of the light from c to v_i and v_t , respectively (p. 92). The resultant wave is the superposition of these wavelets propagating at the appropriate speeds. In Fig. 4.53a an incident wave results in the emission of wavelets successively from scattering centers A and B . These overlap to form the transmitted wave. The reflected wave, which comes back down into the incident medium as usual ($\theta_i = \theta_r$), is not shown. In a time t the incident front travels a distance $v_i t = \overline{CB}$, while the transmitted front moves a distance $v_t t = \overline{AD} > \overline{CB}$. Since one wave moves from A to E in the same time that the other moves from

TABLE 4.2 Critical Angles

n_{it}	θ_c (degrees)	θ_c (radians)	n_{it}	θ_c (degrees)	θ_c (radians)
1.30	50.2849	0.8776	1.50	41.8103	0.7297
1.31	49.7612	0.8685	1.51	41.4718	0.7238
1.32	49.2509	0.8596	1.52	41.1395	0.7180
1.33	48.7535	0.8509	1.53	40.8132	0.7123
1.34	48.2682	0.8424	1.54	40.4927	0.7067
1.35	47.7946	0.8342	1.55	40.1778	0.7012
1.36	47.3321	0.8261	1.56	39.8683	0.6958
1.37	46.8803	0.8182	1.57	39.5642	0.6905
1.38	46.4387	0.8105	1.58	39.2652	0.6853
1.39	46.0070	0.8030	1.59	38.9713	0.6802
1.40	45.5847	0.7956	1.60	38.6822	0.6751
1.41	45.1715	0.7884	1.61	38.3978	0.6702
1.42	44.7670	0.7813	1.62	38.1181	0.6653
1.43	44.3709	0.7744	1.63	37.8428	0.6605
1.44	43.9830	0.7676	1.64	37.5719	0.6558
1.45	43.6028	0.7610	1.65	37.3052	0.6511
1.46	43.2302	0.7545	1.66	37.0427	0.6465
1.47	42.8649	0.7481	1.67	36.7842	0.6420
1.48	42.5066	0.7419	1.68	36.5296	0.6376
1.49	42.1552	0.7357	1.69	36.2789	0.6332

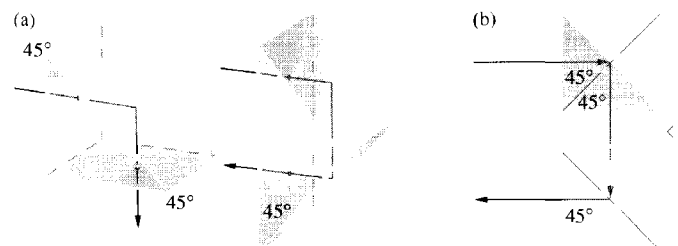


Figure 4.52 Total internal reflection.

C to B , and since they have the same frequency and period, they must change phase by the same amount in the process. Thus the disturbance at point E must be in-phase with that at point B ; both of these points must be on the same transmitted wavefront (remember Section 4.4.2).

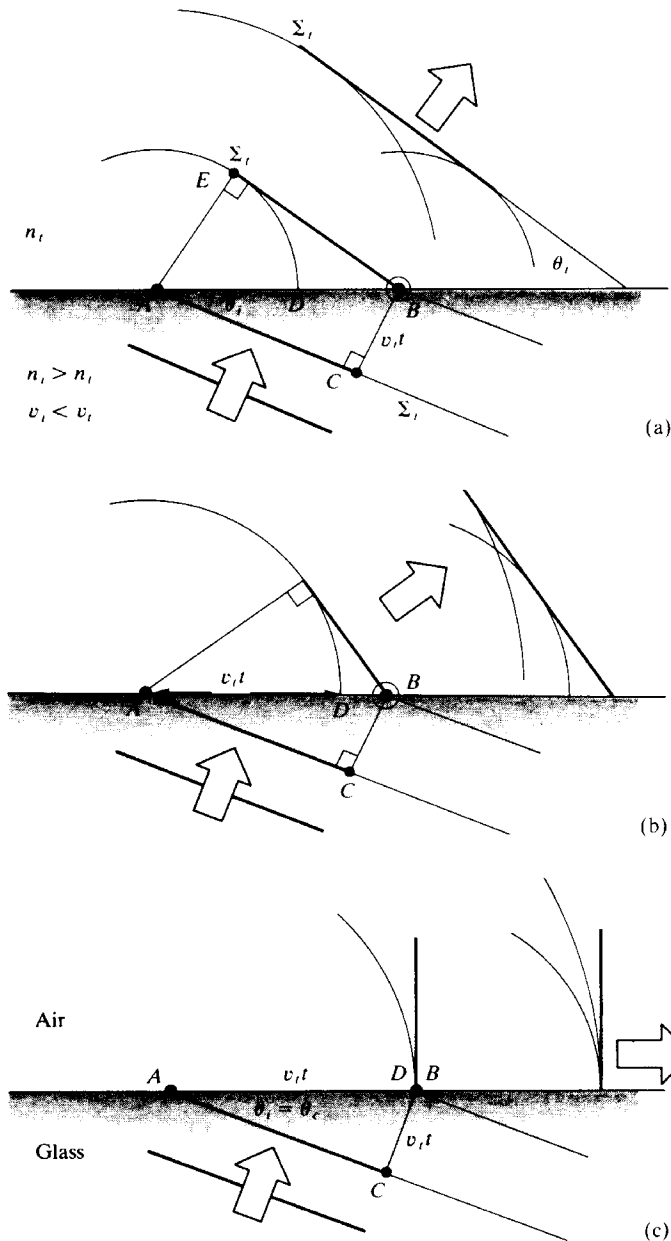
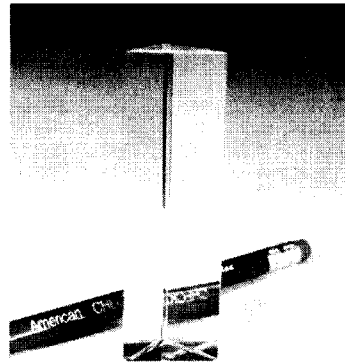


Figure 4.53 An examination of the transmitted wave in the process of total internal reflection from a scattering perspective. Here we keep θ_i and n_i constant and in successive parts of the diagram decrease n_t , thereby increasing v_t . The reflected wave ($\theta_r = \theta_i$) is not drawn.



The prism behaves like a mirror and reflects a portion of the pencil (reversing the lettering on it). The operating process is total internal reflection. (Photo by E.H.)

It can be seen that the greater v_t is in comparison to v_i , the more tilted the transmitted front will be (i.e., the larger θ_t will be). That much is depicted in Fig. 4.53b, where n_{ti} has been taken to be smaller by assuming n_t to be smaller. The result is a higher speed v_t , increasing $\overline{AD}$ and causing a greater transmission angle. In Fig. 4.53c a special case is reached: $\overline{AD} = \overline{AB} = v_t t$, and the wavelets will overlap in-phase *only along the line of the interface*, $\theta_t = 90^\circ$. From triangle ABC , $\sin \theta_i = v_t / v_i = n_i / n_t$, which is Eq. (4.69). For the two given media (i.e., for the particular value of n_{ti}), the direction in which the scattered wavelets will add constructively in the transmitting medium is along the interface. The resulting disturbance ($\theta_t = 90^\circ$) is known as a *surface wave*.

4.7.1 The Evanescent Wave

Because the frequency of X-rays is higher than the resonance frequencies of the atoms of the medium, Eq. (3.70) suggests, and experiments confirm, that the index of refraction of X-rays is less than 1.0. Thus the wave velocity of X-rays (i.e., the phase speed) in matter exceeds its value (c) in vacuum, although it usually does so by less than 1 part in 10 000, even in the densest solids. When X-rays traveling in air enter a dense material like glass, the beam bends ever so slightly *away* from the normal rather than toward it. With the above discussion of total internal reflection in mind, we should expect that X-rays will be **totally “externally” reflected** when, for example, $n_i = n_{\text{air}}$ and $n_t = n_{\text{glass}}$. This is the way it’s often spoken of in the literature, but that’s a misnomer; since for X-rays $n_{\text{air}} > n_{\text{glass}}$ and therefore $n_i > n_t$ (even though glass is physically more dense than air), the process is

actually still *internal* reflection. In any event, because n_t is less than, but very nearly equal to, 1 the index ratio $n_{ti} \approx 1$ and $\theta_c \approx 90^\circ$.

In 1923 A. H. Compton reasoned that even though X-rays incident on a sample at ordinary angles are not specularly reflected, they should be totally “externally” reflected at glancing incidence. He shined 0.128 nm X-rays on a glass plate and got a critical angle of about 10 minutes of arc (0.167°) measured with respect to the surface. That yielded an index of refraction for glass that differed from 1 by -4.2×10^{-6} .

We’ll come back to some important practical applications of both total internal and total “external” reflection later on (p. 200).

If we assume in the case of total internal reflection that there is no transmitted wave, it becomes impossible to satisfy the boundary conditions using only the incident and reflected waves—things are not at all as simple as they might seem. Furthermore, we can reformulate Eqs. (4.34) and (4.40) (Problem 4.75) such that

$$r_\perp = \frac{\cos \theta_i - (n_{ti}^2 - \sin^2 \theta_i)^{1/2}}{\cos \theta_i + (n_{ti}^2 - \sin^2 \theta_i)^{1/2}} \quad (4.70)$$

$$\text{and} \quad r_\parallel = \frac{n_{ti}^2 \cos \theta_i - (n_{ti}^2 - \sin^2 \theta_i)^{1/2}}{n_{ti}^2 \cos \theta_i + (n_{ti}^2 - \sin^2 \theta_i)^{1/2}} \quad (4.71)$$

Since $\sin \theta_c = n_{ti}$ when $\theta_i > \theta_c$, $\sin \theta_i > n_{ti}$, and both $r_\perp$ and $r_\parallel$ become complex quantities. Despite this (Problem 4.76), $r_\perp r_\perp^* = r_\parallel r_\parallel^* = 1$ and $R = 1$, which means that $I_r = I_i$ and $I_t = 0$. Thus, although there must be a transmitted wave, it cannot, on the average, carry energy across the boundary. We shall not perform the complete and rather lengthy computation needed to derive expressions for all the reflected and transmitted fields, but we can get an appreciation of what’s happening in the following way. The wavefunction for the transmitted electric field is

$$\vec{E}_t = \vec{E}_{0t} \exp i(\vec{k}_t \cdot \vec{r} - \omega t)$$

$$\text{where} \quad \vec{k}_t \cdot \vec{r} = k_{tx}x + k_{ty}y$$

there being no z -component of $\vec{k}$. But

$$k_{tx} = k_t \sin \theta_t$$

$$\text{and} \quad k_{ty} = k_t \cos \theta_t$$

as seen in Fig. 4.54. Once again using Snell’s Law,

$$k_t \cos \theta_t = \pm k_i \left(1 - \frac{\sin^2 \theta_i}{n_{ti}^2} \right)^{1/2} \quad (4.72)$$

or, since we are concerned with the case where $\sin \theta_i > n_{ti}$,

$$k_{ty} = \pm i k_i \left(\frac{\sin^2 \theta_i}{n_{ti}^2} - 1 \right)^{1/2} \equiv \pm i\beta$$

and

$$k_{tx} = \frac{k_t}{n_{ti}} \sin \theta_i$$

Hence

$$\vec{E}_t = \vec{E}_{0t} e^{\mp \beta y} e^{i(k_{tx}x \sin \theta_i / n_{ti} - \omega t)} \quad (4.73)$$

Neglecting the positive exponential, which is physically untenable, we have a wave whose amplitude drops off exponentially as it penetrates the less dense medium. The disturbance advances in the x -direction as a surface or **evanescent wave**. Notice that the wavefronts or surfaces of constant phase (parallel to the yz -plane) are perpendicular to the surfaces of constant amplitude (parallel to the xz -plane), and as such the wave is *inhomogeneous* (p. 26). Its amplitude decays rapidly in the y -direction, becoming negligible at a distance into the second medium of only a few wavelengths.

If you are still concerned about the conservation of energy, a more extensive treatment would have shown that energy actually circulates back and forth across the interface, resulting on the average in a zero net flow through the boundary into the second medium. Yet one puzzling point remains, inas-

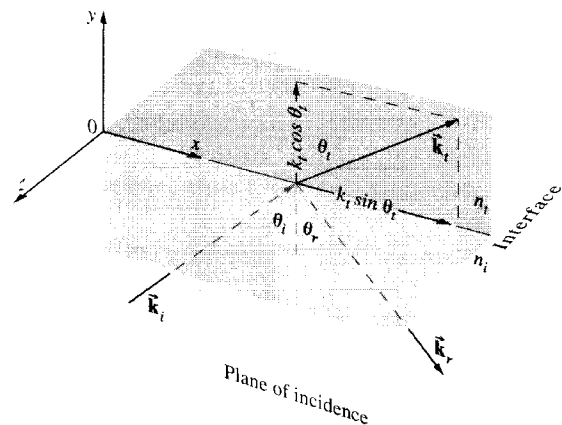


Figure 4.54 Propagation vectors for internal reflection.

much as there is still a bit of energy to be accounted for, namely, that associated with the evanescent wave that moves along the boundary in the plane-of-incidence. Since this energy could not have penetrated into the less dense medium under the present circumstances (so long as $\theta_i \geq \theta_c$), we must look elsewhere for its source. Under actual experimental conditions the incident beam would have a finite cross section and therefore would obviously differ from a true plane wave. This deviation gives rise (via diffraction) to a slight transmission of energy across the interface, which is manifested in the evanescent wave.

Incidentally, it is clear from (c) and (d) in Fig. 4.44 that the incident and reflected waves (except at $\theta_i = 90^\circ$) do not differ in phase by π and cannot therefore cancel each other. It follows from the continuity of the tangential component of $\vec{E}$ that there must be an oscillatory field in the less dense medium, with a component parallel to the interface having a frequency ω (i.e., the evanescent wave).

