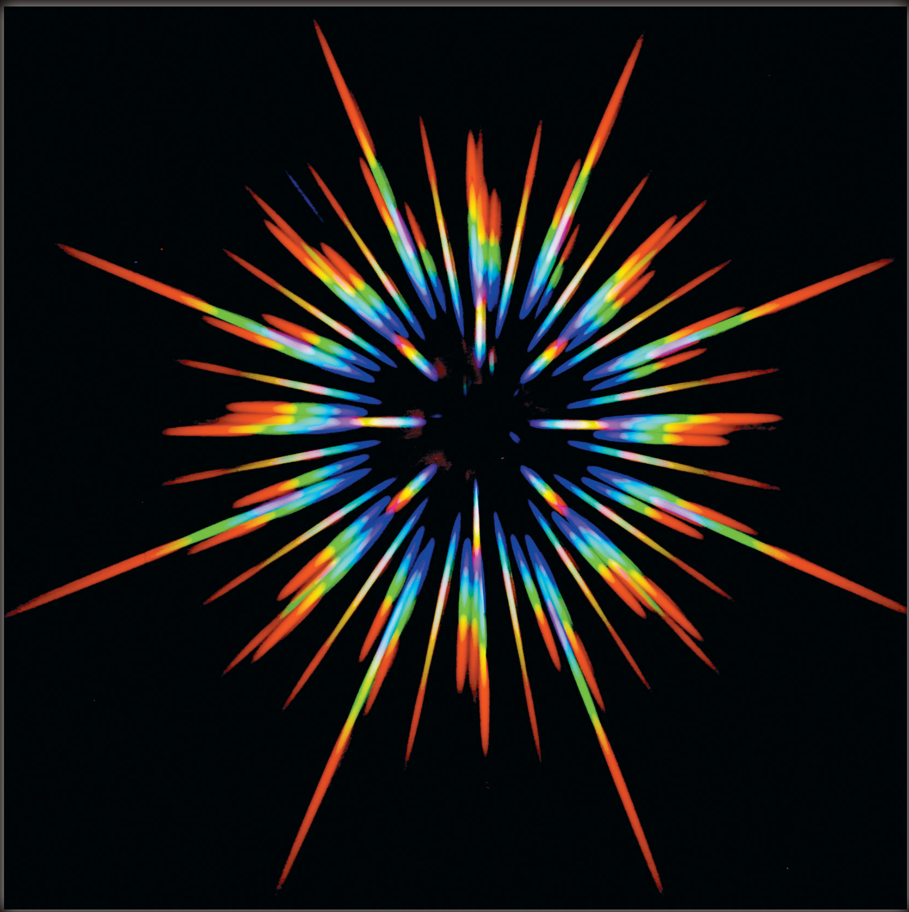
OPTICS

FIFTH EDITION



EUGENE HECHT

5ed OPTICS

Eugene Hecht

Adelphi University

PEARSON

Program Management Team Lead: Kristen Flathman
Compositor: Cenveo® Publisher Services
Cover Designer: Marilyn Perry
Illustrators: Jim Atherton
Rights & Permissions Project Manager: Rachel Youdelman
Manufacturing Buyer: Maura Zaldivar-Garcia
Marketing Manager: Elizabeth Elsworth
Cover Photo Credit: Thomas W. Mossberg, PhD

CIP data is on file with the Library of Congress

1 2 3 4 5 6 7 8 9 10—V311—18 17 16 15 14

PEARSON

www.pearsonhighered.com

ISBN 10: 0-133-97722-6;
ISBN 13: 978- 0-133-97722-6

Preface

To ***Ca, b. w. l.***

The creation of this 5th edition was guided by three overarching imperatives: wherever possible, to improve the pedagogy; to continue to modernize the treatment (e.g., with a bit more on photons, phasors, and Fourier); and to update the content to keep pace with technological advances (e.g., the book now discusses atomic interferometers, and metamaterials). Optics is a fast-evolving field and this edition strives to provide an up-to-date approach to the discipline, all the while focusing mainly on pedagogy.

To that end there are several goals: (1) to sustain an appreciation of the central role played by atomic scattering in almost every aspect of Optics; (2) to establish from the outset, the underlying quantum-mechanical nature of light (indeed, of all quantum particles), even as the book is grounded in traditional methodology. Thus the reader will find electron and neutron diffraction patterns pictured alongside the customary photon images; (3) to provide an early introduction to the powerful perspective of Fourier theory, which has come to be so prevalent in modern-day analysis. Accordingly, the concepts of spatial frequency and spatial period are introduced and graphically illustrated as early as Chapter 2, right along with temporal frequency and period.

At the request of student users, I have dispersed throughout the text over one-hundred completely worked-out EXAMPLES that make use of the principles explored in each Section. More than two hundred problems, sans solutions, have been added to the ends of the chapters to increase the available selection of fresh homework questions. A complete teacher's solutions manual is available upon request. Inasmuch as "a picture is worth a thousand words," many new diagrams and photographs further enhance the text. The book's pedagogical strength lies in its emphasis on actually explaining what is being discussed. This edition furthers that approach.

Having taught Optics every year since the 4th edition was published, I became aware of places in the book where things could be further clarified for the benefit of today's students. Accordingly, this revision addresses dozens of little sticking points, and fills in lots of missing steps in derivations. Every piece of art has been scrutinized for accuracy, and altered where appropriate to improve readability and pedagogical effectiveness.

Substantial additions of new materials can be found: in Chapter 2 (***Wave Motion***), namely, a subsection on *Twisted Light*; in Chapter 3 (***Electromagnetic Theory, Photons, and Light***), an elementary treatment of divergence and curl, additional discussion of photons, as well as subsections on *Squeezed Light*, and *Negative Refraction*; in Chapter 4 (***The Propagation of Light***), a short commentary on optical density, a piece on EM boundary conditions, more on evanescent waves, subsections on *Refraction of Light From a Point Source, Negative Refraction, Huygens's Ray Construction,* and *The Goos-Hänchen Shift*; in Chapter 5 (***Geometrical Optics***), lots of new art illustrating the behavior of lenses and mirrors, along with additional remarks on fiberoptics, as well as subsections on *Virtual Objects*, *Focal-Plane Ray Tracing,* and *Holey/Microstructured Fibers*; in Chapter 6 (***More on Geometrical Optics***), there is a fresh look at simple ray tracing through a thick lens; in Chapter 7 (***The Superposition of Waves***), one can find a new subsection on *Negative Phase Velocity*, a much extended treatment of Fourier analysis with lots of diagrams showing—without calculus—how the process actually works, and a discussion of the optical frequency comb (which was recognized by a 2005 Nobel Prize); in Chapter 8 (***Polarization***), a powerful technique is developed using phasors to analyze polarized light; there is also a new discussion of the transmittance of polarizers, and a subsection on *Wavefronts and Rays in Uniaxial Crystals*; Chapter 9 (***Interference***), begins with a brief conceptual discussion of diffraction and coherence as it relates to Young's Experiment. There are several new subsections, among which are *Near Field/Far Field, Electric Field Amplitude via Phasors, Manifestations of Diffraction, Particle Interference, Establishing The Wave Theory of Light,* and *Measuring Coherence Length*. Chapter 10 (***Diffraction***), contains a new subsection called *Phasors and the Electric-Field Amplitude.* Dozens of newly created diagrams and photographs extensively illustrate a variety of diffraction phenomena. Chapter 11 (***Fourier Optics***), now has a subsection, *Two-Dimensional Images*, which contains a remarkable series of illustrations depicting how spatial frequency components combine to create images. Chapter 12 (***Basics of Coherence Theory***), contains several new introductory subsections among which are *Fringes and Coherence*, and *Diffraction and the Vanishing Fringes*. There are also a number of additional highly supportive illustrations. Chapter 13 (***Modern Optics: Lasers and Other Topics***), contains an enriched and updated treatment of lasers accompanied by tables and illustrations as well as several new subsections, including *Optoelectronic Image Reconstruction*.

This 5th edition offers a substantial amount of new material that will be of special interest to teachers of Optics. For example: in addition to plane, spherical, and cylindrical waves, we can now generate helical waves for which the surface of constant phase spirals as it advances through space (Section 2.11, p. 31).

Beyond the mathematics, students often have trouble understanding what the operations of *divergence* and *curl* correspond to physically. Accordingly, the present revision contains a section exploring what those operators actually do, in fairly simple terms (Section 3.1.5, p. 43).

The phenomenon of *negative refraction* is an active area of contemporary research and a brief introduction to the basic physics involved can now be found in Chapter 4 (p. 106).

Huygens devised a method for constructing refracted rays (p. 108), which is lovely in and of itself, but it also allows for a convenient way to appreciate refraction in anisotropic crystals (p. 350).

When studying the interaction of electromagnetic waves with material media (e.g., in the derivation of the Fresnel Equations), one utilizes the *boundary conditions*. Since some student readers may have little familiarity with E&M, the 5^{th} edition contains a brief discussion of the physical origins of those conditions (Section 4.6.1, p. 114).

The book now contains a brief discussion of the *Goos-Hänchen shift* which occurs in total internal reflection, It's a piece of interesting physics that is often overlooked in introductory treatments (Section 4.7.1, p. 129).

Focal-plane ray tracing is a straightforward way to track rays through complicated lens systems. This simple yet powerful technique, which is new to this edition, works nicely in the classroom and is well worth a few minutes of lecture time (p. 169).

Several fresh diagrams now make clear the nature of virtual images and, more subtly, *virtual objects* arising via lens systems (p. 168–169).

The widespread use of fiberoptics has necessitated an up-to-date exposition of certain aspects of the subject (p. 200–204). Among the new material the reader can now find a discussion of *microstructured fibers* and, more generally, *photonic crystals*, both entailing significant physics (p. 204–206).

In addition to the usual somewhat formulaic, and alas, "dry" mathematical treatment of Fourier series, the book now contains a fascinating graphical analysis that conceptually shows what those several integrals are actually doing. This is great stuff for undergraduates (Section 7.3.1, p. 301–305).

Phasors are utilized extensively to help students visualize the addition of harmonic waves. The technique is very useful in treating the orthogonal field components that constitute the various polarization states (p. 336–3). Moreover, the method provides a nice graphical means to analyze the behavior of wave plates (p. 363).

Young's Experiment and double-beam interference in general, are central to both classical and quantum Optics. Yet the usual introduction to this material is far too simplistic in that it overlooks the limitations imposed by the phenomena of diffraction and coherence. The analysis now briefly explores those concerns early on (Section 9.1.1, p. 394).

The traditional discussion of interference is extended using phasors to graphically represent electric-field amplitudes, giving students an alternative way to visualize what's happening (Section 9.3.1, p. 401).

Diffraction can also be conveniently appreciated via electric-field phasors (p. 462–463). That methodology leads naturally to the classical *vibration curve*, which brings to mind Feynman's probability-amplitude approach to quantum mechanics. In any event, it provides students with a complementary means of apprehending diffraction that is essentially free of calculus.

The reader interested in Fourier optics can now find a wonderful series of illustrations showing how sinusoidal spatial frequency contributions can come together to generate a recognizable two-dimensional image; in this case of a young Einstein (p. 547). This extraordinary sequence of figures should be discussed, even in an introductory class where the material in Chapter 11 might otherwise be beyond the level of the course—it's fundamental to modern image theory, and conceptually beautiful.

To make the advanced treatment of coherence in Chapter 12 more accessible to a wider readership, this edition now contains an essentially non-mathematical introduction (p. 582); it sets the stage for the traditional presentation.

Finally, the material on lasers, though only introductory, has been extended (p. 611) and brought more into line with the contemporary state of affairs.

Over the years since the 4th edition dozens of colleagues around the world have provided comments, advice, suggestions, articles, and photographs for this new edition; I sincerely thank them all. I am especially grateful to Professor Chris Mack of the University of Texas at Austin, and Dr. Andreas Karpf of Adelphi University. I'm also indebted to my many students who have blind tested all the new expositive material, worked the new problems (often on exams), and helped take some of the new photos. Regarding the latter I particularly thank Tanya Spellman, George Harrison, and Irina Ostrozhnyuk for the hours spent, cameras in hand.

I am most appreciative of the support provided by the team at Addison Wesley, especially by Program Manager Katie Conley who has ably and thoughtfully guided the creation of this 5th edition from start to finish. The manuscript was scrupulously and gracefully copy edited by Joanne Boehme who did a remarkable job. Hundreds of complex diagrams were artfully drawn by Jim Atherton of Atherton Customs; his work is extraordinary and speaks for itself. This edition of Optics was developed under the ever-present guidance of John Orr of Orr Book Services. His abiding commitment to producing an accurate, beautiful book deserves special praise. In an era when traditional publishing is undergoing radical change, he uncompromisingly maintained the very highest standards, for which I am most grateful. It was truly a pleasure and a privilege working with such a consummate professional.

Lastly I thank my dear friend, proofreader extraordinaire, my wife, Carolyn Eisen Hecht who patiently coped with the travails of one more edition of one more book. Her good humor, forbearance, emotional generosity, and wise counsel were essential.

Anyone wishing to offer comments or suggestions concerning this edition, or to provide contributions to a future edition, can reach me at Adelphi University, Physics Department, Garden City, NY, 11530 or better yet, at genehecht@aol.com.

Contents

1 A Brief History

1.1 Prolegomenon

In chapters to come we will evolve a formal treatment of much of the science of Optics, with particular emphasis on aspects of contemporary interest. The subject embraces a vast body of knowledge accumulated over roughly three thousand years of the human scene. Before embarking on a study of the modern view of things optical, let's briefly trace the road that led us there, if for no other reason than to put it all in perspective.

1.2 In the Beginning

The origins of optical technology date back to remote antiquity. Exodus 38:8 (ca. 1200 B.C.E.) recounts how Bezaleel, while preparing the ark and tabernacle, recast "the looking-glasses of the women" into a brass laver (a ceremonial basin). Early mirrors were made of polished copper, bronze, and later on of speculum, a copper alloy rich in tin. Specimens have survived from ancient Egypt—a mirror in perfect condition was unearthed along with some tools from the workers' quarters near the pyramid of Sesostris II (ca. 1900 B.C.E.) in the Nile valley. The Greek philosophers Pythagoras, Democritus, Empedocles, Plato, Aristotle, and others developed several theories of the nature of light. The rectilinear propagation of light (p. 91) was known, as was the *Law of Reflection* (p. 97) enunciated by Euclid (300 B.C.E.) in his book *Catoptrics*. Hero of Alexandria attempted to explain both these phenomena by asserting that light traverses the shortest allowed path between two points. The burning glass (a positive lens used to start fires) was alluded to by Aristophanes in his comic play *The Clouds* (424 B.C.E.). The apparent bending of objects partly immersed in water (p. 105) is mentioned in Plato's *Republic*. Refraction was studied by Cleomedes (50 C.E.) and later by Claudius Ptolemy (130 C.E.) of Alexandria, who tabulated fairly precise measurements of the angles of incidence and refraction for several media (p. 100). It is clear from the accounts of the historian Pliny (23–79 C.E.) that the Romans also possessed burning glasses. Several glass and crystal spheres have been found among Roman ruins, and a planar convex lens was recovered in Pompeii. The Roman philosopher Seneca (3 B.C.E.–65 C.E.) pointed out that a glass globe filled with water could be used for magnifying purposes. And it is certainly possible that some Roman artisans may have used magnifying glasses to facilitate very fine detailed work.

After the fall of the Western Roman Empire (475 C.E.), which roughly marks the start of the Dark Ages, little or no scientific progress was made in Europe for a great while. The dominance of the Greco-Roman-Christian culture in the lands embracing the Mediterranean soon gave way by conquest to the rule of Allah. The center of scholarship shifted to the Arab world.

Refraction was studied by Abu Sa`d al-`Ala' Ibn Sahl (940–1000 C.E.), who worked at the Abbasid court in Baghdad, where he wrote *On the Burning Instruments* in 984. His accurate diagrammatical illustration of refraction, the first ever, appears in that book. Ibn Sahl described both parabolic and ellipsoidal burning mirrors and analyzed the hyperbolic plano-convex lens, as well as the hyperbolic biconvex lens. The scholar Abu Ali al-Hasan ibn al-Haytham (965–1039), known in the Western world as Alhazen, was a prolific writer on a variety of topics, including 14 books on Optics alone. He elaborated on the Law of Reflection, putting the angles of incidence and reflection in the same plane normal to the interface (p. 99); he studied spherical and parabolic mirrors and gave a detailed description of the human eye (p. 207). Anticipating Fermat, Alhazen suggested that light travels the fastest path through a medium.