The exponential decay of the surface wave, or *boundary wave*, as it is also sometimes called, has been confirmed experimentally at optical frequencies.*

Imagine that a beam of light traveling within a block of glass is internally reflected at a boundary. Presumably, if you pressed another piece of glass against the first, the air–glass interface could be made to vanish, and the beam would then propagate onward undisturbed. Furthermore, you might expect this transition from total to no reflection to occur gradually as the air film thinned out. In much the same way, if you hold a drinking glass or a prism, you can see the ridges of your fingerprints in a region that, because of total internal reflection, is otherwise mirrorlike. In more general terms, when the evanescent wave extends with appreciable amplitude across the rare medium into a nearby region occupied by a higher-index material, energy may flow through the gap in what is known as **frustrated total internal reflection** (FTIR). The evanescent wave, having traversed the gap, is still strong enough to drive electrons in the “frustrating” medium; they in turn will generate a wave that significantly alters the field configuration, thereby permitting energy to flow. Figure 4.55 is a schematic representation of FTIR: the width of the lines depicting the wavefronts decreases across the gap as a

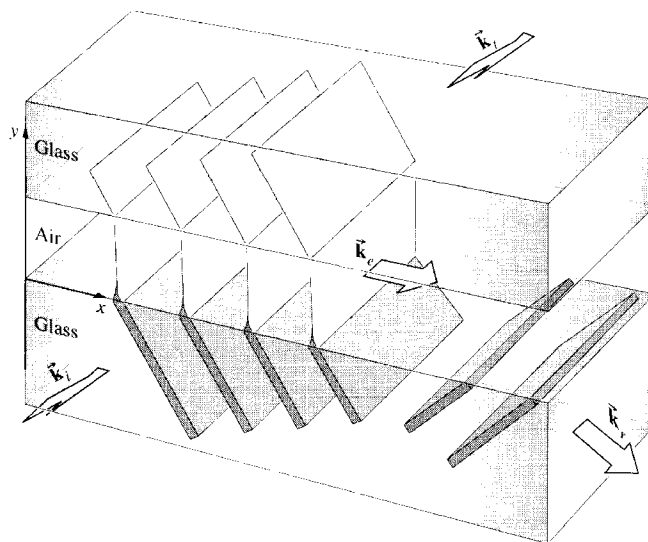


Figure 4.55 Frustrated total internal reflection.

reminder that the amplitude of the field behaves in the same way. The process as a whole is remarkably similar to the quantum-mechanical phenomenon of *barrier penetration* or *tunneling*, which has numerous applications in contemporary physics.

One can demonstrate FTIR with the prism arrangement of Fig. 4.56 in a manner that is fairly self-evident. Moreover, if the hypotenuse faces of both prisms are made planar and parallel, they can be positioned so as to transmit and reflect any desired fraction of the incident flux density. Devices that perform this function are known as **beamsplitters**. A *beamsplitter cube* can be made rather conveniently by using a thin, low-index transparent film as a precision spacer. Low-loss reflectors whose transmittance can be controlled by frustrating internal reflection are of considerable practical interest. FTIR can also be observed in other regions of the electromagnetic spectrum. Three-centimeter microwaves are particularly easy to work with, inasmuch as the evanescent wave will extend roughly 10^5 times farther than it would at optical frequencies. One can duplicate the above optical experiments with solid prisms made of paraffin or hollow ones of acrylic plastic filled with kerosene or motor oil. Any one of these would have an index of about 1.5 for 3-cm waves. It then becomes an easy matter to measure the dependence of the field amplitude on y .

*Take a look at the fascinating article by K. H. Drexhage, "Monomolecular layers and light." *Sci. Am.* **222**, 108 (1970).

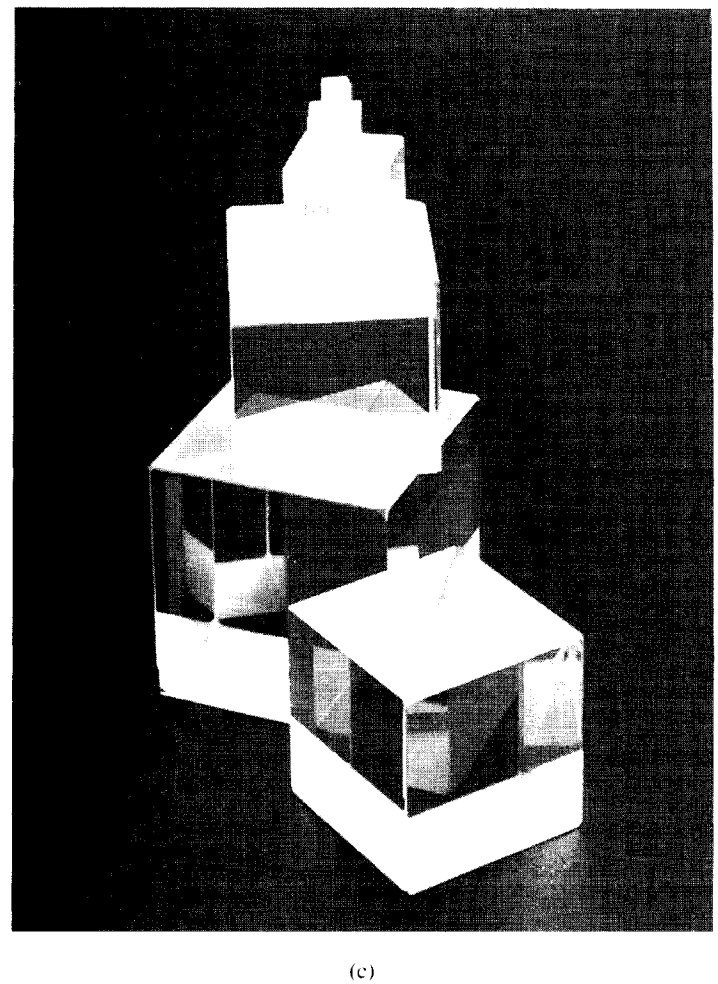
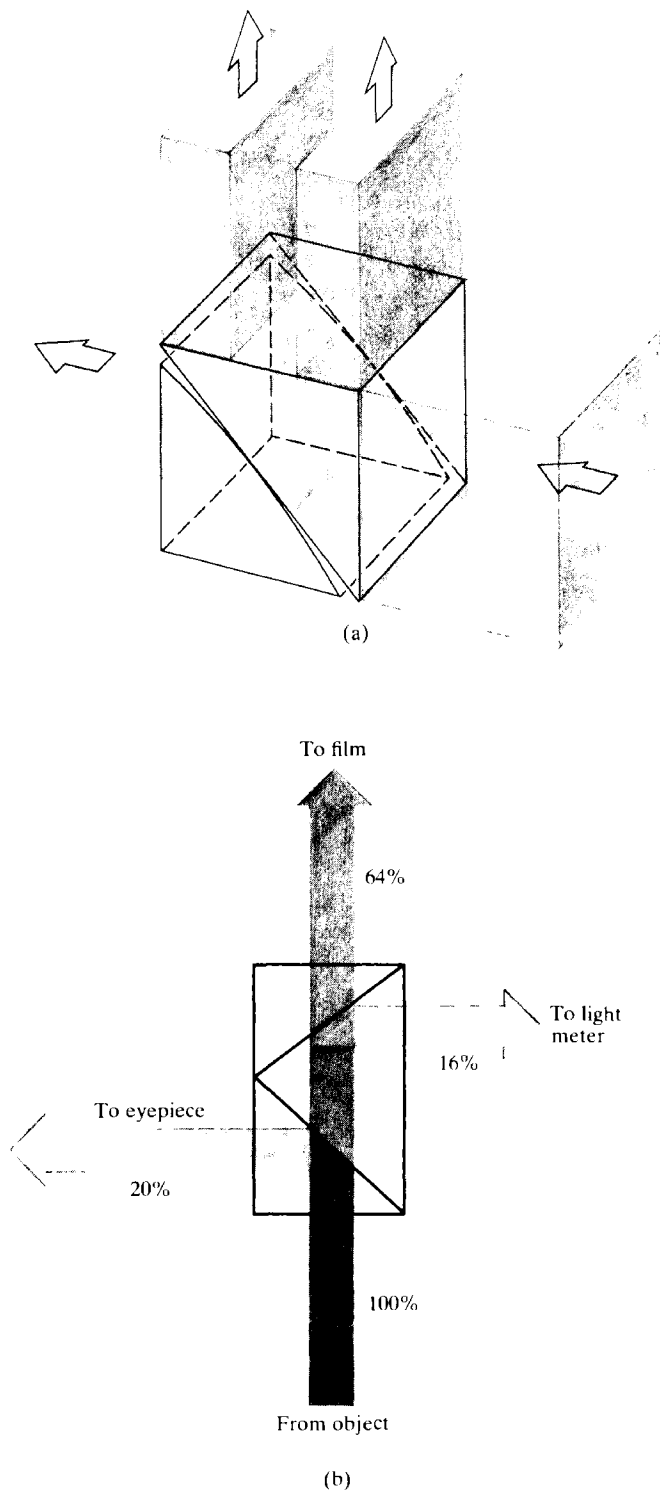


Figure 4.56 (a) A beamsplitter utilizing FTIR. (b) A typical modern application of FTIR: a conventional beamsplitter arrangement used to take photographs through a microscope. (c) Beamsplitter cubes. (Photo courtesy Melles Griot.)

4.8 Optical Properties of Metals

The characteristic feature of conducting media is the presence of a number of free electric charges (free in the sense of being unbound, i.e., able to circulate within the material). For metals these charges are of course electrons, and their motion constitutes a current. The current per unit area resulting from the application of a field $\vec{E}$ is related by means of Eq. (A1.15) to the conductivity of the medium σ . For a dielectric there are no

free or conduction electrons and $\sigma = 0$, whereas for metals σ is nonzero and finite. In contrast, an idealized “perfect” conductor would have an infinite conductivity. This is equivalent to saying that the electrons, driven into oscillation by a harmonic wave, would simply follow the field’s alternations. There would be no restoring force, no natural frequencies, and no absorption, only re-emission. In real metals the conduction electrons undergo collisions with the thermally agitated lattice or with imperfections and in so doing irreversibly convert electromagnetic energy into joule heat. The absorption of radiant energy by a material is a function of its conductivity.

Waves in a Metal

If we visualize the medium as continuous, Maxwell’s Equations lead to

$$\frac{\partial^2 \vec{E}}{\partial x^2} + \frac{\partial^2 \vec{E}}{\partial y^2} + \frac{\partial^2 \vec{E}}{\partial z^2} = \mu\epsilon \frac{\partial^2 \vec{E}}{\partial t^2} + \mu\sigma \frac{\partial \vec{E}}{\partial t} \quad (4.74)$$

which is Eq. (A1.21) in Cartesian coordinates. The last term, $\mu\sigma \partial \vec{E} / \partial t$, is a first-order time derivative, like the damping force in the oscillator model (p. 71). The time rate-of-change of $\vec{E}$ generates a voltage, currents circulate, and since the material is resistive, light is converted to thermal energy—ergo absorption. This expression can be reduced to the unattenuated wave equation, if the permittivity is reformulated as a complex quantity. This in turn leads to a complex index of refraction, which, as we saw earlier (p. 71), is tantamount to absorption. We then need only substitute the complex index

$$\tilde{n} = n_R - in_I \quad (4.75)$$

(where the real and imaginary indices n_R and n_I are both real numbers) into the corresponding solution for a nonconducting medium. Alternatively, we can utilize the wave equation and appropriate boundary conditions to yield a specific solution. In either event, it is possible to find a simple sinusoidal plane-wave solution applicable within the conductor. Such a wave propagating in the y -direction is ordinarily written as

$$\vec{E} = \vec{E}_0 \cos(\omega t - ky)$$

or as a function of n ,

$$\vec{E} = \vec{E}_0 \cos \omega(t - \tilde{n}y/c)$$

but here the refractive index must be taken as complex. Writ-

ing the wave as an exponential and using Eq. (4.75) yields

$$\vec{E} = \vec{E}_0 e^{(-\omega n_I y/c)} e^{i\omega(t - n_R y/c)} \quad (4.76)$$

or
$$\vec{E} = \vec{E}_0 e^{-\omega n_I y/c} \cos \omega(t - n_R y/c) \quad (4.77)$$

The disturbance advances in the y -direction with a speed c/n_R , precisely as if n_R were the more usual index of refraction. As the wave progresses into the conductor, its amplitude, $\vec{E}_0 \exp(-\omega n_I y/c)$, is exponentially attenuated. Inasmuch as irradiance is proportional to the square of the amplitude, we have

$$I(y) = I_0 e^{-\alpha y} \quad (4.78)$$

where $I_0 = I(0)$; that is, I_0 is the irradiance at $y = 0$ (the interface), and $\alpha \equiv 2\omega n_I/c$ is called the *absorption coefficient* or (even better) the **attenuation coefficient**. The flux density will drop by a factor of $e^{-1} = 1/2.7 \approx \frac{1}{3}$ after the wave has propagated a distance $y = 1/\alpha$, known as the **skin** or **penetration depth**. For a material to be transparent, the penetration depth must be large in comparison to its thickness. The penetration depth for metals, however, is exceedingly small. For example, copper at ultraviolet wavelengths ($\lambda_0 \approx 100$ nm) has a miniscule penetration depth, about 0.6 nm, while it is still only about 6 nm in the infrared ($\lambda_0 \approx 10\,000$ nm). This accounts for the generally observed opacity of metals, which nonetheless can become partly transparent when formed into extremely thin films (e.g., in the case of partially silvered two-way mirrors). The familiar metallic sheen of conductors corresponds to a high reflectance, which exists because the incident wave cannot effectively penetrate the material. Relatively few electrons in the metal “see” the transmitted wave, and therefore, although each absorbs strongly, little total energy is dissipated by them. Instead, most of the incoming energy reappears as the reflected wave. The majority of metals, including the less common ones (e.g., sodium, potassium, cesium, vanadium, niobium, gadolinium, holmium, yttrium, scandium, and osmium) have a silvery gray appearance like that of aluminum, tin, or steel. They reflect almost all the incident light (roughly 85 to 95%) regardless of wavelengths and are therefore essentially colorless.

Equation (4.77) is certainly reminiscent of Eq. (4.73) and FTIR. In both cases there is an exponential decay of the amplitude. Moreover, a complete analysis would show that the transmitted waves are not strictly transverse, there being a component of the field in the direction of propagation in both instances.

The representation of metal as a continuous medium works fairly well in the low-frequency, long-wavelength domain of the infrared. Yet we certainly might expect that as the wavelength of the incident beam decreased the actual granular nature of matter would have to be reckoned with. Indeed, the continuum model shows large discrepancies from experimental results at optical frequencies. And so we again turn to the classical atomistic picture initially formulated by Hendrik Lorentz, Paul Karl Ludwig Drude (1863–1906), and others. This simple approach will provide qualitative agreement with the experimental data, but the ultimate treatment requires quantum theory.

The Dispersion Equation

Envision the conductor as an assemblage of driven, damped oscillators. Some correspond to free electrons and will therefore have zero restoring force, whereas others are bound to the atom, much like those in the dielectric media of Section 3.5.1. The conduction electrons are, however, the predominant contributors to the optical properties of metals. Recall that the displacement of a vibrating electron was given by

$$x(t) = \frac{q_e/m_e}{(\omega_0^2 - \omega^2)} E(t) \quad [3.66]$$

With no restoring force, $\omega_0 = 0$, the displacement is opposite in sign to the driving force $q_e E(t)$ and therefore 180° out-of-phase with it. This is unlike the situation for transparent dielectrics, where the resonance frequencies are above the visible and the electrons oscillate in-phase with the driving force (Fig. 4.57). Free electrons oscillating out-of-phase with the incident light will reradiate wavelets that tend to cancel the incoming disturbance. The effect, as we have already seen, is a rapidly decaying refracted wave.