By the latter part of the thirteenth century, Europe was only beginning to rouse from its intellectual stupor. Alhazen's work was translated into Latin, and it had a great effect on the writings of Robert Grosseteste (1175–1253), Bishop of Lincoln, and on the Polish mathematician Vitello (or Witelo), both of whom were influential in rekindling the study of Optics. Their works were known to the Franciscan Roger Bacon (1215–1294), who is considered by many to be the first scientist in the modern sense. He seems to have initiated the idea of using lenses for correcting vision and even hinted at the possibility of combining lenses to form a telescope. Bacon also had some understanding of the way in which rays traverse a lens. After his death, Optics again languished. Even so, by the mid-1300s, European paintings were depicting monks wearing eyeglasses. And alchemists had come up with a liquid amalgam of tin and mercury that was rubbed onto the back of glass plates to make mirrors. Leonardo da Vinci (1452–1519) described the *camera obscura* (p. 220), later popularized by

Giovanni Battista Della Porta (1535–1615). (US National Library of Medicine)

the work of Giovanni Battista Della Porta (1535–1615), who discussed multiple mirrors and combinations of positive and negative lenses in his *Magia naturalis* (1589).

This, for the most part, modest array of events constitutes what might be called the first period of Optics. It was undoubtedly a beginning—but on the whole a humble one. The whirlwind of accomplishment and excitement was to come later, in the seventeenth century.

1.3 From the Seventeenth Century

It is not clear who actually invented the refracting telescope, but records in the archives at The Hague show that on October 2, 1608, Hans Lippershey (1587–1619), a Dutch spectacle maker, applied for a patent on the device. Galileo Galilei (1564–1642), in Padua, heard about the invention and within several months had built his own instrument (p. 227), grinding the lenses by hand. The compound microscope was invented at just about the same time, possibly by the Dutchman Zacharias Janssen (1588–1632). The microscope's concave eyepiece was replaced with a convex lens by Francisco Fontana (1580–1656) of Naples, and a similar change in the telescope was introduced by Johannes Kepler (1571–1630). In 1611, Kepler published his *Dioptrice*. He had discovered total internal reflection (p. 125) and arrived at the small angle approximation to the Law of Refraction, in which case the incident and transmission angles are proportional. He evolved a treatment of first-order Optics for thin-lens systems and in his book describes the detailed operation of both the Keplerian (positive eyepiece) and Galilean (negative eyepiece) telescopes. Willebrord Snel (1591–1626), whose name is usually inexplicably spelled Snell, professor at Leyden, empirically discovered the long-hidden *Law of Refraction* (p. 100) in 1621—this was one of the great moments in Optics. By learning precisely how rays of light are redirected on traversing a boundary between two media, Snell in one swoop swung open the door to modern applied Optics. René Descartes (1596–1650) was the first to publish the now familiar formulation of the Law of Refraction in terms of sines. Descartes deduced the

A very early picture of an outdoor European village scene. The man on the left is selling eyeglasses. (INTERFOTO/Alamy)

Johannes Kepler (1571–1630). (Nickolae/Fotolia)

law using a model in which light was viewed as a pressure transmitted by an elastic medium; as he put it in his *La Dioptrique* (1637)

> recall the nature that I have attributed to light, when I said that it is nothing other than a certain motion or an action conceived in a very subtle matter, which fills the pores of all other bodies. . . .

The universe was a plenum. Pierre de Fermat (1601–1665), taking exception to Descartes's assumptions, rederived the Law of Reflection (p. 109) from his own *Principle of Least Time* (1657).

The phenomenon of diffraction, that is, the deviation from rectilinear propagation that occurs when light advances beyond an obstruction (p. 449), was first noted by Professor Francesco Maria Grimaldi (1618–1663) at the Jesuit College in Bologna. He had observed bands of light within the shadow of a rod illuminated by a small source. Robert Hooke (1635–1703), curator of experiments for the Royal Society, London, later

René Descartes by Frans Hals (1596–1650). (Georgios Kollidas/Shutterstock)

Sir Isaac Newton (1642–1727). (Georgios Kollidas/Fotolia)

also observed diffraction effects. He was the first to study the colored interference patterns (p. 408) generated by thin films (*Micrographia*, 1665). He proposed the idea that light was a rapid vibratory motion of the medium propagating at a very great speed. Moreover, "every pulse or vibration of the luminous body will generate a sphere"—this was the beginning of the wave theory. Within a year of Galileo's death, Isaac Newton (1642–1727) was born. The thrust of Newton's scientific effort was to build on direct observation and avoid speculative hypotheses. Thus he remained ambivalent for a long while about the actual nature of light. Was it corpuscular—a stream of particles, as some maintained? Or was light a wave in an all-pervading medium, the aether? At the age of 23, he began his now famous experiments on dispersion.

> I procured me a triangular glass prism to try therewith the celebrated phenomena of colours.

Newton concluded that white light was composed of a mixture of a whole range of independent colors (p. 193). He maintained that the corpuscles of light associated with the various colors excited the aether into characteristic vibrations. Even though his work simultaneously embraced both the wave and emission (corpuscular) theories, he did become more committed to the latter as he grew older. His main reason for rejecting the wave theory as it stood then was the daunting problem of explaining rectilinear propagation in terms of waves that spread out in all directions.

After some all-too-limited experiments, Newton gave up trying to remove chromatic aberration from refracting telescope lenses. Erroneously concluding that it could not be done, he turned to the design of reflectors. Sir Isaac's first reflecting telescope, completed in 1668, was only 6 inches long and 1 inch in diameter, but it magnified some 30 times.

At about the same time that Newton was emphasizing the emission theory in England, Christiaan Huygens (1629–1695),

Christiaan Huygens (1629–1695). (*Portrait of Christiaan Huygens* (ca. 1680), Abraham Bloteling. Engraving. Rijksmuseum [Object number RP-P-1896-A-19320].)

on the continent, was greatly extending the wave theory. Unlike Descartes, Hooke, and Newton, Huygens correctly concluded that light effectively slowed down on entering more dense media. He was able to derive the Laws of Reflection and Refraction and even explained the double refraction of calcite (p. 344), using his wave theory. And it was while working with calcite that he discovered the phenomenon of *polarization* (p. 330).

> As there are two different refractions, I conceived also that there are two different emanations of the waves of light. . . .

Thus light was either a stream of particles or a rapid undulation of aethereal matter. In any case, it was generally agreed that its speed was exceedingly large. Indeed, many believed that light propagated instantaneously, a notion that went back at least as far as Aristotle. The fact that it was finite was determined by the Dane Ole Christensen Römer (1644–1710). Jupiter's nearest moon, Io, has an orbit about that planet that is nearly in the plane of Jupiter's own orbit around the Sun. Römer made a careful study of the eclipses of Io as it moved through the shadow behind Jupiter. In 1676 he predicted that on November 9 Io would emerge from the dark some 10 minutes later than would have been expected on the basis of its yearly averaged motion. Precisely on schedule, Io performed as predicted, a phenomenon Römer correctly explained as arising from the finite speed of light. He was able to determine that light took about 22 minutes to traverse the diameter of the Earth's orbit around the Sun—a distance of about 186 million miles. Huygens and Newton, among others, were quite convinced of the validity of Römer's work. Independently estimating the Earth's orbital diameter, they assigned values to c equivalent to 2.3×10^8 m/s and 2.4×10^8 m/s, respectively.*

*A. Wróblewski, *Am. J. Phys.* **53,** 620 (1985).

The great weight of Newton's opinion hung like a shroud over the wave theory during the eighteenth century, all but stifling its advocates. Despite this, the prominent mathematician Leonhard Euler (1707–1783) was a devotee of the wave theory, even if an unheeded one. Euler proposed that the undesirable color effects seen in a lens were absent in the eye (which is an erroneous assumption) because the different media present negated dispersion. He suggested that achromatic lenses (p. 272) might be constructed in a similar way. Inspired by this work, Samuel Klingenstjerna (1698–1765), a professor at Uppsala, reperformed Newton's experiments on achromatism and determined them to be in error. Klingenstjerna was in communication with a London optician, John Dollond (1706–1761), who was observing similar results. Dollond finally, in 1758, combined two elements, one of crown and the other of flint glass, to form a single achromatic lens. Incidentally, Dollond's invention was actually preceded by the unpublished work of the amateur scientist Chester Moor Hall (1703–1771) in Essex.

1.4 The Nineteenth Century

The wave theory of light was reborn at the hands of Dr. Thomas Young (1773–1829), one of the truly great minds of the century. In 1801, 1802, and 1803, he read papers before the Royal Society, extolling the wave theory and adding to it a new fundamental concept, the so-called *Principle of Interference* (p. 390):

> When two undulations, from different origins, coincide either perfectly or very nearly in direction, their joint effect is a combination of the motions belonging to each.

Thomas Young (1773–1829). (Smithsonian Institution)

Augustin Jean Fresnel (1788–1827). (US National Library of Medicine)

He was able to explain the colored fringes of thin films and determined wavelengths of various colors using Newton's data. Even though Young, time and again, maintained that his conceptions had their very origins in the research of Newton, he was severely attacked. In a series of articles, probably written by Lord Brougham, in the *Edinburgh Review*, Young's papers were said to be "destitute of every species of merit."

Augustin Jean Fresnel (1788–1827), born in Broglie, Normandy, began his brilliant revival of the wave theory in France, unaware of the efforts of Young some 13 years earlier. Fresnel synthesized the concepts of Huygens's wave description and the interference principle. The mode of propagation of a primary wave was viewed as a succession of spherical secondary wavelets, which overlapped and interfered to re-form the advancing primary wave as it would appear an instant later. In Fresnel's words:

> The vibrations of a luminous wave in any one of its points may be considered as the sum of the elementary movements conveyed to it at the same moment, from the separate action of all the portions of the unobstructed wave considered in any one of its anterior positions.

These waves were presumed to be longitudinal, in analogy with sound waves in air. Fresnel was able to calculate the diffraction patterns arising from various obstacles and apertures and satisfactorily accounted for rectilinear propagation in homogeneous isotropic media, thus dispelling Newton's main objection to the undulatory theory. When finally apprised of Young's priority to the interference principle, a somewhat disappointed Fresnel nonetheless wrote to Young, telling him that he was consoled by finding himself in such good company—the two great men became allies.

Huygens was aware of the phenomenon of polarization arising in calcite crystals, as was Newton. Indeed, the latter in his *Opticks* stated,

> Every Ray of Light has therefore two opposite Sides. . . .

It was not until 1808 that Étienne Louis Malus (1775–1812) discovered that this two-sidedness of light also arose upon reflection (p. 355); the phenomenon was not inherent to crystalline media. Fresnel and Dominique François Arago (1786–1853) then conducted a series of experiments to determine the effect of polarization on interference, but the results were utterly inexplicable within the framework of their longitudinal wave picture. This was a dark hour indeed. For several years Young, Arago, and Fresnel wrestled with the problem until finally Young suggested that the aethereal vibration might be *transverse*, as is a wave on a string. The two-sidedness of light was then simply a manifestation of the two orthogonal vibrations of the aether, transverse to the ray direction. Fresnel went on to evolve a mechanistic description of aether oscillations, which led to his now famous formulas for the amplitudes of reflected and transmitted light (p. 115). By 1825 the emission (or corpuscular) theory had only a few tenacious advocates.

The first terrestrial determination of the speed of light was performed by Armand Hippolyte Louis Fizeau (1819–1896) in 1849. His apparatus, consisting of a rotating toothed wheel and a distant mirror (8633 m), was set up in the suburbs of Paris from Suresnes to Montmartre. A pulse of light leaving an opening in the wheel struck the mirror and returned. By adjusting the known rotational speed of the wheel, the returning pulse could be made either to pass through an opening and be seen or to be obstructed by a tooth. Fizeau arrived at a value of the speed of light equal to 315 300 km/s. His colleague Jean Bernard Léon Foucault (1819–1868) was also involved in research on the speed of light. In 1834 Charles Wheatstone (1802–1875) had designed a rotating-mirror arrangement in order to measure the duration of an electric spark. Using this scheme, Arago had proposed to measure the speed of light in dense media but was never able to carry out the experiment. Foucault took up the work, which was later to provide material for his doctoral thesis. On May 6, 1850, he reported to the Academy of Sciences that the speed of light in water was *less* than that in air. This result was in direct conflict with Newton's formulation of the emission theory and a hard blow to its few remaining devotees.

While all of this was happening in Optics, quite independently, the study of electricity and magnetism was also bearing fruit. In 1845 the master experimentalist Michael Faraday (1791–1867) established an interrelationship between electromagnetism and light when he found that the polarization direction of a beam could be altered by a strong magnetic field applied to the medium. James Clerk Maxwell (1831–1879) brilliantly summarized and extended all the empirical knowledge on the subject in a single set of mathematical equations. Beginning with this remarkably succinct

James Clerk Maxwell (1831–1879). (E.H.)

and beautifully symmetrical synthesis, he was able to show, purely theoretically, that the electromagnetic field could propagate as a transverse wave in the luminiferous aether (p. 46).

Solving for the speed of the wave, Maxwell arrived at an expression in terms of electric and magnetic properties of the medium ($c = 1/\sqrt{\epsilon_0\mu_0}$). Upon substituting known empirically determined values for these quantities, he obtained a numerical result equal to the measured speed of light! The conclusion was inescapable—*light was "an electromagnetic disturbance in the form of waves" propagated through the aether*. Maxwell died at the age of 48, eight years too soon to see the experimental confirmation of his insights and far too soon for physics. Heinrich Rudolf Hertz (1857–1894) verified the existence of long electromagnetic waves by generating and detecting them in an extensive series of experiments published in 1888.

The acceptance of the wave theory of light seemed to necessitate an equal acceptance of the existence of an all-pervading substratum, the luminiferous aether. If there were waves, it seemed obvious that there must be a supporting medium. Quite naturally, a great deal of scientific effort went into determining the physical nature of the aether, yet it would have to possess some rather strange properties. It had to be so tenuous as to allow an apparently unimpeded motion of celestial bodies. At the same time, it could support the exceedingly high-frequency ($\sim 10^{15}$ Hz) oscillations of light traveling at 186 000 miles per second. That implied remarkably strong restoring forces within the aethereal substance. The speed at which a wave advances through a medium is dependent on the characteristics of the disturbed substratum and not on any motion of the source. This is in contrast to the behavior of a stream of particles whose speed with respect to the source is the essential parameter.