Assuming that the average field experienced by an electron moving about within a conductor is just the applied field $\vec{E}(t)$, we can extend the dispersion equation of a rare medium [Eq. (3.72)] to read

$$n^2(\omega) = 1 + \frac{Nq_e^2}{\epsilon_0 m_e} \left[\frac{f_e}{-\omega^2 + i\gamma_e \omega} + \sum_j \frac{f_j}{\omega_{0j}^2 - \omega^2 + i\gamma_j \omega} \right] \quad (4.79)$$

The first bracketed term is the contribution from the free electrons, wherein N is the number of atoms per unit volume. Each of these has f_e conduction electrons, which have no natural fre-

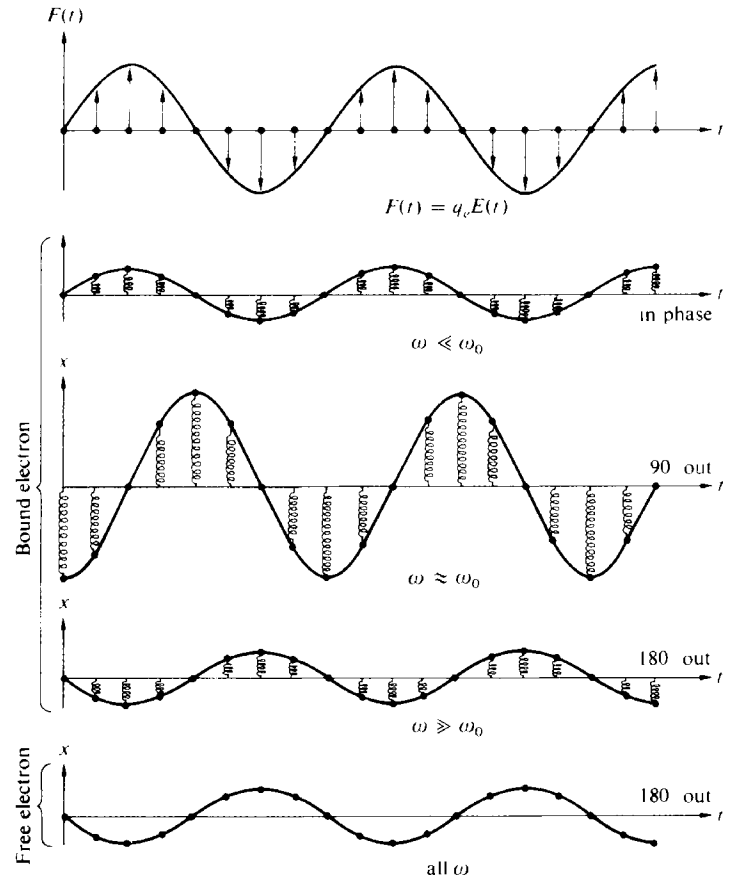


Figure 4.57 Oscillations of bound and free electrons.

quencies. The second term arises from the bound electrons and is identical to Eq. (3.72). It should be noted that if a metal has a particular color, it indicates that the atoms are partaking of selective absorption by way of the bound electrons, in addition to the general absorption characteristic of the free electrons. Recall that a medium that is very strongly absorbing at a given frequency doesn't actually absorb much of the incident light at that frequency but rather *selectively reflects* it. Gold and copper are reddish yellow because n_i increases with wavelength, and the larger values of λ are reflected more strongly. Thus, for example, gold should be fairly opaque to the longer visible wavelengths. Consequently, under white light, a gold foil less than roughly 10^{-6} m thick will indeed transmit predominantly greenish blue light.

We can get a rough idea of the response of metals to light by making a few simplifying assumptions. Accordingly, neglect the bound electron contribution and assume that γ_e is

also negligible for very large ω , whereupon

$$n^2(\omega) = 1 - \frac{Nq_e^2}{\epsilon_0 m_e \omega^2} \quad (4.80)$$

The latter assumption is based on the fact that at high frequencies the electrons will undergo a great many oscillations between each collision. Free electrons and positive ions within a metal may be thought of as a plasma whose density oscillates at a natural frequency ω_p , the **plasma frequency**. This in turn can be shown to equal $(Nq_e^2/\epsilon_0 m_e)^{1/2}$, and so

$$n^2(\omega) = 1 - (\omega_p/\omega)^2 \quad (4.81)$$

The plasma frequency serves as a critical value below which the index is complex and the penetrating wave drops off exponentially [Eq. (4.77)] from the boundary; at frequencies above ω_p , n is real, absorption is small, and the conductor is transparent. In the latter circumstance n is less than 1, as it was for dielectrics at very high frequencies (v can be greater than c —see p. 72). Hence we can expect metals in general to be fairly transparent to X-rays. Table 4.3 lists the plasma frequencies for some of the alkali metals that are transparent even to ultraviolet.

The index of refraction for a metal will usually be complex, and the impinging wave will suffer absorption in an amount that is frequency dependent. For example, the outer visors on the Apollo space suits were overlaid with a very thin film of gold (see photo). The coating reflected about 70% of the incident light and was used under bright conditions, such as low and forward Sun angles. It was designed to decrease the thermal load on the cooling system by strongly reflecting radiant energy in the infrared while still transmitting adequately in the



Edwin Aldrin Jr. at Tranquility Base on the Moon. The photographer, Neil Armstrong, is reflected in the gold-coated visor. (Photo courtesy NASA.)

visible. Inexpensive metal-coated sunglasses which are quite similar in principle are also available commercially, and they're well worth having just to experiment with.

The ionized upper atmosphere of the Earth contains a distribution of free electrons that behave very much like those confined within a metal. The index of refraction of such a medium will be real and less than 1 for frequencies above ω_p . In July of 1965 the *Mariner IV* spacecraft made use of this effect to examine the ionosphere of the planet Mars, 216 million kilometers from Earth.*

If we wish to communicate between two distant terrestrial points, we might bounce low-frequency waves off the Earth's ionosphere. To speak to someone on the Moon, however, we should use high-frequency signals, to which the ionosphere would be transparent.

TABLE 4.3 Critical Wavelengths and Frequencies for Some Alkali Metals

Metal	λ_p (observed) nm	λ_p (calculated) nm	$\nu_p = c/\lambda_p$ (observed) Hz
Lithium (Li)	155	155	1.94×10^{15}
Sodium (Na)	210	209	1.43×10^{15}
Potassium (K)	315	287	0.95×10^{15}
Rubidium (Rb)	340	322	0.88×10^{15}

Reflection from a Metal

Imagine that a plane wave initially in air impinges on a conducting surface. The transmitted wave advancing at some angle to the normal will be inhomogeneous. But if the conductivity of the medium is increased, the wavefronts will become aligned with the surfaces of constant amplitude, whereupon $\mathbf{k}_t$ and $\hat{\mathbf{u}}_n$ will approach parallelism. In other

*R. Von Eshelman, *Sci. Am.* **220**, 78 (1969).

words, in a good conductor the transmitted wave propagates in a direction normal to the interface regardless of θ_i .

Let's now compute the reflectance, $R = I_r/I_i$, for the simplest case of normal incidence on a metal. Taking $n_i = 1$ and $n_t = \tilde{n}$ (i.e., the complex index), we have from Eq. (4.47) that

$$R = \left(\frac{\tilde{n} - 1}{\tilde{n} + 1} \right) \left(\frac{\tilde{n} - 1}{\tilde{n} + 1} \right)^* \quad (4.82)$$

and therefore, since $\tilde{n} = n_R - in_I$,

$$R = \frac{(n_R - 1)^2 + n_I^2}{(n_R + 1)^2 + n_I^2} \quad (4.83)$$

If the conductivity of the material goes to zero, we have the case of a dielectric, whereupon in principle the index is real ($n_I = 0$), and the attenuation coefficient, α , is zero. Under those circumstances, the index of the transmitting medium n_t is n_R , and the reflectance [Eq. (4.83)] becomes identical with that of Eq. (4.67). If instead n_I is large while n_R is comparatively small, R in turn becomes large (Problem 4.81). In the unattainable limit where $\tilde{n}$ is purely imaginary, 100% of the incident flux density would be reflected ($R = 1$). Notice that it is possible for the reflectance of one metal to be greater than that of another even though its n_I is smaller. For example, at $\lambda_0 = 589.3$ nm the parameters associated with solid sodium are roughly $n_R = 0.04$, $n_I = 2.4$, and $R = 0.9$; and those for bulk tin are $n_R = 1.5$, $n_I = 5.3$, and $R = 0.8$; whereas for a gallium single crystal $n_R = 3.7$, $n_I = 5.4$, and $R = 0.7$.

The curves of $R_{\parallel}$ and $R_{\perp}$ for oblique incidence shown in Fig. 4.58 are somewhat typical of absorbing media. Thus, although R at $\theta_i = 0$ is about 0.5 for gold, as opposed to nearly 0.9 for silver in white light, the two metals have reflectances

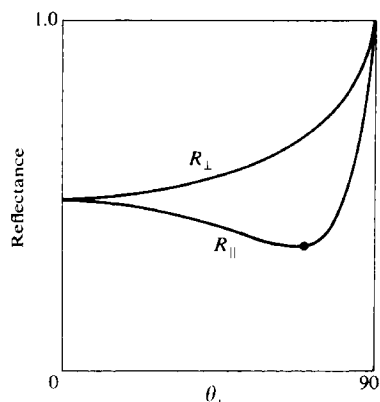


Figure 4.58 Typical reflectance for a linearly polarized beam of white light incident on an absorbing medium.

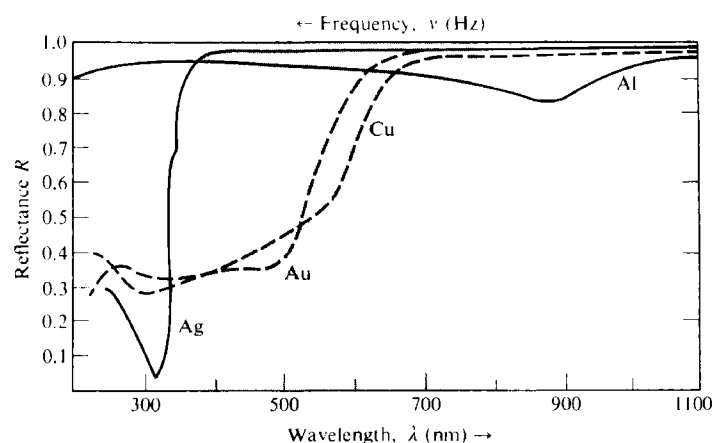


Figure 4.59 Reflectance versus wavelength for silver, gold, copper, and aluminum.

that are quite similar in shape, approaching 1.0 at $\theta_i = 90^\circ$. Just as with dielectrics (Fig. 4.48), $R_{\parallel}$ drops to a minimum at what is now called the *principal angle-of-incidence*, but here that minimum is nonzero. Figure 4.59 illustrates the spectral reflectance at normal incidence for a number of evaporated metal films under ideal conditions. Observe that although gold transmits fairly well in and below the green region of the spectrum, silver, which is highly reflective across the visible, becomes transparent in the ultraviolet at about 316 nm.

Phase shifts arising from reflection off a metal occur in both components of the field (i.e., parallel and perpendicular to the plane-of-incidence). These are generally neither 0 nor π , with a notable exception at $\theta_i = 90^\circ$, where, just as with a dielectric, both components shift phase by 180° on reflection.

4.9 Familiar Aspects of the Interaction of Light and Matter

Let's now examine some of the phenomena that paint the everyday world in a marvel of myriad colors.

As we saw earlier (p. 77), light that contains a roughly equal amount of every frequency in the visible region of the spectrum is perceived as white. A broad source of white light (whether natural or artificial) is one for which every point on its surface can be imagined as sending out a stream of light of every visible frequency. Given that we evolved on this planet, it's not surprising that a source appears white when its emis-

sion spectrum resembles that of the Sun. Similarly, a reflecting surface that accomplishes essentially the same thing will also appear white: a highly reflecting, frequency-independent, *diffusely* scattering object will be perceived as white under white light illumination.

Although water is essentially transparent, water vapor appears white, as does ground glass. The reason is simple enough—if the grain size is small but larger than the wavelengths involved, light will enter each transparent particle, be reflected and refracted, and emerge. There will be no distinction among any of the frequency components, so the reflected light reaching the observer will be white (p. 77). This is the mechanism accountable for the whiteness of things like sugar, salt, paper, cloth, clouds, talcum powder, snow, and paint, each grain or fiber of which is actually transparent.

Similarly, a wadded-up piece of crumpled clear plastic wrap will appear whitish, as will an ordinarily transparent material filled with small air bubbles (e.g., shaving cream or beaten egg white). Even though we usually think of paper, talcum powder, and sugar as each consisting of some sort of opaque white substance, it's an easy matter to dispel that misconception. Cover a printed page with a few of these materials (a sheet of white paper, some grains of sugar, or talcum) and illuminate it from behind. You'll have little difficulty seeing through them. In the case of white paint, one simply suspends colorless transparent particles, such as the oxides of zinc, titanium, or lead, in an equally transparent vehicle, for example, linseed oil or acrylics. Obviously, if the particles and vehicle have the same index of refraction, there will not be any reflections at the grain boundaries. The particles will simply disappear into the conglomeration, which itself remains clear. In contrast, if the indices are markedly different, there will be a good deal of reflection at all wavelengths (Problem 4.74), and the paint will appear white and opaque [take another look at Eq. (4.67)]. To color paint one need only dye the particles so that they absorb all frequencies except the desired range.

Carrying the logic in the reverse direction, if we reduce the relative index, n_{12} , at the grain or fiber boundaries, the particles of material will reflect less, thereby decreasing the overall whiteness of the object. Consequently, a wet white tissue will have a grayish, more transparent look. Wet talcum powder loses its sparkling whiteness, becoming a dull gray, as does wet white cloth. In the same way, a piece of dyed fabric soaked in a clear liquid (e.g., water, gin, or benzene) will lose its whitish haze and become much darker, the colors then being deep and rich like those of a still-wet watercolor painting.

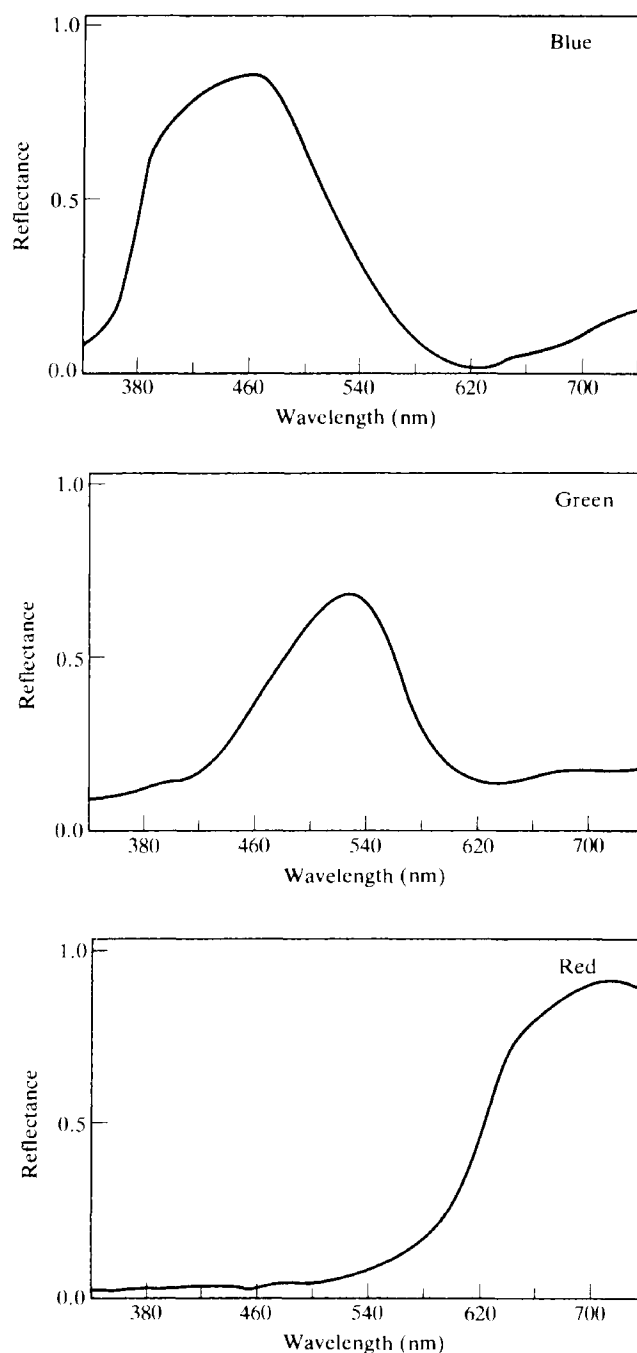


Figure 4.60 Reflection curves for blue, green, and red pigments. These are typical, but there is a great deal of possible variation among the colors.

A diffusely reflecting surface that absorbs somewhat—uniformly across the spectrum—will reflect a bit less than a white surface and so appear mat gray. The less it reflects, the darker

the gray, until it absorbs almost all the light and appears black. A surface that reflects perhaps 70% or 80% or more, but does so specularly, will appear the familiar shiny gray of a typical metal. Metals possess tremendous numbers of free electrons (p. 127) that scatter light very effectively, independent of frequency: they are not bound to the atoms and have no associated resonances. Moreover, the amplitudes of the vibrations are an order of magnitude larger than they were for the bound electrons. The incident light cannot penetrate into the metal any more than a fraction of a wavelength or so before it's canceled completely. There is little or no refracted light; most of the energy is reflected out, and only the small remainder is absorbed. Note that the primary difference between a gray surface and a mirrored surface is one of diffuse versus specular reflection. An artist paints a picture of a polished "white" metal, such as silver or aluminum, by "reflecting" images of things in the room on top of a gray surface.

When the distribution of energy in a beam of light is not effectively uniform across the spectrum, the light appears colored. Figure 4.60 depicts typical frequency distributions for what would be perceived as red, green, and blue light. These curves show the predominant frequency regions, but there can be a great deal of variation in the distributions, and they will still provoke the responses of red, green, and blue. In the early 1800s Thomas Young showed that a broad range of colors could be generated by mixing three beams of light, provided their frequencies were widely separated. When three such beams combine to produce white light, they are called **primary colors**. There is no single unique set of these primaries, nor do they have to be quasimonochromatic. Since a wide range of colors can be created by mixing red (R), green (G), and blue (B), these tend to be used most frequently. They are the three components (emitted by three phosphors) that generate the whole gamut of hues seen on a color television set.

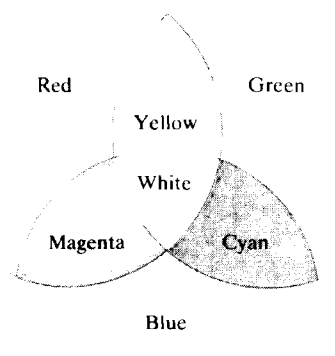


Figure 4.61 Three overlapping beams of colored light. A color television set uses these same three primary light sources—red, green, and blue.

Figure 4.61 summarizes the results when beams of these three primaries are overlapped in a number of different combinations: Red light plus blue light is seen as *magenta* (M), a reddish purple; blue light plus green light is seen as *cyan* (C), a bluish green or turquoise; and perhaps most surprising, red light plus green light is seen as *yellow* (Y). The sum of all three primaries is white:

$$R + B + G = W$$

$$M + G = W, \text{ since } R + B = M$$

$$C + R = W, \text{ since } B + G = C$$

$$Y + B = W, \text{ since } R + G = Y$$

Any two colored light beams that together produce white are said to be **complementary**, and the last three symbolic statements exemplify that situation. Thus

$$R + B + G = W$$

$$R + B = W - G = M$$

$$B + G = W - R = C$$

$$R + G = W - B = Y$$

which means, for example, that a filter that absorbs blue out of white light passes yellow.

Suppose we overlap beams of magenta and yellow light:

$$M + Y = (R + B) + (R + G) = W + R$$

The result is a combination of red and white, or pink. That raises another point: we say a color is **saturated**, that it is deep and intense, when it does not contain any white light. As Fig. 4.62 shows, pink is unsaturated red—red superimposed on a background of white.

The mechanism responsible for the yellowish red hue of gold and copper is, in some respects, similar to the process that causes the sky to appear blue. Putting it rather succinctly, the molecules of air have resonances in the ultraviolet and will be driven into larger-amplitude oscillations as the frequency of the incident light increases toward the ultraviolet. They effectively take energy from and re-emit the blue component of sunlight in all directions, transmitting the complementary red

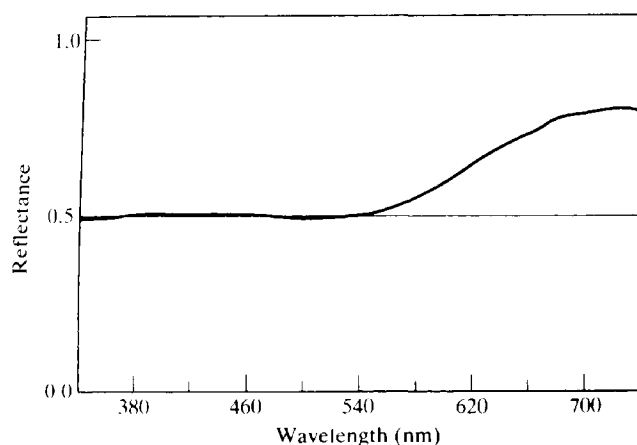


Figure 4.62 Spectral reflection of a pink pigment.

end of the spectrum with little alteration. This is analogous to the selective reflection or scattering of yellow-red light that takes place at the surface of a gold film and the concomitant transmission of blue-green light.

The characteristic colors of most substances have their origin in the phenomenon of **selective** or **preferential absorption**. For example, water has a very faint green-blue tint because of its absorption of red light. That is, the H_2O molecules have a broad resonance in the infrared, which extends somewhat into the visible. The absorption isn't very strong, so there is no accentuated reflection of red light at the surface. Instead it is transmitted and gradually absorbed out until at a depth of about 30 m of seawater, red is almost completely removed from the sunlight. This same process of selective absorption is responsible for the colors of brown eyes and butterflies, of birds and bees and cabbages and kings. Indeed, the great majority of objects in nature appear to have characteristic colors as the result of preferential absorption by pigment molecules. In contrast with most atoms and molecules, which have resonances in the ultraviolet and infrared, the pigment molecules must obviously have resonances in the visible. Yet visible photons have energies of roughly 1.6 eV to 3.2 eV, which, as you might expect, are on the low side for ordinary electron excitation and on the high side for excitation via molecular vibration. Despite this, *there are atoms where the bound electrons form incomplete shells* (gold, for example) and variations in the configuration of these shells provide a mode for low-energy excitation. In addition, there is the large group of organic dye molecules, which evidently also have

resonances in the visible. All such substances, whether natural or synthetic, consist of long-chain molecules made up of regularly alternating single and double bonds in what is called a conjugated system. This structure is typified by the carotene molecule $\text{C}_{40}\text{H}_{56}$ (Fig. 4.63). The carotenoids range in color from yellow to red and are found in carrots, tomatoes, daffodils, dandelions, autumn leaves, and people. The chlorophylls are another group of familiar natural pigments, but here a portion of the long chain is turned around on itself to form a ring. In any event, conjugated systems of this sort contain a number of particularly mobile electrons known as *pi electrons*. They are not bound to specific atomic sites but instead can range over the relatively large dimensions of the molecular chain or ring. In the phraseology of Quantum Mechanics, we say that these are long-wavelength, low-frequency, and therefore low-energy, electron states. The energy required to raise a pi electron to an excited state is comparatively low, corresponding to that of visible photons. In effect, the molecule can be imagined as an oscillator having a resonance frequency in the visible.

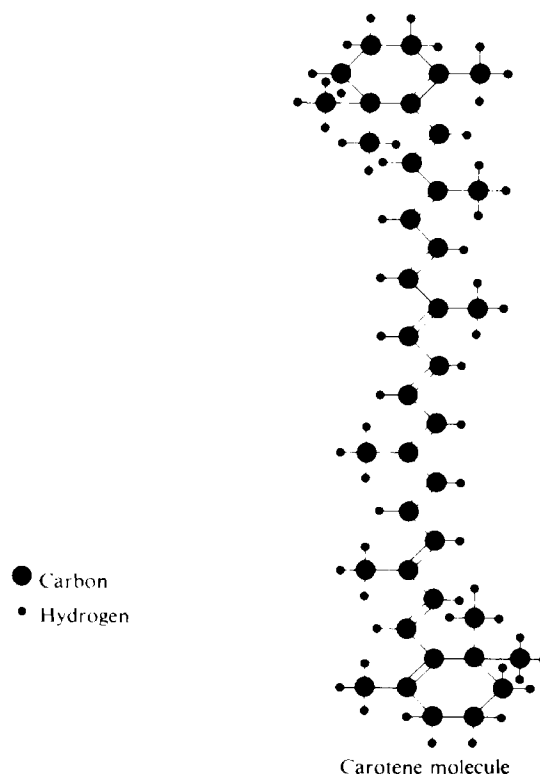


Figure 4.63 The carotene molecule.

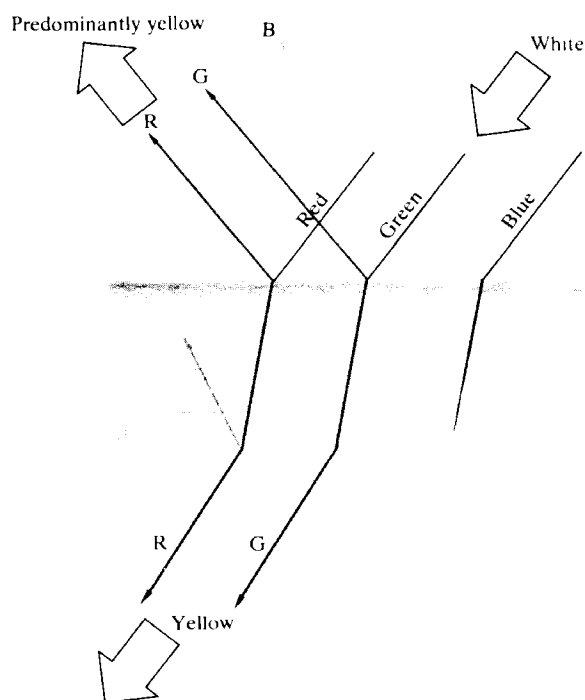


Figure 4.64 Yellow stained glass

The energy levels of an individual atom are precisely defined; that is, the resonances are very sharp. With solids and liquids, however, the proximity of the atoms results in a broadening of the energy levels into wide bands. The resonances spread over a broad range of frequencies. Consequently, we can expect that a dye will not absorb just a narrow portion of the spectrum; indeed if it did, it would reflect most frequencies and appear nearly white.

Imagine a piece of stained glass with a resonance in the blue where it strongly absorbs. If you look through it at a white-light source composed of red, green, and blue, the glass will absorb blue, passing red and green, which is yellow (Fig. 4.64). The glass looks yellow: yellow cloth, paper, dye, paint, and ink all selectively absorb blue. If you peer at something that is a pure blue through a yellow filter, one that passes yellow and absorbs blue, the object will appear black. Here the filter colors the light yellow by removing blue, and we speak of the process as **subtractive coloration**, as opposed to **additive coloration**, which results from overlapping beams of light.

In the same way, fibers of a sample of white cloth or paper

are essentially transparent, but when dyed each fiber behaves as if it were a chip of colored glass. The incident light penetrates the paper, emerging for the most part as a reflected beam only after undergoing numerous reflections and refractions within the dyed fibers. The exiting light will be colored to the extent that it lacks the frequency component absorbed by the dye. This is precisely why a leaf appears green, or a banana yellow.

A bottle of ordinary blue ink looks blue in either reflected or transmitted light. But if the ink is painted on a glass slide and the solvent evaporates, something rather interesting happens. The concentrated pigment absorbs so effectively that it preferentially reflects at the resonant frequency, and we are back to the idea that a strong absorber (larger n_I) is a strong reflector. Thus, concentrated blue-green ink reflects red, whereas red-blue ink reflects green. Try it with a felt marker (overhead projector pens are best), but you must use reflected light, being careful not to inundate the sample with unwanted light from below. The most convenient way to accomplish that is to put colored ink onto a black surface that isn't very absorbant. For example, smear red ink over a black area on a glossy printed page (or better yet, on a black piece of plastic) and it will glow green in reflected light. Gentian violet, which you can buy in any drugstore, works beautifully. Put some on a glass slide and let it dry in a thick coat. Examine both the reflected and transmitted light—they will be complementary.

The whole range of colors (including red, green, and blue) can be produced by passing white light through various combinations of magenta, cyan, and yellow filters (Fig. 4.65). These are the primary colors of subtractive mixing, the primaries of the paint box, although they are often mistakenly spoken of as red, blue, and yellow. They are the basic colors of the dyes used to make photographs and the inks used to print them. Ideally, if you mix all the subtractive primaries together (either by combining paints or by stacking filters), you get no color, no light—black. Each removes a region of the spectrum, and together they absorb it all.

If the range of frequencies being absorbed spreads across the visible, the object will appear black. That is not to say that there is no reflection at all—you obviously can see a reflected image in a piece of black patent leather, and a rough black surface reflects also, only diffusely. If you still have those red and blue inks, mix them, add some green, and you'll get black.