Certain aspects of the nature of aether intrude when studying the optics of moving objects, and it was this area of research, evolving quietly on its own, that ultimately led to the next great turning point. In 1725 James Bradley (1693–1762), then Savilian Professor of Astronomy at Oxford, attempted to measure the distance to a star by observing its orientation at two different times of the year. The position of the Earth changed as it orbited around the Sun and thereby provided a large baseline for triangulation on the star. To his surprise, Bradley found that the "fixed" stars displayed an apparent systematic movement related to the direction of motion of the Earth in orbit and not dependent, as had been anticipated, on the Earth's position in space. This so-called *stellar aberration* is analogous to the well-known falling-raindrop situation. A raindrop, although traveling vertically with respect to an observer at rest on the Earth, will appear to change its incident angle when the observer is in motion. Thus a corpuscular

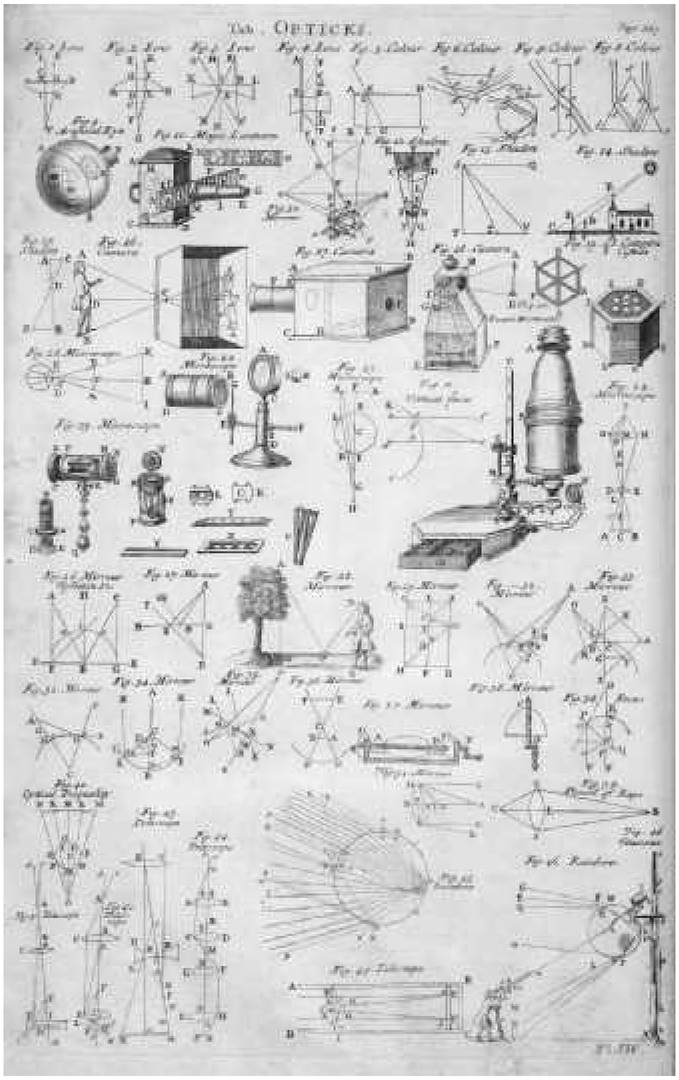


Table of Opticks from Volume 2 of the *Cyclopedia: or, An Universal Dictionary of Arts and Sciences*, edited by Ephraim Chambers, published in London by James and John Knapton in 1728. (University of Wisconsin Digital Collections) model of light could explain stellar aberration rather handily. Alternatively, the wave theory also offers a satisfactory explanation provided that *the aether remains totally undisturbed as the Earth plows through it*.

In response to speculation as to whether the Earth's motion through the aether might result in an observable difference between light from terrestrial and extraterrestrial sources, Arago set out to examine the problem experimentally. He found that there were no such observable differences. Light behaved just as if the Earth were at rest with respect to the aether. To explain these results, Fresnel suggested in effect that light was partially dragged along as it traversed a transparent medium in motion. Experiments by Fizeau, in which light beams passed down moving columns of water, and by Sir George Biddell Airy (1801–1892), who used a water-filled telescope in 1871 to examine stellar aberration, both seemed to confirm Fresnel's drag hypothesis. Assuming an aether at *absolute rest*, Hendrik Antoon Lorentz (1853–1928) derived a theory that encompassed Fresnel's ideas.

In 1879 in a letter to D. P. Todd of the U.S. Nautical Almanac Office, Maxwell suggested a scheme for measuring the speed at which the solar system moved with respect to the luminiferous aether. The American physicist Albert Abraham Michelson (1852–1931), then a naval instructor, took up the idea. Michelson, at the tender age of 26, had already established a favorable reputation by performing an extremely precise determination of the speed of light. A few years later, he began an experiment to measure the effect of the Earth's motion through the aether. Since the speed of light in aether is constant and the Earth, in turn, presumably moves in relation to the aether (orbital speed of 67 000 mi/h), the speed of light measured with respect to the Earth should be affected by the planet's motion. In 1881 he published his findings. There was no detectable motion of the Earth with respect to the aether—the aether was stationary. But the decisiveness of this surprising result was blunted somewhat when Lorentz pointed out an oversight in the calculation. Several years later Michelson, then professor of physics at Case School of Applied Science in Cleveland, Ohio, joined with Edward Williams Morley (1838–1923), a well-known professor of chemistry at Western Reserve, to redo the experiment with considerably greater precision. Amazingly enough, their results, published in 1887, once again were negative:

> It appears from all that precedes reasonably certain that if there be any relative motion between the earth and the luminiferous aether, it must be small; quite small enough entirely to refute Fresnel's explanation of aberration.

Thus, whereas an explanation of stellar aberration within the context of the wave theory required the existence of a relative motion between Earth and aether, the Michelson–Morley Experiment refuted that possibility. Moreover, the findings of Fizeau and Airy necessitated the inclusion of a partial drag of light due to motion of the medium.

1.5 Twentieth-Century Optics

Jules Henri Poincaré (1854–1912) was perhaps the first to grasp the significance of the experimental inability to observe any effects of motion relative to the aether. In 1899 he began to make his views known, and in 1900 he said:

> Our aether, does it really exist? I do not believe that more precise observations could ever reveal anything more than *relative* displacements.

In 1905 Albert Einstein (1879–1955) introduced his *Special Theory of Relativity*, in which he too, quite independently, rejected the aether hypothesis.

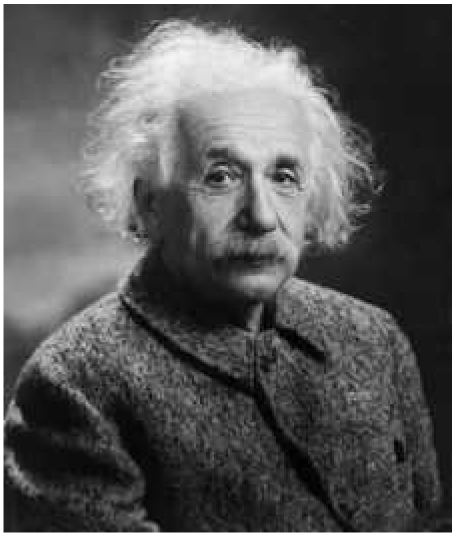
> The introduction of a "luminiferous aether" will prove to be superfluous inasmuch as the view here to be developed will not require an "absolutely stationary space."

He further postulated:

> light is always propagated in empty space with a definite velocity *c* which is independent of the state of motion of the emitting body.

The experiments of Fizeau, Airy, and Michelson–Morley were then explained quite naturally within the framework of Einstein's relativistic kinematics.* Deprived of the aether, physicists simply had to get used to the idea that electromagnetic waves could propagate through free space—there was no alternative. Light was now envisaged as a self-sustaining wave with the conceptual emphasis passing from aether to field. The electromagnetic wave became an entity in itself.

On October 19, 1900, Max Karl Ernst Ludwig Planck (1858–1947) read a paper before the German Physical Society in which he introduced the hesitant beginnings of what was to become yet



Albert Einstein (1879–1955). (Orren Jack Turner/Library of Congress Prints and Photographs Division [LC-USZ62-60242])

*See, for example, *Special Relativity* by French, Chapter 5.

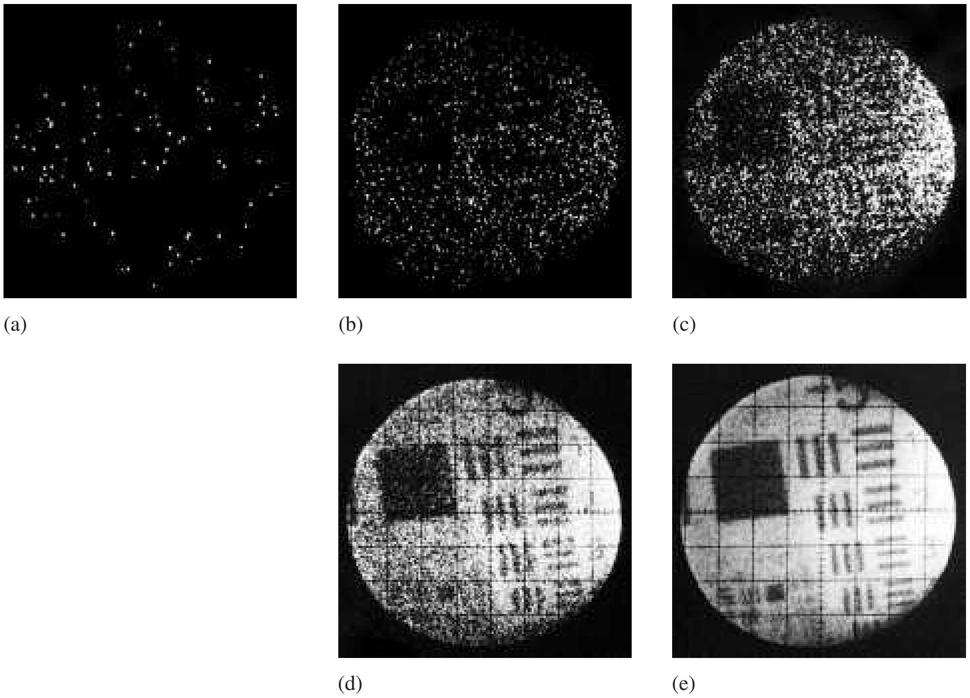


Figure 1.1 A rather convincing illustration of the particle nature of light. This sequence of photos was made using a position-sensing photomultiplier tube illuminated by an (8.5×10^3 count-per-second) image of a bar chart. The exposure times were (*a*) 8 ms, (*b*) 125 ms, (*c*) 1 s, (*d*) 10 s, and (*e*) 100 s. Each dot can be interpreted as the arrival of a single photon. (ITT Electro-Optical Products Division)

another great revolution in scientific thought—*Quantum Mechanics*, a theory embracing submicroscopic phenomena (p. 53). In 1905, boldly building on these ideas, Einstein proposed a new form of corpuscular theory in which he asserted that light consisted of globs or "particles" of energy. Each such quantum of radiant energy or *photon*,† as it came to be called, had an energy proportional to its frequency ν, that is, $\mathscr{E} = h\nu$, where h is known as Planck's constant (Fig. 1.1). By the end of the 1920s, through the efforts of Bohr, Born, Heisenberg, Schrödinger, De Broglie, Pauli, Dirac, and others, Quantum Mechanics had become a well-verified theory. It gradually became evident that the concepts of particle and wave, which in the macroscopic world seem so obviously mutually exclusive, must be merged in the submicroscopic domain. The mental image of an atomic particle (e.g., electrons and neutrons) as a minute localized lump of matter would no longer suffice. Indeed, it was found that these "particles" could generate interference and diffraction patterns in precisely the same way as would light (p. 404). Thus photons, protons, electrons, neutrons, and so forth—the whole lot—have both particle and wave manifestations. Still, the matter was by no means settled. "Every physicist thinks that he knows what a photon is," wrote Einstein. "I spent my life to find out what a photon is and I still don't know it."

Relativity liberated light from the aether and showed the kinship between mass and energy (via $\mathscr{E}_0 = mc^2$). What seemed to be two almost antithetical quantities now became interchangeable. Quantum Mechanics went on to establish that a particle‡ of momentum p had an associated wavelength λ, such that $p = h/\lambda$. The easy images of submicroscopic specks of matter became untenable, and the wave-particle dichotomy dissolved into a duality.

Quantum Mechanics also treats the manner in which light is absorbed and emitted by atoms (p. 66). Suppose we cause a gas to glow by heating it or passing an electrical discharge through it. The light emitted is characteristic of the very structure of the atoms constituting the gas. Spectroscopy, which is the branch of Optics dealing with spectrum analysis (p. 75), developed from the research of Newton. William Hyde Wollaston (1766–1828) made the earliest observations of the dark lines in the solar spectrum (1802). Because of the slit-shaped aperture generally used in spectroscopes, the output consisted of narrow colored bands of light, the so-called *spectral lines*. Working independently, Joseph Fraunhofer (1787–1826) greatly extended the subject. After accidentally discovering the double line of sodium (p. 136), he went on to study sunlight and made the first wavelength determinations using diffraction gratings (p. 488). Gustav Robert Kirchhoff (1824–1887) and Robert Wilhelm Bunsen (1811–1899), working together at Heidelberg, established that each kind of atom had its own signature in a characteristic array of spectral lines. And in 1913 Niels Henrik David Bohr (1885–1962) set forth a precursory quantum theory of the hydrogen atom, which was able to predict the wavelengths of its emission spectrum. The light emitted by an atom is now understood to arise from its outermost electrons (p. 66). The process is the domain of modern quantum theory, which describes the most minute details with incredible precision and beauty.

The flourishing of applied Optics in the second half of the twentieth century represents a renaissance in itself. In the 1950s

†The word *photon* was coined by G. N. Lewis, *Nature*, December 18, 1926.

‡Perhaps it might help if we just called them all *wavicles*.

several workers began to inculcate Optics with the mathematical techniques and insights of communications theory. Just as the idea of momentum provides another dimension in which to visualize aspects of mechanics, the concept of spatial frequency offers a rich new way of appreciating a broad range of optical phenomena. Bound together by the mathematical formalism of Fourier analysis (p. 300), the outgrowths of this contemporary emphasis have been far-reaching. Of particular interest are the theory of image formation and evaluation (p. 544), the *transfer functions* (p. 570), and the idea of *spatial filtering* (p. 320).

The advent of the high-speed digital computer brought with it a vast improvement in the design of complex optical systems. Aspherical lens elements (p. 152) took on renewed practical significance, and the *diffraction-limited* system with an appreciable field of view became a reality. The technique of ion bombardment polishing, in which one atom at a time is chipped away, was introduced to meet the need for extreme precision in the preparation of optical elements. The use of single and multilayer thin-film coatings (reflecting, antireflecting, etc.) became commonplace (p. 435). Fiberoptics evolved into a practical communications tool (p. 196), and thin-film light guides continued to be studied. A great deal of attention was paid to the infrared end of the spectrum (surveillance systems, missile guidance, etc.), and this in turn stimulated the development of infrared materials. Plastics began to be used extensively in Optics (lens elements, replica gratings, fibers, aspherics, etc.). A new class of partially vitrified glass ceramics with exceedingly low thermal expansion was developed. A resurgence in the construction of astronomical observatories (both terrestrial and extraterrestrial) operating across the whole spectrum was well under way by the end of the 1960s and vigorously sustained into the twenty-first century (p. 228).

The first laser was built in 1960, and within a decade laserbeams spanned the range from infrared to ultraviolet. The availability of high-power coherent sources led to the discovery of a number of new optical effects (harmonic generation, frequency mixing, etc.) and thence to a panorama of marvelous new devices. The technology needed to produce a practicable optical communications system developed rapidly. The sophisticated use of crystals in devices such as second-harmonic generators (p. 660), electro-optic and acousto-optic modulators, and the like spurred a great deal of contemporary research in crystal optics. The wavefront reconstruction technique known as *holography* (p. 644), which produces magnificent three-dimensional images, was found to have numerous additional applications (nondestructive testing, data storage, etc.).

The military orientation of much of the developmental work in the 1960s continued into the 2000s with added vigor. Today that technological interest in Optics ranges across the spectrum from “smart bombs” and spy satellites to “death rays” and infrared gadgets that see in the dark. But economic considerations coupled with the need to improve the quality of life have brought products of the discipline into the consumer marketplace as never before. Lasers are in use everywhere: reading videodiscs in living rooms, cutting steel in factories, scanning labels in supermarkets, and performing surgery in hospitals. Millions of optical display systems on clocks and calculators and computers are blinking all around the world. The almost exclusive use, for the last one hundred years, of electrical signals to handle and transmit data is now rapidly giving way to more efficient optical techniques. A far-reaching revolution in the methods of processing and communicating information is quietly taking place, a revolution that will continue to change our lives in the years ahead.

Profound insights are slow in coming. What few we have took over three thousand years to glean, even though the pace is ever quickening. It is marvelous indeed to watch the answer subtly change while the question immutably remains—*what is light?**

*For more reading on the history of Optics, see F. Cajori, *A History of Physics*, and V. Ronchi, *The Nature of Light*. Excerpts from a number of original papers can conveniently be found in W. F. Magie, *A Source Book in Physics*, and in M. H. Shamos, *Great Experiments in Physics*.