In addition to the above processes specifically related to reflection, refraction, and absorption, there are other color-generating mechanisms, which we shall explore later on. For

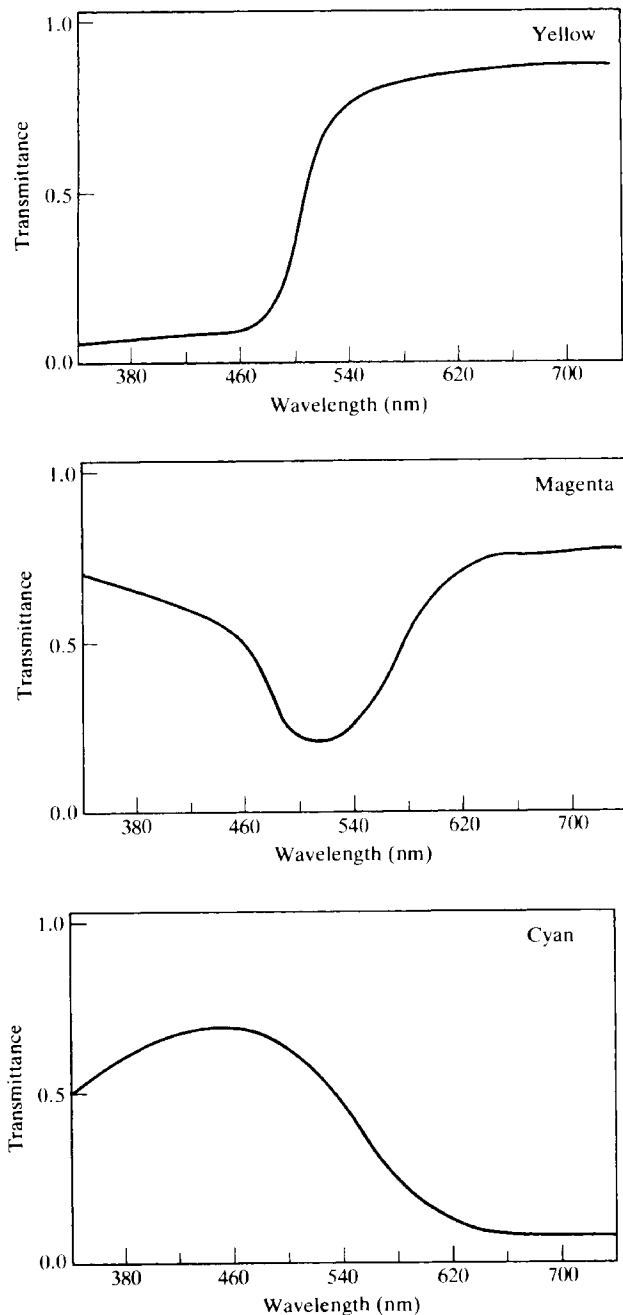


Figure 4.65 Transmission curves for colored filters.

example, scarabaeid beetles mantle themselves in the brilliant colors produced by diffraction gratings on their wing cases, and wavelength-dependent interference effects contribute to the color patterns seen on oil slicks, mother-of-pearl, soap bubbles, peacocks, and hummingbirds.

4.10 The Stokes Treatment of Reflection and Refraction

A rather elegant and novel way of looking at reflection and transmission at a boundary was developed by the British physicist Sir George Gabriel Stokes (1819–1903). Suppose that we have an incident wave of amplitude E_{0i} impinging on the planar interface separating two dielectric media, as in Fig. 4.66a. As we saw earlier in this chapter, because r and t are the fractional amplitudes reflected and transmitted, respectively (where $n_i = n_1$ and $n_t = n_2$), then $E_{0r} = rE_{0i}$ and $E_{0t} = tE_{0i}$. Again we are reminded that Fermat's Principle led to the Principle of Reversibility, which implies that the situation depicted in Fig. 4.66b, where all the ray directions are reversed, must also be physically possible. With the one proviso that there be no energy dissipation (no absorption), a wave's meanderings must be reversible. Equivalently, in the idiom of modern physics one speaks of *time-reversal invariance*, that is, if a process occurs, the reverse process can also occur. Thus if we take a hypothetical motion picture of the wave incident on, reflecting from, and transmitting through the interface, the behavior depicted when the film is run backward must also be physically realizable. Accordingly, examine Fig. 4.66c, where there are now two incident waves of amplitudes $E_{0i}r$ and $E_{0i}t$. A portion of the wave whose amplitude is $E_{0i}t$ is both reflected and transmitted at the interface. Without making any assumptions, let r' and t' be the amplitude reflection and transmission coefficients, respectively, for a wave incident from below (i.e., $n_i = n_2$, $n_t = n_1$). Consequently, the reflected portion is $E_{0i}tr'$, and the transmitted portion is $E_{0i}tt'$. Similarly, the incoming wave whose amplitude is $E_{0i}r$ splits into segments of amplitude $E_{0i}rr$ and $E_{0i}rt$. If the configuration in Fig. 4.66c is to be identical with that in Fig. 4.66b, then obviously

$$E_{0i}tt' + E_{0i}rr = E_{0i} \quad (4.84)$$

$$\text{and} \quad E_{0i}rt + E_{0i}tr' = 0 \quad (4.85)$$

$$\text{Hence} \quad tt' = 1 - r^2 \quad (4.86)$$

$$\text{and} \quad r' = -r \quad (4.87)$$

the latter two equations being known as the Stokes Relations. This discussion calls for a bit more caution than is usually granted it. It must be pointed out that *the amplitude coefficients are functions of the incident angles*, and therefore the

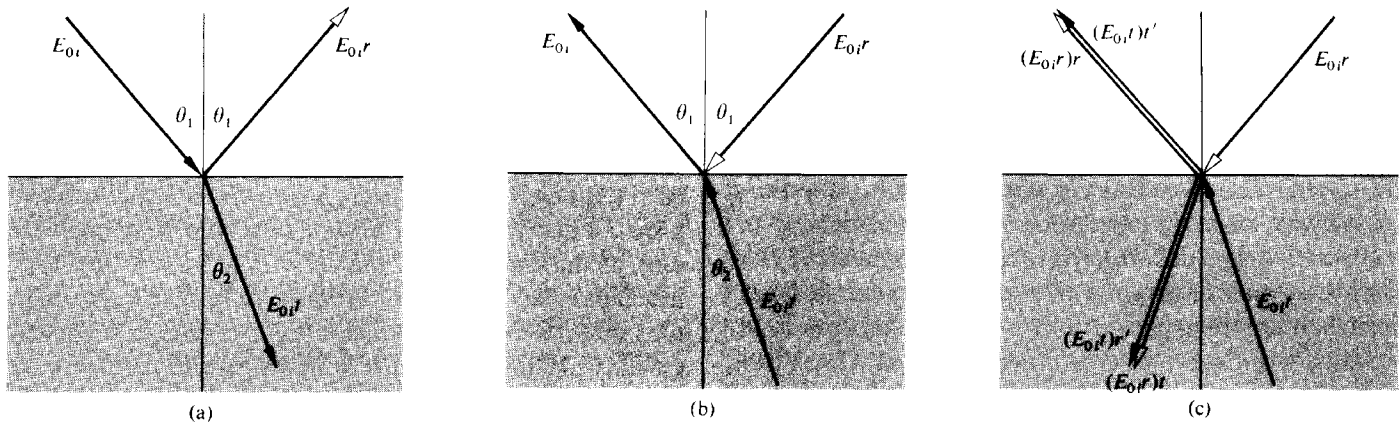


Figure 4.66 Reflection and refraction via the Stokes treatment.

Stokes Relations might better be written as

$$t(\theta_i)t'(\theta_2) = 1 - r^2(\theta_i) \quad (4.88)$$

and
$$r'(\theta_2) = -r(\theta_i) \quad (4.89)$$

where $n_1 \sin \theta_1 = n_2 \sin \theta_2$. The second equation indicates, by virtue of the minus sign, that *there is a 180° phase difference between the waves internally and externally reflected*. It is most important to keep in mind that here θ_i and θ_2 are pairs of angles that are related by way of Snell's Law. Note as well that we never did say whether n_1 was greater or less than n_2 , so Eqs. (4.88) and (4.89) apply in either case. Let's return for a moment to one of the Fresnel Equations:

$$r_\perp = -\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \quad [4.42]$$

If a ray enters from above, as in Fig. 4.66a, and we assume $n_2 > n_1$, $r_\perp$ is computed by setting $\theta_i = \theta_1$ and $\theta_t = \theta_2$ (external reflection), the latter being derived from Snell's Law. If, on the other hand, the wave is incident at that same angle from below (in this instance internal reflection), $\theta_i = \theta_1$ and we again substitute in Eq. (4.42), but here θ_t is not θ_2 , as before. The values of $r_\perp$ for internal and external reflection *at the same incident angle* are obviously different. Now suppose, in this case of internal reflection, that $\theta_i = \theta_2$. Then $\theta_t = \theta_1$, the ray directions are the reverse of those in the first situation, and Eq. (4.42) yields

$$r'_\perp(\theta_2) = -\frac{\sin(\theta_2 - \theta_1)}{\sin(\theta_2 + \theta_1)}$$

Although it may be unnecessary we once again point out that

this is just the negative of what was determined for $\theta_i = \theta_1$ and external reflection, that is,

$$r'_\perp(\theta_2) = -r_\perp(\theta_1) \quad (4.90)$$

The use of primed and unprimed symbols to denote the amplitude coefficients should serve as a reminder that we are once more dealing with angles related by Snell's Law. In the same way, interchanging θ_i and θ_t in Eq. (4.43) leads to

$$r'_\parallel(\theta_2) = -r_\parallel(\theta_1) \quad (4.91)$$

The 180° phase difference between each pair of components is evident in Fig. 4.44, but keep in mind that when $\theta_i = \theta_p$, $\theta_t = \theta'_p$ and vice versa (Problem 4.78). Beyond $\theta_i = \theta_c$ there is no transmitted wave, Eq. (4.89) is not applicable, and as we have seen, the phase difference is no longer 180° .

It is common to conclude that both the parallel and perpendicular components of the externally reflected beam change phase by π radians, while the internally reflected beam undergoes no phase shift at all. This is quite incorrect (compare Figs. 4.45a and 4.46a).

4.11 Photons, Waves, and Probability

Much of the theoretical grounding of Optics is predicated on wave theory. We take for granted both that we understand the phenomenon and that it's "real." As one example out of the many that will be encountered, the process of scattering seems

to be understandable only in terms of interference; classical particles simply do not interfere. When a beam propagates through a dense medium, interference in the forward direction is constructive, whereas in all other directions it's almost completely destructive. Thus nearly all the light energy advances in the forward direction. But this raises interesting questions about the basic nature of interference and the usual interpretation of what's happening. ***Interference is a nonlocalized phenomenon; it cannot happen at only one single point in space,*** even though we often talk about the interference at a point P . The principle of Conservation of Energy makes it clear that if there is constructive interference at one point, the "extra" energy at that location must have come from elsewhere. There must therefore be destructive interference somewhere else. ***Interference takes place over an extended region of space in a coordinated fashion that leaves the total amount of radiant energy unchanged.***

Now imagine a light beam traversing a dense medium, as in Fig. 4.6. Do real energy-carrying electromagnetic wavelets (which are never actually measured) propagate out laterally, only to interfere destructively everywhere beyond the beam? If so, these wavelets cancel and the energy they transport outward is inexplicably returned to the beam since, in the end, there is no net lateral scattering. That's true no matter how far away P is. Moreover, this applies to *all* interference effects (Chapter 9). If two or more electromagnetic waves arrive at point P out-of-phase and cancel, "What does that mean as far as their energy is concerned?" Energy can be redistributed, but *it* doesn't cancel out. We've learned from Quantum Mechanics that at base interference is one of the most fundamental mysteries in physics.

Remembering Einstein's admonition that there are no spherical wavelets emitted by atoms, perhaps we're being too literal in our interpretation of the classical wave field. After all, strictly speaking, the classical electromagnetic wave with its continuous distribution of energy does not actually exist. Perhaps we should think of the wavelets and the overall pattern they produce (rather than being a *real* wave field) as a theoretical device that, wonderfully enough, tells us where the light will end up. In any event, Maxwell's Equations provide a means of calculating the macroscopic distribution of electromagnetic energy in space.

Moving ahead in a semiclassical way, imagine a distribution of light given by some function of the off-axis angle θ . For example, consider the irradiance on a screen placed far beyond a slit-shaped aperture (p. 453) such that $I(\theta) =$

$I(0) \operatorname{sinc}^2 \beta(\theta)$. Suppose that instead of observing the pattern by eye a detector composed of a diaphragm followed by a photomultiplier tube is used. Such a device could be moved around from one point to another, and over a constant time interval, it could measure the number of photons arriving at each location, $N(\theta)$. Taking a great many such measurements, a spatial distribution of the number of photon counts would emerge that would be of the very same form as that for the irradiance, namely, $N(\theta) = N(0) \operatorname{sinc}^2 \beta(\theta)$: the number of photons detected is proportional to the irradiance. A countable quantity like this lends itself to statistical analysis, and we can talk about the probability of detecting a photon at any point on the screen. That is, a probability distribution can be constructed, reminiscent of Fig. 3.23. Because the space variables (θ , x , y , or z) are continuous, it's necessary to introduce a **probability density**, let it be $\varphi(\theta)$. Then $\varphi(\theta) d\theta$ is the probability that a photon will be found in the infinitesimal range from θ to $\theta + d\theta$. In this case $\varphi(\theta) = \varphi(0) \operatorname{sinc}^2 \beta(\theta)$.

The square of the net electric field amplitude at every point in space corresponds to the irradiance (which can be measured directly), and that's equivalent to the likelihood of finding photons at any point. Accordingly, let's tentatively define the **probability amplitude** as that quantity whose absolute value squared equals the probability density. Thus the net E_0 at P can be interpreted as being proportional to a **semiclassical probability amplitude** inasmuch as ***the probability of detecting a photon at some point in space depends on the irradiance at that location and $I \propto E_0^2$*** . This is in accord with Einstein's conception of the light field, which Max Born (who initiated the statistical interpretation of Quantum Mechanics) described as a *Gespensterfeld*, or phantom field. In that view the waves of that field reveal how the photons distribute in space in the sense that the square of the absolute value of the wave amplitude somehow relates to the probability density of arriving photons. In the formal treatment of Quantum Mechanics, *the probability amplitude is generally a complex quantity* whose absolute value squared corresponds to the probability density (e.g., the Schrödinger wavefunction is a probability amplitude). Thus, however reasonable it was to consider E_0 as equivalent to a semiclassical probability amplitude, that usage cannot be carried over, as is, into quantum theory.

Still, all of this suggests that we might take the scattering process, considered in terms of probabilities, as the basis for a computational scheme. Each scattered wavelet is then a measure of the probability amplitude for light taking a particular route from one point to another, and the net electric field at P

is the sum of all the scattered fields arriving via all possible routes. A quantum-mechanical methodology analogous to this was devised by Feynman, Schwinger, Tomonaga, and Dyson in the course of their development of Quantum Electrodynamics. In brief, the final observable outcome of an event is determined by the superposition of all the various probability amplitudes associated with each and every possible way that the event can occur. In other words, each “route” along which an event can take place, each way it can happen, is given an abstract mathematical representation, a complex probability amplitude. All of these then combine—and interfere, as complex quantities are wont to do—to produce a net probability amplitude for the event to take place.

What follows is a greatly simplified version of that analysis.

4.11.1 QED

Feynman was rather unequivocal in his stance regarding the nature of light:

I want to emphasize that light comes in this form—particles. It is very important to know that light behaves like particles, especially for those of you who have gone to school, where you were probably told something about light behaving like waves. I’m telling you the way it *does* behave—like particles.

For him “light is made of particles (as Newton originally thought)”; it’s a stream of photons whose behavior *en masse* can be determined statistically. For example, if 100 photons are incident perpendicularly on a piece of glass in air, on average 4 will be reflected backward from the first surface they encounter. Which 4 cannot be known, and in fact how those particular 4 photons are selected is a mystery. What can be deduced and confirmed experimentally is that 4% of the incident light will be reflected (p. 121).

Feynman’s analysis proceeds from a few general computational rules, with the ultimate justification being that it works; the scheme makes accurate predictions. (1) *The probability amplitude associated with the occurrence of an event is the “sum” of the constituent probability amplitudes corresponding to each and every possible way the event can occur.* (2) *Each such constituent probability amplitude is generally expressible as a complex quantity.* Rather than analytically combining these constituent probability amplitudes, we can use the phasor representation (p. 23) to approximate the sum-

mation and thereby arrive at a resultant probability amplitude. (3) *The probability of occurrence of the event as a whole is proportional to the absolute square of the resultant probability amplitude.*

We can appreciate how all of this comes together by treating the reflection pictured in Fig. 4.67: a point source S illuminates a mirror, and light is subsequently scattered upward in every direction from every point on the mirror. We wish to determine the probability of a detector at P , recording the arrival of a photon. Here the classical perspective, with its familiar wavelet model, can be used as an analogue to provide guidance (and perhaps a little intellectual comfort, if you still believe in classical EM waves).

For simplicity, take the mirror to be a narrow strip (which is essentially one-dimensional); that doesn’t change things conceptually. Divide it into a number of equal-sized lengths

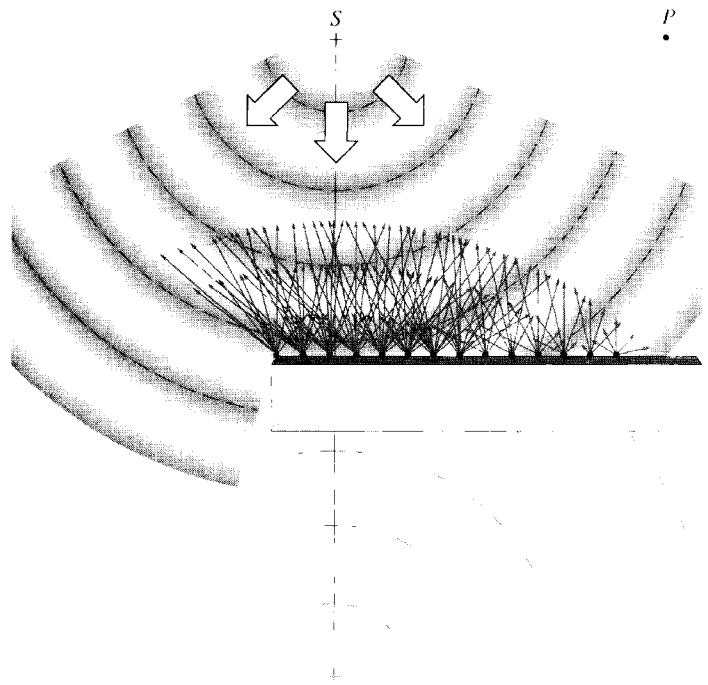


Figure 4.67 A schematic representation of reflection. A wave from S sweeps down and spreads across the surface of the mirror. Every atom on the interface subsequently scatters light back in all upward directions. And some of it ultimately arrives at P having come from every scatterer on the surface.

(Fig. 4.68a), each of which establishes a possible path to P . (Of course, every atom on the surface is a scatterer, and so there are a multitude of paths, but the several we have drawn will do.) Classically, we know that every route from S to the mirror to P corresponds to the path of a scattered wavelet, and that the amplitude (E_{0j}) and phase of each such wavelet at P will determine the net resultant amplitude, E_0 . As we saw with Fermat's Principle (p. 106), the optical path length from S to the mirror to P establishes the phase of each wavelet arriving at P . Moreover, the greater the *path length* is, the more the

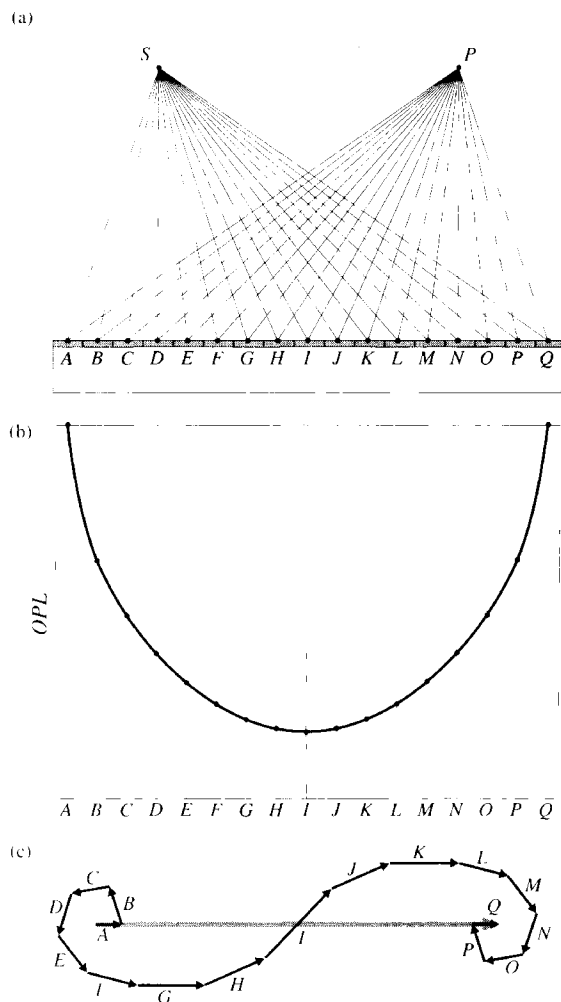


Figure 4.68 (a) Feynman's analysis of the problem of reflection via QED. (a) A number of paths from S to the mirror to P . (b) The OPL for light going from S to P along the paths depicted in (a). Each path has a probability amplitude associated with it. These add to produce a net amplitude.

light spreads out (via the Inverse Square Law) and the smaller is the amplitude of the wavelet arriving at P .

Figure 4.68b is a plot of the OPL with its minimum at the observed path ($S-I-P$), for which $\theta_i = \theta_r$. A large change in OPL , as between ($S-A-P$) and ($S-B-P$) is accompanied by a large phase difference and a correspondingly large rotation of the phasors drawn in Fig. 4.68c. Going from A to B to C and so on to I , the optical path lengths decrease less and less rapidly, and each phasor leads the previous one by a smaller angle (set by the slope of the curve). In effect, the phasors to the left of I rotate counterclockwise from A to I . Since the OPL is a minimum at I , the phasors from that region are large and differ very little in phase angle. Going from I to J to K and so on to Q , the optical path lengths increase more and more rapidly, and each phasor lags the previous one by a larger angle. In effect, the phasors to the right of I rotate clockwise from I to Q .

In Fig. 4.68c the resultant amplitude is drawn from the starting tail to the ending tip, and classically it corresponds to the net electric field amplitude at P . The irradiance, I , is proportional to the square of the net field amplitude, and that, in turn, should be a measure of the likelihood of finding a photon when a detector is placed at P .

Let's move beyond the classical ideas of scattered wavelets and electric fields (nonetheless being guided by them) and construct a quantum-mechanical treatment. Photons can go from S to the mirror to P along each of an innumerable number of distinct paths. It's reasonable to assume that each such path makes a specific contribution to the end result; an exceedingly long route out to the very edge of the mirror and back to P should contribute differently than a more direct route. Following Feynman, we associate some (as yet unspecified) complex quantity, a constituent **quantum-mechanical (QM) probability amplitude**, with each possible path. Each such constituent QM probability amplitude can be represented as a phasor whose angle is determined by the total time of flight from S to the mirror to P , and whose size is determined by the path length traversed. (Of course, this is just what obtained with each phasor in Fig. 4.68c. Still there are convincing reasons why the classical E -field cannot be the QM probability amplitude.) The total QM probability amplitude is the sum of all such phasors corresponding to all possible paths, and that is analogous to the resultant phasor in Fig. 4.68c.

Now re-label Fig. 4.68c so that it represents the quantum-mechanical formulation. Clearly, *most of the length of the resultant QM probability amplitude arises from contributions in the immediate vicinity of path $S-I-P$, where the constituent phasors are large and nearly in-phase*. Most of the

accumulated probability for light to go from S to P via reflection arises along, and immediately adjacent to, path $S-I-P$. The regions at the ends of the mirror contribute very little because the phasors from those areas form tight spirals at both extremes (Fig. 4.68c). Covering the ends of the mirror will have little effect on the length of the resultant amplitude and therefore little effect on the amount of light reaching P . Keep in mind that this diagram is rather crude; instead of 17 routes from S to P , there are billions of possible paths, and the phasors on both ends of the spiral wind around countless times.

QED predicts that light emitted by a point source S reflects out to P from all across the mirror, but that the most likely route is $S-I-P$, in which case $\theta_i = \theta_r$. With your eye at P looking into the mirror, you see one sharp image of S .

4.11.2 Photons and the Laws of Reflection and Refraction

Suppose that light consists of a stream of photons and consider one such quantum that strikes the interface between two dielectric media (e.g., air and glass) at an angle θ_i . That photon is absorbed by an atom (e.g., in the glass), and an identical photon is subsequently transmitted at an angle θ_r . We know that if this were just one out of billions of such quanta in a narrow laserbeam, it would conform to Snell's Law. To explore this behavior, let's examine the dynamics associated with the odyssey of our single photon. Recall Eq. (3.54), namely, $p = h/\lambda$ and so its vector momentum would be

$$\vec{p} = \hbar \vec{k}$$

where $\vec{k}$ is the propagation vector and $\hbar \equiv h/2\pi$. Consequently, the incident and transmitted momenta are $\vec{p}_i = \hbar \vec{k}_i$ and $\vec{p}_r = \hbar \vec{k}_r$, respectively. We assume (without much justification) that

while the material in the vicinity of the interface affects the component of momentum perpendicular to the interface it leaves the parallel component unchanged. Indeed, we know experimentally that linear momentum perpendicular to the interface can be transferred to a medium from a light beam (Section 3.3.4). The statement of conservation of the component of momentum parallel to the interface for a single photon takes the form

$$p_i \sin \theta_i = p_r \sin \theta_r$$

We are at an important juncture here; classically the momentum of a material particle depends on its speed. When $n_r > n_i$ it follows (from the Snell's Law and the above equation) that $p_r > p_i$, and the particles of light must presumably speed up. Indeed, the first derivation of the Law of Refraction, published by René Descartes (1637), wrongly treated light as a stream of particles that sped up as it entered the optically more dense medium (see Problem 4.10). By contrast, the first person to measure the shortening of the wavelength of light as it entered a more dense medium was probably Thomas Young (≈ 1802).^{*} He rightly inferred that the speed of a light beam was thereby actually reduced: $v < c$.

We now know from Quantum Mechanics that the speed of a photon is always c and that its momentum depends on its wavelength and not its speed. Thus

$$\frac{h}{\lambda_i} \sin \theta_i = \frac{h}{\lambda_r} \sin \theta_r$$

Multiplying both sides by c/ν , we get Snell's Law.

Do keep in mind that the above analysis is a bit simplistic, but it is appealing pedagogically.

^{*}Foucault's definitive experiments proving the point were done in 1850.

PROBLEMS

Complete solutions to all problems—except those with an asterisk—can be found in the back of the book.

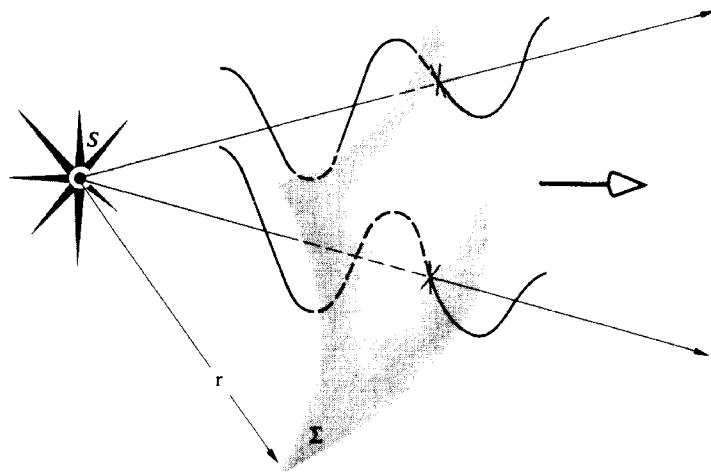
4.1 Work your way through an argument using dimensional analysis to establish the λ^{-4} dependence of the percentage of light scattered in Rayleigh Scattering. Let E_{0i} and E_{0s} be the incident and scattered amplitudes, the latter at a distance r from the scatterer. Assume $E_{0s} \propto E_{0i}$ and $E_{0s} \propto 1/r$. Furthermore, plausibly assume that the scattered

amplitude is proportional to the volume, V , of the scatterer; within limits this is reasonable. Determine the units of the constant of proportionality.

4.2* A white floodlight beam crosses a large volume containing a tenuous molecular gas mixture of mostly oxygen and nitrogen. Compare the relative amount of scattering occurring for the yellow (580 nm) component with that of the violet (400 nm) component.

4.3* Figure P.4.3 depicts light emerging from a point source. It shows three different representations of radiant energy streaming outward. Identify each one and discuss its relationship to the others.

Figure P.4.3 A segment of a spherical wave.



4.4 The equation for a driven damped oscillator is

$$m_e \ddot{x} + m_e \gamma \dot{x} + m_e \omega_0^2 x = q_e E(t)$$

- (a) Explain the significance of each term.
 (b) Let $E = E_0 e^{i\omega t}$ and $x = x_0 e^{i(\omega t - \alpha)}$, where E_0 and x_0 are real quantities. Substitute into the above expression and show that

$$x_0 = \frac{q_e E_0}{m_e} \frac{1}{[(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2]^{1/2}}$$

- (c) Derive an expression for the phase lag, α , and discuss how α varies as ω goes from $\omega \ll \omega_0$ to $\omega = \omega_0$ to $\omega \gg \omega_0$.

4.5 Imagine that we have a nonabsorbing glass plate of index n and thickness Δy , which stands between a source S and an observer P .

- (a) If the unobstructed wave (without the plate present) is $E_u = E_0 \exp i\omega(t - y/c)$, show that with the plate in place the observer sees a wave

$$E_p = E_0 \exp i\omega[t - (n - 1) \Delta y/c - y/c]$$

- (b) Show that if either $n \approx 1$ or Δy is very small, then

$$E_p = E_u + \frac{\omega(n - 1) \Delta y}{c} E_u e^{-i\pi/2}$$

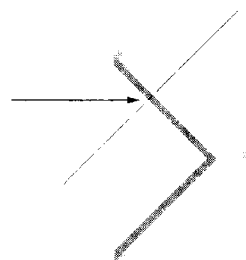
The second term on the right may be envisioned as the field arising from the oscillators in the plate.

4.6* A very narrow laserbeam is incident at an angle of 58° on a horizontal mirror. The reflected beam strikes a wall at a spot 5.0 m away from the point of incidence where the beam hit the mirror. How far horizontally is the wall from that point of incidence?

4.7* On entering the tomb of FRED the Hero of Nod, you find yourself in a dark closed chamber with a small hole in a wall 3.0 m up from the floor. Once a year, on FRED's birthday, a beam of sunlight enters via the hole, strikes a small polished gold disk on the floor 4.0 m from the wall and reflects off it, lighting up a great diamond embedded in the forehead of a glorious statue of FRED, 20 m from the wall. Roughly how tall is the statue?