2 Wave Motion

The issue of the actual nature of light is central to a complete treatment of Optics, and we will struggle with it throughout this work. The straightforward question "Is light a wave phenomenon or a particle phenomenon?" is far more complicated than it might at first seem. For example, the essential feature of a particle is its localization; it exists in a well-defined, "small" region of space. Practically, we tend to take something familiar like a ball or a pebble and shrink it down in imagination until it becomes vanishingly small, and that's a "particle," or at least the basis for the concept of "particle." But a ball interacts with its environment; it has a gravitational field that interacts with the Earth (and the Moon, and Sun, etc.). This field, which spreads out into space—whatever *it* is—cannot be separated from the ball; it is an inextricable part of the ball just as it is an inextricable part of the definition of "particle." Real particles interact via fields, and, in a sense, the field is the particle and the particle is the field. That little conundrum is the domain of Quantum Field Theory, a discipline we'll talk more about later (p. 140). Suffice it to say now that if light is a stream of submicroscopic particles (photons), they are by no means "ordinary" miniball classical particles.

On the other hand, the essential feature of a wave is its nonlocalization. *A classical traveling wave is a self-sustaining disturbance of a medium, which moves through space transporting energy and momentum.* We tend to think of the ideal wave as a continuous entity that exists over an extended region. But when we look closely at real waves (such as waves on strings), we see composite phenomena comprising vast numbers of particles moving in concert. The media supporting these waves are atomic (i.e., particulate), and so the waves are not continuous entities in and of themselves. The only possible exception might be the electromagnetic wave. Conceptually, the classical electromagnetic wave (p. 46) is supposed to be a continuous entity, and *it* serves as the model for the very notion of wave as distinct from particle. But in the past century we found that the energy of an electromagnetic wave is *not* distributed continuously. The classical formulation of the electromagnetic theory of light, however wonderful it is on a macroscopic level, is profoundly wanting on a microscopic level. Einstein was the first to suggest that the electromagnetic wave, which we perceive macroscopically, is the statistical manifestation of a fundamentally granular underlying microscopic phenomenon (p. 53). In the subatomic domain, the classical concept of a physical wave is an illusion. Still, in the large-scale regime in which we ordinarily work, electromagnetic waves seem real enough and classical theory applies superbly well.

Because both the classical and quantum-mechanical treatments of light make use of the mathematical description of waves, this chapter lays out the basics of what both formalisms will need. The ideas we develop here will apply to all physical waves, from a surface tension ripple in a cup of tea to a pulse of light reaching us from some distant galaxy.

2.1 One-Dimensional Waves

An essential aspect of a traveling wave is that it is a self-sustaining disturbance of the medium through which it propagates. The most familiar waves, and the easiest to visualize (Fig. 2.1), are the mechanical waves, among which are waves on strings, surface waves on liquids, sound waves in the air, and compression waves in both solids and fluids. Sound waves are **longitudinal**—*the medium is displaced in the direction of motion of the wave*. Waves on a string (and electromagnetic waves) are **transverse**—*the medium is displaced in a direction perpendicular to that of the motion of the wave*. In all cases, although the energy-carrying disturbance advances through the medium, the individual participating atoms remain in the vicinity of their equilibrium positions: *the disturbance advances*, *not the material medium*. That's one of several crucial features of a wave that distinguishes it from a stream of particles. The wind blowing across a field sets up "waves of grain" that sweep by, even though each stalk only sways in place. Leonardo da Vinci seems to have been the first person to recognize that a wave does not transport the medium through which it travels, and it is precisely this property that allows waves to propagate at very great speeds.

What we want to do now is figure out the form the wave equation must have. To that end, envision some such disturbance ψ moving in the positive x-direction with a constant speed v. The specific nature of the disturbance is at the moment unimportant. It might be the vertical displacement of the string in Fig. 2.2 or the magnitude of an electric or magnetic field associated with an

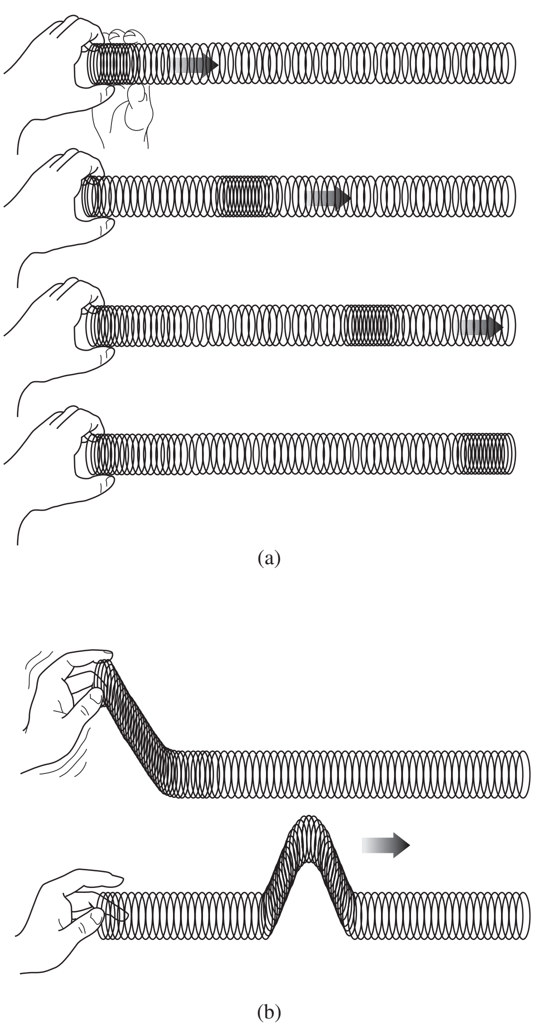


Figure 2.1 (*a*) A longitudinal wave in a spring. (*b*) A transverse wave in a spring.

electromagnetic wave (or even the quantum-mechanical probability amplitude of a matter wave).

Since the disturbance is moving, it must be a function of both position and time;

$$\psi(x, t) = f(x, t) \tag{2.1}$$

where $f(x, t)$ corresponds to some specific function or wave shape. This is represented in Fig. 2.3*a*, which shows a pulse traveling in the stationary coordinate system S at a speed v. The shape of the disturbance at any instant, say, $t = 0$, can be found by holding time constant at that value. In this case,

$$\psi(x, t)\big|_{t=0} = f(x, 0) = f(x) \tag{2.2}$$

represents the **profile** of the wave at that time. For example, if $f(x) = e^{-ax^2}$, where a is a constant, the profile has the shape of a bell; that is, it is a **Gaussian function**. (Squaring the x makes it symmetrical around the $x = 0$ axis.) Setting $t = 0$ is analogous to taking a "photograph" of the pulse as it travels by.

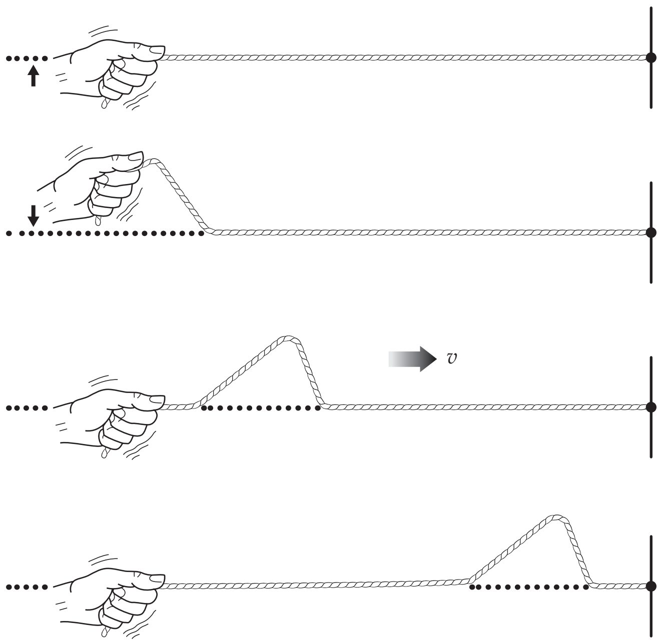


Figure 2.2 A wave on a string.

For the moment we limit ourselves to a wave that *does not change its shape* as it progresses through space. After a time t the pulse has moved along the x-axis a distance vt, but in all other respects it remains unaltered. We now introduce a coordinate system S', that travels along with the pulse (Fig. 2.3*b*) at the speed v. In this system ψ is no longer a function of time, and as we move along with S', we see a stationary constant profile described by Eq. (2.2). Here, the coordinate is x' rather than x, so that

$$\psi = f(x') \tag{2.3}$$

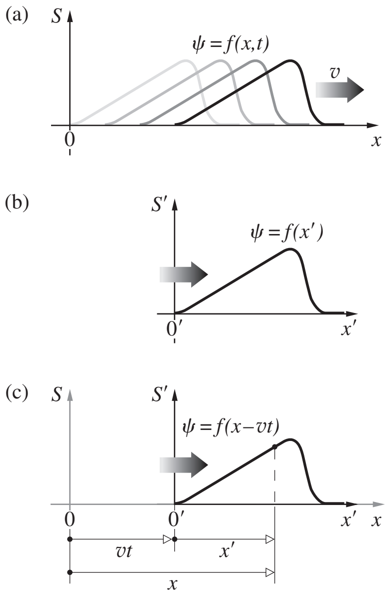


Figure 2.3 Moving reference frame.

The disturbance looks the same at any value of t in S' as it did at $t = 0$ in S when S and S' had a common origin (Fig. 2.3*c*).

We now want to rewrite Eq. (2.3) in terms of x to get the wave as it would be described by someone at rest in S. It follows from Fig. 2.3*c* that

$$x' = x - vt \tag{2.4}$$

and substituting into Eq. (2.3)

$$\psi(x, t) = f(x - vt) \tag{2.5}$$

This then represents the most general form of the one-dimensional **wavefunction**. To be more specific, we have only to choose a shape, Eq. (2.2), and then substitute $(x - vt)$ for x in $f(x)$. ***The resulting expression describes a wave having the desired profile, moving in the positive x-direction with a speed*** v. Thus, $\psi(x, t) = e^{-a(x-vt)^2}$ is a bell-shaped wave, a pulse.

To see how this all works in a bit more detail, let's unfold the analysis for a specific pulse, for example, $\psi(x) = 3/[10x^2 + 1] = f(x)$. That profile is plotted in Fig. 2.4*a*, and if it was a wave on a rope, ψ would be the vertical displacement and we might even replace it by the symbol y. Whether ψ represents displacement or pressure or electric field, we now have the profile of the disturbance. To turn $f(x)$ into $\psi(x, t)$, that is, to turn it into the description of a wave moving in the positive x-direction at a speed v, we replace x wherever it appears in $f(x)$ by $(x - vt)$, thereby yielding $\psi(x, t) = 3/[10(x - vt)^2 + 1]$. If v is arbitrarily set equal to, say, 1.0 m/s and the function is plotted successively at $t = 0$, $t = 1$ s, $t = 2$ s, and $t = 3$ s, we get Fig. 2.4*b*, which shows the pulse sailing off to the right at 1.0 m/s, just the way it's supposed to. Incidentally, had we substituted $(x + vt)$ for x in the profile function, the resulting wave would move off to the left.

If we check the form of Eq. (2.5) by examining ψ after an increase in time of Δt and a corresponding increase of $v\,\Delta t$ in x, we find

$$f[(x + v\,\Delta t) - v(t + \Delta t)] = f(x - vt)$$

and the profile is unaltered.

Similarly, if the wave was traveling in the negative x-direction, that is, to the left, Eq. (2.5) would become

$$\psi = f(x + vt), \quad \text{with} \quad v > 0 \tag{2.6}$$

We may conclude therefore that, regardless of the shape of the disturbance, the variables x and t must appear in the function as a unit, that is, as a single variable in the form $(x \mp vt)$. Equation (2.5) is often expressed equivalently as some function of $(t - x/v)$, since

$$f(x - vt) = F\left(-\frac{x - vt}{v}\right) = F(t - x/v) \tag{2.7}$$

The pulse shown in Fig. 2.2 and the disturbance described by Eq. (2.5) are spoken of as *one-dimensional* because the waves sweep over points lying on a line—it takes only one space variable to specify them. Don't be confused by the fact that in this particular case the rope happens to rise up into a second dimension. In contrast, a two-dimensional wave propagates out across a surface, like the ripples on a pond, and can be described by two space variables.

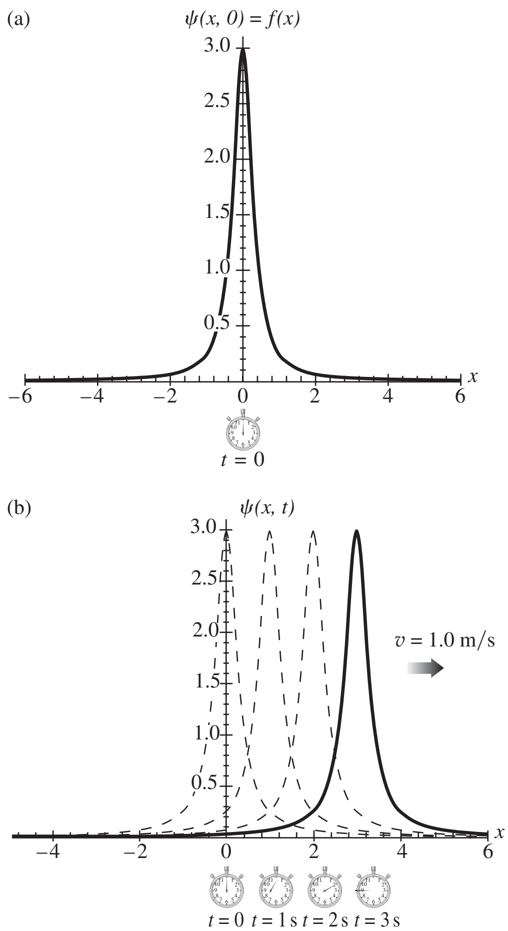


Figure 2.4 (*a*) The profile of a pulse given by the function $f(x) = 3/(10x^2 + 1)$. (*b*) The profile shown in (*a*) is now moving as a wave, $\psi(x, t) = 3/[10(x - vt)^2 + 1]$, to the right. We assign it a speed of 1 m/s and it advances in the positive x-direction.

2.1.1 The Differential Wave Equation

In 1747 Jean Le Rond d'Alembert introduced partial differential equations into the mathematical treatment of physics. That same year, he wrote an article on the motion of vibrating strings in which the so-called *differential wave equation* appears for the first time. This linear, homogeneous, second-order, partial differential equation is usually taken as the defining expression for physical waves in a lossless medium. There are lots of different kinds of waves, and each is described by its own wavefunction $\psi(x)$. Some are written in terms of pressure, or displacement, while others deal with electromagnetic fields, but remarkably all such wavefunctions are solutions of the same differential wave equation. The reason it's a *partial* differential equation is that the wave must be a function of several independent variables, namely, those of space and time. A *linear* differential equation is essentially one consisting of two or more terms, each composed of a constant multiplying a function $\psi(x)$ or its derivatives. The relevant point is that each such term must appear only to the first power; nor can there be any cross products of ψ with its derivatives, or of its derivatives. Recall that the *order* of a differential equation equals the order of the highest derivative in that equation. Furthermore, if a differential equation is of order N, the solution will contain N arbitrary constants.

We now derive the one-dimensional form of the wave equation guided by the foreknowledge (p. 14) that the most basic of waves traveling at a fixed speed requires two constants (amplitude and frequency or wavelength) to specify it, and this suggests second derivatives. Because there are two independent variables (here, x and t) we can take the derivative of $\psi(x, t)$ with respect to either x or t. This is done by just differentiating with respect to one variable and treating the other as if it were constant. The usual rules for differentiation apply, but to make the distinction evident the partial derivative is written as $\partial/\partial x$.

To relate the space and time dependencies of $\psi(x, t)$, take the partial derivative of $\psi(x, t) = f(x')$ with respect to x, holding t constant. Using $x' = x \mp vt$, and inasmuch as

$$\frac{\partial \psi}{\partial x} = \frac{\partial f}{\partial x}$$

$$\frac{\partial \psi}{\partial x} = \frac{\partial f}{\partial x'}\frac{\partial x'}{\partial x} = \frac{\partial f}{\partial x'} \tag{2.8}$$

because $$\frac{\partial x'}{\partial x} = \frac{\partial (x \mp vt)}{\partial x} = 1$$

Holding x constant, the partial derivative with respect to time is

$$\frac{\partial \psi}{\partial t} = \frac{\partial f}{\partial x'}\frac{\partial x'}{\partial t} = \frac{\partial f}{\partial x'}(\mp v) = \mp v\frac{\partial f}{\partial x'} \tag{2.9}$$

Combining Eqs. (2.8) and (2.9) yields

$$\frac{\partial \psi}{\partial t} = \mp v\frac{\partial \psi}{\partial x}$$

This says that the rate of change of ψ with t and with x are equal, to within a multiplicative constant, as shown in Fig. 2.5. The second partial derivatives of Eqs. (2.8) and (2.9) are

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{\partial^2 f}{\partial x'^2} \tag{2.10}$$

and $$\frac{\partial^2 \psi}{\partial t^2} = \frac{\partial}{\partial t}\left(\mp v\frac{\partial f}{\partial x'}\right) = \mp v\frac{\partial}{\partial x'}\left(\frac{\partial f}{\partial t}\right)$$

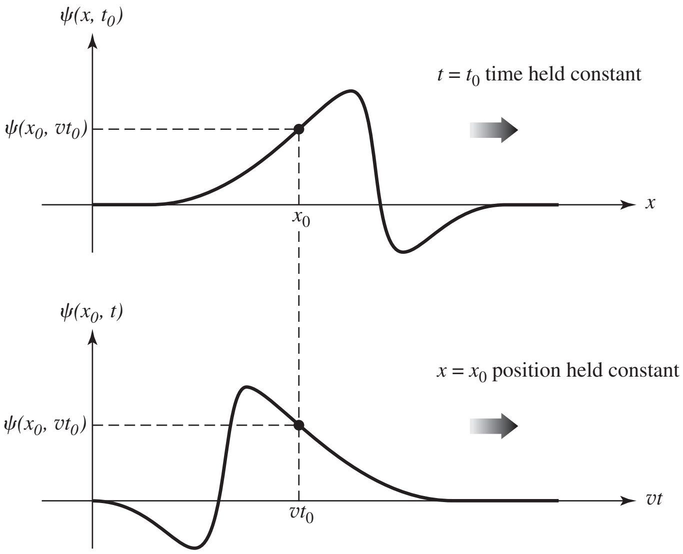


Figure 2.5 Variation of ψ with x and t.

Since

$$\frac{\partial \psi}{\partial t} = \frac{\partial f}{\partial t}$$

$$\frac{\partial^2 \psi}{\partial t^2} = \mp v\frac{\partial}{\partial x'}\left(\frac{\partial \psi}{\partial t}\right)$$

It follows, using Eq. (2.9), that

$$\frac{\partial^2 \psi}{\partial t^2} = v^2\frac{\partial^2 f}{\partial x'^2}$$

Combining this with Eq. (2.10), we obtain

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \psi}{\partial t^2} \tag{2.11}$$

which is the desired one-dimensional **differential wave equation**.

EXAMPLE 2.1

The wave shown in Fig. 2.4 is given by

$$\psi(x, t) = \frac{3}{[10(x - vt)^2 + 1]}$$

Show, using brute force, that this is a solution to the one-dimensional differential wave equation.

SOLUTION

$$\frac{\partial^2 \psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \psi}{\partial t^2}$$

Continued

Differentiating with respect to x:

$$\frac{\partial\psi}{\partial x} = \frac{\partial}{\partial x}\left[\frac{3}{10(x - vt)^2 + 1}\right]$$

$$\frac{\partial\psi}{\partial x} = (-1)\,3[10(x - vt)^2 + 1]^{-2}\,20(x - vt)$$

$$\frac{\partial\psi}{\partial x} = (-1)\,60[10(x - vt)^2 + 1]^{-2}(x - vt)$$

$$\frac{\partial^2\psi}{\partial x^2} = \frac{-60(-2)\,20(x - vt)(x - vt)}{[10(x - vt)^2 + 1]^3} - \frac{60}{[10(x - vt)^2 + 1]^2}$$

$$\frac{\partial^2\psi}{\partial x^2} = \frac{2400(x - vt)^2}{[10(x - vt)^2 + 1]^3} - \frac{60}{[10(x - vt)^2 + 1]^2}$$

Differentiating with respect to t:

$$\frac{\partial\psi}{\partial t} = \frac{\partial}{\partial t}\left[\frac{3}{10(x - vt)^2 + 1}\right]$$

$$\frac{\partial\psi}{\partial t} = (-1)\,3[10(x - vt)^2 + 1]^{-2}\,20(-v)(x - vt)$$

$$\frac{\partial\psi}{\partial t} = 60v(x - vt)\,[10(x - vt)^2 + 1]^{-2}$$

$$\frac{\partial^2\psi}{\partial t^2} = \frac{60v(x - vt)(-2)\,20(x - vt)(-v)}{[10(x - vt)^2 + 1]^3} + \frac{-60v^2}{[10(x - vt)^2 + 1]^2}$$

$$\frac{\partial^2\psi}{\partial t^2} = \frac{2400v^2(x - vt)^2}{[10(x - vt)^2 + 1]^3} - \frac{60v^2}{[10(x - vt)^2 + 1]^2}$$

Hence

$$\frac{\partial^2\psi}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi}{\partial t^2}$$

Note that Eq. (2.11) is a so-called *homogeneous* differential equation; it doesn't contain a term (such as a "force" or a "source") involving only independent variables. In other words, ψ is in each term of the equation, and that means that if ψ is a solution any multiple of ψ will also be a solution. Equation 2.11 is ***the wave equation for undamped systems*** that do not contain sources in the region under consideration. The effects of damping can be described by adding in a $\partial\psi/\partial t$ term to form a more general wave equation, but we'll come back to that later (p. 73).

As a rule, partial differential equations arise when the system being described is continuous. The fact that time is one of the independent variables reflects the continuity of temporal change in the process under analysis. Field theories, in general, treat continuous distributions of quantities in space and time and so take the form of partial differential equations. Maxwell's formulation of electromagnetism, which is a field theory, yields a variation of Eq. (2.11), and from that the concept of the electromagnetic wave arises in a completely natural way (p. 46).

We began this discussion with the special case of waves that have a constant shape as they propagate, even though, as a rule, waves don't maintain a fixed profile. Still, that simple assumption has led us to the general formulation, the differential wave equation. If a function that represents a wave is a solution of that equation, it will at the same time be a function of $(x \mp vt)$—specifically, one that is twice differentiable (in a nontrivial way) with respect to both x and t.

EXAMPLE 2.2

Does the function

$$\psi(x, t) = \exp\,[(-4ax^2 - bt^2 + 4\sqrt{ab}\,xt)]$$

where in a and b are constants, describe a wave? If so, what is its speed and direction of propagation?

SOLUTION

Factor the bracketed term:

$$\psi(x, t) = \exp\,[-a(4x^2 + bt^2/a - 4\sqrt{b/a}\,xt)]$$

$$\psi(x, t) = \exp\,[-4a(x - \sqrt{b/4a}\,t)^2]$$

That's a twice differentiable function of $(x - vt)$, so it is a solution of Eq. (2.11) and therefore describes a wave. Here $v = \frac{1}{2}\sqrt{b/a}$ and it travels in the positive x-direction.

2.2 Harmonic Waves

Let's now examine the simplest waveform, one for which the profile is a sine or cosine curve. These are variously known as sinusoidal waves, simple harmonic waves, or more succinctly as **harmonic waves**. We shall see in Chapter 7 that any wave shape can be synthesized by a superposition of harmonic waves, and they therefore take on a special significance.

Choose as the profile the simple function

$$\psi(x, t)\big|_{t=0} = \psi(x) = A \sin kx = f(x) \tag{2.12}$$

where k is a positive constant known as the **propagation number**. It's necessary to introduce the constant k simply because we cannot take the sine of a quantity that has physical units. The sine is the ratio of two lengths and is therefore unitless. Accordingly, kx is properly in radians, which is not a real physical unit. The sine varies from $+1$ to -1 so that the maximum value of $\psi(x)$ is A. This maximum disturbance is known as the **amplitude** of the wave (Fig. 2.6). To transform Eq. (2.12) into a *progressive wave* traveling at speed v in the

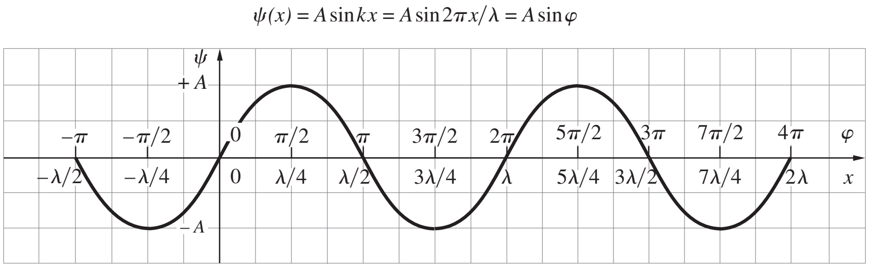


Figure 2.6 A harmonic function, which serves as the profile of a harmonic wave. One wavelength corresponds to a change in phase φ of 2π rad.

positive x-direction, we need merely replace x by $(x - vt)$, in which case

$$\psi(x, t) = A \sin k(x - vt) = f(x - vt) \tag{2.13}$$

This is clearly a solution of the differential wave equation (see Problem 2.24). Holding either x or t fixed results in a sinusoidal disturbance; the wave is periodic in both space and time. The **spatial period** is known as the **wavelength** and is denoted by λ. Wavelength is *the number of units of length per wave*. The customary measure of λ is the nanometer, where 1 nm $= 10^{-9}$ m, although the micron (1 μm $= 10^{-6}$ m) is often used and the older angstrom (1 Å $= 10^{-10}$ m) can still be found in the literature. An increase or decrease in x by the amount λ should leave ψ unaltered, that is,

$$\psi(x, t) = \psi(x \pm \lambda, t) \tag{2.14}$$

In the case of a harmonic wave, this is equivalent to altering the argument of the sine function by $\pm 2\pi$. Therefore,

$$\sin k(x - vt) = \sin k[(x \pm \lambda) - vt] = \sin [k(x - vt) \pm 2\pi]$$

and so

$$|k\lambda| = 2\pi$$

or, since both k and λ are positive numbers,

$$k = 2\pi/\lambda \tag{2.15}$$

Figure 2.6 shows how to plot the profile given by Eq. (2.12) in terms of λ. Here φ is the argument of the sine function, also called the **phase**. In other words, $\psi(x) = A \sin \varphi$. Notice that $\psi(x) = 0$ whenever $\sin \varphi = 0$, which happens when $\varphi = 0, \pi, 2\pi, 3\pi$, and so on. That occurs at $x = 0, \lambda/2, \lambda$, and $3\lambda/2$, respectively.

In an analogous fashion to the above discussion of λ, we now examine the **temporal period**, τ. This is the amount of time it takes for one complete wave to pass a stationary observer. In this case, it is the repetitive behavior of the wave in time that is of interest, so that

$$\psi(x, t) = \psi(x, t \pm \tau) \tag{2.16}$$

and

$$\sin k(x - vt) = \sin k[x - v(t \pm \tau)]$$

$$\sin k(x - vt) = \sin [k(x - vt) \pm 2\pi]$$

Therefore,

$$|kv\tau| = 2\pi$$

But these are all positive quantities; hence

$$kv\tau = 2\pi \tag{2.17}$$

or

$$\frac{2\pi}{\lambda} v\tau = 2\pi$$

from which it follows that

$$\tau = \lambda/v \tag{2.18}$$

The period is *the number of units of time per wave* (Fig. 2.7), the inverse of which is the **temporal frequency** ν, or *the number of waves per unit of time* (i.e., per second). Thus,

$$\nu \equiv 1/\tau$$

in units of cycles per second or Hertz. Equation (2.18) then becomes

$$v = \nu\lambda \tag{2.19}$$

Imagine that you are at rest and a harmonic wave on a string is progressing past you. The number of waves that sweep by per second is ν, and the length of each is λ. In 1.0 s, the overall length of the disturbance that passes you is the product $\nu\lambda$. If, for example, each wave is 2.0 m long and they come at a rate of 5.0 per second, then in 1.0 s, 10 m of wave fly by. This is just what we mean by the speed of the wave (v)—the rate, in m/s, at which it advances. Said slightly differently, because a length of wave λ passes by in a time τ, its speed must equal $\lambda/\tau = \nu\lambda$. Incidentally, Newton derived this relationship in the *Principia* (1687) in a section called "To find the velocity of waves."

Two other quantities are often used in the literature of wave motion. One is the **angular temporal frequency**

$$\omega \equiv 2\pi/\tau = 2\pi\nu \tag{2.20}$$

given in units of radians per second. The other, which is important in spectroscopy, is the **wave number** or **spatial frequency**

$$\kappa \equiv 1/\lambda \qquad (2.21)$$

measured in inverse meters. In other words, κ is *the number of waves per unit of length* (i.e., per meter). All of these quantities apply equally well to waves that are not harmonic, as long as each such wave is made up of a single regularly repeated **profile-element** (Fig. 2.8).

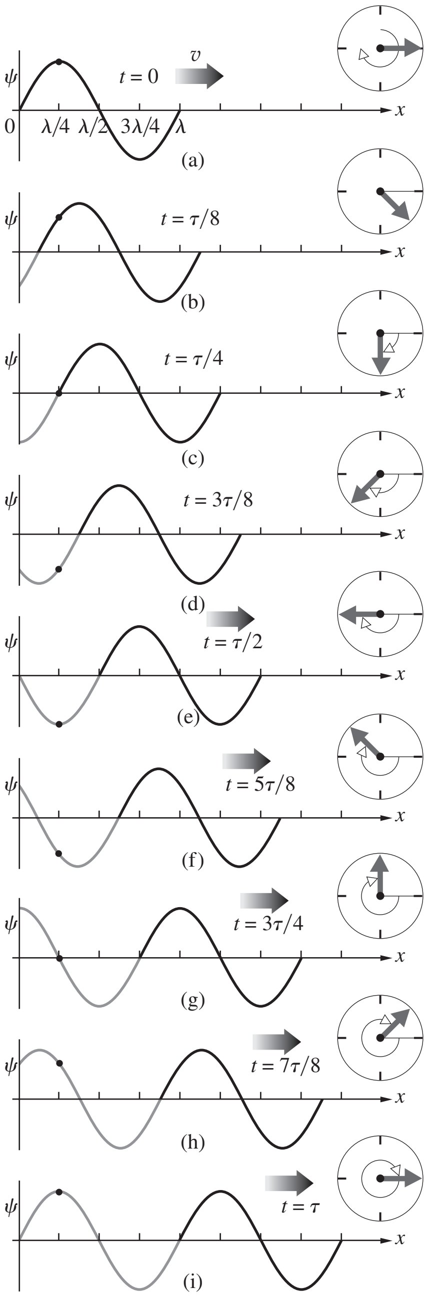


Figure 2.7 A harmonic wave moving along the x-axis during a time of one period. Note that if this is a picture of a rope any one point on it only moves vertically. We'll discuss the significance of the rotating arrow in Section 2.6. For the moment observe that the projection of that arrow on the vertical axis equals the value of ψ at $x = 0$.

EXAMPLE 2.3

A Nd:YAG laser puts out a beam of 1.06 μm electromagnetic radiation in vacuum. Determine (a) the beam's temporal frequency; (b) its temporal period; and (c) its spatial frequency.

SOLUTION

(a) Since $v = \nu\lambda$

$$\nu = \frac{v}{\lambda} = \frac{2.99 \times 10^8 \text{ m/s}}{1.06 \times 10^{-6} \text{ m}} = 2.82 \times 10^{14}\,\text{Hz}$$

or $\nu = 282$ TH. **(b)** The temporal period is $\tau = 1/\nu = 1/2.82 \times 10^{14}\,\text{Hz} = 3.55 \times 10^{-15}\text{s}$, or 3.55 fs. **(c)** The spatial frequency is $\kappa = 1/\lambda = 1/1.06 \times 10^{-6}\,\text{m} = 943 \times 10^3 \text{m}^{-1}$, that is, 943 thousand waves per meter.