4.8* Figure P.4.8 shows what's called a corner mirror. Determine the direction of the exiting ray with respect to the incident ray.

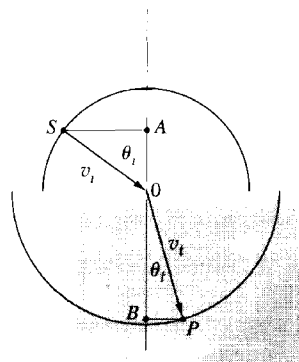
Figure P.4.8



4.9 Calculate the transmission angle for a ray incident in air at 30° on a block of crown glass ($n_g = 1.52$).

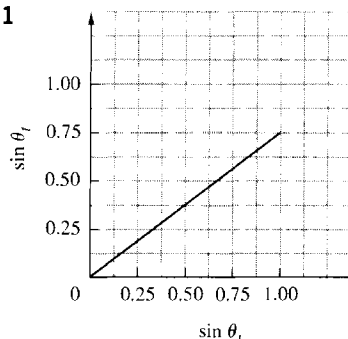
4.10* The construction in Fig. P.4.10 corresponds to Descartes's derivation of the Law of Refraction. Light moves from S to O in the same time it travels from O to P . Moreover, its transverse momentum is unchanged on traversing the interface. Use all of this to "derive" Snell's Law.

Figure P.4.10



4.11* Figure P.4.11 is a plot of the sine of the angle-of-incidence versus the sine of the transmission angle measured as light passed from air into a more optically dense medium. Discuss the curve. What is the significance of the slope of the line? Guess at what the dense medium might be.

Figure P.4.11



4.12* A ray of yellow light from a sodium discharge lamp falls on the surface of a diamond in air at 45° . If at that frequency $n_d = 2.42$, compute the angular deviation suffered upon transmission.

4.13* Given an interface between water ($n_w = \frac{4}{3}$) and glass ($n_g = \frac{3}{2}$), compute the transmission angle for a beam incident in the water at 45° . If the transmitted beam is reversed so that it impinges on the interface, show that $\theta_t = 45^\circ$.

4.14 A beam of 12-cm planar microwaves strikes the surface of a dielectric at 45° . If $n_{di} = \frac{4}{3}$, compute (a) the wavelength in the transmitting medium, and (b) the angle θ_t .

4.15* Light of wavelength 600 nm in vacuum enters a block of glass where $n_g = 1.5$. Compute its wavelength in the glass. What color would it appear to someone embedded in the glass (see Table 3.4)?

4.16* A laserbeam impinges on an air–liquid interface at an angle of 55° . The refracted ray is observed to be transmitted at 40° . What is the refractive index of the liquid?

4.17* An underwater swimmer shines a beam of light up toward the surface. It strikes the air–water interface at 35° . At what angle will it emerge into the air?

4.18 Make a plot of θ_t versus θ_i for an air–glass boundary where $n_{ga} = 1.5$. Discuss the shape of the curve.

4.19* A laserbeam having a diameter D in air strikes a piece of glass (n_g) at an angle θ_i . What is the diameter of the beam in the glass?

4.20* An exceedingly narrow beam of white light is incident at 60.0° on a sheet of glass 10.0 cm thick in air. The index of refraction

for red light is 1.505 and for violet light it's 1.545. Determine the approximate diameter of the emerging beam.

4.21* Prove that to someone looking straight down into a swimming pool, any object in the water will appear to be at $\frac{3}{4}$ of its true depth.

4.22* A laserbeam impinges on the top surface of a 2.00-cm-thick parallel glass ($n = 1.50$) plate at an angle of 35° . How long is the actual path through the glass?

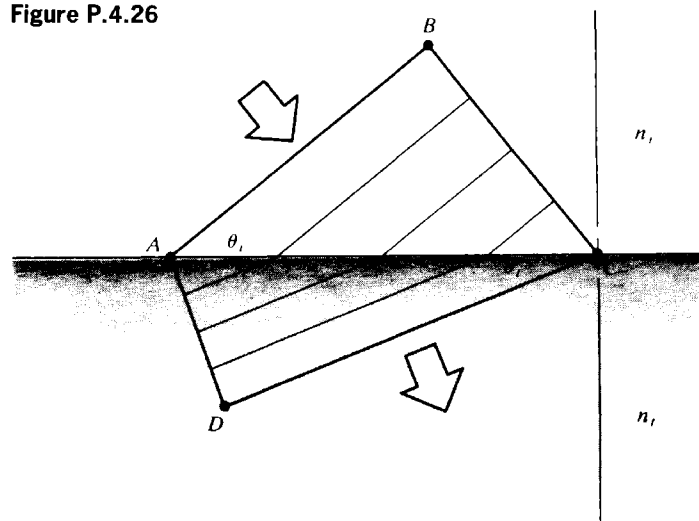
4.23* Light is incident in the air on an air–glass interface. If the index of refraction of the glass is 1.70, find the incident angle such that the transmission angle is to equal $\frac{1}{2}\theta_i$.

4.24* Suppose that you focus a camera with a close-up bellows attachment directly down on a letter printed on this page. The letter is then covered with a 1.00-mm-thick microscope slide ($n = 1.55$). How high must the camera be raised in order to keep the letter in focus?

4.25* A coin is resting on the bottom of a tank of water ($n_w = 1.33$) 1.00 m deep. On top of the water floats a layer of benzene ($n_b = 1.50$), which is 20.0 cm thick. Looking down nearly perpendicularly, how far beneath the topmost surface does the coin appear? Draw a ray diagram.

4.26 In Fig. P.4.26 the wavefronts in the incident medium match the fronts in the transmitting medium everywhere on the interface—a concept known as *wavefront continuity*. Write expressions for the number of waves per unit length along the interface in terms of θ_i and λ_i in one case and θ_t and λ_t in the other. Use these to derive Snell's Law. Do you think Snell's Law applies to sound waves? Explain.

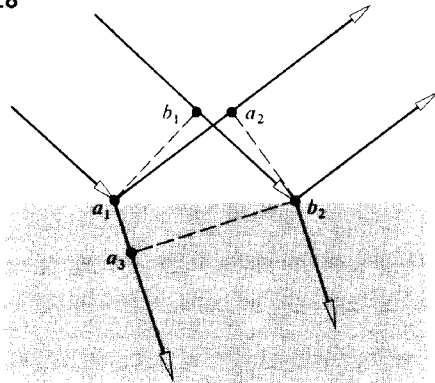
Figure P.4.26



4.27* With the previous problem in mind, return to Eq. (4.19) and take the origin of the coordinate system in the plane-of-incidence and on the interface (Fig. 4.39). Show that that equation is then equivalent to equating the x -components of the various propagation vectors. Show that it is also equivalent to the notion of wavefront continuity.

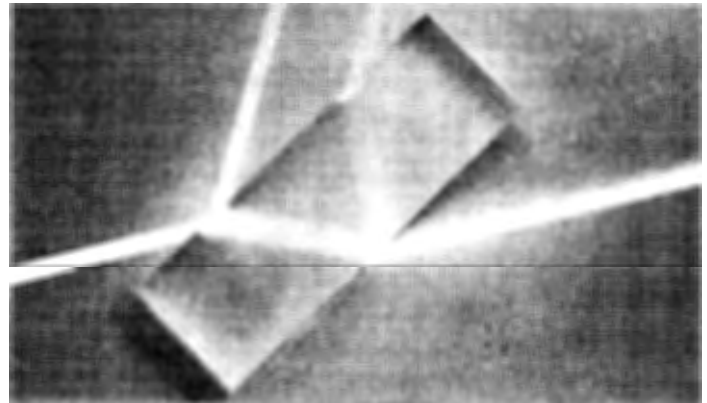
4.28 Making use of the ideas of equal transit times between corresponding points and the orthogonality of rays and wavefronts, derive the law of reflection and Snell's Law. The ray diagram of Fig. P.4.28 should be helpful.

Figure P.4.28



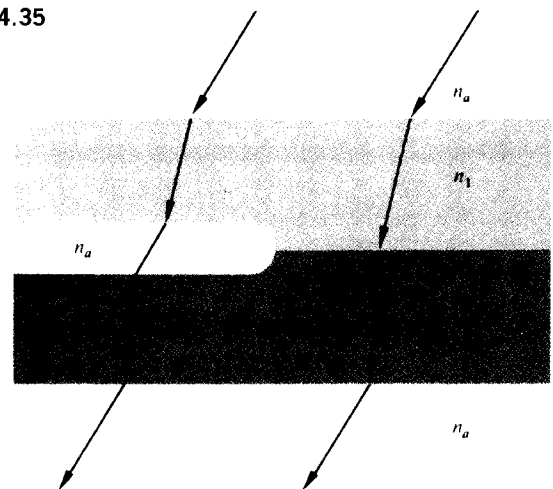
an expression for the lateral displacement of the beam. Incidentally, the incoming and outgoing rays would be parallel even for a stack of plates of different material.

Figure P.4.34



4.35* Show that the two rays that enter the system in Fig. P.4.35 parallel to each other emerge from it being parallel.

Figure P.4.35



4.29 Starting with Snell's Law, prove that the vector refraction equation has the form

$$n_t \hat{\mathbf{k}}_t - n_i \hat{\mathbf{k}}_i = (n_t \cos \theta_t - n_i \cos \theta_i) \hat{\mathbf{u}}_n \quad [4.7]$$

4.30 Derive a vector expression equivalent to the Law of Reflection. As before, let the normal go from the incident to the transmitting medium, even though it obviously doesn't really matter.

4.31 In the case of reflection from a planar surface, use Fermat's Principle to prove that the incident and reflected rays share a common plane with the normal $\hat{\mathbf{u}}_n$, namely, the plane-of-incidence.

4.32* Derive the Law of Reflection, $\theta_i = \theta_r$, by using the calculus to minimize the transit time, as required by Fermat's Principle.

4.33* According to the mathematician Hermann Schwarz, there is one triangle that can be inscribed within an acute triangle such that it has a minimal perimeter. Using two planar mirrors, a laserbeam, and Fermat's Principle, explain how you can show that this inscribed triangle has its vertices at the points where the altitudes of the acute triangle intersect its corresponding sides.

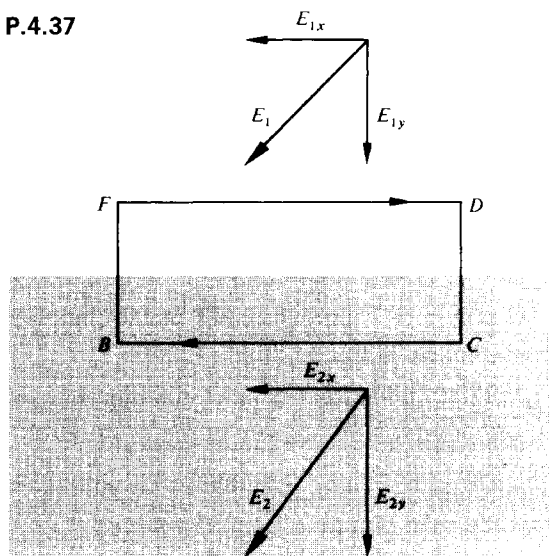
4.34 Show analytically that a beam entering a planar transparent plate, as in Fig. P.4.34, emerges parallel to its initial direction. Derive

4.36 Discuss the results of Problem 4.34 in the light of Fermat's Principle: that is, how does the relative index n_{21} affect things? To see the lateral displacement, look at a broad source through a thick piece of glass ($\approx \frac{1}{4}$ inch) or a stack (four will do) of microscope slides *held at an angle*. There will be an obvious shift between the region of the source seen directly and the region viewed through the glass.

4.37 Show that even in the nonstatic case the tangential component of the electric field intensity $\mathbf{E}$ is continuous across an interface.

[Hint: Using Fig. P.4.37 and Eq. (3.5), shrink sides FB and CD , thereby letting the area bounded go to zero.]

Figure P.4.37



4.38 Suppose a lightwave that is linearly polarized in the plane-of-incidence impinges at 30° on a crown-glass ($n_g = 1.52$) plate in air. Compute the appropriate amplitude reflection and transmission coefficients at the interface. Compare your results with Fig. 4.39.

4.39 Derive Eqs. (4.42) through (4.45) for $r_\perp$, $r_\parallel$, $t_\perp$, and $t_\parallel$.

4.40* A beam of light in air strikes the surface of a smooth piece of plastic having an index of refraction of 1.55 at an angle with the normal of 20.0° . The incident light has component E -field amplitudes parallel and perpendicular to the plane-of-incidence of 10.0 V/m and 20.0 V/m, respectively. Determine the corresponding reflected field amplitudes.

4.41* A laserbeam is incident on the interface between air and some dielectric of index n . For small values of θ_i show that $\theta_t = \theta_i/n$. Use this and Eq. (4.42) to establish that at near-normal incidence $[-r_\perp]_{\theta_i \approx 0} = (n-1)/(n+1)$.

4.42* Light is incident in air perpendicularly on a sheet of crown glass having an index of refraction of 1.522. Determine both the reflectance and the transmittance.

4.43* A beam of quasimonochromatic light having an irradiance of 500 W/m^2 is incident in air perpendicularly on the surface of a tank of water ($n_w = 1.333$). Determine the transmitted irradiance.

4.44* Quasimonochromatic light having an irradiance of 400 W/m^2 is incident normally on the cornea ($n_c = 1.376$) of the human eye. If the person is swimming under the water ($n_w = 1.33$), determine the transmitted irradiance into the cornea.

4.45* Compare the amplitude reflection coefficients for an air–water ($n_w = 4/3$) interface with that of an air–crown glass ($n_g = 3/2$) interface, both at near-normal incidence. What are the corresponding ratios of the reflected to the incident irradiances?

4.46* Use Eq. (4.42) and the power series expansion of the sine function to establish that at near-normal incidence we can obtain a better approximation than the one in Problem 4.41, which is $[-r_\perp]_{\theta_i \approx 0} = (n-1)/(n+1)$, namely

$$[-r_\perp]_{\theta_i \approx 0} = \left(\frac{n-1}{n+1} \right) \left(1 + \frac{\theta_i^2}{n} \right)$$

4.47* Establish that at near-normal incidence the equation

$$[r_\parallel]_{\theta_i \approx 0} = \left(\frac{n-1}{n+1} \right) \left(1 - \frac{\theta_i^2}{n} \right)$$

is a good approximation. [Hint: Use the results of the previous problem, Eq. (4.43), and the power series expansions of the sine and cosine functions.]

4.48* Prove that for a vacuum–dielectric interface at glancing incidence $r_\perp \rightarrow -1$, as in Fig. 4.41.

4.49* In Fig. 4.41 the curve of $r_\perp$ approaches -1.0 as the angle-of-incidence approaches 90° . Prove that if α_i is the angle the curve makes with the vertical at $\theta_i = 90^\circ$, then

$$\tan \alpha_\perp = \frac{\sqrt{n^2 - 1}}{2}$$

[Hint: First show that $d\theta_i/d\theta_i = 0$.]

4.50 Prove that

$$t_\perp + (-r_\perp) = 1 \quad [4.49]$$

for all θ_i , first from the boundary conditions and then from the Fresnel Equations.