Using the above definitions we can write a number of equivalent expressions for the traveling harmonic wave:

$$\psi = A \sin k(x \mp vt) \qquad [2.13]$$

$$\psi = A \sin 2\pi \left(\frac{x}{\lambda} \mp \frac{t}{\tau}\right) \qquad (2.22)$$

$$\psi = A \sin 2\pi (\kappa x \mp \nu t) \qquad (2.23)$$

$$\psi = A \sin (kx \mp \omega t) \qquad (2.24)$$

$$\psi = A \sin 2\pi\nu \left(\frac{x}{v} \mp t\right) \qquad (2.25)$$

Of these, Eqs. (2.13) and (2.24) will be encountered most frequently. Note that all these idealized waves are of infinite extent. That is, for any fixed value of t, there is no mathematical limitation on x, which varies from $-\infty$ to $+\infty$. Each such wave has a single constant frequency and is therefore **monochromatic**

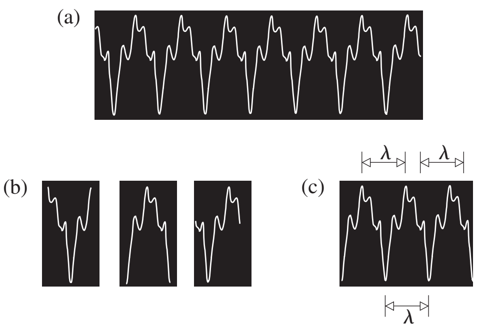


Figure 2.8 (*a*) The waveform produced by a saxophone. Imagine any number of profile-elements (*b*) that, when repeated, create the waveform (*c*). The distance over which the wave repeats itself is called the wavelength, λ.

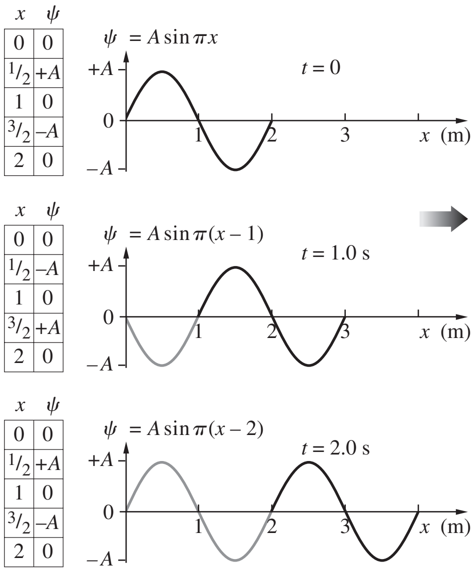


Figure 2.9 A progressive wave of the form $\psi(x, t) = A \sin k(x - vt)$, moving to the right at a speed of 1.0 m/s.

or, even better, **monoenergetic**. Real waves are never monochromatic. Even a perfect sinusoidal generator cannot have been operating forever. Its output will unavoidably contain a range of frequencies, albeit a small one, just because the wave does not extend back to $t = -\infty$. Thus all waves comprise a band of frequencies, and when that band is narrow the wave is said to be **quasimonochromatic**.

Before we move on, let's put some numbers into Eq. (2.13) and see how to deal with each term. To that end, arbitrarily let $v = 1.0$ m/s and $\lambda = 2.0$ m. Then the wavefunction

$$\psi = A \sin \frac{2\pi}{\lambda}(x - vt)$$

in SI units becomes

$$\psi = A \sin \pi(x - t)$$

Figure 2.9 shows how the wave progresses to the right at 1.0 m/s as the time goes from $t = 0$ [whereupon $\psi = A \sin \pi x$] to $t = 1.0$ s [whereupon $\psi = A \sin \pi(x - 1.0)$] to $t = 2.0$ s [whereupon $\psi = A \sin \pi(x - 2.0)$].

EXAMPLE 2.4

Consider the function

$$\psi(y, t) = (0.040) \sin 2\pi\left(\frac{y}{6.0 \times 10^{-7}} + \frac{t}{2.0 \times 10^{-15}}\right)$$

where everything is in appropriate SI units. (a) Does this expression have the form of a wave? Explain. If so, determine its (b) frequency, (c) wavelength, (d) amplitude, (e) direction of propagation, and (f) speed.

SOLUTION

(a) Factor $1/6.0 \times 10^{-7}$ from the term in parentheses and it becomes clear that $\psi(y, t)$ is a twice differentiable function of $(y \pm vt)$, so it does represent a harmonic wave. **(b)** We could also simply use Eq. (2.22)

$$\psi = A \sin 2\pi\left(\frac{x}{\lambda} + \frac{t}{\tau}\right)$$

whereupon it follows that the period $\tau = 2.0 \times 10^{-15}$ s. Hence $\nu = 1/\tau = 5.0 \times 10^{14}$ Hz. **(c)** The wavelength is $\lambda = 6.0 \times 10^{-7}$m. **(d)** The amplitude is $A = 0.040$. **(e)** The wave travels in the negative y direction. **(f)** The speed $v = \nu\lambda = (5.0 \times 10^{14}\text{ Hz})(6.0 \times 10^{-7}\text{m}) = 3.0 \times 10^{8}$ m/s. Alternatively if we factor $1/6.0 \times 10^{-7}$ from the parentheses the speed becomes $6.0 \times 10^{-7}/2.0 \times 10^{-15} = 3.0 \times 10^{8}$ m/s.

Spatial Frequency

Periodic waves are structures that move through space and time displaying wavelengths, temporal periods, and temporal frequencies; they undulate in time. In modern Optics we are also interested in stationary periodic distributions of information that conceptually resemble snapshots of waves. Indeed, later on in Chapters 7 and 11 we'll see that ordinary images of buildings and people and picket fences can all be synthesized using periodic functions in space, utilizing a process called Fourier analysis.

What we need to keep in mind here is that optical information can be spread out in space in a periodic way much like a wave profile. To make the point we convert the sinusoid of Fig. 2.6 into a diagram of smoothly varying brightness, namely, Fig. 2.10. This sinusoidal brightness variation has a *spatial period* of several millimeters (measured, e.g., from bright peak to bright peak). Here a pair of black and white bands corresponds to one "wavelength," that is, so many millimeters (or centimeters) per black and white pair. The inverse of that—one over the

Figure 2.10 A sinusoidal brightness distribution of relatively low spatial frequency.

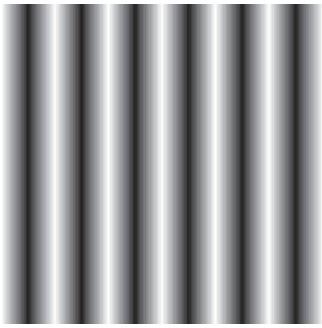


Figure 2.11 A sinusoidal brightness distribution of relatively high spatial frequency.

spatial period—is the *spatial frequency*, the number of black and white pairs per millimeter (or per centimeter). Figure 2.11 depicts a similar pattern with a shorter spatial period and a higher spatial frequency. These are single spatial frequency distributions akin to monochromatic profiles in the time domain. As we go on we'll see how images can be built up out of the superposition of individual spatial frequency contributions just like those of Figs. 2.10 and 2.11.

2.3 Phase and Phase Velocity

Examine any one of the harmonic wavefunctions, such as

$$\psi(x, t) = A \sin (kx - \omega t) \tag{2.26}$$

The entire argument of the sine is the phase φ of the wave, where

$$\varphi = (kx - \omega t) \tag{2.27}$$

At $t = x = 0$,

$$\psi(x, t)\big|_{\substack{x=0 \\ t=0}} = \psi(0, 0) = 0$$

which is certainly a special case. More generally, we can write

$$\psi(x, t) = A \sin (kx - \omega t + \varepsilon) \tag{2.28}$$

where ε is the **initial phase**. To get a sense of the physical meaning of ε, imagine that we wish to produce a progressive harmonic wave on a stretched string, as in Fig. 2.12. In order to generate harmonic waves, the hand holding the string would have to move such that its vertical displacement y was proportional to the negative of its acceleration, that is, in simple harmonic motion (see Problem 2.27). But at $t = 0$ and $x = 0$, the hand certainly need not be on the x-axis about to move downward, as in Fig. 2.12. It could, of course, begin its motion on an upward swing, in which case $\varepsilon = \pi$, as in Fig. 2.13. In this latter case,

$$\psi(x, t) = y(x, t) = A \sin (kx - \omega t + \pi)$$

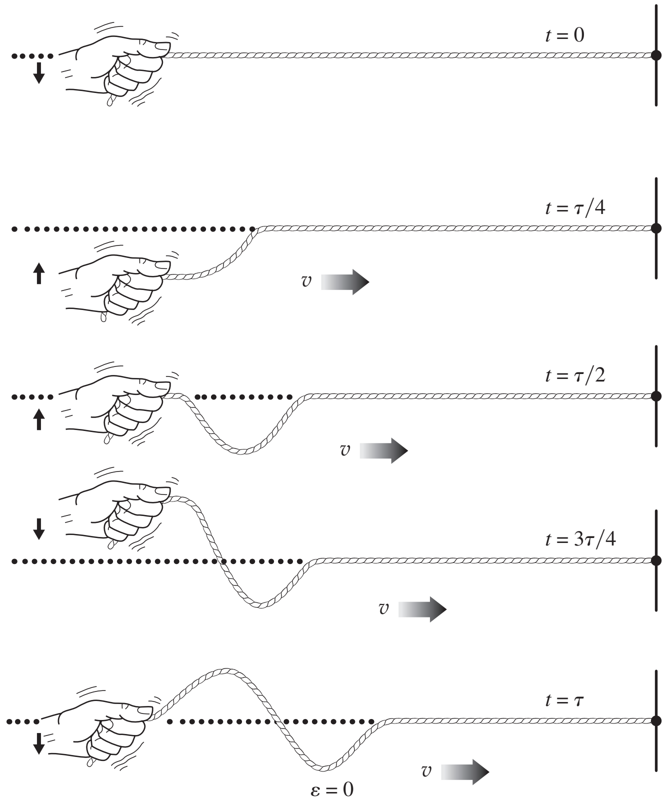


Figure 2.12 With $\varepsilon = 0$ note that at $x = 0$ and $t = \tau/4 = \pi/2\omega$, $y = A \sin(-\pi/2) = -A$.

which is equivalent to

$$\psi(x, t) = A \sin (\omega t - kx) \tag{2.29}$$

or
$$\psi(x, t) = A \cos\left(\omega t - kx - \frac{\pi}{2}\right)$$

The initial phase angle is just the constant contribution to the phase arising at the generator and is independent of how far in space, or how long in time, the wave has traveled.

The phase in Eq. (2.26) is $(kx - \omega t)$, whereas in Eq. (2.29) it's $(\omega t - kx)$. Nonetheless, both of these equations describe waves moving in the positive x-direction that are otherwise identical except for a relative phase difference of π. As is often the case, when the initial phase is of no particular significance in a given situation, either Eq. (2.26) or (2.29) or, if you like, a cosine function can be used to represent the wave. Even so, in some situations one expression for the phase may be mathematically more appealing than another; the literature abounds with both, and so we will use both.

The phase of a disturbance such as $\psi(x, t)$ given by Eq. (2.28) is

$$\varphi(x, t) = (kx - \omega t + \varepsilon)$$

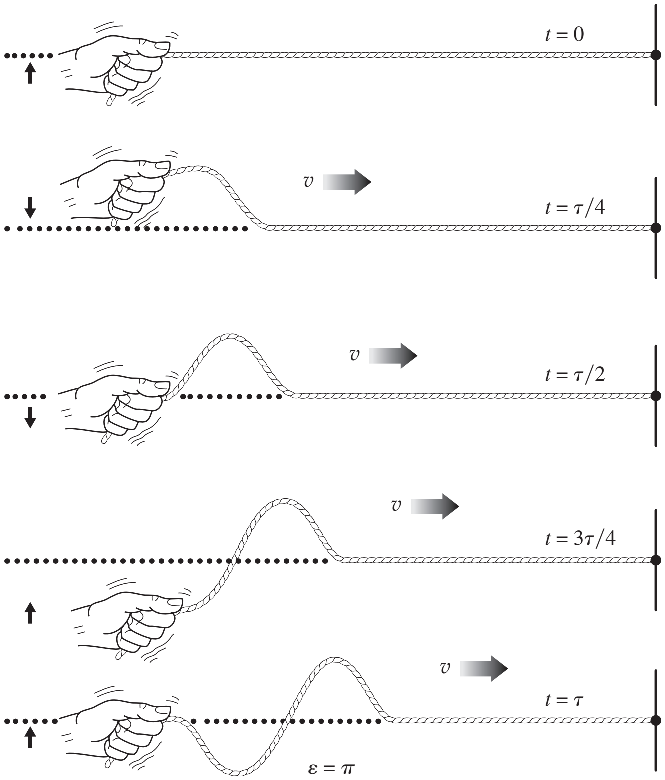


Figure 2.13 With $\varepsilon = \pi$ note that at $x = 0$ and $t = \tau/4$, $y = A \sin(\pi/2) = A$.

and is obviously a function of x and t. In fact, the partial derivative of φ with respect to t, holding x constant, is the *rate-of-change of phase with time*, or

$$\left|\left(\frac{\partial\varphi}{\partial t}\right)_x\right| = \omega \tag{2.30}$$

The rate-of-change of phase at any fixed location is the angular frequency of the wave, the rate at which a point on the rope in Fig. 2.12 oscillates up and down. That point must go through the same number of cycles per second as the wave. For each cycle, φ changes by 2π. The quantity ω is the number of radians the phase sweeps through per second. The quantity k is the number of radians the phase sweeps through per meter.

Similarly, the *rate-of-change of phase with distance*, holding t constant, is

$$\left|\left(\frac{\partial\varphi}{\partial x}\right)_t\right| = k \tag{2.31}$$

These two expressions should bring to mind an equation from the theory of partial derivatives, one used frequently in Thermodynamics, namely,

$$\left(\frac{\partial x}{\partial t}\right)_\varphi = \frac{-(\partial\varphi/\partial t)_x}{(\partial\varphi/\partial x)_t} \tag{2.32}$$

The term on the left represents the *speed of propagation of the condition of constant phase*. Imagine a harmonic wave and choose any point on the profile, for example, a crest of the wave. As the wave moves through space, the displacement y of the crest remains fixed. Since the only variable in the harmonic wavefunction is the phase, it too must be constant for that moving point. That is, the phase is fixed at such a value as to yield the constant y corresponding to the chosen point. The point moves along with the profile at the speed v, and so too does the condition of constant phase.

Taking the appropriate partial derivatives of φ as given, for example, by Eq. (2.29) and substituting them into Eq. (2.32), we get

$$\left(\frac{\partial x}{\partial t}\right)_\varphi = \pm\frac{\omega}{k} = \pm v \tag{2.33}$$

The units of ω are rad/s and the units of k are rad/m. The units of ω/k are appropriately m/s. This is the *speed* at which the profile moves and is known commonly as the **phase velocity** of the wave. The phase velocity is accompanied by a positive sign when the wave moves in the direction of increasing x and a negative one in the direction of decreasing x. This is consistent with our development of v as the magnitude of the wave velocity: $v > 0$.

Consider the idea of the propagation of constant phase and how it relates to any one of the harmonic wave equations, say,

$$\psi = A \sin k(x \mp vt)$$

with
$$\varphi = k(x - vt) = \text{constant}$$

As t increases, x must increase. Even if $x < 0$ so that $\varphi < 0$, x must increase (i.e., become less negative). Here, then, the condition of constant phase moves in the direction of increasing x. As long as the two terms in the phase subtract from each other, the wave travels in the positive x-direction. On the other hand, for

$$\varphi = k(x + vt) = \text{constant}$$

as t increases x can be positive and decreasing or negative and becoming more negative. In either case, the constant-phase condition moves in the decreasing x-direction.

EXAMPLE 2.5

A propagating wave at time $t = 0$ can be expressed in SI units as $\psi(y, 0) = (0.030\,\text{m})\cos(\pi y/2.0)$. The disturbance moves in the negative y-direction with a phase velocity of 2.0 m/s. Write an expression for the wave at a time of 6.0 s.

SOLUTION

Write the wave in the form

$$\psi(y, t) = A\cos 2\pi\left(\frac{y}{\lambda} \pm \frac{t}{\tau}\right)$$

Continued

Here $A = 0.030\,\text{m}$ and

$$\psi(y, 0) = (0.030\,\text{m}) \cos 2\pi\left(\frac{y}{4.0}\right)$$

We need the period and since $\lambda = 4.0\,\text{m}$, $v = \nu\lambda = \lambda/\tau$; $\tau = \lambda/v = (4.0\,\text{m})/(2.0\,\text{m/s}) = 2.0\,\text{s}$. Hence

$$\psi(y, t) = (0.030\,\text{m}) \cos 2\pi\left(\frac{y}{4.0} + \frac{t}{2.0}\right)$$

The positive sign in the phase indicates motion in the negative y-direction. At $t = 6.0\,\text{s}$

$$\psi(y, 6.0) = (0.030\,\text{m}) \cos 2\pi\left(\frac{y}{4.0} + 3.0\right)$$

Any point on a harmonic wave having a fixed magnitude moves such that $\varphi(x, t)$ is constant in time, in other words, $d\varphi(x, t)/dt = 0$ or, alternatively, $d\psi(x, t)/dt = 0$. This is true for all waves, periodic or not, and it leads (Problem 2.34) to the expression

$$\pm v = \frac{-(\partial\psi/\partial t)_x}{(\partial\psi/\partial x)_t} \tag{2.34}$$

which can be used to conveniently provide v when we have $\psi(x, t)$. Note that because v is always a positive number, when the ratio on the right turns out negative the motion is in the negative x-direction.

Figure 2.14 depicts a source producing hypothetical two-dimensional waves on the surface of a liquid. The essentially sinusoidal nature of the disturbance, as the medium rises and falls, is evident in the diagram. But there is another useful way to envision what's happening. *The curves connecting all the points with a given phase form a set of concentric circles*. Furthermore, given that A is everywhere constant at any one distance from the source, if φ is constant over a circle, ψ too must be constant over that circle. In other words, all the corresponding peaks and troughs fall on circles, and we speak of these as *circular waves*, each of which expands outward at the speed v.

Figure 2.14 Circular waves. (E.H.)

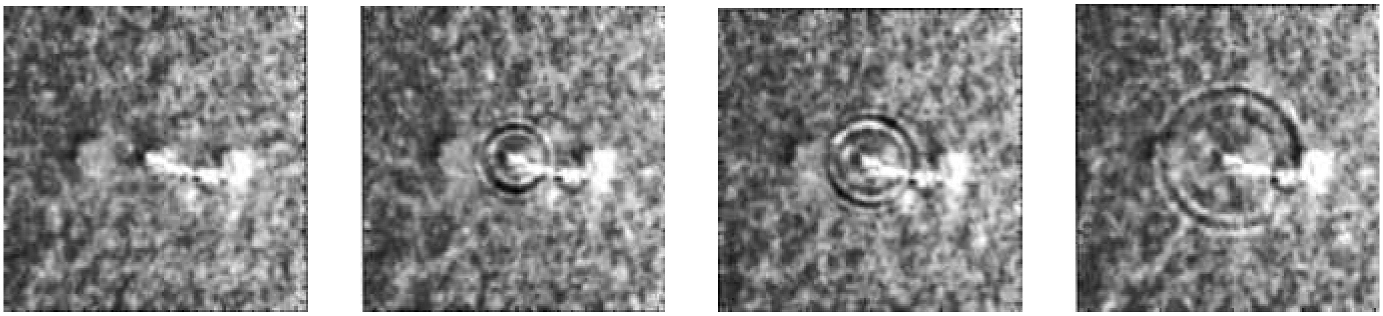
2.4 The Superposition Principle

The form of the differential wave equation [Eq. (2.11)] reveals an intriguing property of waves, one that is quite unlike the behavior of a stream of classical particles. Suppose that the wavefunctions ψ_1 and ψ_2 are each separate solutions of the wave equation; it follows that $(\psi_1 + \psi_2)$ is also a solution. This is known as the **Superposition Principle**, and it can easily be proven, since it must be true that

$$\frac{\partial^2\psi_1}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi_1}{\partial t^2} \quad \text{and} \quad \frac{\partial^2\psi_2}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi_2}{\partial t^2}$$

Adding these yields

$$\frac{\partial^2\psi_1}{\partial x^2} + \frac{\partial^2\psi_2}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2\psi_1}{\partial t^2} + \frac{1}{v^2}\frac{\partial^2\psi_2}{\partial t^2}$$



A solar flare on the Sun caused circular seismic ripples to flow across the surface. (NASA)

and so
$$\frac{\partial^2}{\partial x^2}(\psi_1 + \psi_2) = \frac{1}{v^2}\frac{\partial^2}{\partial t^2}(\psi_1 + \psi_2)$$

which establishes that $(\psi_1 + \psi_2)$ is indeed a solution. What this means is that when two separate waves arrive at the same place in space wherein they overlap, they will simply add to (or subtract from) one another without permanently destroying or disrupting either wave. ***The resulting disturbance at each point in the region of overlap is the algebraic sum of the individual constituent waves at that location*** (Fig. 2.15). Once having passed through the region where the two waves coexist, each will move out and away unaffected by the encounter.

Keep in mind that we are talking about a *linear* superposition of waves, a process that's widely valid and the most commonly encountered. Nonetheless, it is also possible for the wave amplitudes to be large enough to drive the medium in a nonlinear fashion (p. 659). For the time being we'll concentrate on the linear differential wave equation, which results in a linear Superposition Principle.

Much of Optics involves the superposition of waves in one way or another. Even the basic processes of reflection and refraction are manifestations of the scattering of light from countless atoms (p. 88), a phenomenon that can only be treated satisfactorily in terms of the overlapping of waves. It therefore becomes crucial that we understand the process, at least qualitatively, as soon as possible. Consequently, carefully examine the two coexisting waves in Fig. 2.15. At every point (i.e., every value of kx) we simply add ψ_1 and ψ_2, either of which could be positive or negative. As a quick check, keep in mind that wherever either constituent wave is zero (e.g., $\psi_1 = 0$), the resultant disturbance equals the value of the other nonzero constituent wave ($\psi = \psi_2$), and those two curves cross at that location (e.g., at $kx = 0$ and $+3.14$ rad). On the other hand, $\psi = 0$ wherever the two constituent waves have equal magnitudes and opposite signs (e.g., at $kx = +2.67$ rad). Incidentally, notice how a relative *positive* phase difference of 1.0 rad between the two curves shifts ψ_2 to the *left* with respect to ψ_1 by 1.0 rad.

Developing the illustration a bit further, Fig. 2.16 shows how the resultant arising from the superposition of two nearly equal-amplitude waves depends on the *phase-angle difference* between them. In Fig. 2.16*a* the two constituent waves have the same phase; that is, their phase-angle difference is zero, and they are said to be **in-phase**; they rise and fall in-step, reinforcing each other. The composite wave, which then has a substantial amplitude, is sinusoidal with the same frequency and wavelength as the component waves (p. 285). Following the sequence of the drawings, we see that the resultant amplitude diminishes as the phase-angle difference increases until, in Fig. 2.16*d*, it almost vanishes when that difference equals π. The waves are then said to be 180° **out-of-phase**. The fact that waves which are out-of-phase tend to diminish each other has given the name **interference** to the whole phenomenon.

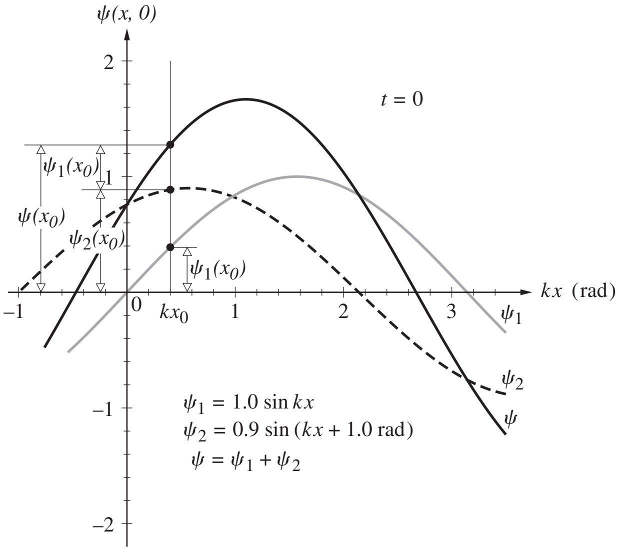


Figure 2.15 The superposition of two equal-wavelength sinusoids ψ_1 and ψ_2, having amplitudes A_1 and A_2, respectively. The resultant, ψ, is a sinusoid with the same wavelength, which at every point equals the algebraic sum of the constituent sinusoids. Thus at $x = x_0$, $\psi(x_0) = \psi_1(x_0) + \psi_2(x_0)$; the magnitudes add. The amplitude of ψ is A and it can be determined in several ways; see Fig. 2.19.

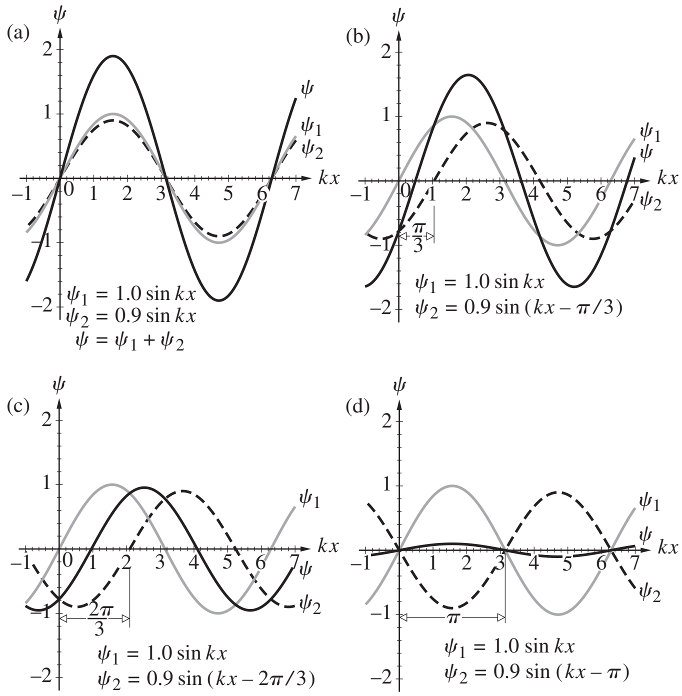


Figure 2.16 The superposition of two sinusoids with amplitudes of $A_1 = 1.0$ and $A_2 = 0.9$. In (*a*) they are in-phase. In (*b*) ψ_1 leads ψ_2 by $\pi/3$. In (*c*) ψ_1 leads ψ_2 by $2\pi/3$. And (*d*) ψ_1 and ψ_2 are out-of-phase by π and almost cancel each other. To see how the amplitudes can be determined, go to Fig. 2.20.

Water waves overlapping and interfering. (E.H.)

2.5 The Complex Representation

As we develop the analysis of wave phenomena, it will become evident that the sine and cosine functions that describe harmonic waves can be somewhat awkward for our purposes. The expressions formulated will sometimes be rather involved and the trigonometric manipulations required to cope with them will be even more unattractive. The complex-number representation offers an alternative description that is mathematically simpler to process. In fact, complex exponentials are used extensively in both Classical and Quantum Mechanics, as well as in Optics.

The complex number $\tilde{z}$ has the form

$$\tilde{z} = x + iy \tag{2.35}$$

where $i = \sqrt{-1}$. The real and imaginary parts of $\tilde{z}$ are, respectively, x and y, where both x and y are themselves real numbers. This is illustrated graphically in the Argand diagram in Fig. 2.17*a*. In terms of polar coordinates (r, θ),

$$x = r\cos\theta \qquad y = r\sin\theta$$

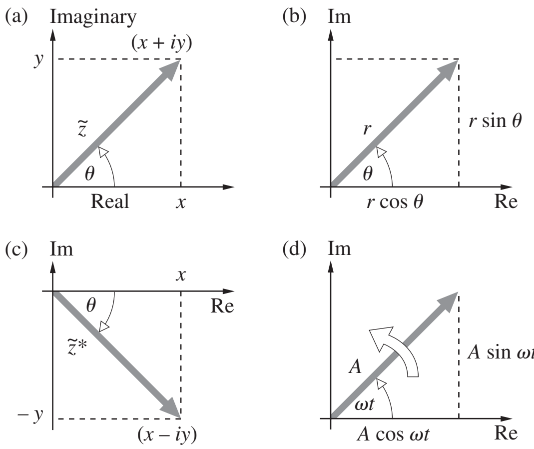


Figure 2.17 An Argand diagram is a representation of a complex number in terms of its real and imaginary components. This can be done using either (*a*) x and y or (*b*) r and θ. Moreover, when θ is a constantly changing function of time (*d*), the arrow rotates at a rate ω.

and $$\tilde{z} = x + iy = r(\cos\theta + i\sin\theta)$$

The *Euler formula**

$$e^{i\theta} = \cos\theta + i\sin\theta$$

leads to the expression $e^{-i\theta} = \cos\theta - i\sin\theta$, and adding and subtracting these two equations yields

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

and $$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

Moreover, the Euler formula allows us (Fig. 2.17*b*) to write

$$\tilde{z} = re^{i\theta} = r\cos\theta + ir\sin\theta$$

where r is the *magnitude* of $\tilde{z}$ and θ is the *phase angle* of $\tilde{z}$, in radians. The magnitude is often denoted by $|\tilde{z}|$ and referred to as the *modulus* or *absolute value* of the complex number. The *complex conjugate*, indicated by an asterisk (Fig. 2.17*c*), is found by replacing i wherever it appears, with $-i$, so that

$$\tilde{z}^* = (x + iy)^* = (x - iy)$$

$$\tilde{z}^* = r(\cos\theta - i\sin\theta)$$

and $$\tilde{z}^* = re^{-i\theta}$$

The operations of addition and subtraction are quite straightforward:

$$\tilde{z}_1 \pm \tilde{z}_2 = (x_1 + iy_1) \pm (x_2 + iy_2)$$

and therefore

$$\tilde{z}_1 \pm \tilde{z}_2 = (x_1 \pm x_2) + i(y_1 \pm y_2)$$

Notice that this process is very much like the component addition of vectors.

Multiplication and division are most simply expressed in polar form

$$\tilde{z}_1\tilde{z}_2 = r_1r_2e^{i(\theta_1+\theta_2)}$$

and $$\frac{\tilde{z}_1}{\tilde{z}_2} = \frac{r_1}{r_2}e^{i(\theta_1+\theta_2)}$$

A number of facts that will be useful in future calculations are well worth mentioning at this point. It follows from the ordinary trigonometric addition formulas (Problem 2.44) that

$$e^{\tilde{z}_1+\tilde{z}_2} = e^{\tilde{z}_1}e^{\tilde{z}_2}$$

*If you have any doubts about this identity, take the differential of $\tilde{z} = \cos\theta + i\sin\theta$, where $r = 1$. This yields $d\tilde{z} = i\tilde{z}\,d\theta$, and integration gives $\tilde{z} = \exp(i\theta)$.

and so, if $\tilde{z}_1 = x$ and $\tilde{z}_2 = iy$,

$$e^{\tilde{z}} = e^{x+iy} = e^x e^{iy}$$

The modulus of a complex quantity is given by

$$r = |\tilde{z}| \equiv (\tilde{z}\tilde{z}^*)^{1/2}$$

and $$|e^{\tilde{z}}| = e^x$$

Inasmuch as $\cos 2\pi = 1$ and $\sin 2\pi = 0$,

$$e^{i2\pi} = 1$$

Similarly,

$$e^{i\pi} = e^{-i\pi} = -1 \quad \text{and} \quad e^{\pm i\pi/2} = \pm i$$

The function $e^{\tilde{z}}$ is periodic; that is, it repeats itself every $i2\pi$:

$$e^{\tilde{z}+i2\pi} = e^{\tilde{z}}e^{i2\pi} = e^{\tilde{z}}$$

Any complex number can be represented as the sum of a real part Re $(\tilde{z})$ and an imaginary part Im (z)

$$\tilde{z} = \text{Re}\,(\tilde{z}) + i\,\text{Im}\,(\tilde{z})$$

such that

$$\text{Re}\,(\tilde{z}) = \frac{1}{2}(\tilde{z} + \tilde{z}^*) \quad \text{and} \quad \text{Im}\,(\tilde{z}) = \frac{1}{2i}(\tilde{z} - \tilde{z}^*)$$

Both of these expressions follow immediately from the Argand diagram, Fig. 2.17*a* and *c*. For example, $\tilde{z} + \tilde{z}^* = 2x$ because the imaginary parts cancel, and so Re $(\tilde{z}) = x$.