4.51* Verify that

$$t_\perp + (-r_\perp) = 1 \quad [4.49]$$

for $\theta_i = 30^\circ$ at a crown–glass and air interface ($n_g = 1.52$).

4.52* Calculate the critical angle beyond which there is total internal reflection at an air–glass ($n_g = 1.5$) interface. Compare this result with that of Problem 4.15.

4.53* Referring back to Problem 4.18, note that as θ_i increases θ_r increases. Prove that the maximum value θ_r may have is θ_c .

4.54* What is the critical angle for total internal reflection for diamond? What, if anything, does the critical angle have to do with the luster of a well-cut diamond?

4.55* Using a block of a transparent, unknown material, it is found that a beam of light inside the material is totally internally reflected at the air–block interface at an angle of 48.0° . What is its index of refraction?

4.56* A prism, ABC , is configured such that angle $BCA = 90^\circ$ and angle $CBA = 45^\circ$. What is the minimum value of its index of refraction if, while immersed in air, a beam traversing face AC is to be totally internally reflected from face BC ?

4.57* A fish looking straight up toward the smooth surface of a pond receives a cone of rays and sees a circle of light filled with the images of sky and birds and whatever else is up there. This bright circular field is surrounded by darkness. Explain what is happening and compute the cone-angle.

4.58* A glass block having an index of 1.55 is covered with a layer of water of index 1.33. For light traveling in the glass, what is the critical angle at the interface?

4.59 Derive an expression for the speed of the evanescent wave in the case of internal reflection. Write it in terms of c , n_i , and θ_i .

4.60 Light having a vacuum wavelength of 600 nm, traveling in a

glass ($n_g = 1.50$) block, is incident at 45° on a glass–air interface. It is then totally internally reflected. Determine the distance into the air at which the amplitude of the evanescent wave has dropped to a value of $1/e$ of its maximum value at the interface.

4.61 Figure P.4.61 shows a laserbeam incident on a wet piece of filter paper atop a sheet of glass whose index of refraction is to be measured—the photograph shows the resulting light pattern. Explain what is happening and derive an expression for n_i in terms of R and d .

4.62 Consider the common mirage associated with an inhomogeneous distribution of air situated above a warm roadway. Envision the bending of the rays as if it were instead a problem in total internal reflection. If an observer, at whose head $n_a = 1.00029$, sees an apparent wet spot at $\theta_i \geq 88.7^\circ$ down the road, find the index of the air immediately above the road.

4.63* Use the Fresnel Equations to prove that light incident at $\theta_p = \frac{1}{2}\pi - \theta_i$ results in a reflected beam that is indeed polarized.

4.64 Show that $\tan \theta_p = n_t/n_i$ and calculate the polarization angle for external incidence on a plate of crown glass ($n_g = 1.52$) in air.

4.65* Beginning with Eq. (4.38), show that for two dielectric media, in general $\tan \theta_p = [\epsilon_i(\epsilon_i\mu_t - \epsilon_t\mu_i)/\epsilon_t(\epsilon_i\mu_t - \epsilon_i\mu_i)]^{1/2}$.

4.66 Show that the polarization angles for internal and external reflection at a given interface are complementary, that is, $\theta_p + \theta'_p = 90^\circ$ (see Problem 4.64).

4.67 It is often useful to work with the *azimuthal angle* γ , which is

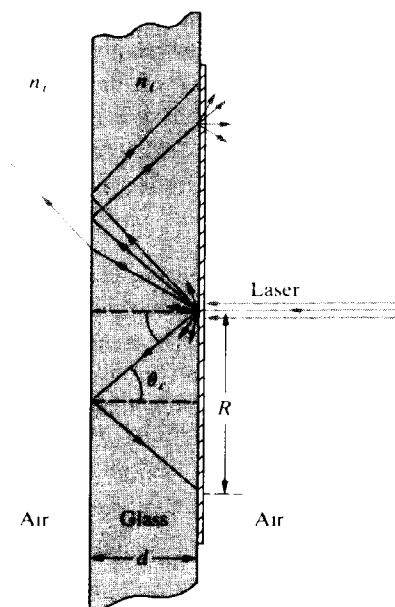
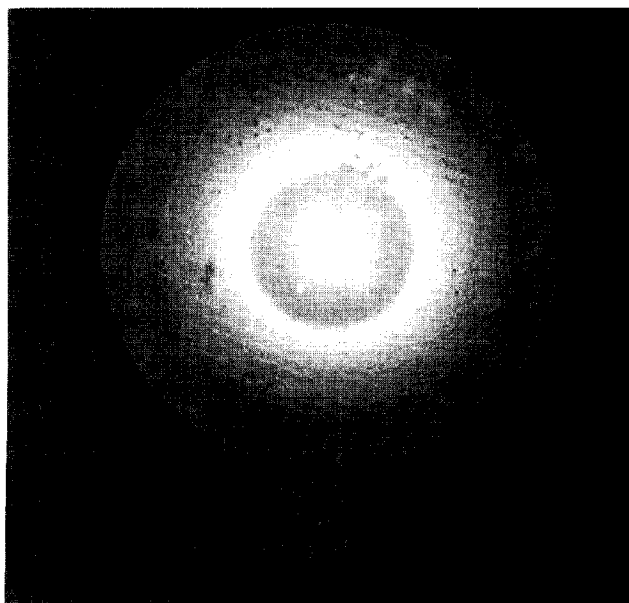


Figure P.4.61 (Photo and diagram courtesy S. Reich, The Weizmann Institute of Science, Israel.)

defined as the angle between the plane-of-vibration and the plane-of-incidence. Thus for linearly polarized light,

$$\tan \gamma_i = [E_{0i}]_{\perp} / [E_{0i}]_{\parallel} \quad (4.92)$$

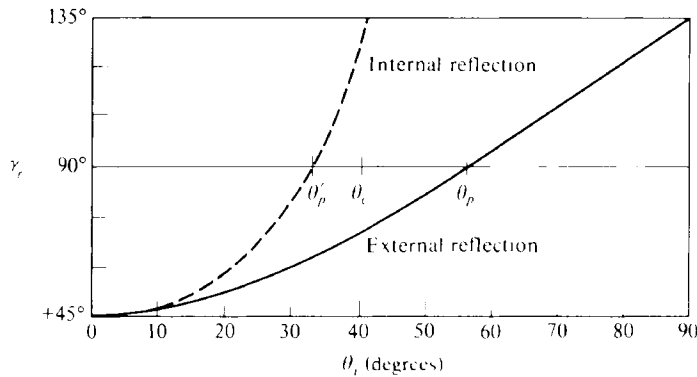
$$\tan \gamma_r = [E_{0r}]_{\perp} / [E_{0r}]_{\parallel} \quad (4.93)$$

and
$$\tan \gamma_t = [E_{0t}]_{\perp} / [E_{0t}]_{\parallel} \quad (4.94)$$

Figure P.4.67 is a plot of γ_r versus θ_i for internal and external reflection at an air-glass interface ($n_{ga} = 1.51$), where $\gamma_i = 45^\circ$. Verify a few of the points on the curves and in addition show that

$$\tan \gamma_r = -\frac{\cos(\theta_i - \theta_t)}{\cos(\theta_i + \theta_t)} \tan \gamma_i \quad (4.95)$$

Figure P.4.67



4.68* Making use of the definitions of the azimuthal angles in Problem 4.67, show that

$$R = R_{\parallel} \cos^2 \gamma_i + R_{\perp} \sin^2 \gamma_i \quad (4.96)$$

and

$$T = T_{\parallel} \cos^2 \gamma_i + T_{\perp} \sin^2 \gamma_i \quad (4.97)$$

4.69 Make a sketch of $R_{\perp}$ and $R_{\parallel}$ for $n_t = 1.5$ and $n_i = 1$ (i.e., internal reflection).

4.70 Show that

$$T_{\parallel} = \frac{\sin 2\theta_i \sin 2\theta_t}{\sin^2(\theta_i + \theta_t) \cos^2(\theta_i - \theta_t)} \quad (4.98)$$

and

$$T_{\perp} = \frac{\sin 2\theta_i \sin 2\theta_t}{\sin^2(\theta_i + \theta_t)} \quad (4.99)$$

4.71* Using the results of Problem 4.70, that is, Eqs. (4.98) and (4.99), show that

$$R_{\parallel} + T_{\parallel} = 1 \quad [4.65]$$

and
$$R_{\perp} + T_{\perp} = 1 \quad [4.66]$$

4.72 Suppose that we look at a source perpendicularly through a stack of N microscope slides. The source seen through even a dozen slides will be noticeably darker. Assuming negligible absorption, show that the total transmittance of the stack is given by

$$T_t = (1 - R)^{2N}$$

and evaluate T_t for three slides in air.

4.73 Making use of the expression

$$I(y) = I_0 e^{-\alpha y} \quad [4.78]$$

for an absorbing medium, we define a quantity called the *unit transmittance* T_1 . At normal incidence, Eq. (4.55), $T = I_t/I_i$, and thus when $y = 1$, $T_1 \equiv I(1)/I_0$. If the total thickness of the slides in the previous problem is d and if they now have a transmittance per unit length T_1 , show that

$$T_t = (1 - R)^{2N} (T_1)^d$$

4.74 Show that at normal incidence on the boundary between two dielectrics, as $n_t \rightarrow 1$, $R \rightarrow 0$, and $T \rightarrow 1$. Moreover, prove that as $n_t \rightarrow 1$, $R_{\parallel} \rightarrow 0$, $R_{\perp} \rightarrow 0$, $T_{\parallel} \rightarrow 1$, and $T_{\perp} \rightarrow 1$ for all θ_i . Thus as the two media take on more similar indices of refraction, less and less energy is carried off in the reflected wave. It should be obvious that when $n_t = 1$ there will be no interface and no reflection.

4.75* Derive the expressions for $r_{\perp}$ and $r_{\parallel}$ given by Eqs. (4.70) and (4.71).

4.76 Show that when $\theta_i > \theta_c$ at a dielectric interface, $r_{\parallel}$ and $r_{\perp}$ are complex and $r_{\perp} r_{\perp}^* = r_{\parallel} r_{\parallel}^* = 1$.

4.77 Figure P.4.77 depicts a ray being multiply reflected by a transparent dielectric plate (the amplitudes of the resulting fragments are indicated). As in Section 4.5, we use the primed coefficient notation because the angles are related by Snell's Law.

(a) Finish labeling the amplitudes of the last four rays.

(b) Show, using the Fresnel Equations, that

$$t_{\parallel} t'_{\parallel} = T_{\parallel} \quad (4.100)$$

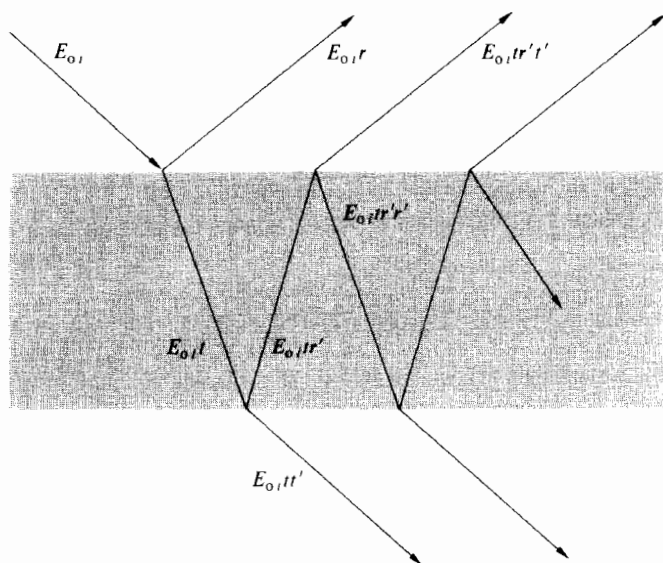
$$t_{\perp} t'_{\perp} = T_{\perp} \quad (4.101)$$

$$r_{\parallel}^2 = r'_{\parallel}{}^2 = R_{\parallel} \quad (4.102)$$

and
$$r_{\perp}^2 = r'_{\perp}{}^2 = R_{\perp} \quad (4.103)$$

4.78* A wave, linearly polarized in the plane-of-incidence, impinges on the interface between two dielectric media. If $n_i > n_t$ and $\theta_i = \theta'_p$, there is no reflected wave, that is, $r'_{\parallel}(\theta'_p) = 0$. Using Stokes's tech-

Figure P.4.77

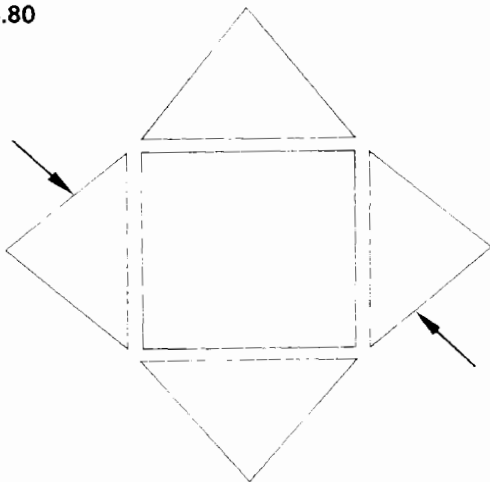


nique, start from scratch to show that $t_{\parallel}(\theta_p)t'_{\parallel}(\theta'_p) = 1$, $r_{\parallel}(\theta_p) = 0$, and $\theta_t = \theta_p$ (Problem 4.66). How does this compare with Eq. (4.100)?

4.79 Making use of the Fresnel Equations, show that $t_{\parallel}(\theta_p)t'_{\parallel}(\theta'_p) = 1$, as in the previous problem.

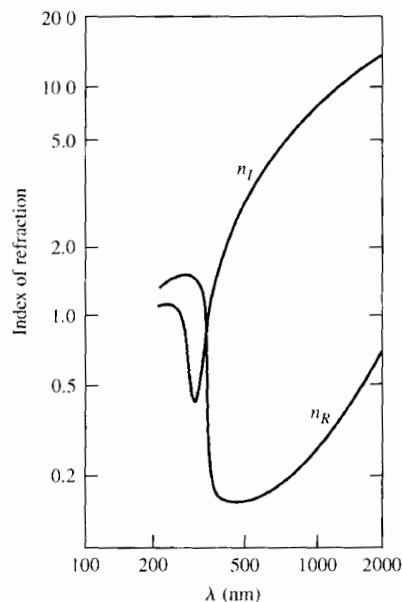
4.80 Figure P.4.80 depicts a glass cube surrounded by four glass prisms in very close proximity to its sides. Sketch in the paths that will be taken by the two rays shown and discuss a possible application for the device.

Figure P.4.80



4.81 Figure P.4.81 is a plot of n_I and n_R versus λ for a common metal. Identify the metal by comparing its characteristics with those considered in the chapter and discuss its optical properties.

Figure P.4.81



4.82 Figure P.4.82 shows a prism-coupler arrangement developed at the Bell Telephone Laboratories. Its function is to feed a laserbeam into a thin (0.000 01-inch) transparent film, which then serves as a sort of waveguide. One application is that of thin-film laserbeam circuitry—a kind of integrated optics. How do you think it works?

Figure P.4.82