From the polar form where

$$\text{Re}\,(\tilde{z}) = r\cos\theta \quad \text{and} \quad \text{Im}\,(\tilde{z}) = r\sin\theta$$

it is clear that either part could be chosen to describe a harmonic wave. It is customary, however, to choose the real part, in which case a harmonic wave is written as

$$\psi(x,t) = \text{Re}\,[Ae^{i(\omega t - kx + \varepsilon)}] \tag{2.36}$$

which is, of course, equivalent to

$$\psi(x,t) = A\cos(\omega t - kx + \varepsilon)$$

Henceforth, wherever it's convenient, we shall write the wavefunction as

$$\psi(x,t) = Ae^{i(\omega t - kx + \varepsilon)} = Ae^{i\varphi} \tag{2.37}$$

and utilize this complex form in the required computations. This is done to take advantage of the ease with which complex exponentials can be manipulated. Only after arriving at a final result, and then only if we want to represent the actual wave, must we take the real part. It has, accordingly, become quite common to write $\psi(x, t)$, as in Eq. (2.37), where it is understood that the actual wave is the real part.

Although the complex representation is commonplace in contemporary physics, it must be applied with caution: ***after expressing a wave as a complex function and then performing operations with/on that function, the real part can be recovered only if those operations are restricted to addition, subtraction, multiplication and/or division by a real quantity, and differentiation and/or integration with respect to a real variable.*** Multiplicative operations (including vector dot and cross products) must be carried out exclusively with real quantities. Wrong results can arise from multiplying complex quantities and then taking the real part (see Problem 2.47).

2.6 Phasors and the Addition of Waves

The arrow in the Argand diagram (Fig. 2.17*d*) is set rotating at a frequency ω by letting the angle equal ωt. This suggests a scheme for representing (and ultimately adding) waves that we will introduce here qualitatively and develop later (p. 286) quantitatively. Figure 2.18 depicts a harmonic wave of amplitude A traveling to the left. The arrow in the diagram has a length A and revolves at a constant rate such that the changing angle it makes with the reference x-axis is ωt. This rotating arrow and its associated phase angle together constitute a **phasor**, which tells us

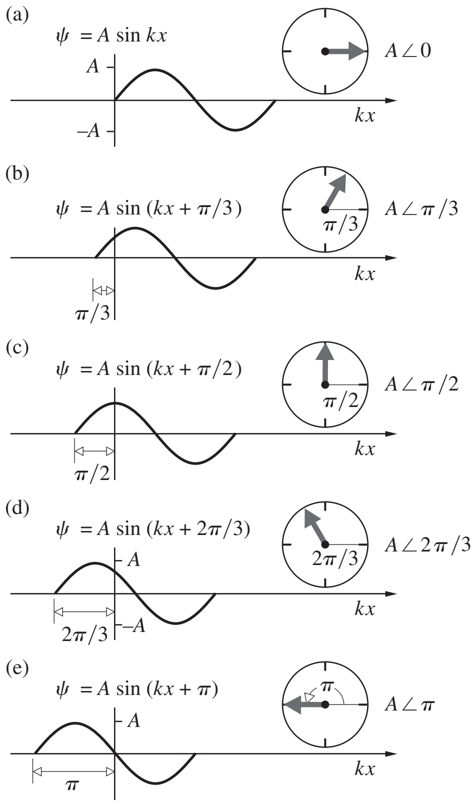


Figure 2.18 A plot of the function $\psi = A\sin(kx + \omega t)$ and the corresponding phasor diagrams. In (*a*), (*b*), (*c*), (*d*), and (*e*), the values of ωt are 0, $\pi/3$, $\pi/2$, $2\pi/3$, and π, respectively. Again the projection of the rotating arrow on the vertical axis equals the value of ψ on the $kx = 0$ axis.

everything we need to know about the corresponding harmonic wave. It's common to express a phasor in terms of its amplitude, A, and phase, φ, as $A\angle\varphi$.

To see how this works, let's first examine each part of Fig. 2.18 separately. The phasor in Fig. 2.18*a* has a zero phase angle; that is, it lies along the reference axis; the associated sine function can also serve as a reference. In Fig. 2.18*b* the phasor has a phase angle of $+\pi/3$ rad, and the sine curve is shifted to the left by $\pi/3$ rad. That sine curve reaches its first peak at a smaller value of kx than does the reference curve in part (*a*), and therefore it *leads* the reference by $\pi/3$ rad. In parts (*c*), (*d*), and (*e*) of Fig. 2.18, the phase angles are $+\pi/2$ rad, $+2\pi/3$ rad, and $+\pi$ rad, respectively. The entire sequence of curves can be seen as a wave, $\psi = A\sin(kx + \omega t)$, traveling to the left. It is equivalently represented by a phasor rotating counterclockwise such that its phase angle at any moment is ωt. Much the same thing happens in Fig. 2.7, but there the wave advances to the right and the phasor rotates clockwise.

When wavefunctions are combined, we are usually interested in the resulting amplitude and phase. With that in mind, reexamine the way waves add together in Fig. 2.16. Apparently, for disturbances that are in-phase (Fig. 2.16*a*) the amplitude of the resultant wave, A, is the sum of the constituent amplitudes: $A = A_1 + A_2 = 1.0 + 0.9 = 1.9$. This is the same answer we would get if we added two colinear vectors pointing in the same direction. Similarly (Fig. 2.16*d*), when the component waves are 180° out-of-phase, $A = A_1 - A_2 = 1.0 - 0.9 = 0.1$ as if two colinear oppositely directed vectors were added. Although phasors are not vectors, they do add in a similar way. Later, we'll prove that two arbitrary phasors, $A_1\angle\varphi_1$ and $A_2\angle\varphi_2$, combine tip-to-tail, as vectors would (Fig. 2.19), to produce a resultant $A\angle\varphi$. Because both phasors rotate together at a rate ω, we can simply freeze them at $t = 0$ and not worry about their time dependence, which makes them a lot easier to draw.

The four phasor diagrams in Fig. 2.20 correspond to the four wave combinations taking place sequentially in Fig. 2.16. When the waves are in-phase (as in Fig. 2.16*a*), we take the phases of both wave-1 and wave-2 to be zero (Fig. 2.20*a*) and position the corresponding phasors tip-to-tail along the zero-φ reference axis. When the waves differ in phase by $\pi/3$ (as in Fig. 2.16*b*), the phasors have a relative phase (Fig. 2.20*b*) of $\pi/3$. The resultant, which has an appropriately reduced amplitude, has a phase φ that is between 0 and $\pi/3$, as can be seen in both Figs. 2.16*b* and 2.20*b*. When the two waves differ in phase by $2\pi/3$ (as in Fig. 2.16*c*), the corresponding phasors almost form an equilateral triangle in Fig. 2.20*c* (but for the fact that $A_1 > A_2$), and so A now lies between A_1 and A_2. Finally, when the phase-angle difference for the two waves (and the two phasors) is π rad (i.e., 180°), they almost cancel and the resulting amplitude is a minimum. Notice (in Fig. 2.20*d*) that the resultant phasor points along the reference axis and so has the same phase (i.e., zero) as $A_1\angle\varphi_1$. Thus it is 180° out-of-phase with $A_2\angle\varphi_2$; the same is true of the corresponding waves in Fig. 2.16*d*.

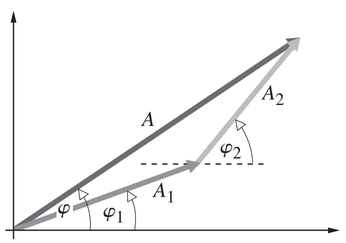


Figure 2.19 The sum of two phasors $A_1\angle\varphi_1$ and $A_2\angle\varphi_2$ equals $A\angle\varphi$. Go back and look at Fig. 2.13, which depicts the overlapping of two sinusoids having amplitudes of $A_1 = 1.0$ and $A_2 = 0.9$ and phases of $\varphi_1 = 0$ and $\varphi_2 = 1.0$ rad.

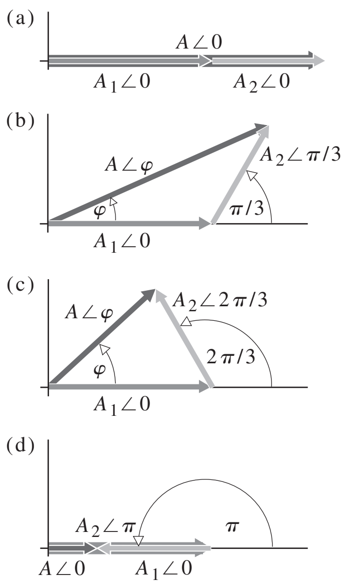


Figure 2.20 The addition of phasors representing two waves having amplitudes of $A_1 = 1.0$ and $A_2 = 0.9$ with four different relative phases, as shown in Fig. 2.16.

This was just the briefest introduction to phasors and phasor addition. We will come back to the method in Section 7.1, where it will be applied extensively.

2.7 Plane Waves

A light wave can be described at a given time at a point in space by its frequency, amplitude, direction of propagation, and so forth, but that doesn't tell us much about the optical disturbance existing over an extended area of space. To find out about that we introduce the spatial concept of a **wavefront**. Light is vibratory, it corresponds to harmonic oscillations of some sort, and the one-dimensional sine wave is an important element in beginning to envision the phenomenon. Figure 2.14 shows how radially traveling sinusoids, fanned out in two dimensions, can be understood to form a unified expanding disturbance, a *circular wave*. Each crest, from every one-dimensional wavelet traveling outward, lies on a circle and that's true of the troughs as well—indeed, it's true for any specific wave magnitude. For any particular phase (say, $5\pi/2$) the component sinusoids have a particular magnitude (e.g., 1.0) and all points with that magnitude lie on a circle (of magnitude 1.0). In other words, the loci of all the points where the phase of each one-dimensional wavelet is the same form a series of concentric circles, each circle having a particular phase (for crests that would be $\pi/2$, $5\pi/2$, $9\pi/2$, etc.).

Quite generally, ***at any instant a wavefront in three dimensions is a surface of constant phase***, sometimes called a *phase front*. In actuality wavefronts usually have extremely complicated configurations. The light wave reflected from a tree or a face is an extended, irregular, bent surface full of bumps and depressions moving out and away, changing as it does. In the remainder of this chapter we'll study the mathematical representations of several highly useful idealized wavefronts, ones that are uncomplicated enough to write easy expressions for.

The plane wave is perhaps the simplest example of a three-dimensional wave. It exists at a given time, when all the surfaces on which a disturbance has constant phase form a set of planes, each generally perpendicular to the propagation direction. There are quite practical reasons for studying this sort of disturbance, one of which is that by using optical devices, we can readily produce light resembling plane waves.

The mathematical expression for a plane that is perpendicular to a given vector $\vec{\mathbf{k}}$ and that passes through some point (x_0, y_0, z_0) is rather easy to derive (Fig. 2.21). First we write the position vector in Cartesian coordinates in terms of the unit basis vectors (Fig. 2.21*a*),

$$\vec{\mathbf{r}} = x\hat{\mathbf{i}} + y\hat{\mathbf{j}} + z\hat{\mathbf{k}}$$

It begins at some arbitrary origin O and ends at the point (x, y, z), which can, for the moment, be anywhere in space. Similarly,

$$(\vec{\mathbf{r}} - \vec{\mathbf{r}}_0) = (x - x_0)\hat{\mathbf{i}} + (y - y_0)\hat{\mathbf{j}} + (z - z_0)\hat{\mathbf{k}}$$

By setting

$$(\vec{\mathbf{r}} - \vec{\mathbf{r}}_0)\cdot\vec{\mathbf{k}} = 0 \tag{2.38}$$

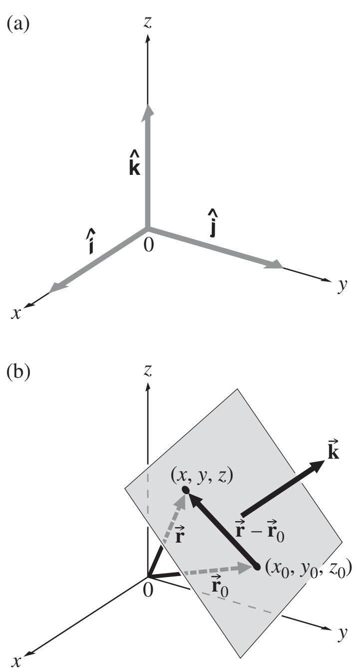


Figure 2.21 (*a*) The Cartesian unit basis vectors. (*b*) A plane wave moving in the $\vec{\mathbf{k}}$-direction.

we force the vector $(\vec{\mathbf{r}} - \vec{\mathbf{r}}_0)$ to sweep out a plane perpendicular to $\vec{\mathbf{k}}$, as its endpoint (x, y, z) takes on all allowed values. With

$$\vec{\mathbf{k}} = k_x\hat{\mathbf{i}} + k_y\hat{\mathbf{j}} + k_z\hat{\mathbf{k}} \tag{2.39}$$

Equation (2.38) can be expressed in the form

$$k_x(x - x_0) + k_y(y - y_0) + k_z(z - z_0) = 0 \tag{2.40}$$

or as

$$k_x x + k_y y + k_z z = a \tag{2.41}$$

where

$$a = k_x x_0 + k_y y_0 + k_z z_0 = \text{constant} \tag{2.42}$$

The most concise form of the equation of a plane perpendicular to $\vec{\mathbf{k}}$ is then just

$$\vec{\mathbf{k}}\cdot\vec{\mathbf{r}} = \text{constant} = a \tag{2.43}$$

The plane is the locus of all points whose position vectors each have the same projection onto the $\vec{\mathbf{k}}$-direction.

We can now construct a set of planes over which $\psi(\vec{r})$ varies in space sinusoidally, namely,

$$\psi(\vec{r}) = A\sin(\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}) \tag{2.44}$$

$$\psi(\vec{r}) = A\cos(\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}) \tag{2.45}$$

or

$$\psi(\vec{r}) = Ae^{i\vec{\mathbf{k}}\cdot\vec{\mathbf{r}}} \tag{2.46}$$

For each of these expressions $\psi(\vec{r})$ is constant over every plane defined by $\vec{\mathbf{k}}\cdot\vec{\mathbf{r}} =$ constant, which is a surface of constant phase (i.e., a wavefront). Since we are dealing with harmonic functions, they should repeat themselves in space after a displacement of λ in the direction of $\vec{\mathbf{k}}$. Figure 2.22 is a rather humble representation of this kind of expression. We have drawn only a few of the infinite number of planes, each having a different $\psi(\vec{r})$. The planes should also have been drawn with an infinite spatial extent, since no limits were put on $\vec{\mathbf{r}}$. The disturbance clearly occupies all of space.

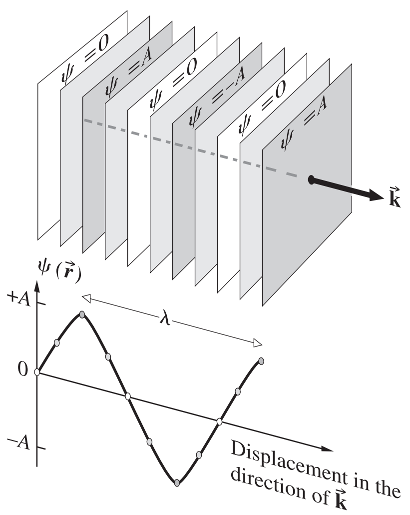


Figure 2.22 Wavefronts for a harmonic plane wave.